DOUGLAS SELF

THE DESIGN OF ACTIVE CROSSOVERS

Focal Press

The Design of Active Crossovers

This page intentionally left blank

The Design of Active Crossovers

Douglas Self

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Focal Press is an Imprint of Elsevier

Focal Press

Focal Press is an imprint of Elsevier
225 Wyman Street, Waltham, MA 02451, USA
The Boulevard, Langford Lane, Kidlington, Oxford, OX5 1GB, UK

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data
Self, Douglas.
The design of active crossovers / Douglas Self.
p. cm.
Includes bibliographical references.
ISBN 978-0-240-81738-5
1. Electric filters, Active–Design and construction. I. Title.
TK7872.F5S447 2011
621.3815'324–dc22 2011013096

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

For information on all Focal Press publications
visit our website at *www.elsevierdirect.com*

Printed in the United States of America
11 12 13 14 5 4 3 2 1

To Julie with all my love

This page intentionally left blank

Contents

This page intentionally left blank

Preface

> *"Unprovided with original learning, unformed in the habits of thinking, unskilled in the arts of composition, I resolved to write a book."*
>
> **Edward Gibbon**

This is believed to be the first book concentrating wholly on active crossovers—in fact, it is probably the first book that focuses exclusively on crossovers of any sort. In a field where any sort of consensus is rare, there is general agreement that audio systems with active crossovers and multiple amplifiers to drive the loudspeakers sound unquestionably better than their passive crossover counterparts. I think that the use of active crossovers may well be the next big step in hi-fi, and that was part of my motivation for writing this book. The use of active crossovers in sound reinforcement is also very fully covered.

You might think that active crossovers are a narrow field for a book, but actually the scope of the subject is rather wider than it sounds. There is a lot of material on active filters, low-noise design, and generally on the intelligent use of opamps, which has much wider applications. More broadly than that, there are many examples of how to tackle a problem logically and get results which are both optimal and economical.

This book will also be useful for passive crossover design. It gives comprehensive details of all the basic crossover alignments, plus some more exotic types such as Gaussian and Legendre crossovers. Having chosen one, whether you use an active crossover or a passive one is up to you. Passive crossover design is greatly complicated by reactive drive unit impedances, the problems of matching drive unit sensitivities, the need to present a reasonable load impedance to an amplifier, etc, so a given crossover alignment is something to be aimed at hopefully. With an active crossovers, be it analogue or digital, the required alignment can be realised precisely.

As to the analogue versus DSP issue, I have tried to structure this book so that at least two-thirds of it is applicable to either the analogue or the DSP approach. You study the concepts, you choose a crossover type, and when you find that at some point you need, say, a second-order Butterworth filter, it is then your choice whether to break out the 5532s or start cutting code for a DSP.

However, when it comes down to the implementation, the focus of this book is analogue, and to be specific, opamp-based analogue. Instructions on DSP implementation would need to discuss topics like the bilinear transform, word-length effects (Fixed-point or floating point? Single or double precision?), the different IIR filter structures, convolution methods for FIR, noise shaping, and so on. All good stuff, but none of it crossover-specific. It can be found in many DSP textbooks. In contrast, a lot of the material on analogue implementation, *is* crossover-oriented, because the use of Sallen & Key filters puts opamps in particularly demanding situations.

I have nothing whatsoever against the DSP approach. It was I that did the initial digital mixing console work at Soundcraft on the Motorola 56001 processor back in the mid 1980s. Clearly some functions such as pure time-delay are much easier in the digital world, and there is unquestionably the advantage of very great precision in setting filter characteristics with no worries about capacitor tolerances and so on. It is doubtful if it is practical to make an eighth-order crossover any other way. Nonetheless, there was just not space to give any meaningful attention to DSP issues.

In this book I have not included the underlying mathematics of filter design. Active crossovers inevitably involve a lot of filter design, and if you have ever looked into a filter textbook you will probably have found it bristling with long equations spattered with Laplace variables (those enigmatic s's), and some heavyweight complex algebra; this does not just mean *complicated* algebra, but algebra that incorporates j, the square root of minus-one. There will be much talk of placing poles and zeros on the complex plane.

This is all very well, and you will unquestionably have a better feel for filter design if you understand it, but it is my contention that today it is not necessary for practical filter design. I say today, because nowadays it is possible to download very capable simulation software for free. There is no need to evaluate unwieldy equations to plot your filter response—you just enter the circuit and in a minute or two you can have the frequency response, phase response, and group delay curve in as much detail as you like. This is quite a significant development, as for a start it means that this book can be produced in a reasonable length. Putting in all the equations would have doubled its size, as well as discouraging a lot of people.

That does not mean there is no mathematics at all; I have included a large number of design equations for specific circuits so the component values can be calculated very rapidly. You don't need to be able to handle anything worse than a square root. There are many fully worked-out practical circuits with component values that can be simply scaled to give the characteristics required. There are tables that allow you to design a wide range of filters very quickly indeed.

This book goes all the way from the basic concepts of crossovers, however implemented, to the details of making a non-standard resistance value by putting two resistors in series or parallel. You might think nothing could be simpler—but actually combining two resistors to get the most accurate result is a surprisingly subtle business.

There is a lot of new material here that has never been published before. To pick a few examples at random: using capacitance multipliers in biquad equalisers, opamp output biasing to reduce distortion, the design of NTM™ notch crossovers, the design of special filters for filler-driver crossovers, the use of mixed capacitors to reduce filter distortion, differentially elevated internal levels to reduce noise, and so on.

What you will not find in this book is any homage to Subjectivism—the cult of thousand-pound cables, audio homeopathy, and faith-based audio in general. Truck with that I will have none of.

I don't claim that this book contains all knowledge on crossovers. If it tried to do that, I think it might have come out as a ten-volume set. I have however given, both in the main text and consolidated in two big appendices, a very large number of references so that any topic can be pursued pretty much to the limit of published knowledge. It is a very long time since any one person could credibly claim to know all of science. We may have already reached the point where it is impossible for any one person to know all about crossover design. I hope this book will prove useful.

Douglas Self

London

Jan 2011

This page intentionally left blank

Acknowledgments

Thanks to Julie for her patience and very practical support as I spent long hours slaving over a hot simulator.

Thanks to Gareth Connor for practical help, unfailing encouragement, and providing the facilities with which some of the experiments in this book were done.

Thanks to Don Morrison for generously providing sample components.

Thanks to Rod Elliot for many useful discussions on notched crossovers.

Thanks to all those on the diyAudio forum who made constructive suggestions on the topics to include in this book.

CHAPTER 1

Crossover Basics

I do hope you've read the Preface. It is not mere amiable meanderings, but gives an oversight as to how this book is constructed and what it is intended to do.

1.1 What a Crossover Does

The basic function of any crossover, be it passive or active, analogue or digital, is to take the audio spectrum that stretches roughly from 20 Hz to 20 kHz and split it into two, three, or sometimes more bands so they can be applied to loudspeaker drive units adapted for those frequencies. In hi-fi use the crossover frequencies are usually fixed and intended for work with one particular loudspeaker design, but for sound reinforcement applications the crossover frequencies are normally variable by front panel controls.

There are also other functions that are sometimes but not always performed by crossovers, and we may list them all as follows, roughly in order of importance:

1. Splitting the audio spectrum into two, three, or occasionally more bands
2. Equalisation to correct drive unit frequency responses
3. Correction for unmatched drive unit sensitivities
4. The introduction of time delays into the crossover outputs to correct for the physical alignment of drive units
5. Equalisation to correct for interactions between drive units and the enclosure, for example, diffraction compensation
6. Equalisation to correct for loudspeaker-room interactions, such as operating in half-space as opposed to quarter-space
7. Enhancement of the natural LF response of the bass drive unit and enclosure by applying controlled bass boost

Most of these functions are equally applicable to both passive and active crossovers, but the time-delay function is rarely implemented in passive crossovers because it requires a lot of expensive components and involves significant power losses.

Some of these functions will probably appear wholly opaque if you are just starting this book, but stay with me. All will be revealed.

The Design of Active Crossovers

1.2 Why a Crossover Is Necessary

The need for any crossover at all is rooted in the impracticability of making a drive unit that can handle the whole ten-octave audio spectrum satisfactorily. This is not merely because the technology of loudspeaker construction is inadequate, but is also based on some basic physics. Ideally the acoustic output of a loudspeaker would come from a single point; with such a source the sound field is uniform, because there can be no interference effects that result from multiple sources, or from a source of finite size.

A tweeter has a small physical size, with a dome usually around an inch (2 or 3 cm) in diameter, and approximates fairly well to a point source. This technology works very well for high frequencies, say down to 1 kHz, but is hopelessly inadequate for bass reproduction because such a small area cannot move much air, and to reproduce bass frequencies you need to move a lot of it. Low-frequency drive units are therefore of much greater diameter, up to 12 inches for domestic hi-fi and up to 18 inches or more for sound-reinforcement applications. As cone area is proportional to the square of diameter, a 12-inch drive unit has 144 times the area of a typical tweeter, and an 18-inch unit has 324 times the area.

It is not at present technically possible to make a big low-frequency drive unit that works accurately up to 20 kHz, because as the frequency increases the cone ceases to move as a unit-it is not one of those most desirable "infinitely rigid pistons" that are always cropping up in loudspeaker theory but never in manufacturer's catalogues. This effect is often called "cone break-up" not because the cone physically falls apart but because, due to its finite stiffness, with rising frequency its surface divides up into separate areas of vibration. This unhappy state of affairs is put to advantage in so-called "full-range" loudspeakers which have a "parasitic tweeter" in the form of a small cone, attached to the voice coil. The idea is that at higher frequencies the main cone does its own thing and is effectively decoupled from the voice coil and the tweeter cone, allowing the latter to radiate high frequencies without being restrained by the much greater mass of the main cone; what you might call a mechanical crossover. As you might imagine, there are many compromises involved in such a simple arrangement and the response is generally much inferior to a good two-way loudspeaker with separate bass unit and tweeter.

Nonetheless, the field of audio being what it is, there are a certain number of hi-fi enthusiasts who advocate full-range speakers for various reasons. Eliminating a passive crossover naturally increases power efficiency (as none is lost in the crossover components) and reduces cost.

1.3 Beaming and Lobing

Even if it was possible to make a large-diameter drive unit which covered the whole audio spectrum perfectly, there is a powerful reason why it is far from certain that this would be a good idea. If a radiating surface is of finite size, then if you stand to the side of the central axis, sound from one area of the drive unit will arrive at your ear at a slightly different time because of the differing path lengths through the air. This will cause interference between the two signals, and, given the right amount of path difference, complete cancellation. There will therefore be major irregularities in the frequency response anywhere but on the central axis. This is called “beaming” or “lobing” and it occurs when the diameter of the radiating object is comparable with the wavelength of the sound. It is obviously to be avoided as much as possible; the variation in response as the angle between the listener and the centre axis changes is called the polar response, and a “uniform polar response” is much sought after in loudspeaker design.

The beaming phenomenon is why a tweeter has to be of small diameter if it is to approach having a uniform polar response. Deciding when beaming becomes significant depends on the application, but the following figures in Table 1.1 [1] have been put forward:

Table 1.1: Onset of Beaming versus Drive Unit Diameter

Driver Diameter (Inches)	Beaming Onset Frequency (Hz)
1	13,680
2	6840
5	3316
6.5	2672
8	2105
10	1658
12	1335
15	1052
18	903

These frequencies are approximately those at which the wavelength in air equals the driver diameter. The whole business of beaming, lobing, and polar response generally is obviously much too complex to be summed up in a single table, but it does give some indication of when you need to start worrying about it.

There is of course much more to a crossover than simply splitting the audio signal into separate frequency bands. The vital point to understand is that the splitting has then to be followed by summation. the frequency bands have to be joined together again seamlessly. This requires the acoustic signals be summed to be correct not only in amplitude but also in phase. The crossover and speaker system can only create the exactly correct signal at one point in space, which is unfortunate, as we have two ears and each listener therefore needs the signal to be correct at *two* points in space. Crossover design is always a matter of compromise to some degree.

It is not sufficient to get a perfect response on-axis, even if one interprets this as being capable of summing correctly at both ears. The off-axis output from the loudspeaker will not only be heard by those in the room unfortunate enough to not get the best seat on the sofa, but it also creates the ambient sound environment through room reflections and reverberation. If it has serious response irregularities then these will detract from the listening experience, even if the direct on-axis sound is beyond reproach.

The term “lobing” is also used to describe the reinforcements and cancellations that occur when two separate drive units are radiating; in this case their size is relatively unimportant because interference would still occur even if both were point sources. When the radiation is shifted at the crossover frequency because the signals to the two drive units are not in phase, this is called “lobing error.” There is much more on this in Chapters 3 and 4.

1.4 Active Crossover Applications

The main fields of application for active crossovers in association with multi-way loudspeakers are high-end hi-fi, sound reinforcement, automotive audio, sound recording studios, cinema theatres, and film studios. In hi-fi, active crossover technology offers better and more consistent quality than passive crossovers. We shall look closely at why this is so later in this chapter.

In the area of sound reinforcement the use of active crossovers is virtually mandated by the need to use banks of loudspeakers with different characteristics, especially sub-woofers. The size and number of the loudspeaker cabinets used mean that it is physically impossible to put them close together, and hence sophisticated control of time delays is essential to obtain the desired coverage and polar responses. The large amount of power used in a typical sound reinforcement system means that the losses inherent in the use of passive crossovers cannot be tolerated. The high-power requirement also means that multiple power amplifiers are always used, and the extra cost of an active crossover system is very small by comparison.

Automotive audio marches to its own drummer, so to speak, the priority of most of its exponents being the maximum possible level of bass at all costs. This is perhaps not the place to speculate on whether this is driven by an appreciation of musical aesthetics or macho territorial aggression, but the result is that subwoofer systems are very popular, and so naturally some sort of crossover system is required. This is usually an active crossover, because the high power levels once again make the losses in a passive crossover unacceptable. This is particularly true because 4 Ω loudspeakers are used, so the current levels in inductors are doubled and I^2R losses are quadrupled, compared with the 8 Ω situation.

Active crossovers do have other important applications besides driving multiway loudspeakers. They are also used in multi-band signal processing, of which the most common example is multi-band compression. A multi-band compressor uses a set of filters, working on exactly the same principle as a loudspeaker crossover, to split the audio signal into two, three, four or even more frequency bands; three or four-band compressors are the most popular. On emerging from the crossover, each band is fed to a separate compressor, after which the signals are recombined, usually by simple summation. This is delightfully simple compared with the acoustic summation that recombines the outputs of the different drive units of a multiway loudspeaker, because there are no problems with polar response or time delays.

The great advantage of multi-band compression is that a peak in level in one frequency band will not cause any gain reduction in the other bands; a high-level transient from bass guitar or kick-drum will not depress the level of the whole mix. Another feature is the ability to use different attack/decay times for different frequency bands. You may be thinking at this point that it would have been more sensible to compress the kick drum *before* you mixed it in with everything else, and you are of course quite right. However, in many situations you are not doing the mixing but dealing with fully mixed audio as it comes along. Radio stations (not excluding the BBC) make considerable use of multi-band compression and limiting on existing stereo material to maximise the impact of their transmissions.

Other applications for multiband processing include multiband distortion, where splitting the distortion operation into separate bands prevents intermodulation turning the sound into an unpleasant muddle. A simple example of this is the "high frequency exciter" or "psychoacoustic enhancer" technology where a filter selects some part of the upper reaches of the audio spectrum and applies distortion to it in order to replace missing or under-strength harmonics. Multiband techniques have also been used in noise reduction techniques, notably in the Dolby A noise reduction system.

The active crossovers used for multi-band signal processing do essentially the same job as those for loudspeakers, and this book will be most useful in their design.

1.5 Bi-Amping and Bi-Wiring

The use of two separate amplifiers driven by an active crossover is sometimes called bi-amping. This is nothing to do with the use of two amplifier channels for stereo. If there are three separate amplifiers powering three drive units, this is called tri-amping or simply multi-amping. Bi-amping should not be confused with bi-wiring, which is a completely different idea. It is worth taking a quick look at it because the alleged benefits of bi-wiring are very relevant to crossover operation.

The most common method of wiring between an amplifier and a multi-way loudspeaker is to use a single cable, as shown in Figure 1.1a. Note that the passive crossover networks shown are the simplest possible, and real passive crossovers will have more components.

In bi-wiring (or tri-wiring, or "multi-cabling") only one amplifier is used, but separate cables run from it to separate sections of a passive crossover that is mounted as usual inside

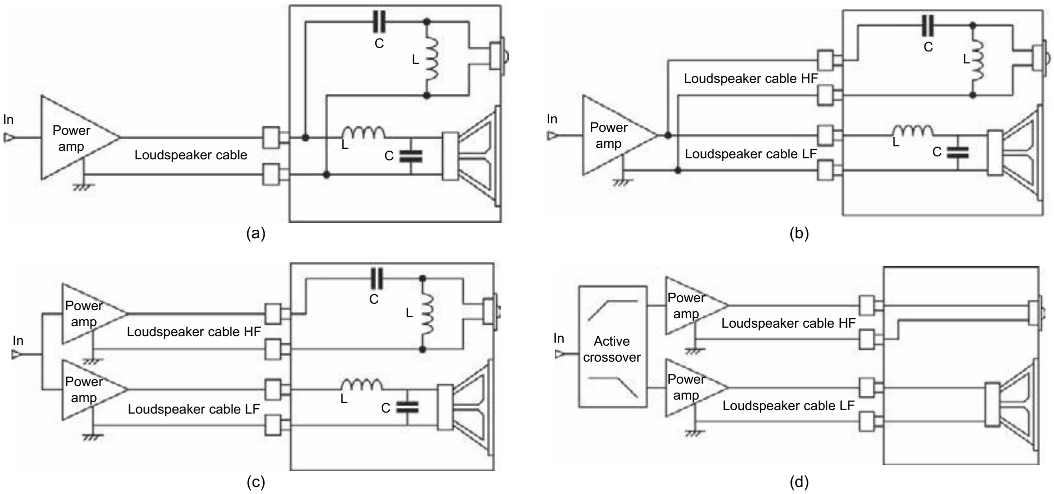


Figure 1.1: (a) Normal loudspeaker wiring (b) bi-wiring (c) full-band multi-amplification, and (d) active-crossover multi-amplification.

the loudspeaker enclosure; this is illustrated in Figure 1.1b. Bi-wiring requires individual access to the passive crossover sections for the LF unit and tweeter, which is arranged for by having two sets of terminals on the rear of the loudspeaker connected together with shorting links that can be easily removed.

Arranging things for tri-wiring is a little more expensive and difficult, as it requires six terminals and four shorting links, plus the facility of connecting two shorting links to one terminal in a tidy manner. This is probably why loudspeakers with three drive units are more often connected for bi-wiring rather than tri-wiring. Normally the tweeter and mid drive units are connected to one pair of terminals and the bass drive unit to the other. There is also the point that if your taste runs to expensive loudspeaker cables of undefined superiority, having three sets of them is significantly more costly than two.

The question is, just what is gained by bi-wiring? The hard fact is: not much. Its enthusiasts usual claim that separating the high and low frequencies into different cables prevents "intermodulation" of some sort in the cabling. I must point out at once that even the cheapest copper cable is absolutely linear and can generate no distortion whatsoever, even when passing heavy currents. The same absolute linearity is shown at the smallest signal levels that can be measured [2]. From time to time some of our less bright audio commentators speculate in the vaguest of terms that a metallic conductor is actually some sort of swamp of "micro-diodes," and that non-linearity might just conceivably be found if the test signal levels were made low enough. This is categorically untrue and quite impossible with it. The physics of conduction in metals has been completely understood for a long time, and there simply is no threshold effect for metallic conduction. The only way that a cabling system can introduce non-linear distortion is if the connectors are defective, which in this context pretty much means "about to fall apart altogether."

Actually, a lot of audiophiles are much less definite about he advantages of bi-wiring. They tend to say something pitifully vaporous like: "If only one pair of conductors delivers the whole signal, there is a danger that the bass frequencies can affect the transmission of the more delicate and subtle treble frequencies." Affect the transmission? Exactly how, pray? And how much danger is there? Not much? Immediate extreme peril? Danger, Will Robinson!

Where bi-wiring can be of use is in minimising the interaction of loudspeaker impedance variations with cable resistances. Let's suppose that the combination of LF crossover and bass drive unit has a dip in its impedance: (the drive unit alone cannot show an impedance lower than its voice coil resistance); this dip will cause a greater voltage-drop across the cable resistance, giving a dip in driver output that would not have existed if the cable (and the amplifier) had zero resistance. If we suppose again, and this is a bit more of a stretch, that impedance dips in the HF and LF impedance might overlap, then applying them to separate cables will reduce the level variations. However, the same effect could be got by doubling up a non-bi-wired loudspeaker cable, as this would halve its resistance.

Bi-wiring has a certain enduring appeal because it enables the experimenter to feel he has done something sophisticated to his system, but requires little effort, relatively little cost, (though this of course depends on how much you think you need to pay for acceptable loudspeaker cables) and the chances of anything going wrong are minimal. This is not of course the case if you start delving about inside a power amplifier.

Bi-amping means a separate amplifier for each crossover section. This almost always implies the use of an active crossover to split the audio spectrum and feed different sections to different amplifiers. Occasionally, however, you hear advocated what might be called full-range multi-amplification. This is shown in Figure 1.1c, where two power amplifiers are fed with identical signals and drive the passive crossover sections separately. This is an expensive way to gain relatively small advantages. The loading on each amplifier will be less, and you might get some of the advantages of bi-amping, such as less stress on each amplifier, and reduced intermodulation of LF and HF signals in the amplifiers. The advantage of the latter should be negligible with well-designed solid-state amplifiers, but I suppose there are valve amplifiers to consider. There is also the possibility of using a smaller power amplifier for the HF path. In general, however, this approach combines most of the disadvantages of passive crossovers with the extra cost of active crossovers.

Figure 1.1d shows standard bi-amping, with two power amplifiers fed with the appropriate frequency bands by an active crossover, the passive crossover no longer being present. This is the approach dealt with in this book.

1.6 Loudspeaker Cables

It would be somewhat off-topic to go into a lot more detail about loudspeaker cable design, but this might be a good place to point out that there is absolutely nothing magical or mysterious about them. My findings [3] are that, looking at the amplifier–cable–loudspeaker system as a whole, the amplifier and cable impedances have the following effects:

- Frequency response variation due to the cable resistance forming the upper arm of a potential divider with the loudspeaker load as the lower arm. The effect of the resistive component from the amplifier output impedance is usually negligible with a solid-state design, but this is unlikely to be the case for valve amplifiers. This is at least potentially audible and is a good reason for using thick cables. It is not a good reason for using anything other than ordinary copper cable.
- A high-frequency roll-off due to the cable inductance forming an LR lowpass filter with the loudspeaker load. The amplifier's series output inductor (almost always present to give stability with capacitive loads) adds directly to this to make up the total series inductance. The shunt capacitance of any normal speaker cable is trivially small, and can have no significant effect on frequency response or anything else.

The main factors in speaker cable selection are therefore series resistance and inductance. If these parameters are less than 100 mΩ for the round-trip resistance and less than 3 μH for the total inductance, any effects will be imperceptible [3]. These conditions can be met by standard 13-Amp mains cable. (I'm not quite sure how the equivalent cable is labelled in the United States.) This cable has three conductors (Live, Neutral, and Earth) each of 1.25 sq mm cross-section, made up of 40 × 0.2 mm strands. Using just two of the three conductors, a 100 mΩ round-trip resistance allows 3.7 metres of cable. The lowest cable resistance is obtained if all three conductors are used, normally by paralleling the Neutral and Earth conductor on the cold (grounded) side of the cable; the maximum length for 100 mΩ is now 5.0 metres, which should do for most of us. This three-conductor method does give what I suppose you might call an "asymmetric" cable, which could offend some delicate sensibilities, but I can assure you that it works nicely.

The loudspeaker cables that I have in daily use are indeed made of such 13-Amp mains cable, bought from an ordinary hardware shop nearly 40 years ago. Should a passing audiophile query the propriety of using such humble cabling, I usually tell them that with so much passage of time in regular use, the electrons have been thoroughly shaken loose and move about with the greatest possible freedom. I do hope nobody reading this book is going to take that seriously.

1.7 The Advantages and Disadvantages of Active Crossovers

Here I have tried to put down all the advantages and disadvantages of the active crossover approach. Some of them may not be very comprehensible until you have read the relevant chapter of this book. My initial plan was to attempt to put them in order of importance, but this is not an easy thing to do. The order here is therefore to some extent subjective, if I may use the term…

1.7.1 The Advantages of Active Crossovers

The advantages of active crossovers are:

1. The over-riding advantage of an active crossover is it offers ultimate freedom of design as virtually any frequency or phase response that can be imagined can be used. The filter slopes of the crossover can be made as steep as required without using large numbers of big and relatively expensive components. Any increase in passive crossover complexity means a significant increase in cost.
2. The design of passive crossovers is restricted by the need to keep the loading on the power amplifier within reasonable limits. With an active crossover, correction of the response for each driver is much simpler as it can be undertaken without having to worry about the combined load becoming too low in impedance for the average amplifier.
3. The design of passive crossovers is further complicated by the need to keep the power losses in the crossover within reasonable limits. The losses in the resistors and in the inductors (because of their inevitable series resistance) of a passive crossover, especially a complex design employing high-order filters or time delay compensation, can be very serious. In a big sound-reinforcement system the losses would be measured in tens of kilowatts. Not only does this seriously degrade the power efficiency of the overall system by wasting power that could be better applied to the loudspeakers, but it also means that the crossover components have to be able to dissipate a significant amount of heat, and are correspondingly big, heavy, and expensive. It is far better to do the processing at the small-signal level; the power used by even a sophisticated active crossover is trivial.
4. If one of the power amplifiers is driven into clipping, that clipping is confined to its own band. Clipping is usually less audible in the bass, so long as there is no intermodulation with high frequencies. It has been stated that an active crossover system can be run 4 dB louder for the same subjective impairment. This is equivalent to more than twice the power, but less than twice the perceived volume, which would require a 10 dB increase in SPL.
5. Delays can be added to compensate for differing acoustic centres for the drive units planes quite easily. Passive delay-lines can be built but are prodigal in their use of expensive, lossy, and potentially non-linear inductors, and as a result have high overall losses.
6. Tweeters and mid drive units can have resonances outside their normal operating range, which are not well suppressed by a passive crossover because it does not put a very low

source impedance across the voice coil. The presence of a series capacitor can greatly reduce the damping of the main resonance [4], and it is also possible for a series capacitor to resonate with the tweeter voice-coil inductance [5], causing an unwanted rise in level above 10 kHz or thereabouts.

7. Drivers of very different sensitivities can be used, if they happen to have the best characteristics for the job, without the need for large power-wasting resistances or expensive and potentially non-linear transformers or auto-transformers. If level controls for the drive units are required, these are very straightforward to implement in an infinitely-variable fashion with variable resistors. When passive crossovers are fitted with level controls (typically for the mid unit or tweeter, or both) these have to use tapped auto-transformers or resistor chains, because the power levels are too high for variable resistors, and so control is only possible in discrete steps.
8. The distortion of the drive unit itself may be reduced by direct connection to a low-impedance amplifier output [6]. It is generally agreed that the current drawn by a moving coil drive unit may be significantly non-linear, so if it is taken from a non-zero impedance, the voltage applied to the drive unit will also be distorted. This may be related to out-of-band tweeter resonance; Jean-Claude Gaertner states that tweeters can have increased distortion below 1 kHz [7]. I do not know for sure, but I strongly suspect, that when drive units are being developed they are driven from amplifiers with effectively zero output impedance, and that linearity is optimised under this condition. Any other approach would mean guessing at the source impedance, which given the number of ways in which it could vary, would be a quite hopeless exercise.
9. With modern opamps and suitable design techniques, an active crossover can be essentially distortion-free, though care must be taken with the selection of capacitors in the filters. It will not however be noise-free, though the noise levels can be made very low indeed by the use of appropriate techniques; these are described later in this book. A passive crossover contains inductors, which if ferrite or iron-cored will introduce distortion. It also contains capacitors, often in the form of non-polar electrolytics, which are not noted for their linearity, or the stability of their value over time. I haven't been able to find any published data on either of these problems. Capacitor linearity is very definitely an issue because they are being used in filters and therefore have significant voltage across them. It is possible for capacitor distortion to occur in active crossovers too, but the signal voltages are much lower and one can expect the amount of distortion generated to be much less. See Chapter 12, where using the worst sort of capacitor increases the distortion from 0.0005% to 0.005%, with a signal level of 10 Vrms. In contrast, the distortion from a passive crossover can easily exceed 1%.
10. With the rise of AV there is more experience in making multi-channel amplifiers economically. The separate-module-for-each-channel approach, where each module has a small toroid mounted right up at one end, while the input circuitry is at the other, is more expensive to manufacture but can give an excellent hum and crosstalk

performance. The main alternative is the huddle-around-the-big-central-toroid approach, which has some serious and intractable hum issues.

11. If a protection system is fitted that is intended to guard the drive units against excessive levels, then it can be closely tailored for each drive unit.
12. Voice coil heating will increase the resistance of the wire in its windings, reducing the output. This is known as thermal compression. It also increases the impedance of a drive unit, and if it is part of a complex passive crossover, the interaction can be such that there are much greater effects on the response than that of thermal compression alone. In one set of tests conducted by Phil Ward [8], the voice coil temperatures of four different loudspeakers showed a maximum of 195°C and a rise in resistance of 176%. That sort of variation has got to cause interaction with almost any sort of passive crossover.
13. It has been proposed that active crossovers can allow the modelling of voice-coil heating by calculations based on signal level, frequency, and known thermal time-constants. Thus the effects of thermal compression (the reduction in output as the voice coil resistance rises with temperature) could be compensated for. It does however imply relatively complex computation that would be better carried out by digital processing rather than in analogue circuitry. There would have to be A to D conversion of the signal and perhaps D to A conversion of to control parameters, even if the actual crossover function was kept in the analogue domain. Controlling the active crossover parameters with analogue switches or VCAs, without compromising signal quality, is going to be hard to do. Modern volume control chips have excellent linearity but they are not really adapted to general control, and using a lot of them would be rather expensive. If you are undertaking this sort of complex stuff then it's probably going to be best to do all the processing in the digital domain.

    Clearly this plan can only work if the crossover is programmed with the thermal parameters for a given loudspeaker and its set of drive units; this information would have to be provided by the loudspeaker manufacturers, and once again we see the need for the active crossover to be matched to the loudspeaker.
14. Drive unit production tolerances can be trimmed out. It has been said that changes in driver characteristics due to aging can also be trimmed out, but since aging is not likely to be an absolutely predictable process, this would ideally require some sort of periodic acoustic testing. For a reference loudspeaker in a laboratory or monitors in a recording studio this is entirely practical, but it is less so in the home environment because of the need for an accurate measuring microphone, or, more likely, one whose response deviations are sufficiently predictable for them to be allowed for. Extra electronics is of course required to implement the testing procedure.
15. The active filter crossover components will have stable values. The inductors in a passive crossover should be stable with time, but the non-polar electrolytic capacitors often used have a bad reputation for shifting value over time. The stability of these

components has improved in recent years but it is still a cause for concern. It has been stated that electrolytics in high-end passive crossovers should be regarded as having a lifetime of no more than 10 years [9]. Plastic film crossover capacitors such as polypropylene show better stability but are very expensive. A fashion has grown up recently for bypassing big passive crossover capacitors with smaller ones; whether this has any beneficial effects is very questionable.

16. The active filter crossover components will not change in the short-term due to internal heating. In a passive crossover the capacitors will have large voltages across them and large currents through them; dielectric losses and ohmic losses in the ESR (equivalent series resistance) may cause these capacitors to heat up with sustained high power and change in value. Non-polar electrolytic capacitors (basically two ordinary electrolytics back-to-back) are considered particularly susceptible to this effect because their relatively small size for a given capacitance-voltage product means they have less area to dissipate heat and so the temperature rise will be greater.
17. The relatively small capacitors used in active crossover filters can be economically chosen to be types that do not exhibit non-linear distortion– polystyrene and polypropylene capacitors have this useful property. Non-polar electrolytic capacitors when used in passive crossovers are known to generate relatively large amounts of distortion.
18. No inductors are required in active crossover circuitry (apart perhaps for a few small ones at inputs and outputs for EMC filtering). Inductors are notorious for being awkward and expensive. If they have ferromagnetic cores they are heavy and generate large amounts of non-linear distortion. If they are air-cored distortion is not a problem, but many more turns of copper wire are needed to get the same inductance, and the result is a bulky and expensive component. Martin Colloms has stated [10] that if an inadequate ferromagnetic core is pushed into saturation by an large transient, the resulting sudden drop in inductance can cause a drastic drop in the impedance seen at the loudspeaker terminals, and this sort of thing does not make life easier for power amplifiers.
19. Passive crossovers typically use a number of inductors, and it may be difficult to mount these so there is no magnetic coupling between them; unwanted coupling is likely to lead to frequency response irregularities. (It should be said that some types of passive crossover use transformers or auto-transformers, where the coupling is of course entirely deliberate)
20. When a passive crossover is designed, it is absolutely not permissible to treat a drive unit as if its impedance was simply that of an 8 Ω resistor. The peaky impedance rise at resonance, and the gentler rise at HF due to the voice coil inductance have to be taken into account to get even halfway acceptable results. This naturally complicates the filter design process considerably, and one way of dealing with this is to attempt to compensate for these impedance variations by placing across the drive unit terminals a

series-resonant LCR circuit (to cancel the resonance peak) and an RC Zobel network (to cancel the voice coil inductance rise) [5]. This means there are at least five more components associated with just one drive unit, and they all have to be big enough to cope with large signals. There may also be changes in impedance due to changes in acoustic loading across the drive unit's passband. In an active crossover system the drive unit is simply connected directly to its power amplifier, and assuming that amplifier has an adequate ability to drive reactive loads, the details of the drive unit impedance curve can be ignored.

21. The presence of an active crossover in the system makes it easy to add bass-end equalisation of the loudspeaker. An equaliser circuit is added which provides a carefully controlled amount of bass boost to counteract the natural roll-off of the bass drive unit in conjunction with its enclosure, thus extending the level bass response to lower frequencies. This sort of thing always has to be approached with care, as too much equalisation will lead to excessive drive unit displacements and permanent damage. This is a particular problem with ported enclosures that put no loading on the drive unit cone at low frequencies.
22. An active crossover system also opens up the possibility of motional feedback, where the position, velocity, etc. of the bass drive unit cone are monitored by an accelerometer or other method, and the information is used to correct frequency response irregularities and non-linear distortion. This is obviously more practical where the crossover and power amplifiers are built into the loudspeaker enclosure. Motional feedback is a big subject and this is not the place to get into it in detail, but it may be remarked in passing that while it sounds like a first-class idea, it has never achieved much success in the marketplace.

That is a pretty formidable list of real advantages for active crossovers. There are also some claimed advantages that are pretty much bogus, and we had better have a quick look at these before we move on:

Some Illusory Advantages of Active Crossovers

1. A commonly quoted "advantage" is that there is less intermodulation distortion in the power amplifiers because each one handles a restricted frequency range. This may once have been a significant issue, but nowadays designing highly linear power amplifiers is not hard when you know what pitfalls to avoid. I have demonstrated this many times [11]. I suppose this might well be a valid advantage if you insist on using valve amplifiers, which have generally poor linearity, at its worst at low frequencies due to the well-known limitations of transformers.
2. Similarly, it has been claimed that the danger of slew-rate limiting is much reduced in the more powerful LF and MF power amplifiers, as they do not handle high frequencies. This is a specious argument, as there is no difficulty whatsoever in

designing a power amplifier that has a faster slew-rate than could ever conceivably be needed for an audio signal [12].

3. It is sometimes stated that an active crossover system results in simpler loading for each power amplifier, as it is connected to only one drive unit. This is true, as regards human understanding of what the impedance curves look like—it is often difficult to see what causes the impedance ups and downs of a complex passive crossover, whereas for a single moving-coil drive unit the effects of the resonance, the voice coil resistance, and the voice coil inductance are all readily identifiable. However, amplifiers are not sentient, and no amount of anthropomorphic thinking will make them so. An impedance curve that is hard to understand, with a cross-section like the Pyrenees, is not necessarily hard to drive. What puts most stress on a power amplifier are impedance dips and high levels of reactance; these increase the peak and average dissipation in the amplifier output devices. It is not a question of the amplifier having to think harder about it, but the amplifier designer may have to.
4. Following from the previous point, an amplifier connected to a single moving-coil drive unit sees basically a mixture of resistance and inductance. Purely capacitative loading is not possible, and it has been said that amplifier stability is therefore easier to obtain. In reality, ensuring amplifier stability into capacitive loads is not a significant problem—you just put the traditional inductor in series with the amplifier—so this is not a big deal. The inductors used have only a few turns and are not expensive. If you are constructing a system where the amplifiers are *permanently* connected to the drive units—for example, if the amplifiers are built into the loudspeaker enclosure—then there is the possibility of omitting the power amplifier output inductors as a single drive unit is essentially resistive and inductive, but I'll say right now that I've never tried this.
5. It has sometimes been claimed that it is easier or cheaper to design a power amplifier that only has to handle a limited bandwidth. This is not true. It is absolutely straightforward to design a good power amplifier that not only covers the whole audio spectrum, but also several octaves on either side of it. Likewise, a limited bandwidth does not reduce costs much. You could make the capacitor in the power amplifier negative-feedback smaller in the MID and HF channels, but that is a matter of a penny or two; if the amplifiers have separate power supplies then those for the MID and HF could have smaller reservoir capacitors, which might save tens of pennies. It is possible to save a significant amount of money by designing the HF power amplifier for a lower power output, but that is a different matter from the bandwidth issue, dealt with in Chapter 14 on crossover system design.
6. If we are dealing with an integrated crossover/power–amplifier/loudspeaker system with everything installed in one enclosure, it is often claimed there is no need to fit short-circuit protection to the amplifiers as their outputs are not externally accessible, and so money can be saved. This is a dubious approach, because it assumes that voice-coils

will never fail short-circuit; if one does there could be serious safety issues. It also assumes that there will be no accidents when testing the amplifiers. The cost saving is very small and is not recommended.

It is still of course very necessary to provide all the power amplifiers with effective DC-offset protection, for safety reasons if for no other. This is a special challenge when driving tweeters, as described below. This is a good deal more expensive than short-circuit protection, as it usually requires an output relay for each power amplifier.

1.7.2 The Disadvantages of Active Crossovers

The disadvantages of active crossovers, once again in what I think to be their order of importance, are:

1. Much greater electronic complexity. The number of power amplifiers is doubled, or more likely tripled, and quite possibly quadrupled. This could potentially lead to lower reliability, as the failure of any one of the power amplifiers or of the active crossover itself renders the whole system unusable. However, with proper design techniques and the use of adequate safety margins, the failure-rate of a modern solid-state power amplifier or active crossover should be so low that reliability is not an issue.
2. As a direct result of the first statement, there is much more hardware to fit into the living room. Not just more electronic boxes, but all the cabling between them. Ways of tackling this issue are considered later in this chapter.
3. Also flowing from (1) is the issue of greater cost. The high-frequency amplifiers can be of lower power output than the low-frequency amplifiers but this definitely does not reduce their cost proportionally. There may be economies of scale to be made by making all the power amplifiers identical, but with the high-frequency ones fed from lower supply voltages. This is perfectly feasible without any compromise on amplifier performance.
   Against this must be set the fact that precise, stable, and generally high-quality components for passive crossovers are not cheap, and a top-end passive crossover can easily end up costing more than an active crossover; this does not however take into account the extra power amplifiers.
4. The vast majority of active crossover loudspeakers have been built as one unit. A mono active crossover, and the two or three power amplifiers required, are put inside into the loudspeaker enclosure. While this makes a very convenient package, which requires only a mains lead to each loudspeaker as extra wiring, it has a deadly disadvantage. The simple truth is that people want to be able to choose their own power amplifiers. Not everybody is prepared to believe that a company that specialises in loudspeakers would be able to come up with a good power amplifier, and is it entirely understandable that people prefer to buy each audio component from specialists in the relevant field.

5. The active crossover must be matched to the loudspeaker and cannot be bought off the shelf. Many speakers allow for bi-wiring but this still leaves the passive crossover components in place. Few if any loudspeakers allow direct access to the drive units without some serious dismantling, and so any domestic system must be either home-built, or custom-built with a commensurately high cost.
6. Tweeters, and to a lesser extent, mid-range drive units are much more exposed to amplifier DC-offset faults. When they are connected to an amplifier via a passive crossover, there will be at least one capacitor between the amplifier and the voice coil. This will be of modest size as it is not intended to pass low frequencies, and so provides very good protection for tweeters and mid-range drive units which do not have the displacement capability or thermal inertia of bass drivers. Tweeters in particular can be destroyed in an instant by a large DC offset. This seems to me a very good, if not irrebuttable, argument for using a low-power amplifier to drive the tweeter in an active crossover system.

   Any DC-coupled amplifier connected to a loudspeaker should of course have DC-offset protection, but since such a fault is usually detected by putting the output signal through a lowpass filter with a very low cutoff frequency to remove all the audio, and then applying it to a comparator, there is inevitably some delay in its operation; an output muting relay also takes a significant amount of time to open its contacts and break the circuit. If the amplifier driving the tweeter is specialised for the task then it can be given DC protection that reacts more quickly, as its lowpass filter will not need to reject high amplitude bass signals; it can therefore have a higher cutoff frequency and will respond faster.
7. There will be a greater power consumption due to quiescent current flowing in more amplifiers. In a solid-state Class-B power amplifier, how much power is involved depends on the details of the output stage design. Class-A amplifiers, built with any technology, uselessly dissipate almost all the power so trustingly fed into them, showing efficiencies of around 1% with musical signals, as opposed to sinewave testing [13]. There is also the power consumption of the active crossover itself, but this is not likely to be significant, even for a very complex analogue active crossover.
8. Passive crossovers cannot be radically maladjusted by the user (though some allow vernier adjustments, typically of HF output level). In the sound-reinforcement business there is always the possibility that some passing “expert” may decide to try and use the crossover controls as a sort of equaliser. This is unlikely to be effective and can result in severe damage to loudspeakers. Lockable security covers solve this problem.
9. Active crossovers add extra electronics to the signal path which would otherwise not be there, and so must be beyond reproach, or at any rate not cause any significant signal degradation. With the intelligent use of either opamp or discrete circuitry this should not be a serious problem.

1.8 The Next Step in Hi-Fi

As we have just seen, active crossovers have a long and convincing list of technical advantages. The score is 22 very real advantages and 5 not-too-convincing advantages, as opposed to 9 advantages for passive crossovers. It is generally accepted that active crossover hi-fi systems sound obviously better than their passive-crossover counterparts. Any sort of consensus is rare in the wide field of audio, so this is highly significant. I strongly suspect that the widespread adoption of active crossovers, suitably matched to their loudspeakers, would be The Next Big Step in Hi-Fi, and possibly even The Last Big Step possible with current technology.

Nonetheless, it is undeniable that active crossovers, despite their compelling advantages, have made very little headway in the domestic market so far, though they are used in all but the smallest sound-reinforcement systems, and extensively in automotive audio. The big question is how to make active crossover technology more acceptable in the marketplace. The first thing we shall do is look at ways of solving the "too many boxes and wires" problem.

1.9 Active Crossover Systems

We will consider the various ways in which an active crossover system can be configured, with an especially hard look at making it acceptable in a domestic environment. Sound-reinforcement systems are a separate issue. One of the significant disincentives to the active crossover approach is the sheer amount of hardware required, in terms of electronic boxes and cables. This can be hard to fit into a minimalist decoration scheme- and indeed often hard to fit into any sort of decoration scheme at all. It is desirable to package the technical functionality as neatly as possible. I am assuming here that a high-quality system is intended, and so three-way active crossovers will be used.

In terms of cabling and equipment the tidiest set-up is undoubtedly achieved with a mono active crossover and its three power amplifiers built into each loudspeaker enclosure, but it is vital to realise that this is not an acceptable approach for people who take their power amplifiers at all seriously.

Figure 1.2 shows an active crossover system using six separate monobloc power amplifiers. This could be configured with all six amplifiers in one location, in which case putting them all near one of the loudspeakers reduces the length of at least one set of loudspeaker cables to a minimum. The amplifiers may be all of the same type, but all that is really required of them is that they have the same gain. Even if we have six nominally identical amplifiers made at the same time by the same manufacturer, somewhere in each amplifier will be at

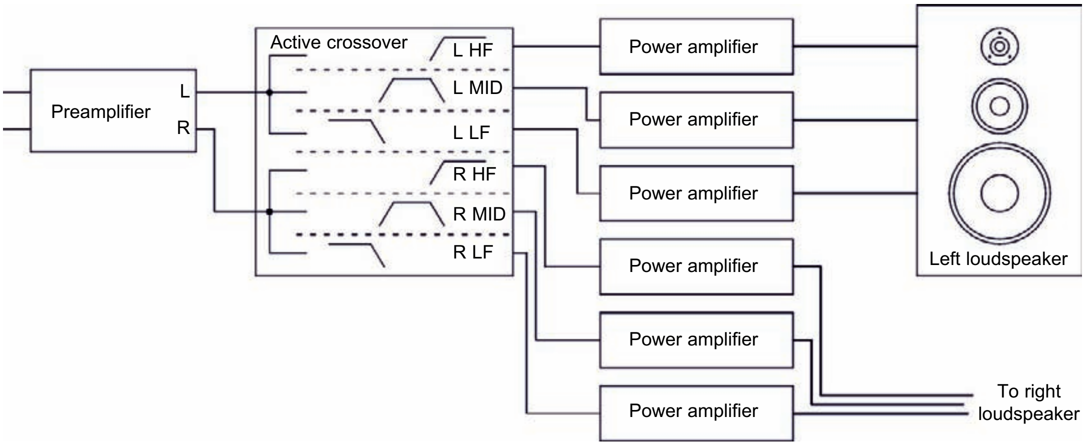


Figure 1.2: Active crossover system using six monobloc power amplifiers.

least two gain-setting resistors, each with a tolerance; nonetheless, in a competent design the variation should be comfortably less than variations in the drive units.

Alternatively, three of the power amplifiers could be placed adjacent to each loudspeaker, considerably shortening the total length of what might be expensive loudspeaker cable, as three of the connections are now very short. Three of the line level cables to the power amplifiers naturally become correspondingly longer, but since their resistance is of much less importance than that of the loudspeaker cables, this is overall a good thing. The increased resistance of long line cables makes the link more susceptible to voltages induced by currents flowing through the ground connection, but I think it is fair to assume that a hi-fi system with active crossovers would use balanced connections to cancel such noise. Ultimately the placing of the various parts of the system is going to be influenced by furniture arrangement and the availability of handy (and hopefully well-ventilated) cupboards in which to stash the boxes.

Adding it up, a system configured in this way consists of 7 electronic boxes (not including the preamplifier), line-level cables, and 6 loudspeaker cables. If the connection from the preamp to the crossover is a two-way cable (i.e., two parallel cables joined together along their length), that is reduced to 7 boxes, 7 line cables, and 6 loudspeaker cables, which is hardly a great improvement.

However, it must be said that there are excellent technical reasons for using two-way cables when you can. Their construction keeps the grounds for the two links physically close together, and prevents them forming a loop that could pick up magnetic fields that would induce current flow; this current would cause voltage drops in the ground resistance and degrade the signal. The use of balanced connections greatly reduces the effects of ground currents but it is of course much sounder to prevent the currents arising in the first place.

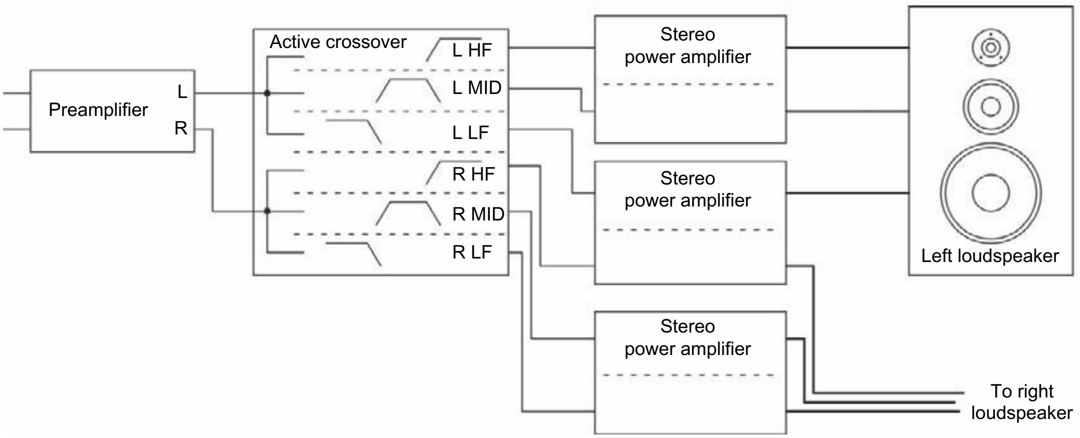


Figure 1.3: Active crossover system using three stereo power amplifiers.

In Figure 1.3, the system is configured with three stereo power amplifiers. This has the advantage that stereo amplifiers are the most common sort and give the greatest choice. The ones used here are assumed to be identical; if a lower-power stereo amplifier is chosen to drive the tweeters then Figure 1.3 would need to have its connections rearranged. Figure 1.3 uses 5 electronic boxes, 8 line cables, and 6 loudspeaker cables. Using two-way cables between the preamplifier and crossover and also to the power amplifiers simplifies this to 5 boxes, 4 line cables, and 6 loudspeaker cables. There are always going to be 6 loudspeaker cables.

Figure 1.4 shows a variation on this approach which puts one of the stereo amplifiers adjacent to the right loudspeaker, instead of piling them all up on the left side. This cuts down the total length of loudspeaker cable required, but there is still a long run from one amplifier on the left to the right loudspeaker on the other side of the room.

The opinion is held in some quarters that very high degrees of isolation between left and right channels is essential to obtain an optimal stereo image. This is wholly untrue, but audio is not a field in which rational argument can be relied upon to convince everybody. The configuration of Figure 1.4 could be criticised on the grounds that left and right channels pass through one stereo power amplifier and this might compromise the crosstalk figures. It is not actually harder to get a good crosstalk performance from a stereo power amplifier than from a stereo preamplifier; in fact it is usually easier because the preamplifier has more complex signal routing for source selection and the like. Nonetheless it is only fair to point out that there might be objections to the 3× stereo amplifier arrangement because however it is configured, at least one amplifier will have to handle both right and left signals.

We will now take a radical step and assume the ready availability of three-channel power amplifiers.

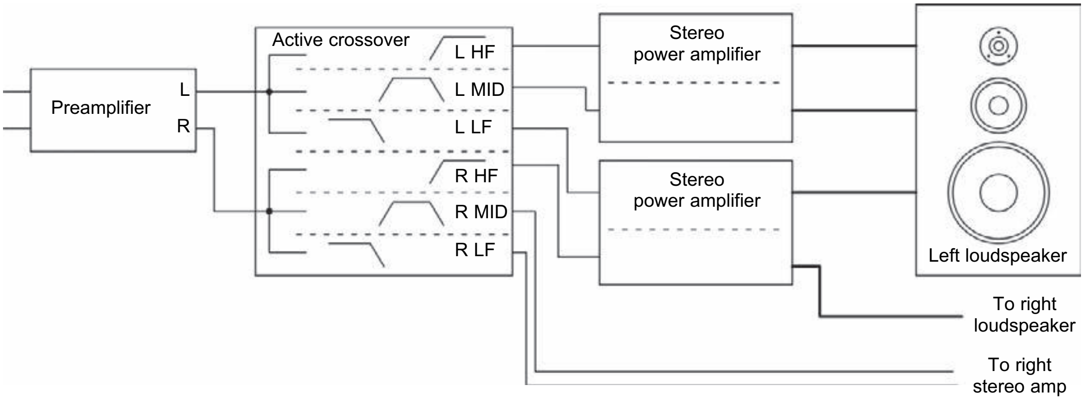


Figure 1.4: Alternative setup of three stereo power amplifiers, with one placed on right speaker side.

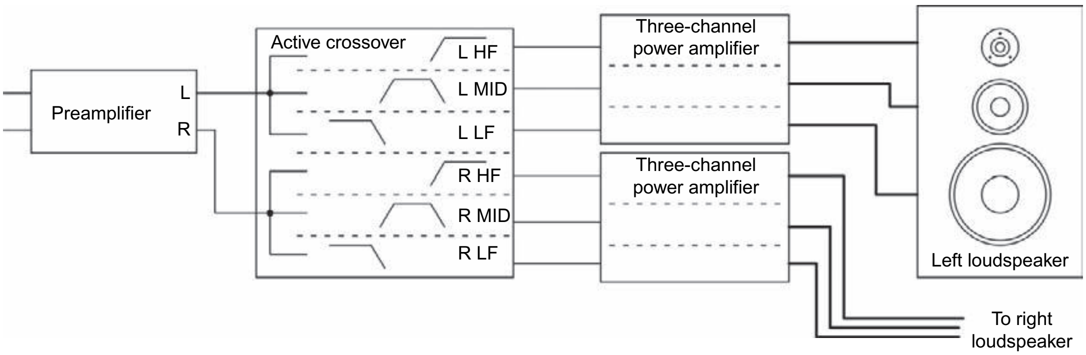


Figure 1.5: Active crossover system using two three-channel power amplifiers.

Multi-channel power amplifiers at a reasonable cost have been available for surround-sound systems for many years, to deal with 5:1 formats and so on. The last multi-channel power amplifier I designed (the TAG 100 × 5R:10) could be configured for ten channels of 80 W/8 Ω each. A three-channel power amplifier of high quality presents absolutely no new technical challenges at all.

As you can see from Figure 1.5, using two three-channel power amplifiers simplifies things considerably. There are now 4 electronic boxes, 8 line cables, and 6 loudspeaker cables. Using a two-way cable from preamp to crossover and three-way cables between the crossover and amplifiers is well worthwhile and reduces the parts count to 4 boxes, 3 line cables, and 6 loudspeaker cables. This important configuration is shown in Figure 1.6. One of the three-channel power amplifiers could be sited over by the right loudspeaker, and this is much to be preferred as it minimises total loudspeaker cable length. I think it is fairly

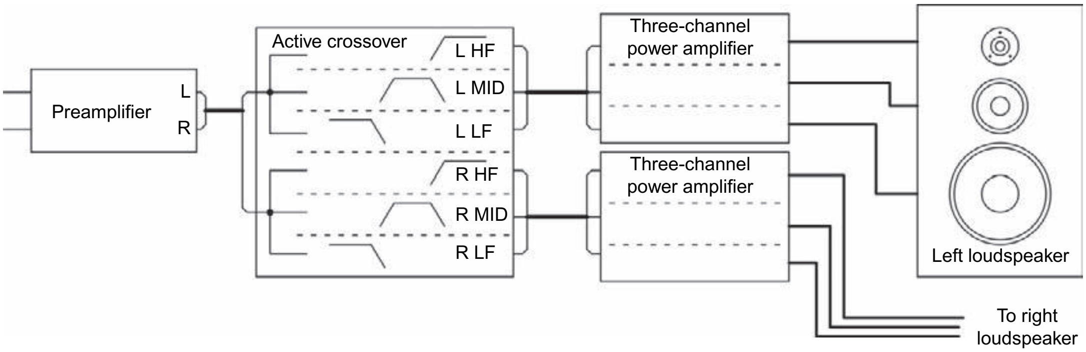


Figure 1.6: Active crossover system using two three-channel power amplifiers and multi-way cables.

clear that this is the best way to configure things, with the least number of separate parts and the possibility of keeping the loudspeaker cables very short indeed if each power amplifier is sited right behind its loudspeaker.

The only real difficulty is those three-channel power amplifiers. Some do exist, evidently intended for multi-channel AV use rather than in active crossover systems; two current examples are the Classé CA-3200 three-channel power amplifier [14] and the Teac A-L700P three channel Amplifier. There appear, however to be no three-channel amplifiers specifically designed for our application here. Such an amplifier would be able to economise on its total power output by having a big output for the LF driver, a medium output for the mid drive unit, and a smaller output again for the tweeter. The downside to that plan is that it would be less versatile than a three-channel amplifier with equal outputs, which could be pressed into stereo or multi-channel service if required.

Three-way cables should present no problems; in the UK, the widespread Maplin chain sells four-way audio line cables with individually lap-screened cores at a very reasonable price. Individual screened cores are of course highly desirable to prevent capacitive crosstalk.

To take things to their logical conclusion, we can use a six-channel power amplifier. These are used extensively in AV applications so are not hard to obtain. If we assume the use of multi-way cables from the start, we get Figure 1.7. This cuts down the separate pieces of hardware to the minimum, giving 3 electronic boxes, 2 line cables, and the usual 6 loudspeaker cables. Unfortunately this configuration brings back the need for long and possibly expensive loudspeaker cables to drive the right loudspeaker, and it also raises again the question of left-right crosstalk in the power amplifier.

In my view, the configuration using two three-channel amplifiers is clearly the best approach. There is one more box but the loudspeaker cables can be of minimal length.

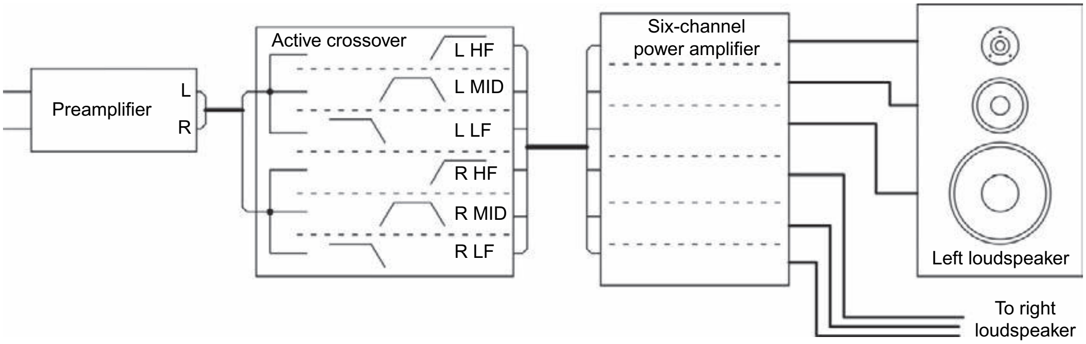


Figure 1.7: Active crossover system using a single six-channel power amplifier and multi-way cables.

1.10 Matching Crossovers and Loudspeakers

Having looked at the best way to package and configure an active crossover system, there are some other issues to deal with. In the sound reinforcement business, active crossovers are sold with wide-ranging control over crossover frequencies, time delay compensation, and so on. This is feasible in a situation where skilled operators set up the system. In a hi-fi situation it is more common to have fixed crossover parameters that are carefully matched to the loudspeaker characteristics. These parameters include not only crossover frequencies and delays, but also things like drive unit frequency response equalisation, diffraction compensation equalisation, and so on. The crossover has to be set up for one model of loudspeaker only. This is fine if the crossover and amplifiers are built into the loudspeaker, but as we have seen, this is not going to work for potential customers who want to choose their own power amplifiers, and that is most of them.

The solution to this problem is to sell active crossovers and loudspeakers as a matched package, but leave the power amplifiers to be bought separately. All that would be required of the amplifiers is that they should all have the same gain. For the higher reaches of the high end, each active crossover unit could be individually calibrated by acoustic testing to match each loudspeaker with its drive units at the factory, authoritatively solving the problem of variation in drive unit parameters. The loudspeakers and crossover would then be tied together by their serial numbers. Obviously it would be essential to not swap over the Left and Right channels at any point. The calibration could be done by plugging in resistors until the optimal result is obtained, and then soldering them in permanently. This kind of hand tuning would not of course be cheap, but with the right measuring gear and an intelligent algorithm for adjusting the various crossover parameters, I am sure it could be done at a price some folk would be prepared to pay.

What it would not do is compensate for drift in drive unit parameters over time. This could only be addressed by regular re-calibration, which would be unpopular with most people if it involved sending the crossover and loudspeakers back to the factory—and handing over some more money, no doubt.

For this scheme it would be necessary to construct loudspeakers so that direct access could be gained to the drive units, optionally by passing the internal passive crossover, if one is fitted at all. Providing facilities for bi-wiring on the back of a loudspeaker is straightforward; you just put four terminals on the back and add a pair of shorting links that can be easily removed. Tri-wiring is a little more difficult but still thoroughly doable. Bypassing a passive crossover is however more complex, as it is necessary to disconnect it not only from the input terminals but also from two, three, or possibly four drive units.

1.11 A Modest Proposal: Popularising Active Crossovers

Let us suppose that it becomes accepted practice to sell crossovers and speakers as a package, but the amplifiers are bought separately. As we have just seen the best and neatest way to configure the complexities of a 3-way active crossover system is to use three- or six-channel power amplifiers, and one might hope that a market for these would develop. One wonders if there might be some sort of psychological resistance to buying two parts of the audio chain with a gap (the power amplifier) between them. We can, I think, confidently predict that some enthusiasts would favour different makes of power amplifier for the LF, MID, and HF channels. There would be much room for entertaining debate if you like that sort of thing. We might even speculate that third-party active crossovers might be markted to replace those from the speaker manufacturer.

We also saw that it is highly desirable to use multiway cables between the active crossover and the power amplifiers, for tidiness as well as the best technical performance. A market could emerge for 3-way or 6-way leads. Ideally each audio line should be individually screened to prevent capacitive crosstalk between them, especially on long cable runs, but this puts the cost of the cable up substantially. Conventional line outputs have a series resistor in the output to ensure stability when faced with cable capacitance from signal to ground, resulting in a typical output impedance of 50–100 Ω. This is enough to allow the capacitance between adjacent and unshielded signal conductors to significantly degrade the crosstalk performance. A most effective solution to this problem is the use of so-called "Zero-impedance" line outputs on the crossover. This technology typically achieves an output impedance of a fraction of an Ohm, and can reduce crosstalk by 40 dB or more; it is equally stable into long cable runs, and the extra cost is trivial. Zero-impedance outputs are fully described in Chapter 17 on line output stages.

An unbalanced 6-way interconnection uses 6 signal feeds and at least one ground wire, and so requires at least 7 pins in the connector used.

Since an active crossover system is expected to reach high levels of quality, the use of balanced interconnections needs examination. The hot and cold signals from each channel can be wrapped by a single grounded screen, but ideally there would be separate assemblies of this sort for each channel, which makes for rather non-standard cable. Conventional balanced operation naturally increases the number of contacts on the connectors used to at least $(6 \times 2) + 1 = 13$, though multiple ground connections would be good, not only to reduce voltage drops due to ground currents, but also to avoid alarming the superstitious end of the audio market.

This assumes that the outputs are truly balanced, in that there are two outputs in anti-phase. While this gives a handy 6 dB increase in level over the link, the potential 6 dB improvement in signal-to-noise ratio is likely to be rendered irrelevant by the noise from the electronics. If we could drop the requirement for anti-phase outputs then we could have a quasi-balanced cable with 6 signal feeds, one cold line to sense the source-end ground, and one ground wire. The single cold line would be distributed to the cold inputs of six balanced input amplifiers at the receiving (power amplifier) end. The danger is, of course, that people would condemn it as " not a real balanced connection" though it would almost certainly give an equally good performance as regards common-mode rejection. This plan would require a minimum of 8 pins in the connector.

1.12 Multi-Way Connectors

A multi-way cable requires multi-way connectors. It would be nice if active crossovers became so popular that a committee designed a special connector for us (like the HDMI connector, for example), but realistically that isn't likely to happen soon. It is therefore worth looking at what existing connectors are capable of meeting our needs. What we must strive to avoid is a connector configuration already in common use, because in the real world people are prone to plug stuff into any socket that will physically accept it.

XLR connectors have the benefit of being fairly familiar, but they only go up to 7 pins. This is fine for 3-way balanced use, but only allows unbalanced 6-way operation (six audio feeds plus a common ground). Since we are dealing with higher-end systems, it seems inappropriate to rule out balanced operation.

The familiar standard DIN connectors have a 13.2 mm diameter metal body and go up to 8 pins, so we can have fully balanced 3-way operation, or 6-way quasi-balanced operation, but not fully balanced 6-way usage. These connectors have some unfortunate associations with the low-quality DINs in the past, but today reliable high-quality versions are freely available and are used for MIDI links and stage lighting control. They are not popular with those who have to solder cables into them, but then this is probably true of any small multi-way connector.

The smaller Mini-DIN connectors are 9.5 mm in diameter and officially come in seven patterns, with the number of pins from 3 to 9, though there are at least two non-standard 10-pin versions which are not approved by the Deutsches Institut für Normung, (DIN) the German standards body. Once more we can have fully balanced 3-way operation, or 6-way quasi-balanced operation, but not a fully balanced 6-way mode. Mini-DINs are sometimes called "Video camera connectors" though they have in fact been used for a very wide variety of uses, including computer power supply connectors, so there is an element of risk there.

The 9-way D-type connector is inexpensive, and offers screw retention, but it is hard to argue that it has anything of a high-end audio air about it. It is also likely to get plugged into the wrong place, whereas standard-format connectors like XLRs or DINs with an unusually large number of pins are much safer. Once more we can have fully balanced 3-way operation, or 6-way quasi-balanced operation, but not a fully balanced 6-way mode.

There are various proprietary connector systems; for example the Neutricon by Neutrik goes up to 8-pin. It is stocked by Farnell so is easy to source.

The 15-way D-type connector has the same problem of lack of glamour as the 9-way, but it is the only easily sourced connector which will allow fully balanced 6-way operation. The two spare ways can be usefully pressed into service as extra ground connections.

1.13 Subjectivism

As I warned you in the preface, this book has no truck with faith-based audio. There is no discussion of oxygen-free copper, signal cables that only work one way, magic capacitors hand-rolled on the thighs of Burmese virgins under a full moon, or loudspeaker cables that cost more than a decent car. Valve technology is ignored because it is inefficient and obsolete, and despite much ill-informed special pleading, it has absolutely no magical redeeming features. You will find here no gas-fired pentodes, nor superheated triodes fed with the best Welsh steam coal and still bespattered with the mud of the Somme. It would of course be possible to design a complex active crossover using valves, but even if you accepted mediocre performance as regards noise and distortion, the result would be very expensive, very hot, and very heavy.

You will, I think, find enough real intriguing intellectual challenges in crossover design to make it unnecessary to seek out non existent ones.

References

[1] http://www.talkbass.com/wiki/index.php/Driver_diameter_&_pistonic_beaming_frequencies (accessed 12.10).
[2] D. Self, Ultra-low-noise amplifiers & granularity distortion, JAES 35 (1987) 907–915.
[3] D. Self, Audio Power Amplifier Handbook, fifth ed., Newnes, Boston, MA, p. 235 (cable effects), ISBN: 978-0-240-52162-6.
[4] M. Colloms, High Performance Loudspeakers, third ed., Pentech Press, UK, 1985, p. 174, ISBN: 0-7273-0806-8.
[5] M. Colloms, High Performance Loudspeakers, third ed., Pentech Press, UK, 1985, p. 178, ISBN: 0-7273-0806-8.
[6] J. Borwick (Ed.), Loudspeaker & Headphone Handbook, second ed., Focal Press, Boston, MA, 1994, p. 234, ISBN: 0-240-51371-1.
[7] J.-C. Gaertner, Project 21 Part 1, Linear Audio, vol. 0, pub J Didden, 2010, p. 132, ISBN: 9-789490-919015.
[8] P. Ward, http://www.soundonsound.com/sos/jul02/articles/monitors2.asp (thermal compression) (accessed 12.10).
[9] P. Newell, Recording Studio Design, second ed., Focal Press, Boston, MA, p. 511, ISBN: 978-0-240-52086-5.
[10] M. Colloms, High Performance Loudspeakers, third ed., Pentech Press, UK, 1985, ISBN: 0-7273-0806-8.
[11] D. Self, Audio Power Amplifier Handbook, fifth ed., Newnes, Boston, MA, p. 190 (power amp distortion), ISBN: 978-0-240-52162-6.
[12] D. Self, Audio Power Amplifier Handbook, fifth ed., Newnes, Boston, MA, p. 255 (slew-rate limiting), ISBN: 978-0-240-52162-6.
[13] D. Self, Self On Audio (Audio Power Analysis), second ed., Newnes, Boston, MA, p. 458 (Class-A efficiency), ISBN: 0-7506-8166-7.
[14] K. Rubinson, http://www.stereophile.com/solidpoweramps/207classe/ (review) (accessed 12.10).

CHAPTER 2

How Loudspeakers Work

This chapter is not in any way a guide to how to design loudspeakers, nor is it a complete explanation of how they work; what it does do is look at their functioning from the point of view of the crossover designer. The specific topic of time-aligning the drive units is dealt with in Chapter 10 on delay compensation.

There are many ways to make a loudspeaker. Some of the less common types are electrostatics, ribbon loudspeakers, electromagnetic planar loudspeakers, Heil air-motion transformers, and the rather worrying Ionophone, but here I must concentrate on the most popular concepts: sealed boxes, ported reflex boxes, ABR systems, transmission line loudspeakers, and horns.

2.1 Sealed-Box Loudspeakers

A sealed enclosure, as shown in Figure 2.1a, is the simplest. It gives a good transient response, assuming a correct choice of Q, good low-frequency power handling because the drive unit cone is always loaded, and has less sensitivity to misaligned parameters than other types. On the other hand, sealed enclosures have higher low-frequency cutoff points and lower sensitivity than the other loudspeaker types for the same box volume.

A loudspeaker in a sealed box has two important factors working on it—the mass of the moving parts of the driver and the compliance ("springiness" of the air in the box). These two factors make up a second-order system, like a weight bouncing on a spring, and the effect is that the loudspeaker response is that of a second-order lowpass filter with a 12 dB/octave rolloff as frequency falls. Driver cone excursion increases at 12 dB/octave with decreasing frequency for constant SPL, so cone excursion tends to be constant below the cutoff frequency.

Such filters can have different values of Q; the higher the Q the more peaked the response (as shown in Figure 2.2). The Q is an important choice in the design, and is often referred to as the "alignment" by reference to a filter type. Thus a Butterworth alignment uses a Q of 0.707 ($1/\sqrt{2}$) and gives a maximally flat response, just like a second-order Butterworth filter. This is not necessarily the ideal Q; higher Qs give significantly more LF output below the cutoff frequency, but at the expense of response peaking, and too high a Q gives a boom-box or "one-note bass" effect, which is not usually what is wanted. A Q of 0.5 gives a good

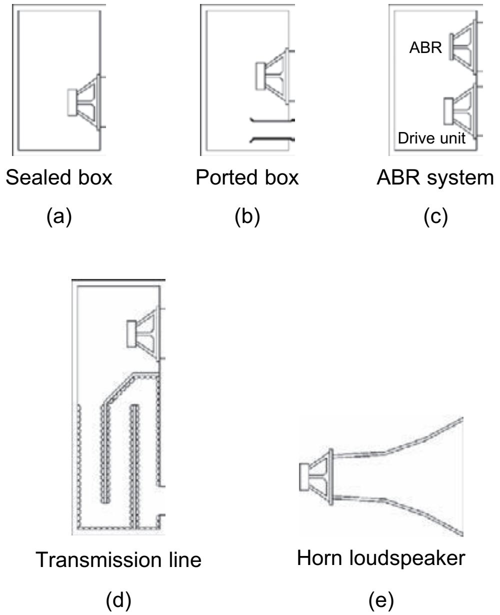


Figure 2.1: The basic loudspeaker arrangements: sealed box, ported reflex, ABR, transmission line, and horn.

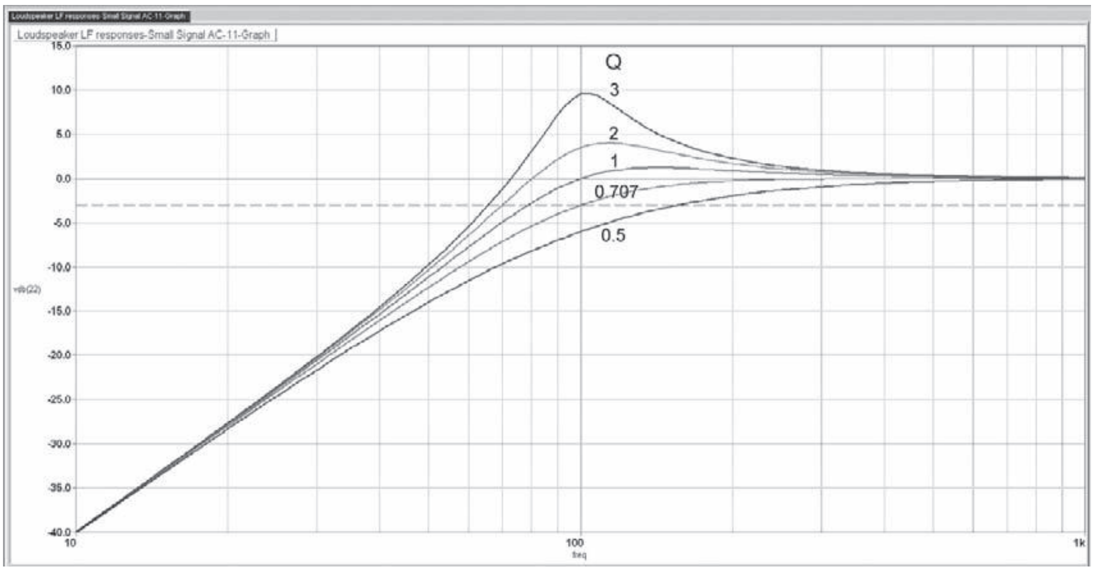


Figure 2.2: Sealed-box loudspeaker LF responses with a cutoff frequency of 100 Hz and various Qs.

transient response ($Q = 0.707$ gives overshoot on square waves) but is often criticised for being "too taut" or over-damped, and the LF output clearly suffers.

For a given drive unit, the system Q is determined by the box volume. The bigger the box, the lower the Q; the responses in Figure 2.2 correspond to a range of box volumes of about 50 times.

The box may be left completely empty, or it may be lined or wholly filled with acoustic damping material, such as bonded acetate fibre (BAF) or long fibre wool. This not only absorbs internal reflections, but for the "fill-'er-up" approach, alters the thermodynamic properties of the enclosed air so that the enclosure behaves as though slightly larger than its physical size.

Sealed box loudspeakers are normally classified into two kinds: the infinite baffle type and the air suspension type. If you have a large enclosure where the compliance of the air inside is greater than the compliance of the drive unit suspension, so that most of the restoring force comes from the latter, it is regarded as an infinite baffle type. A small enclosure where the compliance of the air inside is less than the compliance of the drive unit suspension by a factor of three or more, is regarded as an air suspension type.

The efficiency of a drive unit in a sealed box is proportional to the cube of the drive unit resonance frequency (the Thiele–Small parameter F_s) [1] so attempts to get a lot of low-frequency extension with a given drive unit and a small box volume inevitably result in low efficiency. Nowadays amplifier power is relatively cheap, but there may be problems with large cone excursions and voice-coil heating. Other sorts of loudspeaker enclosures that reinforce the LF output with the rear cone radiation are often used to increase the efficiency, as we shall now see.

2.2 Reflex (Ported) Loudspeakers

The alternative loudspeaker types all, in different ways, return low-frequency energy from the rear of the drive unit in the correct phase to reinforce that radiated from the front.

Reflex or ported loudspeakers, (sometimes called vented boxes) as shown in Figure 2.1b, have a pipe-like port which allows the passage of air in and out of the box. The mass of air in the port resonates with the compliance of the air in the box, causing the port output to be in phase with the forward radiation of the drive unit at low frequencies, and allows the bass response to be extended without the lower efficiency that results when that is done with a sealed box. At frequencies below the port resonance the port output is in anti-phase, causing the output to fall at 24 dB/octave rather than 12 dB/octave as for the sealed box. The response looks like Figure 2.2, but with steeper rolloffs. This does not result in a notch in the response, as occurs in the ABR system, but the increased slope makes things more difficult if you are planning low-frequency response extension by equalisation in the crossover, as the matching of equaliser and driver/box responses is more critical.

While a ported loudspeaker can have lower drive unit distortion, greater power handling, and a lower cutoff frequency than a sealed box system using the same drive unit, there are snags. Below the cutoff frequency the drive unit quickly becomes unloaded and its excursion increases dramatically, worsening non-linear distortions and threatening mechanical damage. To guard against this a subsonic filter should always be used when the source is vinyl. It has been persuasively argued that the arrival of CDs, with extended bass but no subsonic disturbances, caused a major increase in the popularity of ported loudspeakers. The transient response of a ported loudspeaker is considered to be not so good as for a sealed box system with the same drive unit. Ported box systems are significantly more sensitive to misaligned parameters than are sealed box systems, and require more careful design.

At low frequencies large amounts of air are moving through the port; to minimise chuffing noises from turbulent air flow it should have the largest diameter possible consistent with the resonance required and be flared at both ends.

2.3 Auxiliary Bass Radiator (ABR) Loudspeakers

Auxiliary bass radiator loudspeakers (also called passive radiator loudspeakers) work in a similar way to ported loudspeakers. The mass of air in the port is replaced by the mass of the auxiliary bass radiator, which is essentially an LF drive unit with the usual cone and suspension but no voice coil or magnet, as shown in Figure 2.1c. The response of an auxiliary bass radiator loudspeaker is therefore very similar to that of a ported loudspeaker using the same driver. However, the LF cutoff frequency will be somewhat higher, and the rolloff slope will be steeper, due to the presence of a notch in the frequency response which corresponds to the free air resonant frequency of the ABR. This causes a steeper rolloff below the system's tuned frequency Fb, (the resonant frequency of LF driver and box together) and a poorer transient response. Normally the notch is well below the loudspeaker passband (say below 20 Hz), where the response has already dropped considerably; so despite the rapid phase changes associated with the notch, it is normally considered to be of little or no audible significance; it is not practical to equalise it away. The larger the ABR, the more mass its cone will have, and the lower its resonance frequency will be for the same target Fb, pushing the notch further away from the bottom of the loudspeaker passband. There is much more flexibility in design as the ABR cone mass can be altered. The ABR unit may be mounted above or below the main LF drive unit. Two ABR units are sometimes combined with a single active LF driver unit; if so, they are mounted on either side so that the inertial forces cancel out and put less excitation into the front baffle.

ABRs eliminate port turbulence noises and port resonances and give good LF extension for a given box volume. They are however more complex to design than ported types, and are more costly due to the extra auxiliary bass radiator.

2.4 Transmission Line Loudspeakers

A true transmission line or "acoustic labyrinth" loudspeaker sends the energy from the rear of the drive unit down an infinite pipe, where it is gradually absorbed and never heard of again. Since infinite pipes rarely fit in well with domestic surroundings, practical "transmission line" loudspeakers are a compromise. The rear radiation passes down a length of duct chosen so that it is in the correct phase to reinforce the forward radiation at low frequencies. To get worthwhile reinforcement the duct needs to be a quarter of a wavelength long at the frequency of interest, so it shifts the phase of the rear output by 90°. At 40 Hz the wavelength of sound is 8.58 metres, and a quarter of that is 2.145 metres, which is an impractically long duct unless it is folded at least twice, as shown in Figure 2.1d, where the bobbles represent absorbent material lining the duct. Transmission line loudspeakers are often described as non-resonant, but without the quarter-wave resonance (which could be suppressed by using enough absorbent stuffing in the duct) there will be no enhancement of the bass response.

The low-frequency reinforcement works in the same way as a ported box, and the response falls at 24 dB/octave, once again making low-frequency response extension by equalisation in the crossover more difficult. Transmission line loudspeakers are not used in sound reinforcement because the duct greatly increases the volume and weight of the box, without offering any advantages over a ported system; they have the same disadvantage in the domestic environment. Compared with sealed-box, ported, and ABR designs, they are not popular.

2.5 Horn Loudspeakers

An acoustic horn is used to improve the coupling efficiency between a drive unit and the air. It acts as an acoustic transformer that provides mechanical impedance matching, converting large pressure variations over a small area into low pressure variations over a large area, giving a greater acoustic output from a given driver. Horns have a low-frequency limit set by the flare rate and the mouth size; the slower the flare rate, the lower the frequencies a horn can reproduce for a given length. A horn with an area flare rate of 30% per foot is considered to be effective down to about 30 Hz. The flare may have an exponential or tractrix flare; conical flares, as used on old phonographs, have a poorer low frequency response. The flare in Figure 2.1e is illustrative only.

For these reasons horns with an extended low-frequency response are physically large, and only practical if folded in some way. This is not quite as clever as it appears because bouncing the sound round corners can cause frequency response anomalies at the upper end of the working range, due to reflections and resonances. Horn loudspeakers are widely used in sound reinforcement because of their high efficiency.

Constant directivity horns are a relatively recent development where an initial exponential section is combined with a final conical flare, dispersing the shorter high-frequency wavelengths more effectively. They require special equalisation, and this is covered in Chapter 11.

2.6 Diffraction

Diffraction affects loudspeaker operation for the simple reason that high frequencies have shorter wavelengths than long frequencies. At high frequencies, a loudspeaker cone radiates sound mostly in a forward direction, into what is called "half-space"; in other words a hemisphere facing forwards. At low frequencies, the longer wavelengths of the sound bend, or diffract, around the sides and rear of the enclosure, so radiation occurs into "full-space," which is a sphere with the loudspeaker at its centre. In other words the loudspeaker becomes omnidirectional. The difference between these two modes appears as a lower output on the forward axis at low frequencies. In the quite unrealistic conditions of an anechoic chamber the low frequency loss is 6 dB, halving the SPL. In the special case where the loudspeaker enclosure is spherical, the transition between the two modes is smooth and can be approximated by a first-order shelving response, as shown in Figure 2.3, which is derived from the classic paper on the subject by Olson in 1969 [2].

This effect is sometimes called the "6 dB baffle step" or the "diffraction loss" of the enclosure ("step" is actually not a very good description—it is actually a very gentle rise). The frequency at which this occurs is proportional to the enclosure dimensions, and the effect can be compensated for by simple shelving equalisation.

While spherical loudspeaker enclosures are perfectly practical (a notable example being the Cabasse La Sphère [3], which claims to be "a true 4-way co-axial point source". they are

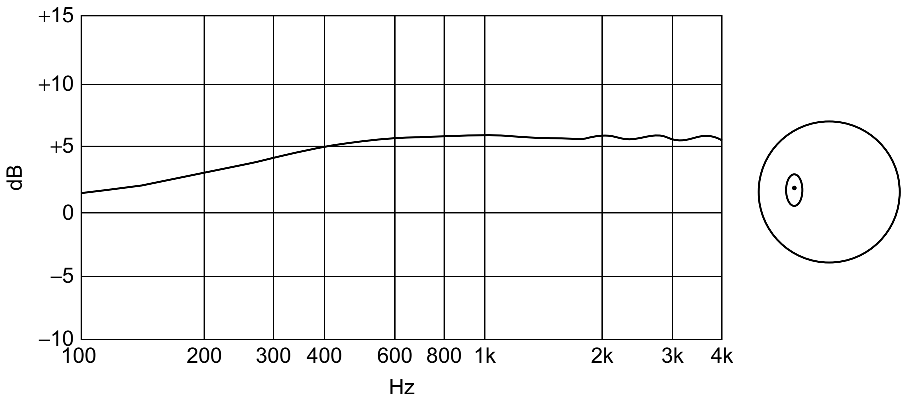


Figure 2.3: HF rise of 6 dB due to diffraction around a spherical loudspeaker enclosure 24 inches in diameter (after Olson, 1969).

never likely to be very popular. Apart from anything else, they need special stands to stop them rolling around the room. There are also problems because the internal reflection paths are all the same length. (I once knew a chap who built a pair of spherical loudspeakers out of GRP, using full-range drive units. He was unable to resist the temptation to paint them like a giant pair of eye-balls, which did not create a restful effect.)

For any shape other than spherical, the diffraction business becomes more complicated. Any other shape has some sort of edge or edges, and when the sound waves from the drive unit strike them, a negative pressure wave is formed which radiates forward and sums with the direct output from the drive unit. This creates an interference effect that adds wobbles to the basic frequency response with its 6 dB rise; these wobbles can be as large as the step itself. This edge effect is illustrated in Figure 2.4.

About the worst thing you can do when choosing a loudspeaker enclosure is to mount it in the end of a cylinder, as in Figure 2.5. This has an equal distance from the drive unit to every part of the edge at the front, and so produces the most accentuated response deviations, exceeding 10 dB. You will note that the horizontal distance between peaks or between dips on the graph decreases as the frequency increases; this is simply because the graph has a logarithmic frequency scale but the diffraction effect, being based on multiples

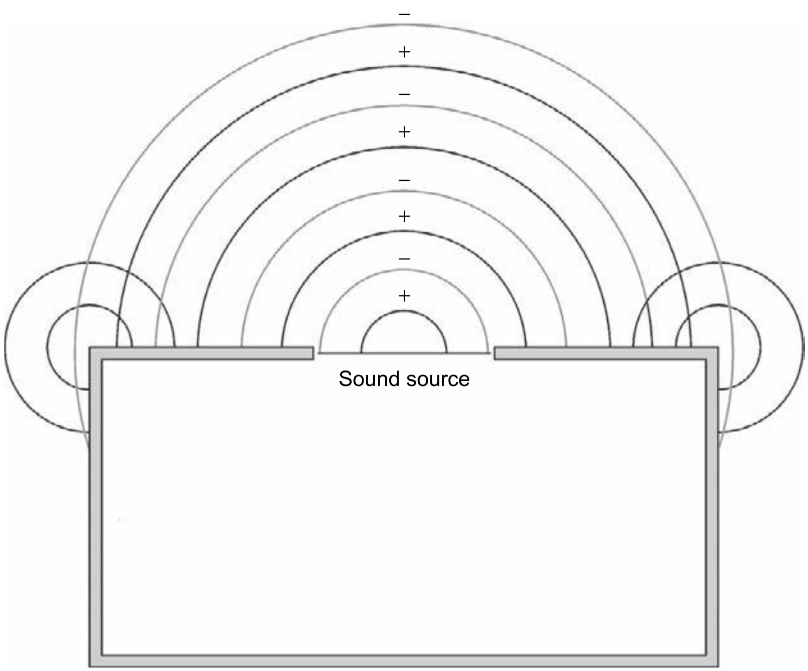


Figure 2.4: Extra sound waves generated by diffraction at the corners of a loudspeaker enclosure.

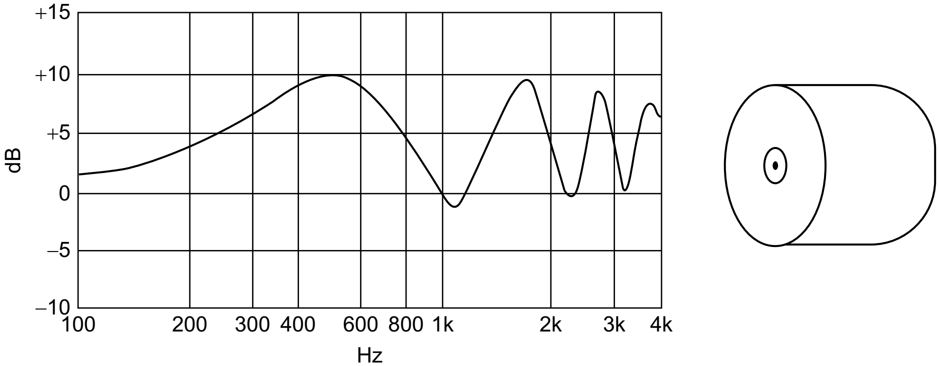


Figure 2.5: Serious response disturbances due to corners, superimposed on 6 dB rise; cylindrical loudspeaker enclosure 24 inches in diameter and 24 inches long (after Olson, 1969).

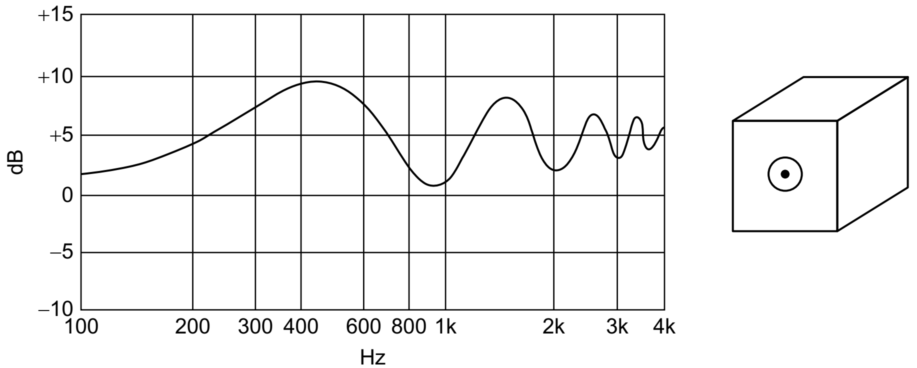


Figure 2.6: Reduced response disturbances due to corners, superimposed on 6 dB rise; cubic loudspeaker enclosure 24 inches in on a side (after Olson, 1969).

of wavelength, is linear. Although the response to be corrected should be constant, because it depends on the mechanical dimensions of the enclosure, any attempt to equalise away these variations would take quite a bit of doing. The obvious conclusion is that this shape of enclosure is of no use to man or beast, and should be shunned.

Rectangular (or to be strictly accurate, rectangular parallelepiped, i.e., with all edges parallel) boxes are of course much easier to fabricate than spheres or cylinders, and are correspondingly more popular. The simplest shape is the cube, which has the advantage over the cylinder that the path length from the drive unit to the front edges varies. This helps to smooth out the response wobbles, but as Figure 2.6 shows, they are still substantial, the difference between the first peak and the first dip being about 8 dB.

Figure 2.7 shows the result of converting the flat face of the forward baffle into a truncated pyramid, to make the edges less sharp. The frequency response deviations are now much less pronounced, not exceeding 2 dB, apart from the inherent 6 dB rise. Olson gives the following information about its dimensions: "The length of the edges of the truncated surface is 1 ft. The height of the truncated pyramid is 6 in. The lengths of the edges of the rectangular parallelepiped are 1 ft. and 2 ft. The loudspeaker mechanism is mounted in the centre of the truncated surface. The lengths of the edges of the rectangular parallelepiped are 2 ft. and 3 ft. The loudspeaker mechanism is mounted midway between two long edges and 1 ft. from one short edge."

The vast majority of loudspeakers are of course rectangular boxes, as shown in Figure 2.8. These give a much worse response than the previous example because of the sharp edges of

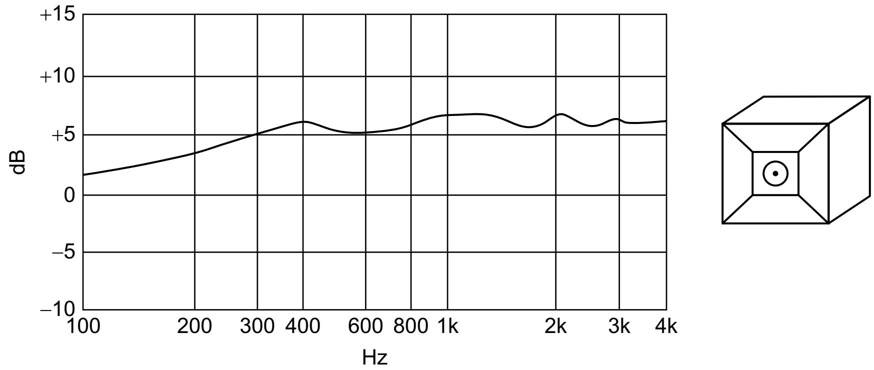


Figure 2.7: Much reduced response disturbances superimposed on 6 dB rise; truncated pyramid and rectangular parallelepiped loudspeaker enclosure; dimensions given in text (after Olson, 1969).

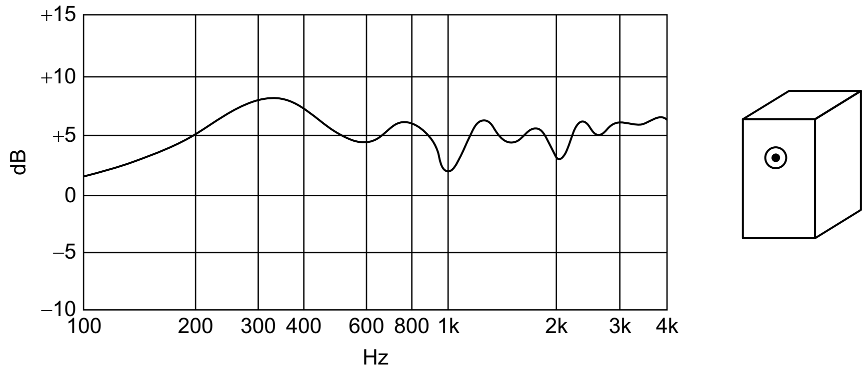


Figure 2.8: Response disturbances due to sharp corners, superimposed on 6 dB rise; rectangular loudspeaker enclosure; dimensions given in text (after Olson, 1969).

the front baffle, with deviations of up to 6 dB. The lengths of the edges of the rectangular box were 2 ft. and 3 ft. The drive unit was mounted midway between two long edges and 1 ft. from the top short edge.

Using the information derived from these enclosure shapes, Olson suggested an enclosure that would give good results while still fitting into a domestic setting rather better than a sphere. The results are given in Figure 2.9, and the response deviations are about 2 dB; not as good as the sphere, but far better than the simple rectangular box. Olson say about its dimensions: “A rectangular truncated pyramid is mounted upon a rectangular parallelepiped. The lengths of the edges of the rectangular parallelepiped are 1, 2, and 3 ft. The lengths of the edges of the truncated surface are 1 ft. and 2½ ft. The height of the truncated pyramid is 6 in. One surface of the pyramid and one surface of the parallelepiped lie in the same plane.”

There are a few interesting points about the Olson tests that are not normally quoted. Firstly, the drive unit he used was specially designed and had a cone diameter of only 7/8 of an inch, so it was small compared with the wavelengths in question and could be treated as a perfect piston; that is why it looks so small on the enclosure sketches. All the measurements were done on-axis in an anechoic chamber. Secondly, while the reference that is always given is the 1969 JAES paper, because it is accessible, the original paper was presented at the Second Annual Convention of the AES in 1950. This accounts for the somewhat steampunk look of the measuring equipment pictured in the paper.

Other workers in this field have reported that further reduction in the response ripples can be obtained by optimising the driver position and rounding the enclosure edges.

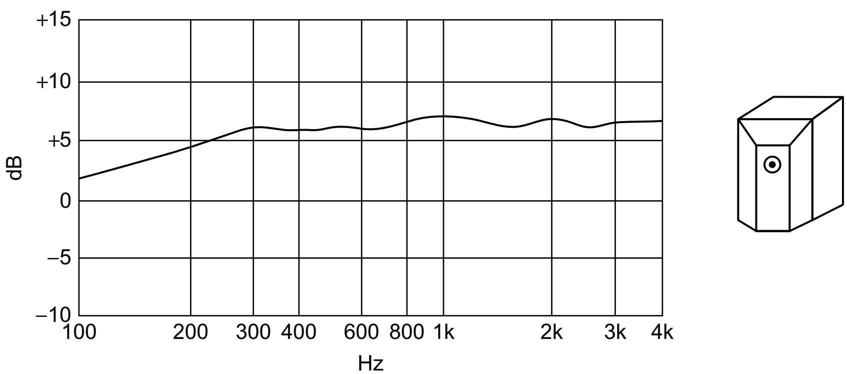


Figure 2.9: Much reduced response disturbances due to corners, added to 6 dB rise; Olson optimal loudspeaker enclosure; dimensions given in text (after Olson, 1969).

2.7 Modulation Distortion

With any crossover the drive units have to handle a range of frequencies, and this leads to the problem known as modulation distortion. What has been called "total modulation distortion" [4] is made up of two components: amplitude modulation distortion and frequency modulation distortion. The first is intermodulation distortion caused by non-linearities in the drive unit. The second is a result of the Doppler effect, a low-frequency movement of the cone causing frequency modulation of the higher frequencies present. Both modulation distortion components are reduced by limiting the range of frequencies that the drive units handle and by using steep crossover slopes.

Further Reading

If you want to get deeper into loudspeaker design issues, two excellent books are *Loudspeakers* by Newell and Holland, [5] and *The Loudspeaker Design Cookbook* by Vance Dickason [6].

References

[1] http://en.wikipedia.org/wiki/Thiele/Small (accessed Dec 2010).
[2] H.F. Olson, Direct radiator loudspeaker enclosures, JAES 17 (1) (1969) 22–29.
[3] Cabasse, http://www.cabasse.com/us/ (accessed Dec 2010).
[4] P.W. Klipsch, Modulation Distortion in Loudspeakers, AES Convention 34, 1968, Paper No. 562.
[5] Newell, Holland, Loudspeakers for Music Recording and Reproduction, Focal Press, USA, 2007, ISBN-13: 978-0-240-52014-8, ISBN-10: 0-2405-2014-9.
[6] V. Dickason, The Loudspeaker Design Cookbook, seventh ed., Audio Amateur Press, USA, 2006, ISBN: 1882580478.

This page intentionally left blank

CHAPTER 3

Crossover Requirements

3.1 General Crossover Requirements

The desirable characteristics of a crossover system are easy to state, but not so easy to achieve in practice. There is a general consensus that there are five principal requirements that apply to all crossovers, be they active or passive, and that they should be ranked in order of importance thus:

1. Adequate flatness of summed amplitude/frequency response on-axis
2. Sufficiently steep rolloff slopes between the output bands
3. Acceptable polar response
4. Acceptable phase response
5. Acceptable group delay behaviour

To some extent, the amount of space devoted in this book to each requirement is dependent on their relative importance as set out in this list.

3.1.1 Adequate Flatness of Summed Amplitude/Frequency Response On-Axis

This requires that the output of each filter is appropriate in both amplitude and phase over a sufficient frequency range so that when summation occurs the overall amplitude/frequency response is flat. Note that this requirement does *not* place restrictions on the phase of the summed result; in many cases this will show considerable phase shift that varies across the audio band. This issue is addressed in requirement 4 below.

Crossovers that sum to a completely flat amplitude response include the first-order crossover, the second-order Linkwitz–Riley crossover, the third-order Butterworth crossover and the fourth-order Linkwitz–Riley crossover. A new addition to this list is the Neville Thiele Method notch crossover.

There are many crossover types that can be made to sum very nearly flat by tweaking the filter cutoff frequencies. For example, the second-order Bessel crossover can be made flat to within ±0.07 dB by using a frequency offset ratio of 1.45 times, and the third-order Linkwitz–Riley crossover can be made flat to within ±0.33 dB by applying a frequency offset ratio of 0.872 times. There is much more on this in Chapter 4.

3.1.2 Sufficiently Steep Rolloff Slopes between the Filter Outputs

The filter rolloff slopes must be fast enough to prevent driver damage. Even a small amount of LF energy can rapidly wreck a tweeter. The slopes must be steep enough to not excite areas of poor drive unit frequency response, such as resonances outside normal band of usage; this applies only to mid-range drive units and tweeters, as the LF drive unit resonance will always be used. In addition, the linearity of drive units is very often worse outside their intended frequency range, so steeper slopes will give less non-linear distortion, and that has got to be a good thing.

Restricting the frequency range sent to each drive unit will also minimise frequency modulation distortion caused by the Doppler effect (see Chapter 2). It is also desirable to make the frequency range over which crossover occurs as narrow as possible to minimise the band over which lobing occurs due to two drive units radiating simultaneously.

Unless specially designed drive units are used, the minimum practical slope is usually considered to be 12 dB/octave which requires a second-order crossover. Steeper slopes such as 18 dB/octave (third-order) and 24 dB/octave (fourth-order) are generally considered to be very desirable; 48 dB/octave (eighth-order) slopes are sometimes used in sound reinforcement.

3.1.3 Acceptable Polar Response

An even and well-spread polar response is desirable because it increases the amount of space in which a good sound is obtained. It is also desirable to avoid a large amount of radiated energy from being directed at the floor in front of the loudspeaker, from which it will reflect and cause unwanted comb-filtering effects by interference. Loudspeakers normally give a good polar response in the horizontal on-axis plane, assuming the drive units are mounted in a vertical line as usual. This, however, causes problems in the vertical plane, for in the crossover region two drive units separated in position are radiating simultaneously, and their outputs will interfere, giving reinforcements and cancellations in the radiation pattern at different angles, known as lobing. This is a result of having two drive units separated in space and there is nothing the crossover designer can do about this except make the crossover frequency range as small as possible.

However, it gets worse. If the crossover outputs to each drive unit are in phase, then the main lobe points forward on the horizontal axis, and stays there. If, however, there is a constant phase shift between the outputs, as for first-order crossovers (90° phase shift) or third-order crossovers (270° phase shift) then the main lobe is tilted toward the drive unit that is phase lagging. This is usually the LF unit, so the main energy is being unhelpfully directed towards the floor. This is a frequency-dependent effect because it only occurs in the crossover region and is at its greatest at the crossover frequency itself. It is sometimes simply called “lobing error.”

It is therefore clear that if we are going to keep the main lobe of the summed acoustic output on the axis it is highly desirable that the lowpass and highpass outputs are in phase in the crossover region [1]. This property is possessed only by second-order crossovers (assuming one output is inverted to get a flat response, otherwise the phase-shift is 180°) and fourth-order crossovers with no inversion. This is one reason for the popularity of the fourth-order Linkwitz–Riley crossover.

All the above depends on the drive unit time-delay compensation being correct; the drive units must be either physically mounted or electrically compensated so that the direct sound from each one arrives at the listener's ear at the same time over the whole of the crossover frequency range. Otherwise, the main lobe will have a frequency-dependent tilt toward the driver with the longest air path to the ear.

A good polar response therefore requires that the crossover outputs be in phase and that the time-delay compensation be correct.

3.1.4 Acceptable Phase Response

An acceptable phase response for the *combined* output is also required. Most crossovers are not linear-phase or minimum phase but have the phase response of a first-order allpass filter, with the phase changing by 180° over the audio band. The best-known of these are the first-order (inverted), the second-order Linkwitz–Riley, and third-order Butterworth crossovers. The fourth-order Linkwitz–Riley crossover has the phase response of a second-order allpass filter, with the phase changing by 360° over the audio band. These phase responses are generally agreed to be inaudible with music signals so the fourth requirement is not too onerous. There is a separate section on the issue of phase perception later in this chapter.

3.1.5 Acceptable Group Delay Behaviour

Group delay is simply a measure of how much a signal is delayed. This is directly connected with an acceptable phase response for the combined output, and in fact the group delay is completely determined by the phase shift. Group delay is mathematically the rate of change of the total phase shift with respect to angular frequency (i.e., frequency measured in radians per second rather than Hertz).

Group delay would be of little interest if it was constant, but as the rate of change of phase varies across the audio band, with the phase response of an allpass filter, the group delay also varies. The change is sometimes smooth, but may show a pronounced peak near the crossover frequency. This variation would, if it was sufficiently severe, cause a time-smearing of acoustical events and would sound truly dreadful, but you must not mistake this with the use of the word "smearing" in hi-fi reviews, where it is purely imaginary.

Table 3.1: Variation of Group Delay Threshold with Frequency

Frequency	Group Delay
500 Hz	3.2 ms
1 kHz	2.0 ms
2 kHz	1.0 ms
4 kHz	1.5 ms
8 kHz	2.0 ms

The thresholds for the perception of group delay variation are well known because of their historical importance on long telephone lines. The most accepted thresholds were given by Blauer and Laws in 1978 [2], and are shown in Table 3.1.

These times are given in milliseconds, and a typical group delay for a 1 kHz crossover would be something like 10 times less. The section on phase perception later in this chapter is concerned with the audibility of allpass filters, and the conclusion is firm that neither their phase-shift nor their group delay can be heard on normal musical signals.

The word "group" is derived from "group velocity" in wave-propagation theory, but for our purposes it is simply the amount by which a signal at a given frequency is delayed.

3.2 Further Requirements for Active Crossovers

In addition to the general requirements for all crossovers given above, there are further special requirements for active crossovers. As explained in Chapter 1, if you are adding an extra unit into the signal path it must be as transparent as possible if overall quality is not to take a step backwards. The ultimate goal is total transparency so that the introduction of the crossover causes no degradation at all.

Some further requirements for active crossovers, in no particular order, are:

1. Negligible extra noise
2. Negligible impairment of system headroom
3. Negligible extra distortion
4. Negligible impairment of frequency response
5. Negligible impairment of reliability

It would be easy to add further requirements to this list, such as no degradation in EMC immunity, or of safety, or a modest power consumption, but these apply to any piece of electrical equipment. Those listed above are pretty clearly the most important.

3.2.1 Negligible Extra Noise

While passive crossovers have many limitations, as described in Chapter 1, they do not add noise to the signal. I am sure someone will now point out that crossover inductors do have some resistance and thus must generate Johnson noise, but I am quite certain that −152 dBu of noise (which is what you would get from a 1 Ω resistance) added to a loudspeaker-level signal is one of the lesser problems facing the audio business.

In contrast, the average active crossover performs its processing at line level, or possibly below, and the relative complex structure of a high-quality crossover may allow lots of opportunities for the signal-to-noise ratio to be degraded. Since active crossovers are placed after the main system volume control, turning down the volume will not turn down its noise contribution and with poor design the result could easily be unwelcome levels of hiss from the loudspeakers.

For this reason I have deemed it important to consider noise performance at every step while describing the varied internal circuit blocks of an active crossover. Chapter 14 describes how to minimise noise by using low-impedance design, by using active gain controls, and by adopting the best order for the stages in the crossover. It also deals with the very important possibility of running the crossover at higher nominal internal levels than the input and output signals, provided the placement of gain controls in the whole sound system makes this feasible, so the circuit noise is relatively lower and we get better signal/noise ratios. The intriguing possibility of running the HF crossover path at a still higher level than the others because of the relatively small amount of energy at the top end of the audio spectrum is also looked at in detail. Still another technique is optimising the order of stages in each crossover path; making sure the lowpass filters come last, after the highpass and allpass delay-compensation filters in the path will mean that the noise from upstream is lowpass filtered and may be reduced by several dB. Finally, Chapter 16 on line inputs demonstrates that the conventional balanced line input stage is a rather noisy beast, and shows a number of ways in which to improve it.

I have attempted to bring all these low-noise techniques together in a demonstration crossover design in Chapter 19 at the end of this book. The measured results for the all−5532 version show a signal/noise ratio of 117.5 dB for the HF path output, 122.2 dB for the MID output, and 127.4 dB for the LF output, which I think proves that using all the above noise reduction techniques together can give some pretty stunning results.

Note however, that this crossover will still be a few dB noisier than a naked power amplifier where the input goes straight to the input transistor pair; such power amplifiers have an Equivalent Input Noise (EIN) of about −120 dBu if well designed [3]. However, almost any sort of balanced input in the power amplifier will reverse this situation and the crossover will produce less noise than the amplifier. This point is examined in more depth at the end of Chapter 19.

3.2.2 Negligible Impairment of System Headroom

As described in the previous section, it can be very advantageous to the noise performance to run an active crossover with elevated internal levels, so long as the placement of the gain controls in the whole system permits this. The scenario you are trying to avoid is having a level control between the crossover and the power amplifier that somehow (not me, guv!) gets turned down so it attenuates excessively. Then somebody turns up the volume control on the preamp or pushes up the mixing desk output faders to compensate. This means there is an excessive level inside the crossover, an unexpected signal peak comes along, and… crunch.

This is not a hard situation to avoid. If the active crossover has output level controls that are essentially gain trims with a limited range then it should not be possible to introduce excessive attenuation. In other cases headroom problems are avoided simply by having one competent person with control over the whole system. In hi-fi applications the maximum input levels are fairly well defined by the FSD of the digital source. In other cases, the mechanical limits of the wax cylinder or the vinyl disc impose a less well-defined but still very real limit. In sound reinforcement applications the input levels are much less predictable, but a combination of control from the mixing desk and the use of compressor/limiters should prevent excessive levels from getting as far as the active crossover.

3.2.3 Negligible Extra Distortion

The most significant source of distortion in the average sound system is either the loudspeakers, or if the signal source is vinyl disc, the process of cutting grooves and then wiggling needles about in them. This has never been and never will be accepted as justification for giving up on the design of very linear preamplifiers and power amplifiers. While progress has been made toward making power amplifiers as distortion-free as small-signal circuitry, there are still major technical challenges to be overcome and at present the most significant source of distortion in the electronic domain is almost always the power amplifier. For this reason it may not be too hard to design a signal path that is significantly more free of distortion, especially the crossover distortion that we all abhor, than the average power amplifier. Nevertheless, as I have said before, we are inserting an extra signal-processing box into the signal path, and it behoves us to make the degradation it introduces as small as economically possible. This is not too hard from the cost point of view as the active crossover, even if built to the highest standards, is likely to be much cheaper than the extra power amplifiers required for multi-way operation.

3.2.4 Negligible Impairment of Frequency Response

This may seem like a strange requirement in a piece of equipment whose whole *raison d'être* is radical modification of the spectrum of the signals it handles, but here I want to distinguish between the frequency response modifications you want and those you don't. Even if we assume that an active crossover has a well-conceived filter structure, correct filter characteristics, suitably low component sensitivities, and is built with components of accurate value, it would still be possible to degrade the frequency response by using an over-aggressive ultrasonic filter or an inappropriate subsonic filter, and these stages should get their full share of attention.

3.2.5 Negligible Impairment of Reliability

There is nothing that upsets the paying customer more than equipment that stops working (OK, if it sets fire to the house that would probably annoy them rather more), so I make no apology for putting this rather general requirement in. We are dealing with opamp circuitry using modest supply rail voltages, the general levels of voltage and current are low, and the opamps even have internal overload protection. So as long as the designer knows what they're about there is no reason why any components should be much stressed. There are a couple of not-quite-obvious things that could go wrong in the power supply, such as regulator heatsinks that are normally OK but prove to be too small when the mains is 10% high in a hot country, or ill-conceived decoupling capacitors that cause the supply to fail to start on an unpredictable basis. These design landmines are dealt with in Chapter 18 on power supply design.

3.3 Linear Phase

A linear-phase crossover has a combined output phase-shifted by an amount proportional to frequency; in other words it introduces a pure time-delay only, and the group delay is constant. Non-linear-phase crossovers have phase-shifts that change non-linearly with frequency and act like allpass filters; for this reason they are often called allpass crossovers. Now while linear phase is clearly desirable on a purely theoretical basis, in that it makes the crossover more transparent and closer to perfection, it is difficult to achieve. There is certainly no settled consensus that it is necessary for a good acoustic performance, and the bulk of evidence is that it is simply not necessary for satisfactory results on normal music. There is more on this issue in the section below on phase perception.

The best-known crossover types with linear phase are first-order non-inverted crossovers, filler-driver crossovers, and subtractive crossovers with time delay such as those put forward by Lipshitz and Vanderkooy in 1983 [1].

3.4 Minimum Phase

Minimum phase, or minimal phase, is a term that is sometimes confused with linear phase. In fact they are not only not the same thing, but almost opposites. A minimum-phase system, in our case a filter, has the minimum phase shift possible to get the amplitude/frequency response it shows. A minimum-phase filter is also one where the phase/frequency response can be mathematically derived from the amplitude/frequency response, and vice-versa. Furthermore, the effect of a minimum-phase filter can be completely cancelled out in both phase and amplitude by using a reciprocal filter that has the opposite effect. A good example of this is given in Figure 11.1 in Chapter 11 on equalisation, where it is demonstrated that a peaking equaliser can be exactly cancelled out by a dip equaliser, and a square-wave put through them both is reconstructed.

There are, as you might expect, several much more precise mathematical ways of defining the minimum-phase condition [4], but they are not helpful here. In general you cannot say that a minimum-phase filter is better than a non-minimum-phase filter, as it depends what you are trying to do with said filter.

Most crossover filters, such as highpass and lowpass types in their various kinds are inherently minimum phase. The classic exception to this is the allpass filter used for time-delay correction. Since an accurate allpass filter has a completely flat amplitude/frequency response, you cannot deduce anything at all from it about the phase/frequency response. The phase-shift of an allpass filter in fact varies strongly with frequency, as described in Chapter 10 on time-delay compensation, but the amplitude/frequency response gives you no clue to that. It is also impossible to undo the effect of an allpass filter because its particular phase-shift characteristics give a constant delay at suitably low frequencies, and you cannot make a filter that has a negative delay. That would mean foretelling the future, and on the whole would probably not be a good thing.

The first-order crossover is minimum phase when its outputs are summed normally; it has a flat phase plot at 0°. If one output is inverted, however, while the SPL and power responses are still flat, the summed output has a first-order allpass phase response, the phase swinging from 0° to −180° over the frequency range. It is therefore no longer minimum-phase.

As we just noted a linear-phase crossover acts as a pure time delay, and so cannot be minimum-phase. You cannot, however, say that a crossover which includes allpass filters for time-delay correction can never be minimum phase, because they are correcting for physical misalignments and what counts is the summed signal at the ear of the listener.

3.5 Absolute Phase

Another phase issue is the perception of absolute phase. In other words, if the polarity of a signal is inverted (it is not relevant whether it is heard via a single or multi-way loudspeaker) does it sound different? The answer is yes, providing you use a single tone with a markedly asymmetrical waveforms, such as a half-wave rectified sinewave or a single unaccompanied human voice. Otherwise, with more complex signals such as music, no difference is heard.

Obviously an active crossover must have all its outputs in the correct phase with each other (in some cases correct means phase-inverted) or dire response errors will result, but it is also necessary to make sure that the outputs are in the correct phase relationship to the crossover input signal.

Almost all hi-fi equipment such as preamplifiers and power amplifiers are now designed to preserve absolute phase. Mixing consoles have always been so designed to prevent unwanted cancellation effects on mixing signals.

3.6 Phase Perception

Some of the crossovers described in this book have quite dramatic phase changes in the summed output around the crossover points, so the sensitivity of human hearing to phase shift is an important consideration. If the phase-shift is proportional to frequency then the group delay is constant with frequency and this is a linear-phase system, as described above; we just get a pure time-delay with no audible consequences. However, in most cases the phase-shift is not remotely proportional to frequency, and so the group delay varies with frequency. This is sometimes called group delay distortion, which is perhaps not ideal as 'distortion' implies non-linearity to most people, while here we are talking about a linear process.

Most of the components in the microphone-recording-loudspeaker chain are minimum-phase; they impose only the phase-shift that would be expected and can be predicted from their amplitude/frequency response. The great exception to this is… the multi-way loudspeaker. The other great exception was the analogue magnetic tape-recorder, which showed rapid phase-changes at the bottom of the audio spectrum, usually going several times round the clock [5]. Fortunately we don't need to worry about *that* any more.

We are, however, going to have multi-way loudspeaker systems around for the foreseeable future, and most of them have allpass crossovers. Clearly an understanding of what degradation, if any, this allpass behaviour causes is vital. Much experimentation has been done and there is only space for a summary here.

One of the earliest findings on phase perception was Ohm's Law. No, not that one, but Ohm's *Other* Law, which is usually called Ohm's Acoustic Law, and was proposed in 1843 [6]. In its original form it simply said that a musical sound is perceived by the ear as a set of sinusoidal harmonics. The great researcher Hermann von Helmholtz extended it in the 1860s into what today is known as Ohm's Acoustic Law, by stating that the timbre of musical tone depends solely on the number and relative level of its harmonics, and *not* on their relative phases. This is a good start, but does not ensure the inaudibility of an allpass response.

An important paper on the audibility of midrange phase distortion was published by Lipshitz, Pocock and Vanderkooy in 1982 [7] and they summarised their conclusions as follows:

1. Quite small phase non-linearities can be audible using suitable test signals.
2. Phase audibility is far more pronounced when using headphones instead of loudspeakers.
3. Simple acoustic signals generated in an anechoic environment show clear phase audibility when headphones are used.
4. On normal music or speech signals phase distortion is not generally audible.

At the end of the abstract of their paper the authors say: "It is stressed that none of these experiments thus far has indicated a present requirement for phase linearity in loudspeakers for the reproduction of music and speech." James Moir also reached the same conclusion [8].

An interesting paper on the audibility of second-order allpass filters was published in 2007 [9], which describes a perception of "ringing" due to the exponentially decaying sinewave in the impulse response of high Q all-pass filters (For example Q = 10). It was found that isolated clicks show this effect best, while it was much more difficult to detect, if audible at all, with test signals such as speech, music, or random noise. That is the usual finding in this sort of experiment—that only isolated clicks show any audible difference. While we learn that high-Q allpass filters should be avoided in crossover design, I think most people would have thought that was the case anyway.

Siegfried Linkwitz has done listening tests where either a first-order allpass filter, a second-order allpass filter (both at 100 Hz), or a direct connection could be switched into the audio path [10]. These filters have similar phase characteristics to allpass crossovers and cause gross visible distortions of a square waveform, but are in practice inaudible. He reports "I have not found a signal for which I can hear a difference. This seems to confirm Ohm's Acoustic Law that we do not hear waveform distortion."

If we now consider the findings of neurophysiologists, we note that the auditory nerves do not fire in synchrony with the sound waveform above 2 kHz; so unless some truly subtle encoding is going on (and there is no reason to suppose that there is), then perception of phase above this frequency would appear to be inherently impossible.

Having said this, it should not be supposed that the ear operates simply as a spectrum analyser. This is known not to be the case. A classic demonstration of this is the phenomenon of "beats." If a 1000-Hz tone and a 1005-Hz tone are applied to the ear together, it is common knowledge that a pulsation at 5 Hz is heard. There is no actual physical component at 5 Hz, as summing the two tones is a linear process. (If instead the two tones were multiplied, as in a radio mixer stage, there *would* be new components generated) Likewise non-linearity in the ear itself can be ruled out if appropriate levels are used.

What the brain is actually responding to is the envelope or peak amplitude of the combined tones, which does indeed go up and down at 5 Hz as the phase relationship between the two waveforms continuously changes. Thus the ear is in this case acting more like an oscilloscope than a spectrum analyser.

It does not however seem to work as a phase-sensitive detector.

The conclusion we can draw is that a crossover whose summed phase response is that of a first-order or second-order allpass filter is wholly acceptable. This obviously implies that a group delay characteristic that emulates a first- or second-order allpass filter is also completely acceptable.

3.7 Target Functions

A target function for a loudspeaker system is the combined crossover and loudspeaker response that you are aiming for. Drive units are hopefully fairly flat over the frequency range that we hope to use them, but if this is not the case, then their response obviously has to be taken into account. Response irregularities may be corrected by equalisation (see Chapter 11), performed either by adding dedicated equalisation stages to the relevant crossover path or by modifying the characteristics of the crossover filters. The latter uses less hardware but is much more difficult to understand unless properly documented.

In some cases the inherent properties of the drive unit and the enclosure may form part of the target function. For example, a suitably damped LF unit and enclosure will have a second-order Butterworth-type maximally flat rolloff at 12 dB/octave. If this is combined with a second-order Butterworth highpass filter in the crossover, then this makes up a Linkwitz–Riley fourth-order alignment which can be used to crossover to a separate subwoofer.

References

[1] S.P. Lipshitz, J. Vanderkooy, A family of linear-phase crossover networks of high slope derived by time delay, JAES 31 (1983) 1.

[2] Blauert, Laws, Group delay distortions in electroacoustical systems, J. Acoust. Soc. Am. 63 (5) (1978) 1478–1483.

[3] D. Self, Audio Power Amplifier Handbook, fifth ed., Elsevier Inc., Boston, MA, p. 104 (power amp input noise), ISBN: 978-0-240-52162-6.

[4] Anon, http://en.wikipedia.org/wiki/Minimum_phase, 2010 (accessed 12.10).

[5] L. Fincham, The subjective importance of uniform group delay at low frequencies, JAES 33 (6) (1985) 436–439.

[6] http://en.wikipedia.org/wiki/Ohm%27s_acoustic_law, 2010 (accessed 12.10).
[7] Lipshitz, Pocock, Vanderkooy, On the audibility of midrange phase distortion in audio systems, JAES (1982) 580–595.
[8] J. Moir, Comment on “Audibility of midrange phase distortion in audio systems,” JAES (1983) 939.
[9] Møller, Minnaar, Olesen, Christensen, Plogsties, On the audibility of all-pass phase in electroacoustical transfer functions, JAES 55 (3) (2007) 113–134.
[10] http://www.linkwitzlab.com/phs-dist.htm (allpass audibility), 2010 (accessed 12.10).

CHAPTER 4

Crossover Types

There are many types of crossover, classified in various ways. The categories listed here define what the crossover does, not how it does it. It is essentially a catalogue of target functions. It does not matter how the filter responses are obtained; that may be by filtering alone or a combination of filtering and drive unit responses.

4.1 All-Pole and Non-All-Pole Crossovers

In discussions of crossover design, and in filter design generally, the terms "pole" and "zero" tend to be freely scattered through the text. These terms relate to the complex mathematical equation that describes the response of a filter or other frequency dependent system such as a servomechanism. A complex equation is one involving *j*, the square root of −1, and not just a *complicated* equation. This equation is often called the transfer function; it usually comes in the form of a fraction, and the poles derive from the bottom half (the denominator), while the zeros derive from the top half (the numerator). Both are distinguished by their frequencies, so you might say: "There is a pole at 1 kHz."

If that is as clear as mud, don't panic. As I said in the preface, this book attempts to avoid getting involved in the whole complex algebra business. For our purposes, a pole causes the frequency response to turn downwards by 6 dB per octave, as in a simple first-order RC lowpass filter. A zero causes the frequency response to turn upwards by 6 dB per octave, but it is *not* equivalent to a first-order CR highpass filter; that is represented by the combination of a pole and a zero. You can see why this stuff is hard to understand.

Poles sometimes occur in pairs (called conjugate pairs) and such a pair acts as a resonator, creating a response peak whose height depends upon the Q of the circuit. Zeros can also occur in conjugate pairs; a notch filter has a conjugate pair of poles combined with a conjugate pair of zeros.

An all-pole crossover is composed entirely of all-pole filters. These are filters such as Butterworth, Bessel, and Linkwitz–Riley types which have a monotonic rolloff. In other words, once the response starts going down it does not come back up again.

Non-all-pole crossovers contain filters such as the inverse–Chebyshev and elliptical (Cauer) types that have zeros as well as poles in their responses and their rolloffs are not monotonic.

For example, an inverse–Chebyshev filter has notches in the stopband. Notch-type crossovers contain either inverse–Chebyshev filters, or notch filters as such, like the Bainter filter. These filters incorporate zeros as well as poles, so they are also non-all-pole crossovers. Notch crossovers are dealt with in Chapter 5.

4.2 Symmetric and Asymmetric Crossovers

Symmetric crossovers have filters with the same slope at the crossover point. For example, a two-way second-order crossover has 12 dB/octave slopes for both HF and LF filters.

Asymmetric crossovers have differing slopes at the crossover point; the HF filter might have a 18 dB/octave slope while the LF filter has a 12 dB/octave slope. Asymmetric crossovers do not in general sum to flat unless a rolloff in the response of one of the drivers is being used to make the slopes equal in terms of acoustic output.

Subtractive crossovers in their straightforward (non-delayed) form always give asymmetric slopes for orders greater than one because the output derived by subtraction always has a 6 dB/octave slope, whatever the slope of the filter. A slope of 6 dB/octave is usually inadequate unless especially capable wide-range drive units are employed. Adding a time delay in the path going to the subtraction stage can give equal slopes, but the delay needs to be very precise for this to extend over an adequate frequency range, and this appears to be a major problem with the concept. This problem is explored in Chapter 6 on subtractive crossovers.

4.3 All-Pass and Constant-Power Crossovers

Crossovers are classified as either "All-Pass Crossovers" (APC) or "Constant-Power crossovers" (CPC) in accordance with the way that the outputs recombine.

An All-Pass Crossover has filter outputs that sum in the air in front of the loudspeaker to create a sound pressure level (SPL) with a flat amplitude/frequency response. The human ear is only sensitive to the pressure changes at the ear hole, and has no way to integrate the acoustic power bouncing around a room. If the filter outputs are summed electrically instead of acoustically, as a test of proper recombination, then the filter outputs should sum to a flat voltage response, voltage being equivalent to SPL. This is sometimes called the amplitude response. Summation is by vector addition (i.e., phase must be taken into account) of the highpass output V_{HP} with the lowpass output V_{LP}:

$$V_{SUM} = V_{HP} + V_{LP} \tag{4.1}$$

APC crossovers are the usual choice because they give the loudspeaker a flat on-axis frequency response, and this is usually where the person who pays for the audio system sits; however, this certainly does not guarantee a flat off-axis response. The phrase "all-pass," rather than something like "flat response," is used to emphasise that while the summed amplitude/frequency response may be flat, the summed phase response is that of an allpass filter. The phase response of most crossovers is that of a first-order allpass filter, with the phase changing by 180° over the audio band; the best-known of these being the first-order (inverted), the second-order Linkwitz–Riley, and third-order Butterworth types. The fourth-order Linkwitz–Riley crossover has the phase response of a second-order allpass filter, with the phase changing by 360°. There is more information on allpass filters in Chapter 10.

A Constant-Power Crossover has filter outputs that sum in the air in front of the loudspeaker to create a flat frequency response in terms of power rather than sound pressure level. The power response of a loudspeaker is the sum of all its off-axis and on-axis amplitude/frequency responses; it is the frequency response of the total acoustical power radiated into a given listening space. Because the signals are not phase-correlated, phase is ignored and summation is by RMS addition of the highpass output V_{HP} with the lowpass output V_{LP}:

$$V_{SUM} = \sqrt{(V_{HP})^2 + (V_{LP})^2} \tag{4.2}$$

CPC crossovers are not popular because they do not in general give a flat on-axis frequency response. However, they are sometimes said to be beneficial in reverberant environments where the off-axis output makes a significant contribution to the sound arriving at the listener. This is usually an undesirable (but sometimes unavoidable) situation as the characteristics of the listening space then have a greater effect on sound quality. For this reason the principles of CPC crossover design constitute a relatively small part of this book.

4.4 Constant-Voltage Crossovers

It sounds as though a constant-voltage crossover would be the same thing as an allpass crossovers, but in fact they are quite different. Constant-voltage crossovers operate by subtracting the output of a filter (which is one output) from the unfiltered input to generate the other output; in this book they are called "subtractive crossovers" as I think that term is a bit clearer. If the output from a lowpass filter is subtracted from the unfiltered input you get a highpass output. Constant-voltage crossovers have the property, rare amongst crossover types, of being linear-phase and so able to "reconstruct the waveform"; in other words, when you sum the two outputs you get back the waveform you put in. This process is described in detail below and also in Chapter 6 on subtractive crossovers. The big snag is that the output obtained by subtraction has a slope of only 6 dB/octave, whatever the slope of the filter, and this is usually inadequate.

The constant-voltage crossover was first properly described by Dick Small, one of the great pioneers of scientific loudspeaker design, in 1971 [1].

4.5 First-Order Crossovers

A first-order crossover has the great advantage that it is the only type that can produce linear-phase and minimum-phase response and a flat amplitude response. As a consequence, it is the only type that allows a waveform to be reconstructed when an HF and LF path are added together. It is potentially attractive for passive crossovers as it is the simplest crossover and so requires the smallest number of expensive crossover components and has the lowest power losses.

There is only one type of first-order crossover, just as there is only one sort of first-order filter. It is only when you go to second-order filters that you get a choice of characteristic, such as Butterworth, Bessel, and so on. It is however possible to use different cutoff frequencies for the highpass and lowpass filters to manipulate the crossover behaviour—see the Solen split first-order crossover described below.

A first-order crossover has some serious disadvantages. The slopes of a first-order crossover are a gentle 6 dB/octave and this does not normally provide enough separation of the frequencies sent to the drive units. Tweeters are likely to be damaged by excessive coil excursions when fed with inappropriately low frequencies. LF drive units are unlikely to be damaged by high frequencies, but if operated outside their frequency design range they are likely to show an irregular response and poor linearity. Likewise, the 6 dB/octave slope is usually too shallow to prevent activation of a tweeter's resonance frequency (which is usually below the crossover frequency), giving rise to an unwanted peak in the response. It is generally considered that if a first-order crossover is used, the drive units must be well-behaved for at least two and preferably three octaves on either side of the crossover frequency.

Actually achieving the maximally flat amplitude response and minimum-phase operation requires very careful driver alignment, and also requires the listener to be exactly the same distance from each drive unit. These problems are consequences of the large overlap in operation of the drive units.

A basic first-order crossover is shown in Figure 4.1, where the first-order crossover filters are shown as simple RC networks. The crossover frequency, which is the –3 dB cutoff frequency for each filter, is 1 kHz; this can be obtained by using 220 nF capacitors for C and 723.4 Ω resistors for R. This crossover frequency is used for all the examples in this chapter, not because it is a good crossover frequency—very often it isn't—but because it's a convenient way of displaying the results, putting the crossover point nicely in the middle of the plot. Using other crossover frequencies gives exactly the same behaviour, but the plots are shifted sideways. The circular summing element to the right represents how the two acoustic outputs from the HF and LF drive units add linearly in the air in front of the loudspeaker. It represents pure mathematical addition and does not load or affect the two filters in any way.

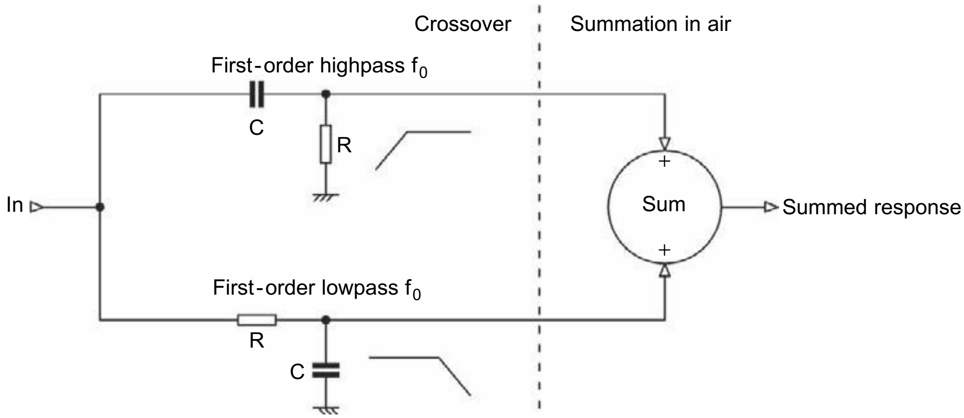


Figure 4.1: A 1 kHz crossover using first-order lowpass and highpass filters.

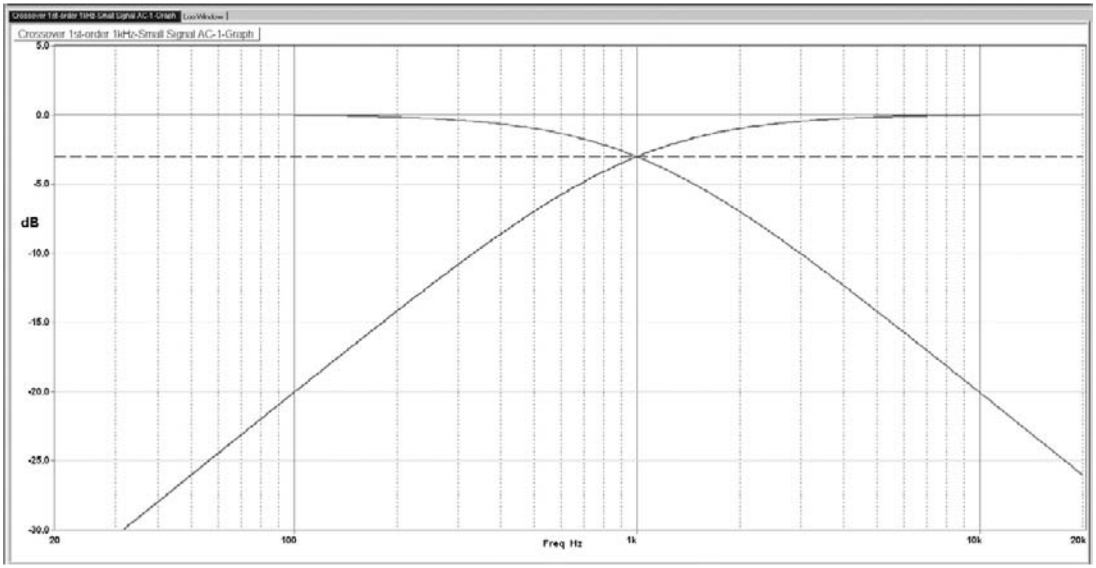


Figure 4.2: Frequency response of first-order 1 kHz crossover; both filter outputs plus their sum (the straight line at 0 dB).

Figure 4.2 shows the amplitude response of the two outputs and their sum. The sum is completely flat with frequency, laying on top of the 0 dB line in the plot. Neither output has been phase-inverted before summation. If one output is phase-inverted, the summed response is still dead flat—this is a property that no higher-order crossover possesses.

The power response is shown in Figure 4.3. Since the two contributions to the total power are uncorrelated (because of multiple room reflections and so on), they add in an RMS-fashion; in

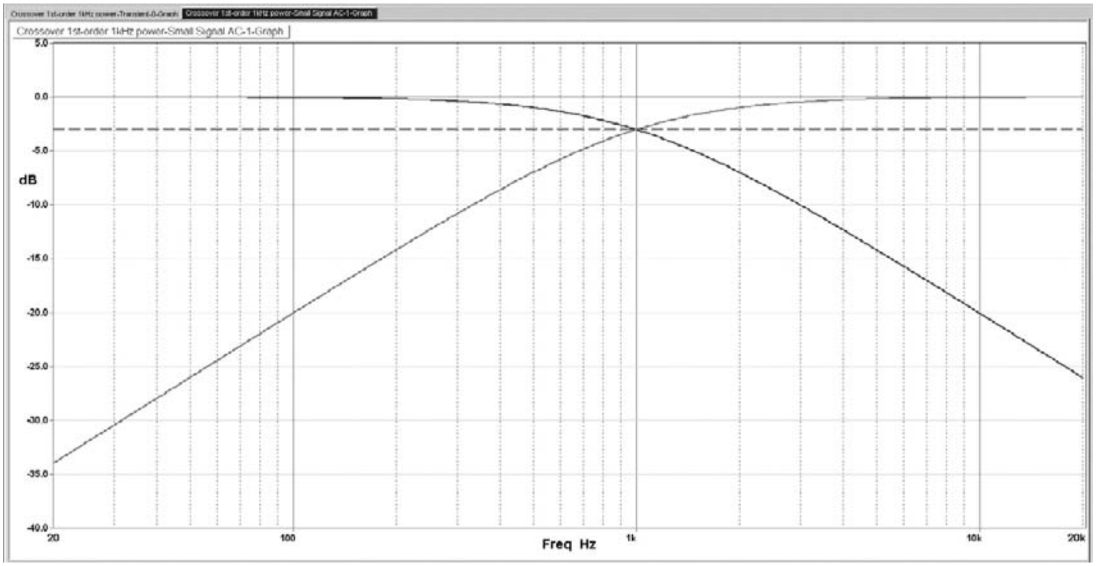


Figure 4.3: Power response of first-order 1 kHz crossover; both filter outputs plus their sum (the straight line at 0 dB).

other words you take the square root of the sum of the squares of the two levels. For the first-order crossover, the power response is also absolutely flat. Figure 4.3 therefore looks identical to Figure 4.2, though it was obtained by RMS summation rather than simple addition.

The two outputs have a constant 90° phase-shift between them, as shown in Figure 4.4. The phase of the summed outputs is always zero, as shown by the horizontal line at 0°. The crossover is minimum phase.

If however the highpass output is phase-inverted before summation, the summed response has a phase that swings from 0° to −180°, as shown in Figure 4.5. This is precisely the phase response of a first-order allpass filter, as described in Chapter 10, and demonstrates why the phrase “allpass crossover” crops up so often in this subject. The crossover is no longer minimum phase. Note that in Figure 4.5, the phase of the highpass output is shown *before* it is inverted. Please note that for this and all other phase plots in this chapter the 0° reference is the lowpass output at 20 Hz.

While a phase response that moves through 180° over the audio band may appear to be highly questionable, the consensus is that it is completely inaudible with normal music, and can only be detected by the use of special test signals such as isolated clicks. This is discussed in detail in Chapter 3.

With one output phase-inverted, the two outputs always have a phase difference of 90°, and this affects the radiation pattern of the two drive units in the frequency range where they

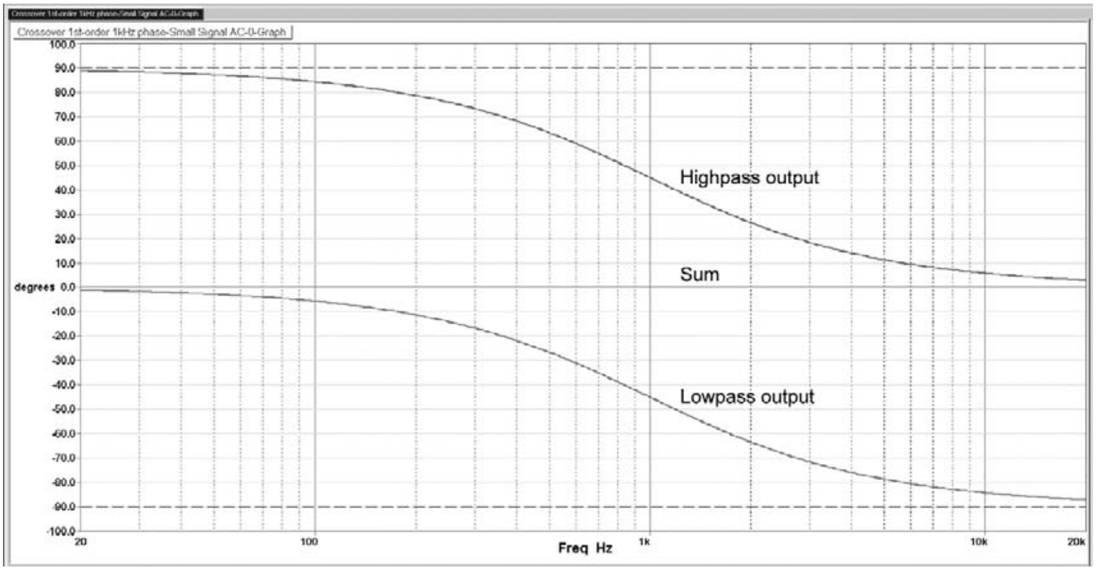


Figure 4.4: Phase response of first-order 1 kHz crossover; both filter outputs plus their sum (straight line at 0°).

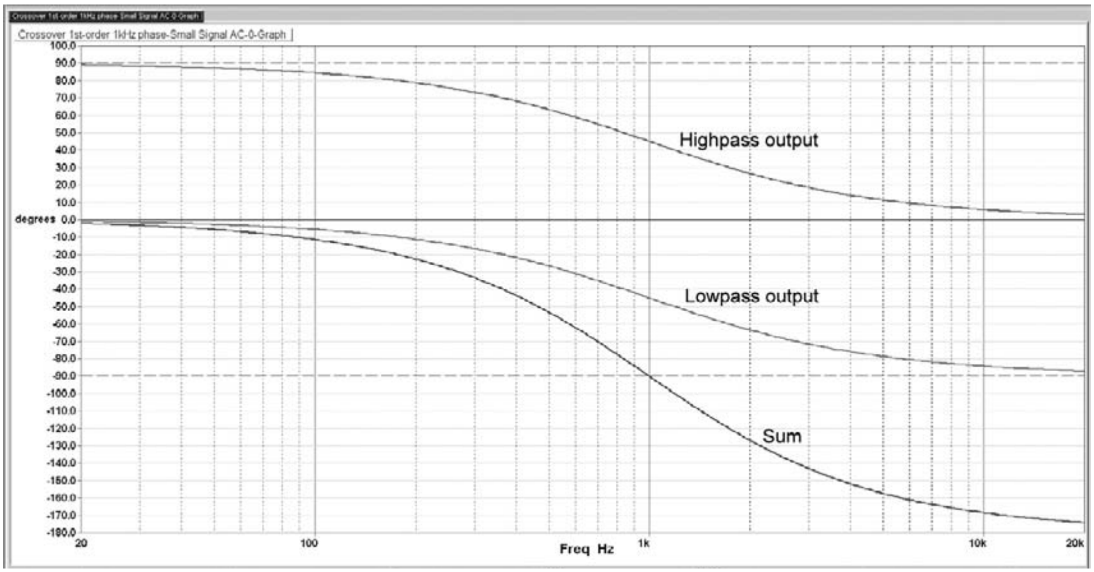


Figure 4.5: Phase response of first-order 1 kHz crossover; both filter outputs plus allpass sum when highpass output is inverted.

are both delivering a significant output. Because of the gentle filter slopes, this range is wider than in higher-order crossovers. With the normal (non-inverted) version, the result is “lobing error” which is a tilting of the vertical coverage pattern, pointing it downwards by 15° from the horizontal; when the crossover has one output inverted the tilt is 15° upwards. This assumes that the time alignment of the tweeter and woofer are correct at the crossover frequency to prevent differing time delays to the listener’s ear; first order crossovers are particularly sensitive to this because of the broad driver overlap. The tilt will increase and lobing will become more severe if the drivers are unduly separated on the baffle face. For a first-order crossover this effect is considered to be significant over at least two octaves.

The normal (non-inverted) connections show a flat group delay because the phase response is flat at zero. The inverted connection, however, has the group delay versus frequency response shown in Figure 4.6, with a flat section at 318 usec, rolling off slowly to zero over the crossover region.

The wide frequency overlap of a first-order crossover means that quite small time-alignment errors can cause large anomalies in the amplitude/frequency response. Vance Dickason, in his excellent *Loudspeaker Design Cookbook* [2], shows how a mere 0.5 inch misalignment can cause a peak or dip of 2.5 dB in a broad band centred on 3 kHz; the peak or dip effect depends on whether one output is inverted or not. A 1-inch misalignment causes amplitude ripples of up to ±4 dB, while a 2-inch misalignment causes errors of up to 10 dB. This is a serious argument against first-order crossovers.

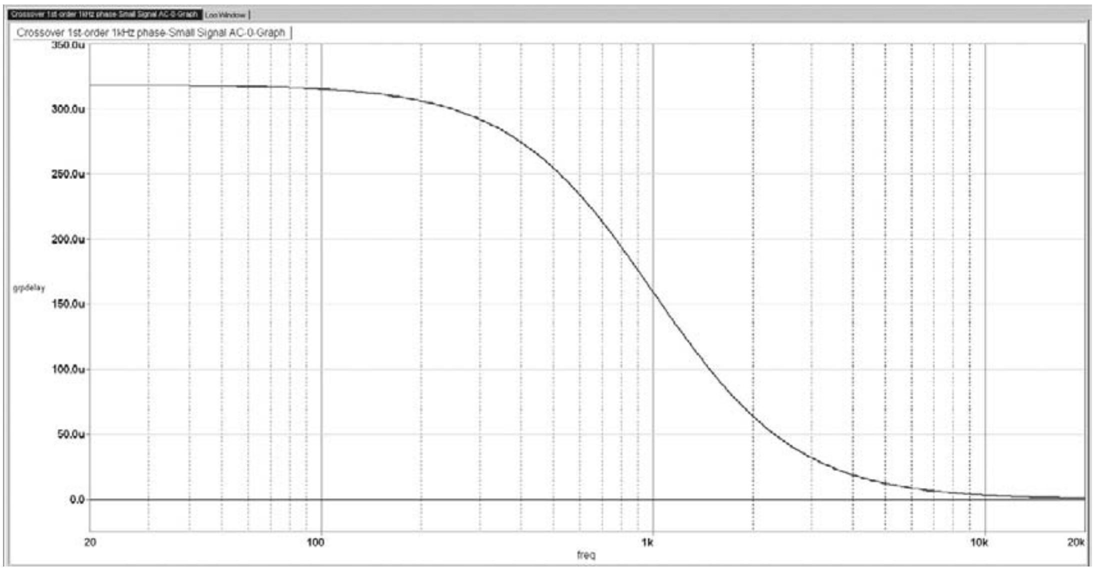


Figure 4.6: The group delay response of a first-order 1 kHz crossover, one output inverted.

A first-order crossover has the unique property of being able to reconstruct the waveform put into it. If you put a square wave into the crossover, and then sum the LF and HF outputs, you get a squarewave back; see Figure 4.7. The input waveform has levels of 0 V and +1.0 V; this is purely to make the middle plot clearer, and input levels of −0.5 V and +0.5 V, symmetrical about zero, give exactly the same reconstruction of the waveform. No other type of all-pole crossover has this ability.

As you read through this chapter, you will see that in some ways first-order crossovers have some uniquely desirable features, such as the ability to reconstruct the waveform, but also

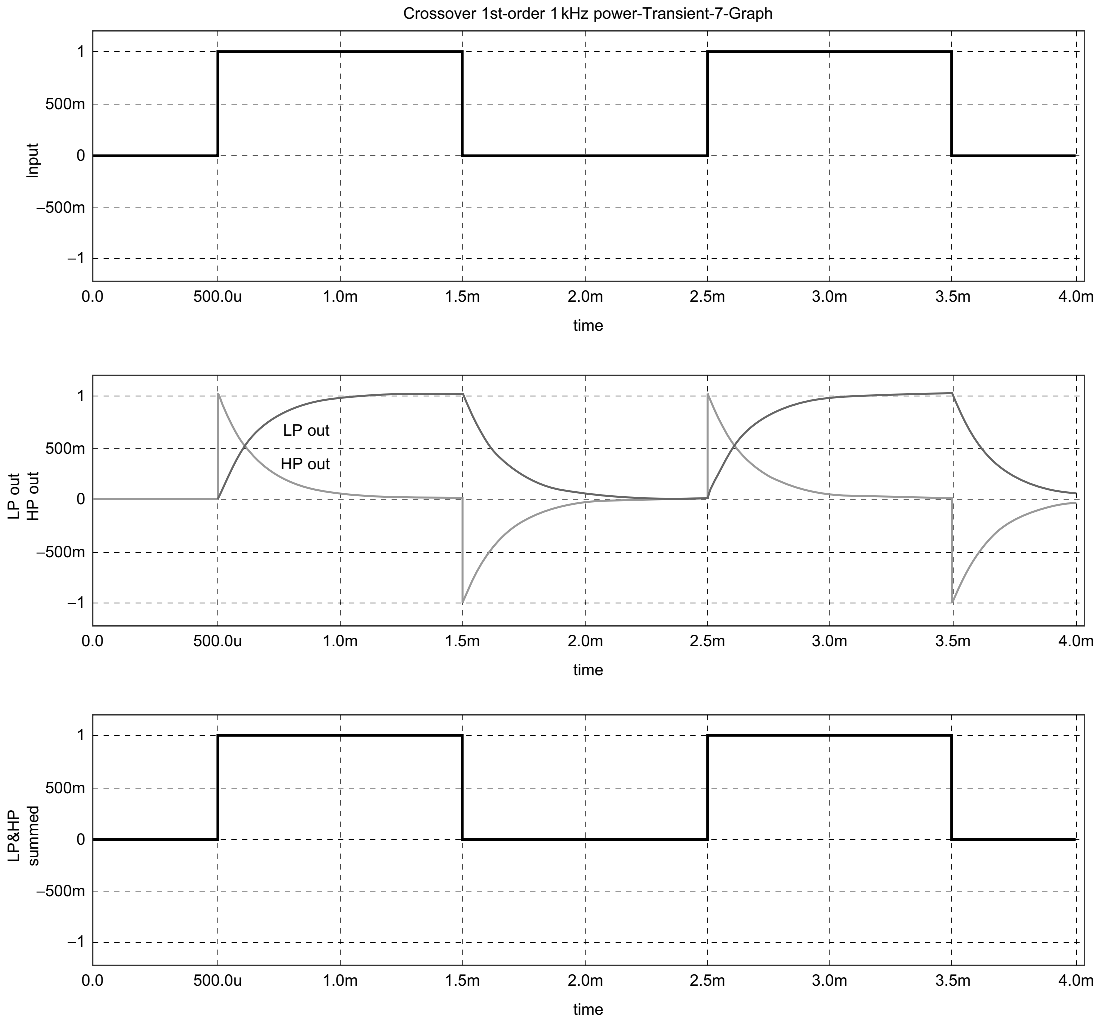


Figure 4.7: Reconstruction of a square-wave by summing the highpass and lowpass output of a first-order crossover. Non-inverted connection.

some serious disadvantages. The 6 dB/octave slopes are too gentle for use with normal drive units, and time alignment is frighteningly critical. Despite these disadvantages the positive features of a first-order crossover are so powerful that there does seem to be a distinct trend toward investing in driver technology in order to create units that are suitable for first-order operation. A recent example is the KEF Q-series of loudspeakers [3].

4.5.1 First-Order Solen Split Crossover

The so-called Solen split first-order crossover is a variation on the standard first-order crossover (Solen is a company that supplies passive crossover parts.) This scheme attempts to improve the slow crossover from one drive unit to the other by putting the crossover point for each filter at −6 dB rather than the usual −3 dB (as in Figure 4.2). This applies a frequency offset to the cutoff (−3 dB) frequencies of each filter, so that the lowpass filter now has a cutoff of 579 Hz and the highpass filter now has a cutoff of 1.726 kHz, pulling apart them apart by a factor of 2.98 times or 1.68 octaves. The pulling-apart process is presumably where the term "split" comes from. It may somewhat ease the demands on the drive units, but what of the amplitude response?

Figure 4.8 shows that, as we might expect, pulling apart the two cutoff frequencies has caused the summed response to sag in the middle, by 6 dB in fact. This is obviously going to sound like rubbish, but we get a better result if we reverse the phase to one of the drive units—a common manoeuvre in crossover design.

In passive crossover design phase-inverting one of the outputs is extremely simple; just swap over the two wires to the drive unit in question; everything is done in the box and no one is any the wiser. Active crossovers, however, will need to use some kind of phase-inverting stage.

Figure 4.9 demonstrates that with one of the phases reversed we still get a dip, but is now only 1.2 dB deep; information on this crossover scheme is scanty, but that is presumably how it is supposed to work. It might be possible to reduce the deviation from perfect flatness by partial cancellation of the dip with a response irregularity in one of the drivers. It is however still difficult to get enthusiastic about a crossover with 6 dB/octave slopes.

4.5.2 First-Order Crossovers: 3-Way

It is difficult to make a first-order 3-way crossover because the slow 6 dB/octave slopes do not provide adequate separation into three bands across the audio spectrum. There is in any case little point because drive units capable of handling the wide frequency ranges inherent in such a crossover would probably be equally suitable for a first-order 2-way crossover setup.

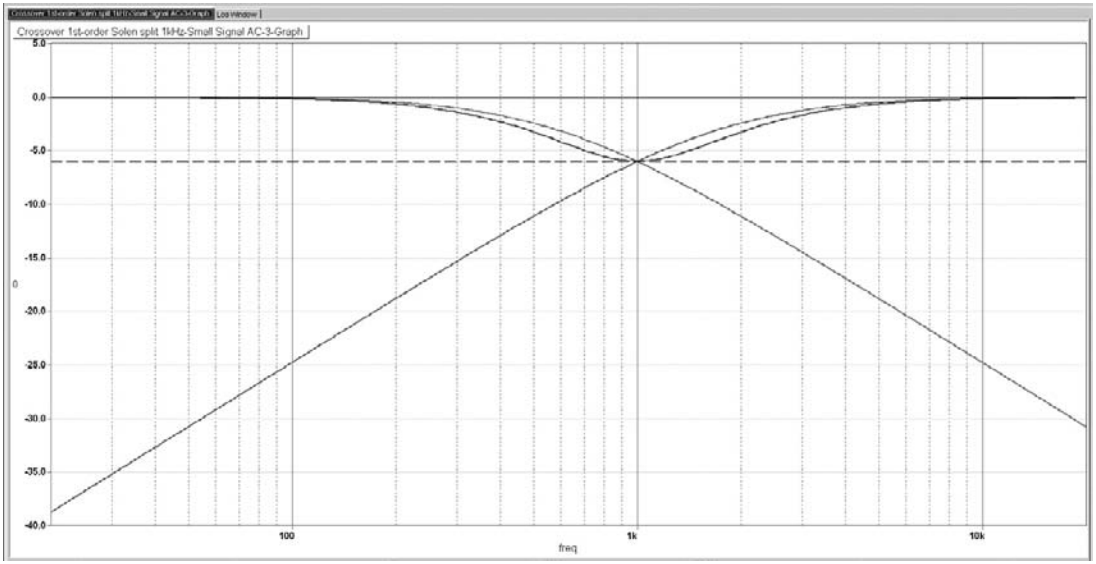


Figure 4.8: Frequency response of Solen split first-order 1 kHz crossover; both filter outputs plus their in-phase sum. Dotted line at −6 dB.

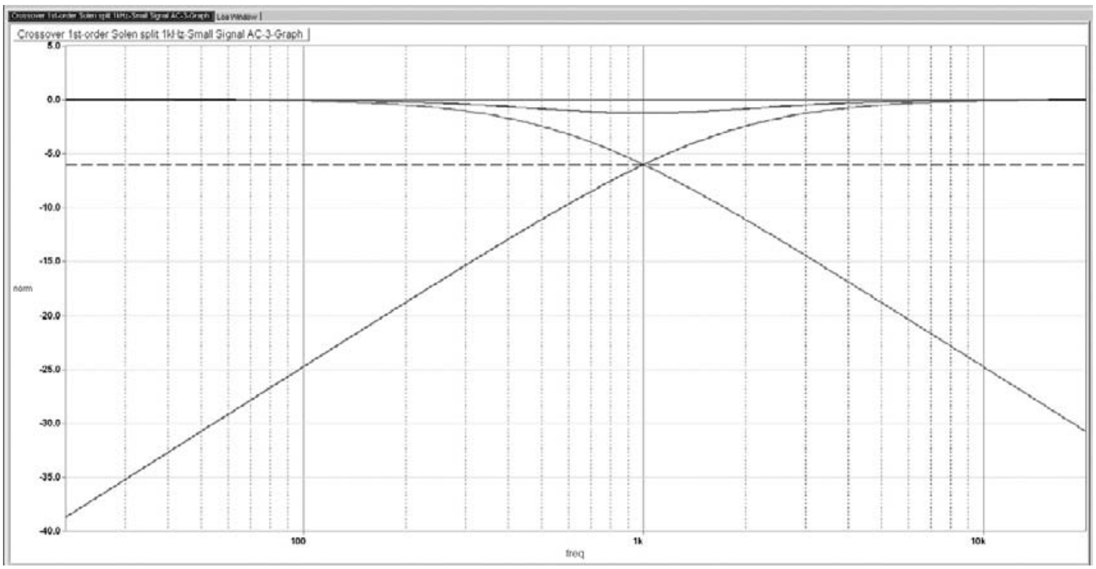


Figure 4.9: Frequency response of Solen split first-order 1 kHz crossover; both filter outputs plus the sum with one output reversed (phase-inverted).

4.6 Second-Order Crossovers

The use of a second-order crossover promises relief from the lobing, tilting, and time alignment criticality of first-order crossovers, because the filter slopes are now twice as steep at 12 dB/octave. A very large number of passive crossovers are second-order, because they are still relatively simple, and this simplicity is very welcome as it reduces power losses and cuts the total cost of the large crossover components required. Neither of these factors applies to active crossovers; the extra power consumption and the extra cost of making a fourth-order crossover rather than a second-order crossover are very small.

All second-order filters have a 180° phase-shift between the two outputs, which causes a deep cancellation notch in the response at the crossover frequency when the HF and LF outputs are summed. Such a response is of no use whatever and the standard cure is to invert the polarity of one of the outputs. With a second-order crossover using Butterworth filters, this gives not a flat response but a +3 dB hump at the crossover frequency. As we shall see, the size of the hump can be much reduced by using a frequency offset; in other words the highpass and lowpass filters are given different cutoff frequencies. An offset factor of 1.30 turns the +3 dB hump into symmetrical amplitude ripples of ±0.45 dB, which is the flattest response that can be achieved for the Butterworth crossover by this method.

Second-order crossovers have much less sensitivity to driver time misalignments because of their 12 dB/octave slopes. Vance Dickason [2] has shown that for a second-order Butterworth crossover, a 1-inch time misalignment gives errors of only fractions of a dB, while a 2-inch misalignment gives maximal errors of 2 dB. The corresponding figures for a first-order crossover are 4 dB and 10 dB respectively. The frequency-offset technique can also be used to reduce the effect of time-alignment errors on the amplitude/frequency response.

A second-order crossover gives better, though by no means stunning, protection of the drive units against inappropriate frequencies, less excitement of unwanted behaviour outside their intended frequency range, and less modulation distortion. Since one output has to be inverted to get a usable amplitude response, the outputs are in phase instead of 180° phase-shifted and so there should be no lobing error, ie tilt in the vertical coverage pattern.

4.6.1 Second-Order Butterworth Crossover

The Second-order Butterworth crossover is perhaps the best-known type, despite the fact that it is far from satisfactory. A classic bit of crossover misdesign that has been published in circuit ideas columns and the like a thousand times is shown in Figure 4.10. You take two second-order Butterworth filters with the same cutoff frequency, one highpass and one lowpass, and there you have your two outputs. As before, the summing device represents how the two outputs add linearly in the air in front of the loudspeaker.

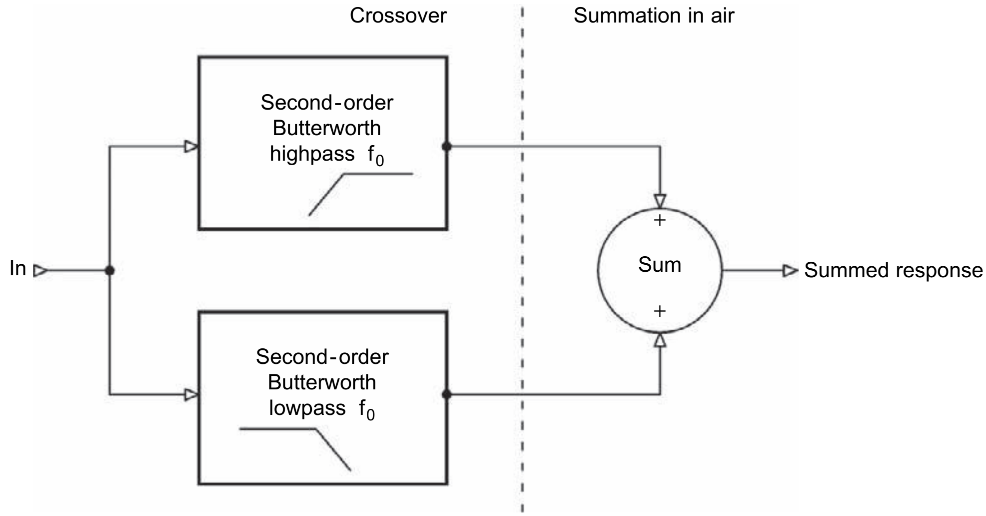


Figure 4.10: A doomed attempt to make a crossover using two second-order Butterworth filters.

As Figure 4.11 shows all too clearly, this does not work well; in fact "catastrophic" would be a more accurate description. Each filter gives a 90° phase-shift at the crossover frequency, one leading and one lagging. The signals being summed are therefore a total of 180° out of phase and cancel out completely. This causes the deep notch at the crossover frequency seen in Figure 4.11.

Since a 180° phase-shift is the root of the problem, that at least can be eliminated by the simple expedient of reversing the connections to one of the drivers, normally the high-frequency one of the pair. Figure 4.12 shows the result—the yawning gulf is transformed into a much less frightening +3 dB hump centred at the crossover frequency.

In a passive crossover this reversed connection can be hidden inside the speaker enclosure along with the crossover components, but in an active crossover the issue is more exposed. You will need to either build a phase-inversion into the active crossover, which again effectively hides the phase-reversal from the user, or specify that one of the power amplifier-speaker cables be reversed. A lot of users are going to feel that there is something not right about such an instruction, and building the inversion into the crossover is strongly recommended.

Clearly our +3 dB hump is much better than an audio grand canyon, but accepting that much deviation from a flat response is clearly not a good foundation for a crossover design. That hump is going to be very audible. It might be cancelled out by an equalisation circuit with a corresponding dip in its response, but there is a simpler approach. Looking at Figure 4.12, it may well occur to you, as it has to many others, that something might be done by pulling apart the two filter cutoff frequencies so the response sags a bit in the

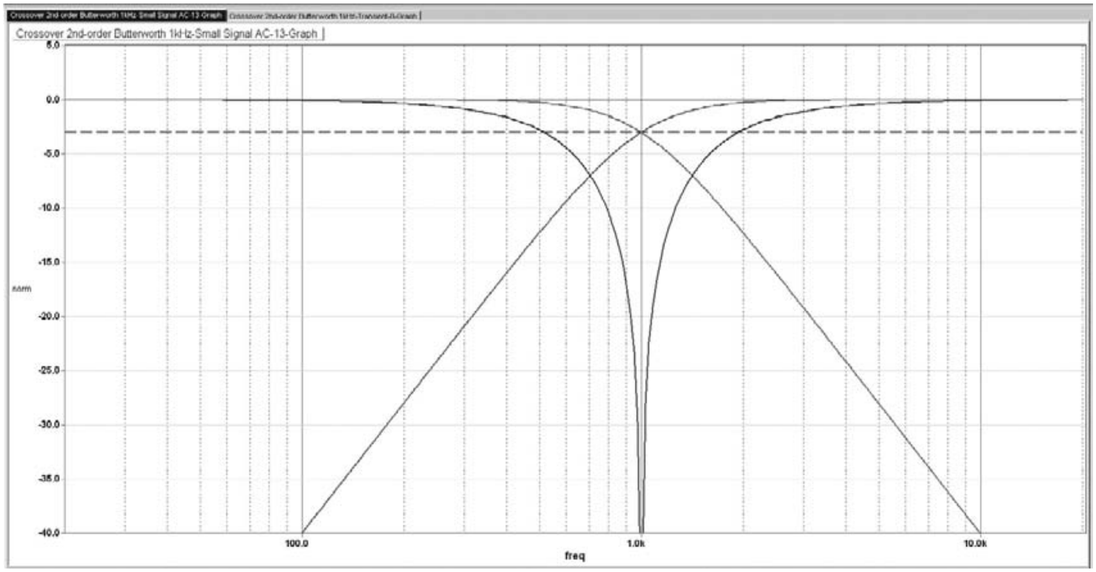


Figure 4.11: The frequency response resulting from the in-phase summation of two second-order Butterworth filters: a disconcerting crevasse in the combined response. The dashed line is at −3 dB.

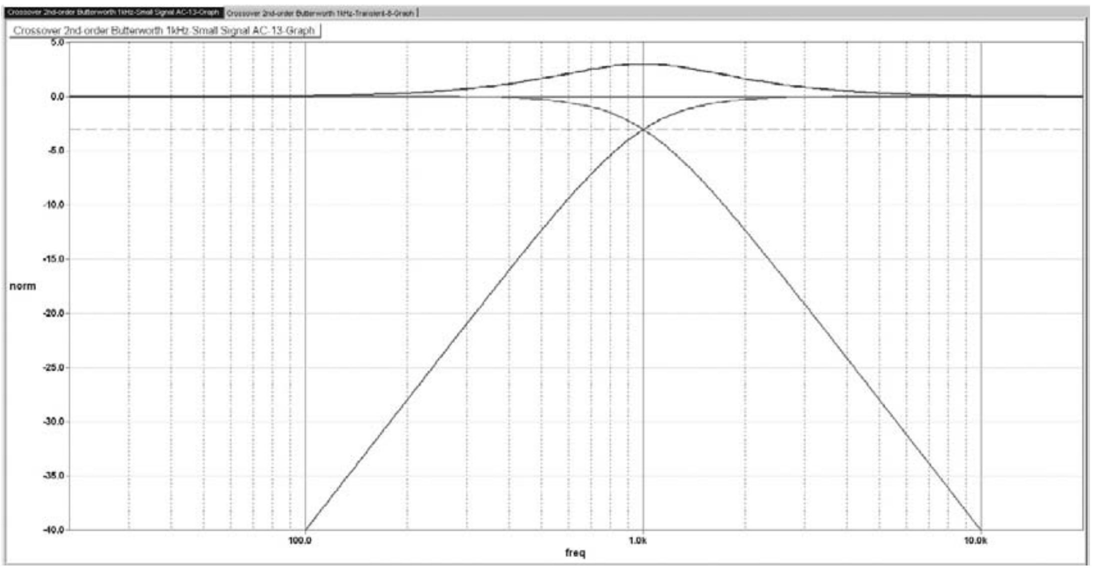


Figure 4.12: Second-order Butterworth crossover, with the phase of one output reversed; the crevasse has become a more usable +3 dB hump. The dashed line is at −3 dB.

middle, as it were. There is no rule in crossover design, be it active or passive, that requires the two halves of the filtering to have the same cutoff frequency.

Figure 4.13 shows the result of offsetting each of the filter cutoff frequencies by a factor of 1.30 times. This means that the highpass filter cutoff frequency is changed from 1.00 kHz to 1.30 kHz, while the lowpass cutoff becomes 1.00/1.30 = 0.769 kHz. Crossover now occurs at −6 dB. With the phase inversion, the offset factor of 1.30 turns the hump into symmetrical amplitude ripples of ±0.45 dB above and below the 0 dB line; this represents the minimum possible response deviation obtainable in this way. Now the amplitude response is looking a good deal more respectable, if not exactly mathematically perfect, and it is very questionable whether response ripples of this size could ever be audible. You may wonder if the frequency-offset process has rescued the response with outputs in-phase; the answer is that it is not much better. There is no longer a notch with theoretically infinite depth, but there is a great big dip 9 dB deep at the bottom, and such a response is still of no use at all.

The frequency offsets required for maximal flatness with various types of crossover are summarised in Table 4.1 at the end of this chapter.

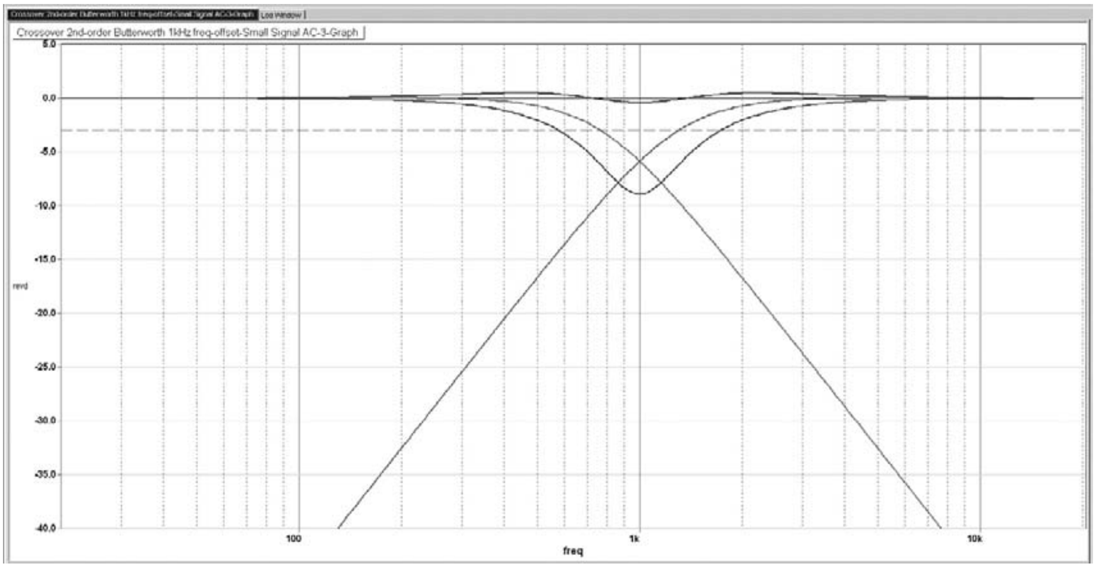


Figure 4.13: Two second-order Butterworth filters with 1.30 times frequency offset, in normal and reversed connection. With the phase of one output reversed; the +3 dB hump is smoothed out to a mere ±0.45 dB ripple. The normal-phase connection still has a serious dip 9 dB deep. The dashed line is at −3 dB; the two filter cutoff (−3 dB) frequencies are now different.

The power response for the second-order Butterworth crossover with no frequency offset is shown in Figure 4.14; it is a perfect straight line. This qualifies it as a Constant Power Crossover (CPC); it is a pity that the power response is much less important than the pressure response. The power response is the same whether or not one of the outputs is phase-reversed. When the two outputs are squared and added to get the total power, the negative sign of the reversed output disappears in the squaring process.

When looking at this power response plot, it is important to appreciate that each crossover output is still shown at −3 dB at the crossover frequency, meaning the power output from it is halved. The two lots of half-power sum to unity, in other words 0 dB.

Earlier we saw that a frequency offset of 1.30 times was required to get near-flat amplitude response. How is that going to affect the perfectly flat power response of Figure 4.14? The answer, predictably, is that any change is going to be for the worse, and Figure 4.15 shows that there is now an 8 dB dip in the power response.

As mentioned earlier, second-order filters have a 180° phase-shift between the two outputs. This is shown in Figure 4.16, where the phase of the sum lies exactly on top of the trace for the lowpass output. The phase of the sum is that of a first-order allpass filter, the same phase characteristic as that of the first-order crossover. This is inaudible with normal music

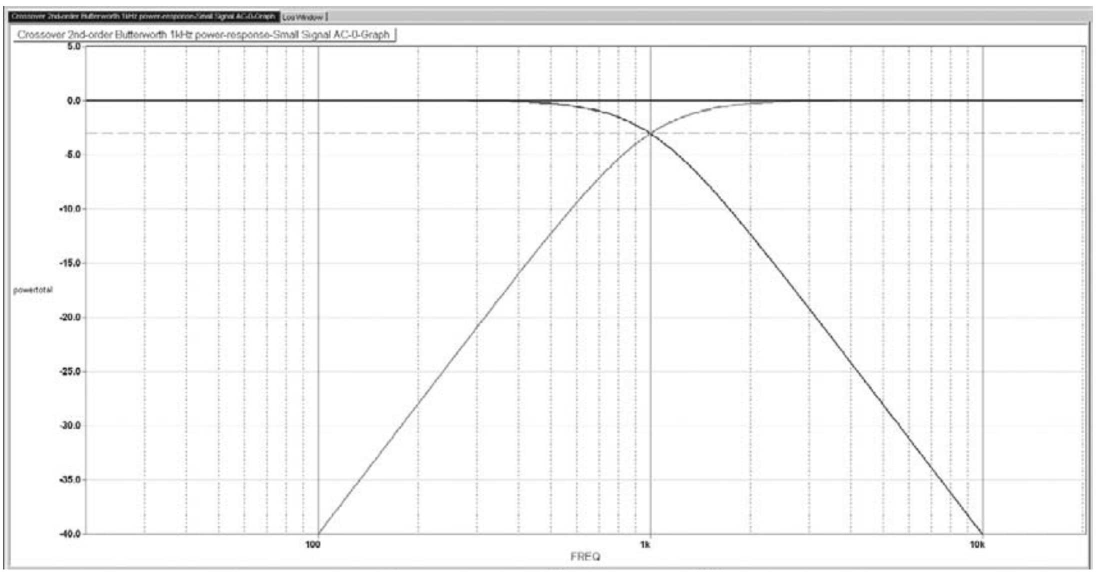


Figure 4.14: Power response of second-order Butterworth crossover with no frequency offset—the sum is perfectly flat at 0 dB. Each crossover output is still at −3 dB at the crossover frequency, meaning the power is halved. The two half-powers sum to 0 dB. Outputs in phase.

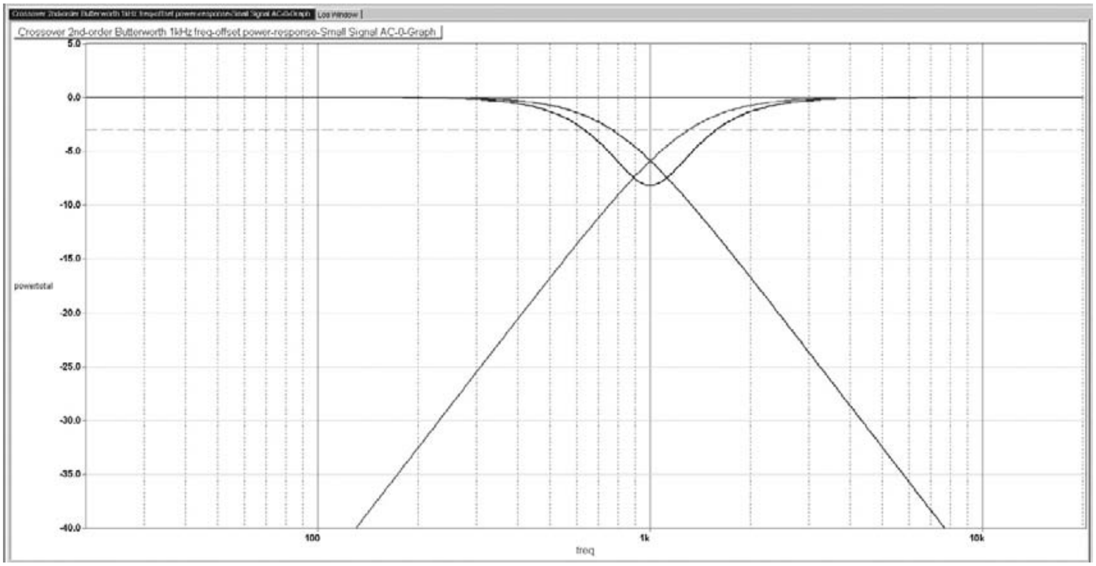


Figure 4.15: Power amplitude response of second-order Butterworth crossover with 1.30x frequency offset—rather less than perfect, showing a −8 dB dip at the crossover frequency.

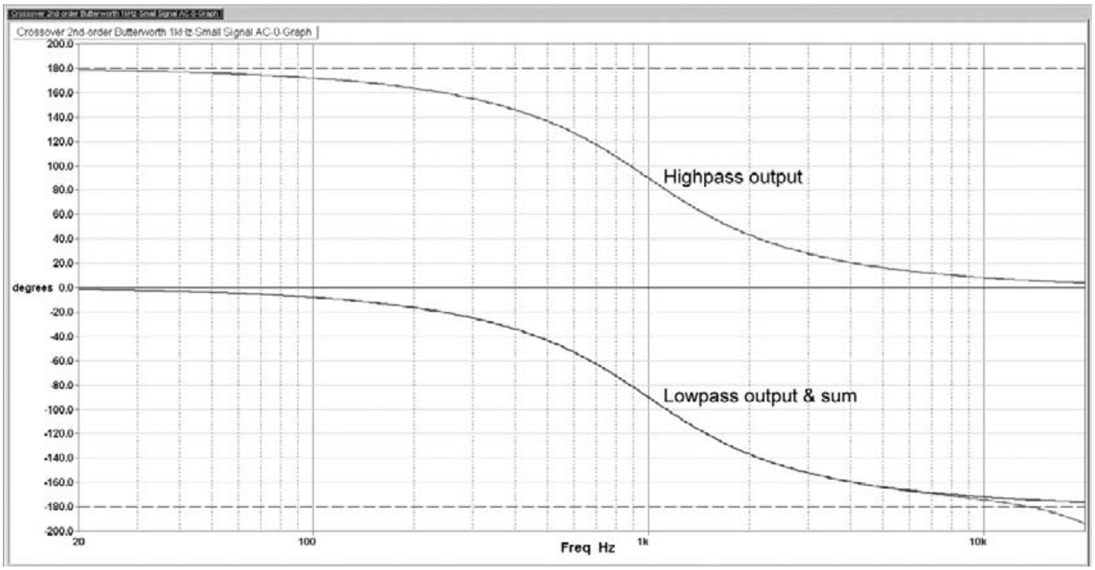


Figure 4.16: Phase response of second-order Butterworth crossover with one output inverted; the outputs are always 180° out of phase. The phase of the sum lies exactly on top of that of the lowpass output. Highpass output shown *before* inversion.

signals. For reasons of space, phase and group delay plots are from here on only given for the more interesting crossover types.

The summed group delay for the inverted connection is shown in Figure 4.17. It has a level section at 226 usec, and a gentle peak of 275 usec just below the crossover frequency. Note that the level section shows less group delay than the first-order crossover.

When we were looking at first-order crossovers, you will recall that it was said that no other crossover could solve the waveform reconstruction problem, that is, to get out the waveform that we put in after summing the outputs. How does a second-order Butterworth cope with the square-wave reconstruction problem?

The answer from Figure 4.18 is that it fails completely, and inverting one of the filter outputs makes thing even worse. The phase-shifts introduced by the second-order filters make it impossible for reconstruction to occur. While it is not very obvious from Figure 4.18, the first and second cycles of the simulation are not quite identical; since we have a circuit with energy storage elements (capacitors) and we are starting from scratch, it takes a little time for things to settle down so you obtain the result for continuous operation. In some cases it is not uncommon for 20 cycles to be required to reach equilibrium. Not every writer on the subject of audio has appreciated this fact, and major embarrassment has resulted.

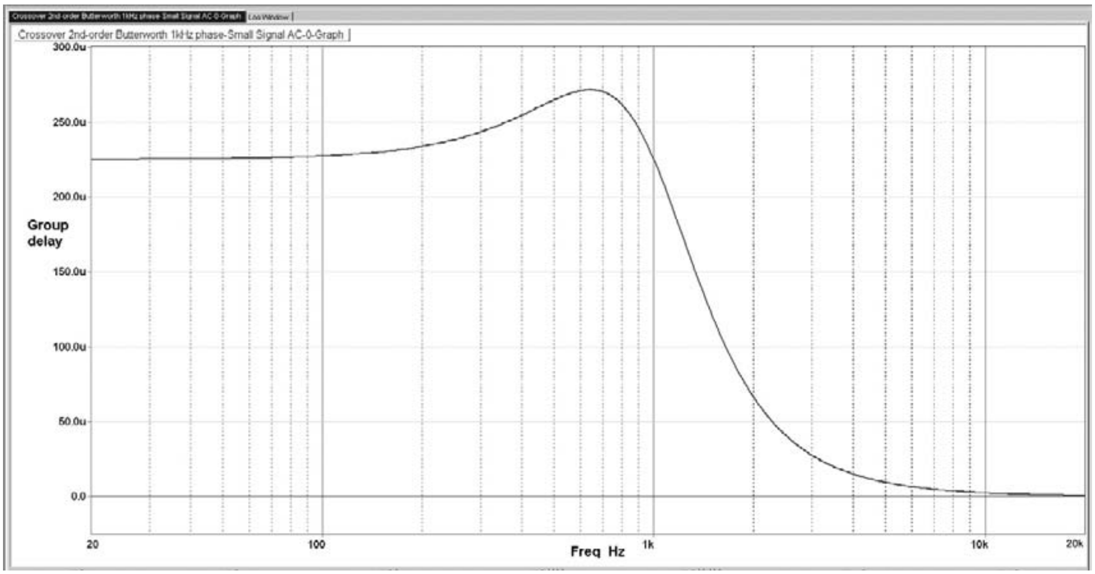


Figure 4.17: The group delay response of a second-order Butterworth crossover, one output inverted.

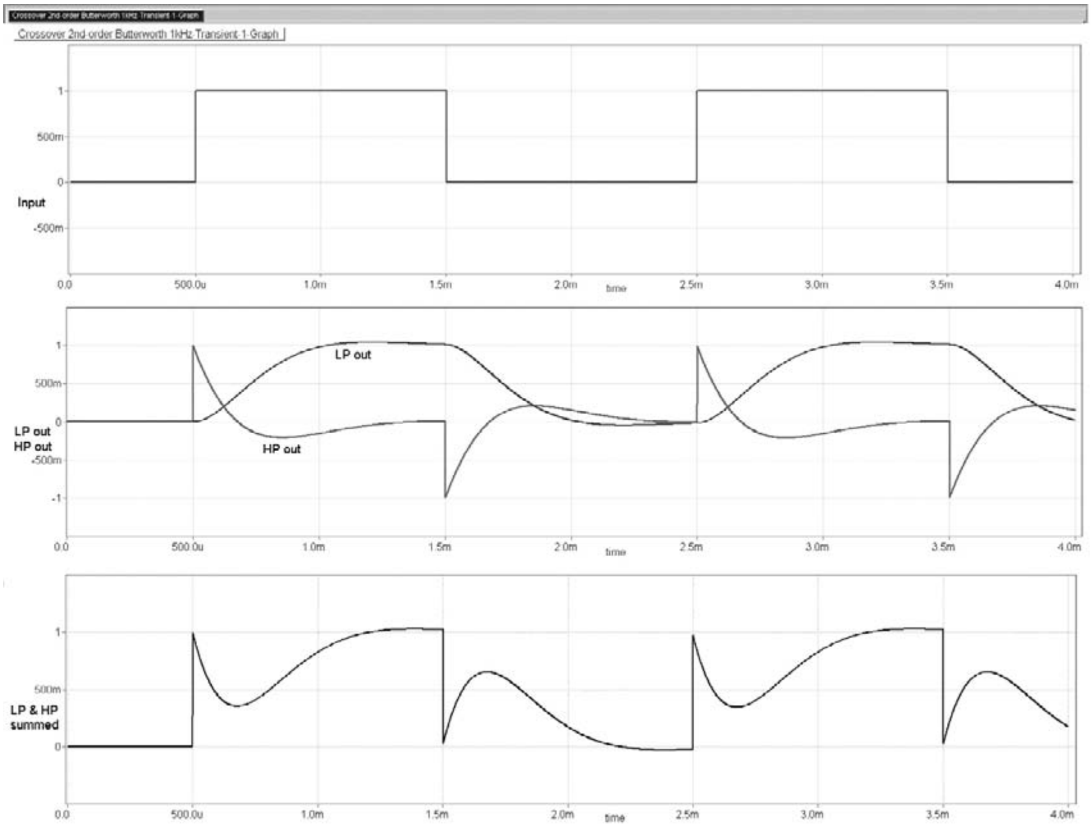


Figure 4.18: Attempted reconstruction of a square-wave by a second-order Butterworth crossover without offset. Total failure!

The second-order Butterworth crossover does not sum to flat even with one output inverted, though frequency offsetting can give a big improvement. The 12 dB/octave slopes are usually considered inadequate.

4.6.2 Second-Order Linkwitz–Riley Crossover

The Butterworth crossover filter can be made very nearly flat by tweaking the cutoff frequencies of the two filters. An alternative and much better approach in the second-order case is to alter the Qs of the filters. Setting the Q of each filter to 0.5, with identical cutoff frequencies, turns the second-order Butterworth crossover into a second-order Linkwitz–Riley crossover with each output −6 dB at the crossover point, as seen in Figure 4.19. The flat response qualifies it as an allpass crossover network.

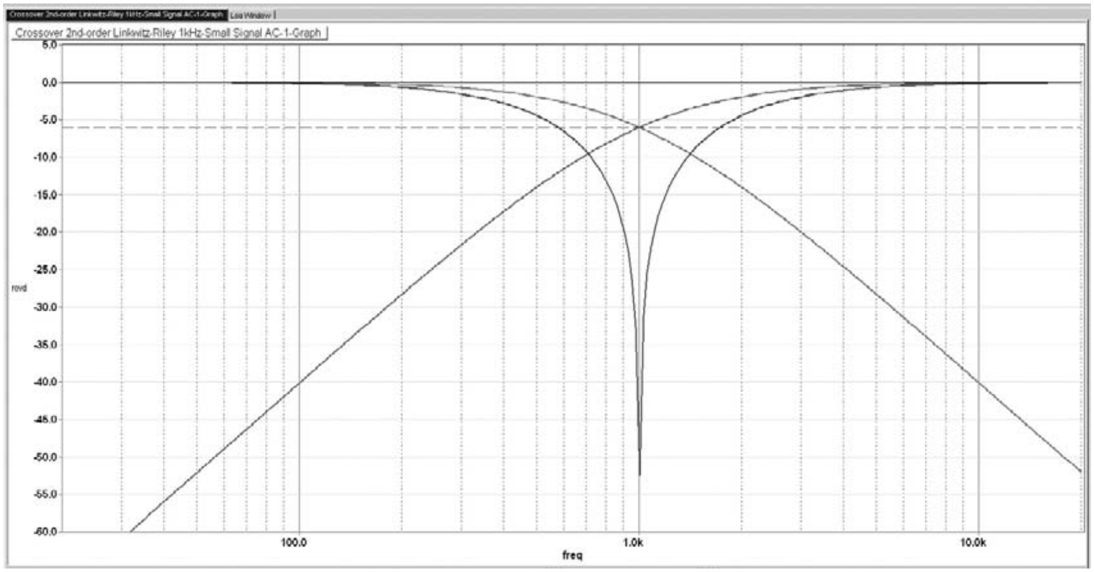


Figure 4.19: The frequency response of a second-order Linkwitz–Riley crossover. The in-phase summation of two filters still has a cancellation crevasse, but with one output reversed the sum is exactly flat at 0 dB. The dashed line is at the −6 dB crossover level.

The two outputs are still 180° out of phase for the normal connection, and give the same yawning gulf in the response as the Butterworth. With the reversed connection the signals add as for the Butterworth case, but they are now 3 dB lower, so there is no hump. The summation gives a completely flat response, without the ripple you get with a frequency-offset second-order Butterworth crossover.

The reversed connection has a −3 dB dip in the power response at the crossover frequency. In this respect it is not as good as the first-order crossover, which has a flat power response as well as a flat voltage (or SPL) response. However, as we have seen, the power response of a crossover is usually a minor consideration.

The second-order Linkwitz–Riley crossover is neither linear phase nor minimum phase. The phase-response plot looks indistinguishable from that of the second-order Butterworth crossover, though there are minor differences. The summed group delay does not peak but rolls off slowly around the crossover frequency.

4.6.3 Second-Order Bessel Crossover

The Bessel filter has a much slower rolloff than the Butterworth, but also has a maximally flat group delay; in other words it stays flat as long as possible before it rolls off, while the

Butterworth group delay has a peak in it. This makes the Bessel filter an interesting possibility for crossovers with flat group delay characteristics.

The second-order Bessel crossover without any frequency offset gives a −8 dB dip for the in-phase connection, as in Figure 4.20, but a more promising broad +2.7 dB hump with one output reversed, as in Figure 4.21. This looks very like the Butterworth hump in Figure 4.9, so it seems very likely we can also reduce this one by applying frequency offset to the filter cutoffs.

The first attempt, a frequency offset of 1.30 times, as used in the second-order Butterworth case, reduces the size of the hump to +1.1 dB at its centre, but this time no dips below the 0 dB line have appeared. This differing behaviour looks ominous, and suggests that the hump may get flatter and flatter with increasing offset, but never actually reach a definite maximally flat condition.

However, this is a great example of a situation where you should not give up too soon. If we keep increasing the offset, then the shape of the summed response changes, until at an offset ratio of 1.45 we obtain a dip and two flanking peaks, as in Figure 4.22. The deviation from 0 dB is less than ±0.07 dB, so this result is actually much better than the second-order Butterworth crossover, which at maximal flatness had deviations of ±0.45 dB. Such small deviations as ±0.07 dB will be utterly lost in drive-unit tolerances. This offset ratio gives crossover at −6.0 dB.

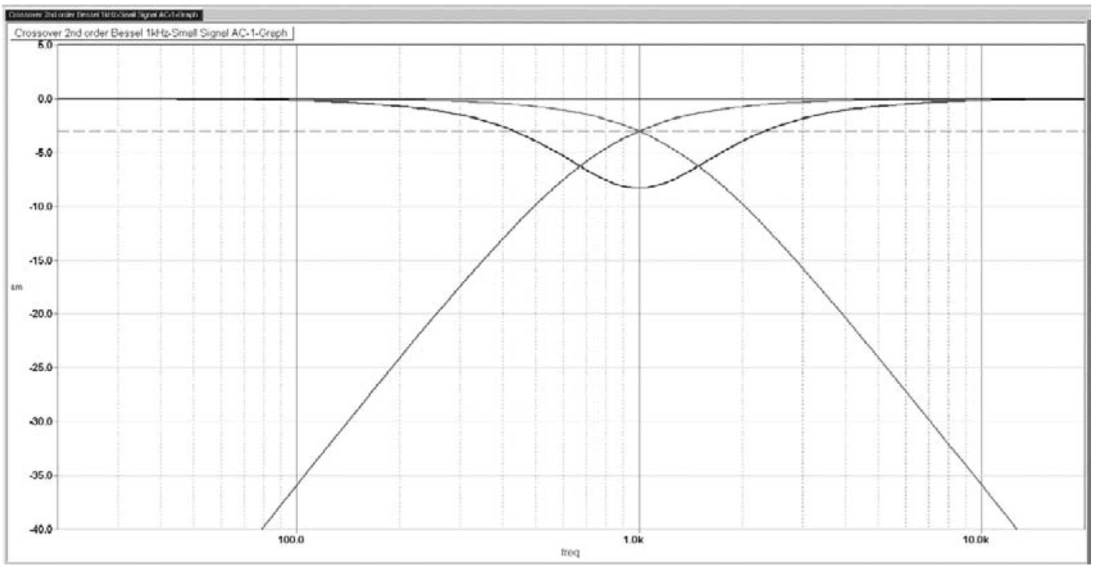


Figure 4.20: The frequency response of a second-order Bessel crossover summed in-phase has a dip going down to −8 dB. The dashed line is at −3 dB.

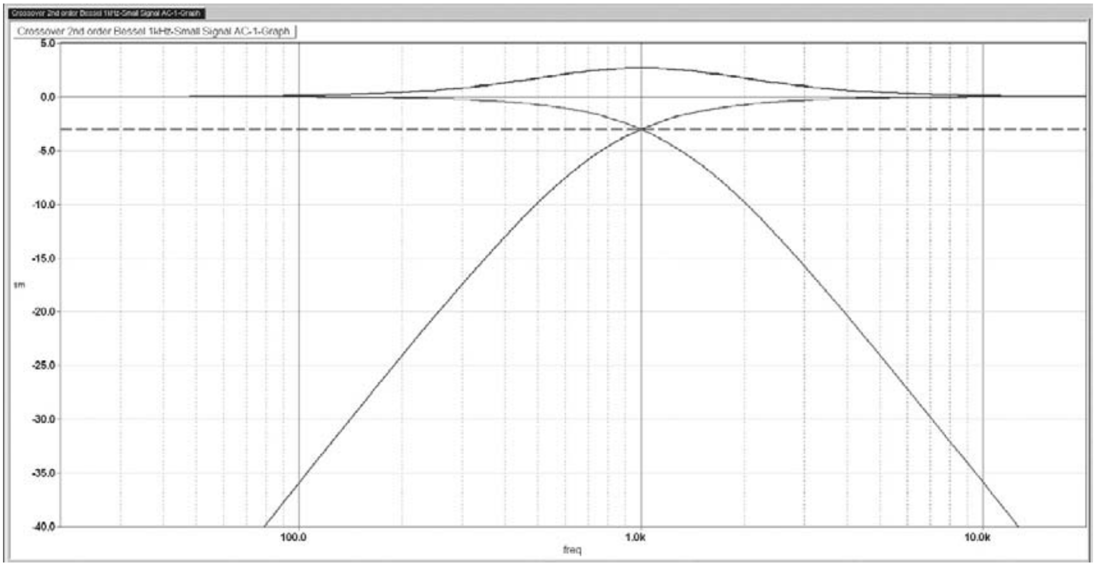


Figure 4.21: The frequency response of a second-order Bessel crossover summed with one output phase-reversed has a +2.5 dB hump. The dashed line is at −3 dB.

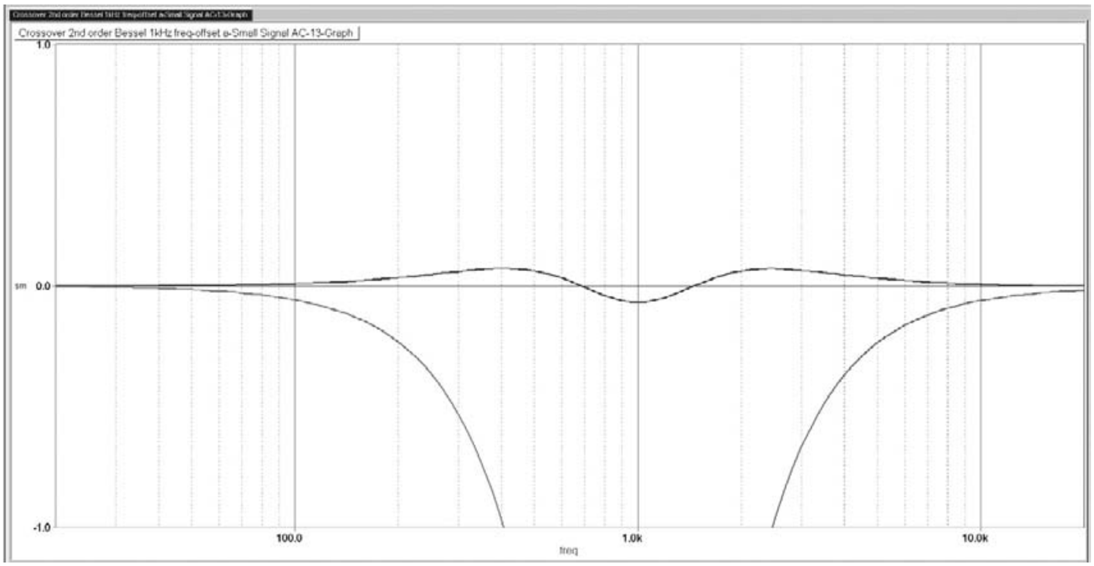


Figure 4.22: The frequency response of a second-order Bessel crossover summed with one output phase-reversed and a frequency offset of 1.449 times has deviations from the 0 dB line of less than 0.07 dB. Note much enlarged vertical scale covering only ±1 dB.

The phase response plot looks very similar to that of the second-order Butterworth crossover, but the rate of change around the crossover region is slightly slower due to the lower Q of the filters. The summed group delay does not peak but rolls off slightly more slowly than the Butterworth around the crossover frequency.

The second-order Bessel is not linear phase, though it deviates from it less than do the second-order Butterworth or Linkwitz-Riley types. It is not minimum phase.

A very good discussion of Bessel crossovers is given in [4].

4.6.4 Second-Order 1.0 dB-Chebyshev Crossover

All Chebyshev filters have ripples in their passband response, and given the problems we have had achieving a near-flat response when we were using filters without such ripples, things don't look too hopeful. Using 1.0 dB-Chebyshev filters, which in second-order form peak by 1 dB just before rolloff, we get Figure 4.23, which shows the in-phase result. There is a deep central dip rather like that of the second-order Bessel crossover, except that this one is even deeper at −14 dB.

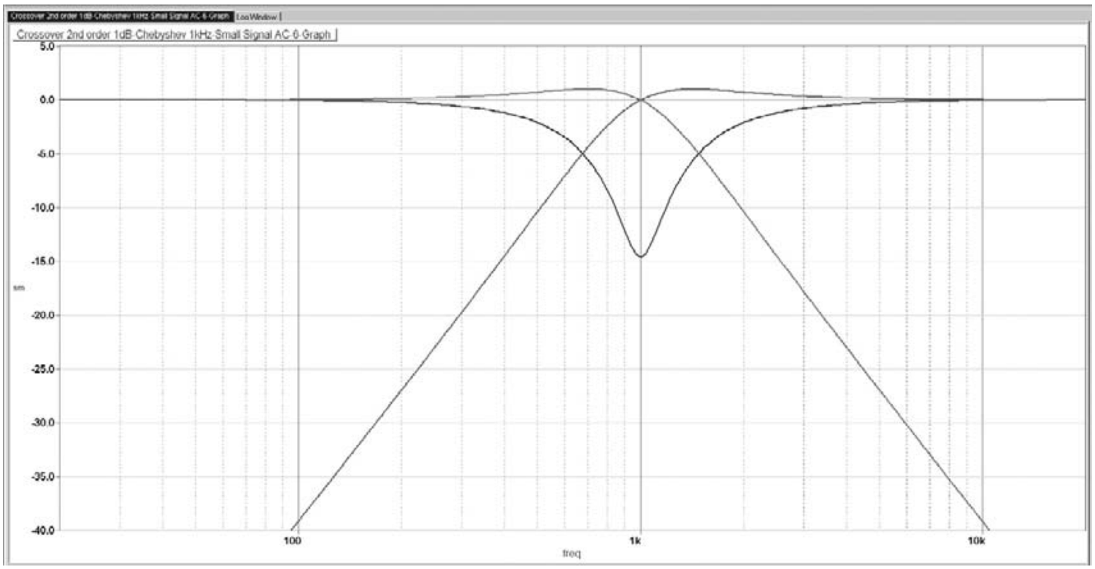


Figure 4.23: The frequency response of a second-order 1.0 dB-Chebyshev crossover. In-phase summation gives a deep dip of −14 dB. The crossover is at 0 dB, because the cutoff frequency of this filter is defined as the point in the rolloff where the response returns to 0 dB after the 1.0 dB peak.

Reversing the phase of one of the outputs gives us a 6 dB hump, substantially higher than that of the reversed-phase second-order Bessel crossover; see Figure 4.24. Flattening this by using frequency offset is a tall order, but we will have a go.

As the frequency offset ratio is increased, a dip develops in the centre of the hump and moves below the 0 dB line, until we reach the optimally flat condition, which unfortunately is not that flat. The deviations are ±1.6 dB at an offset ratio of 1.53 times, and look too big to be a basis for sound crossover design; see Figure 4.25. The large 1.53 times offset ratio causes the crossover point to be at −5.6 dB.

The power response has significantly more ripple than the optimally flat amplitude response.

The phase response plot looks very similar to that of the second-order Butterworth crossover, but changes faster around the crossover frequency. The summed group delay has a very big peak just below the crossover frequency; while it is probably not audible, it is certainly not desirable.

It would be possible to try other types of Chebyshev filters as second-order crossovers; Chapter 7 gives details on how to design Chebyshev filters with 0.5 dB, 1 dB, 2 dB, and 3 dB of passband ripple. There seems to be no reason to think that the versions with greater passband ripple would be any better than the 1.0 dB version, and every reason to think that

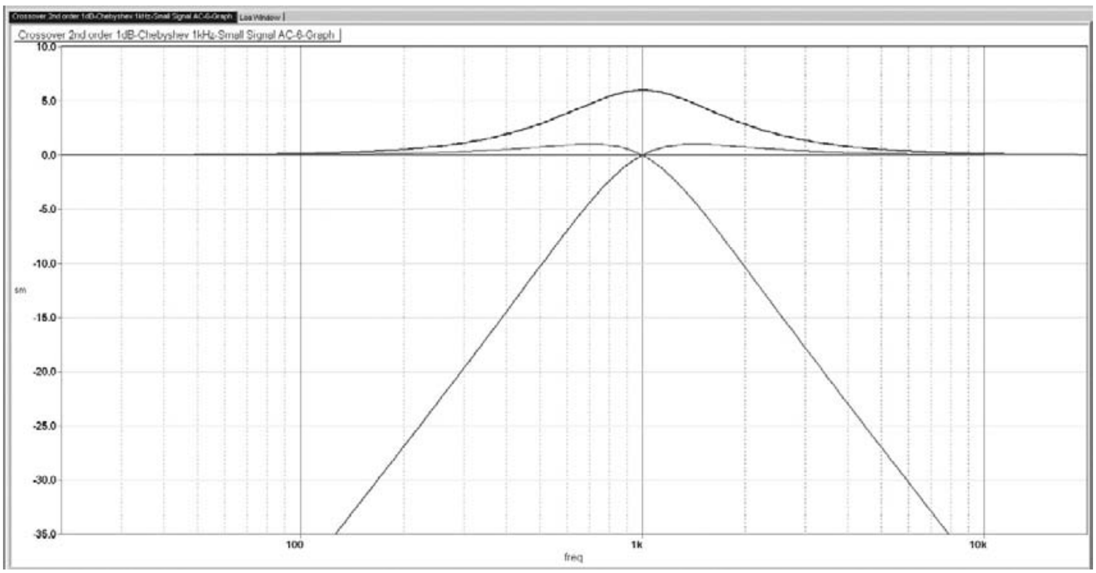


Figure 4.24: The frequency response of a second-order 1.0 dB-Chebyshev crossover. Summation with one output phase-reversed gives a peak of +6 dB. Vertical scale has been moved up by 5 dB to accommodate the peak.

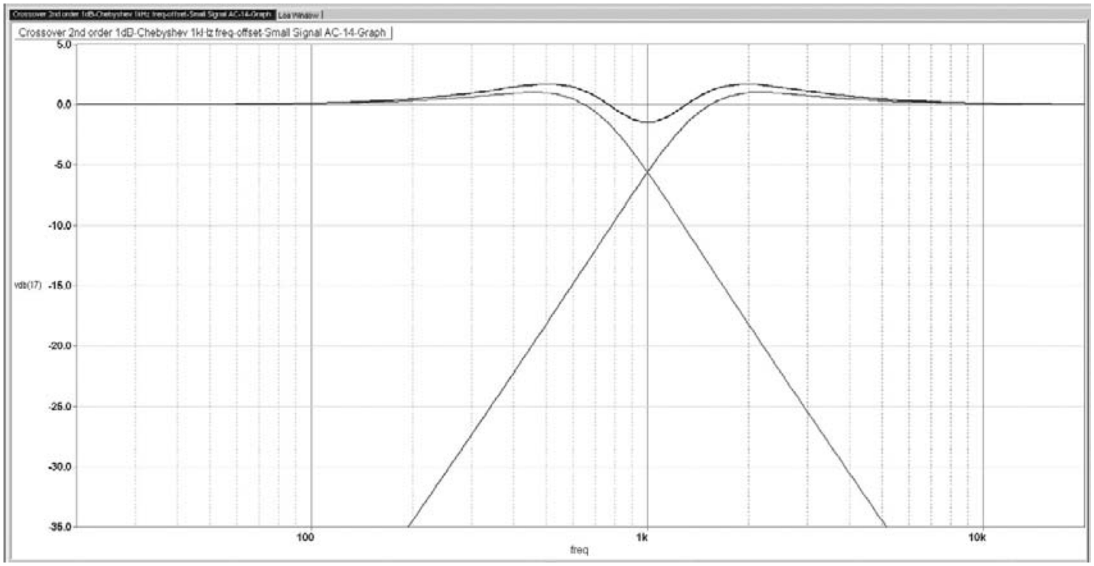


Figure 4.25: The frequency response of a second-order 1.0 dB-Chebyshev crossover with one output phase-reversed, and a frequency offset of 1.53 times. The deviation is ±1.6 dB about the 0 dB line.

they would be worse. The high filter Qs required to realise the 2 dB and 3 dB filters imply very poor group delay characteristics with serious peaking.

4.7 Third-Order Crossovers

Third-order crossovers have the advantage of greater separation between the drive units than either first- or second-order designs can give. The steeper 18 dB/octave slopes make it less likely that drive unit irregularities such as the tweeter resonance will be excited, and allow modulation distortion to be reduced. Third-order crossovers are also less sensitive to driver time-delay misalignments because there is less frequency overlap in the filter outputs. The third-order Butterworth crossover gives a flat amplitude response *and* a flat power response.

On the other side of the ledger, the filter outputs are always 270° apart in phase, which can result in lobing and tilting of the coverage pattern in the range where the drivers overlap, but this range is narrower than for second-order crossovers. Inverting one output reduces this to 90°, and as with the first-order crossover, there is a −15° downward tilt in the crossover region.

Third-order crossovers have further reduced sensitivity to driver time misalignments because of their steeper 18 dB/octave slopes. According to Vance Dickason [2], a third-order

Butterworth crossover and a 2-inch time misalignment gives a maximal error of 2.5 dB, while the corresponding error for a first-order crossover is 10 dB. As before, the frequency-offset technique can also be used to reduce the effect of time-alignment errors on the amplitude response.

Third-order crossovers are generally the most complicated passive types that are popular, though fourth-order and fifth-order passive crossovers have been used.

4.7.1 Third-Order Butterworth Crossover

Figure 4.26 shows the amplitude response of a third-order Butterworth crossover with outputs in phase and with no frequency offset. The response is ruler-flat, and it remains ruler-flat when one of the outputs is reversed. The power response is also flat, as it was for the first-order crossover; all odd-order Butterworth crossovers have a flat power response. It is thus both an APC and a CPC crossover; it is not linear phase or minimum phase. The crossover point is at −3 dB.

The flat power response is illustrated in Figure 4.27.

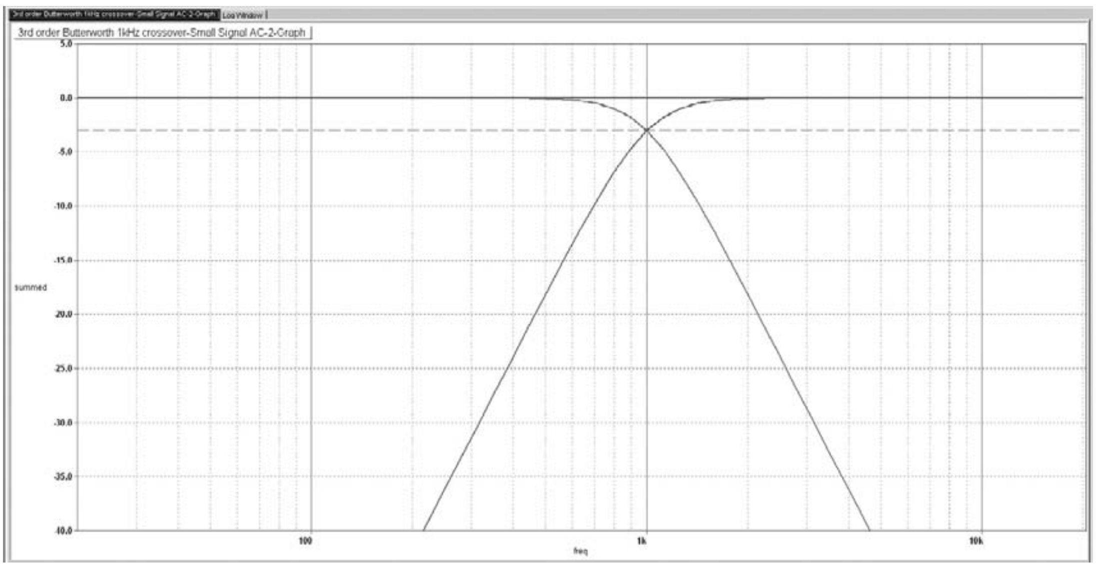


Figure 4.26: The frequency response of a third-order Butterworth crossover. In-phase or phase-reversed summation of the two outputs gives a flat line at 0 dB. The dashed line is at the −3 dB crossover level.

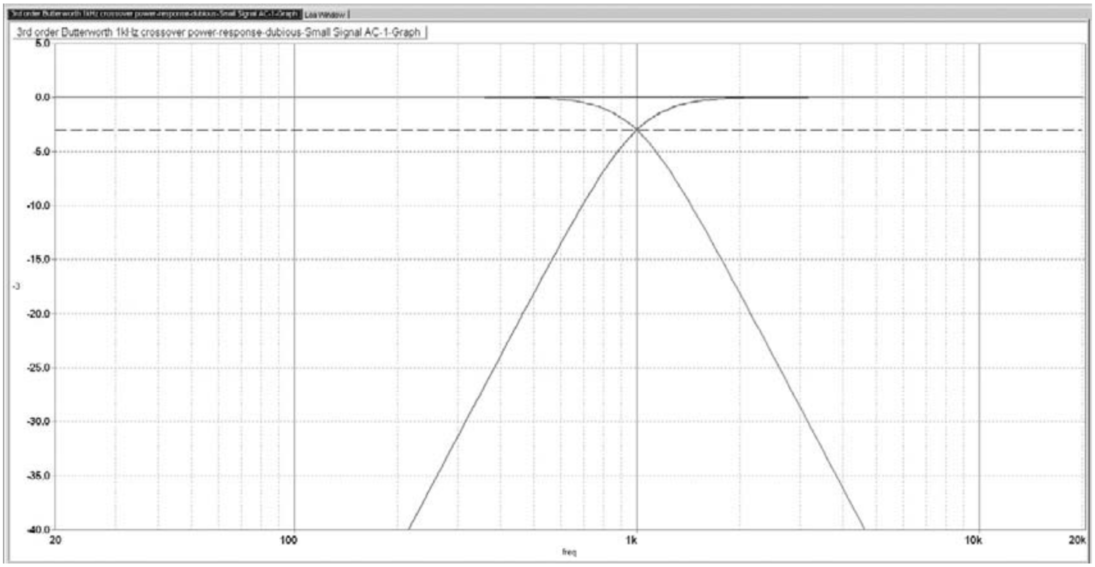


Figure 4.27: The power response of a third-order Butterworth crossover. The RMS summation of the two filter outputs gives a straight line at 0 dB. The dashed line is at the −3 dB level.

The summed phase response of the normal (non-inverted) connection is that of a second-order allpass filter, with the phase changing by 360° over the audio band. If the polarity of one output is inverted, the amplitude response remains perfectly flat but the summed phase response is improved to that of a first-order allpass filter, with the phase changing by only 180° over the audio band. This is shown in Figure 4.28

With outputs normal (not inverted) their phase difference is a constant 270°, but with one output inverted this is reduced to 90°, the same as the first-order crossover. It therefore tilts the polar pattern in the same way, but the frequency range over which this is significant is much reduced because of the steeper slopes of a third-order crossover.

Figure 4.29 shows the how the faster phase changes for the normal connection cause a big peak in the group delay. The inverted connection has no peak and looks much more satisfactory. Note that the group delay at low frequencies is 318 usec, exactly the same as for the first-order crossover.

The third-order Butterworth crossover is one of the better ones. It has flat amplitude and power responses, and a first-order phase response when one output is inverted. On the downside, it has the lobe-tilting problem of all odd-order crossovers, and many people feel that the 18 dB/octave slopes are not really steep enough.

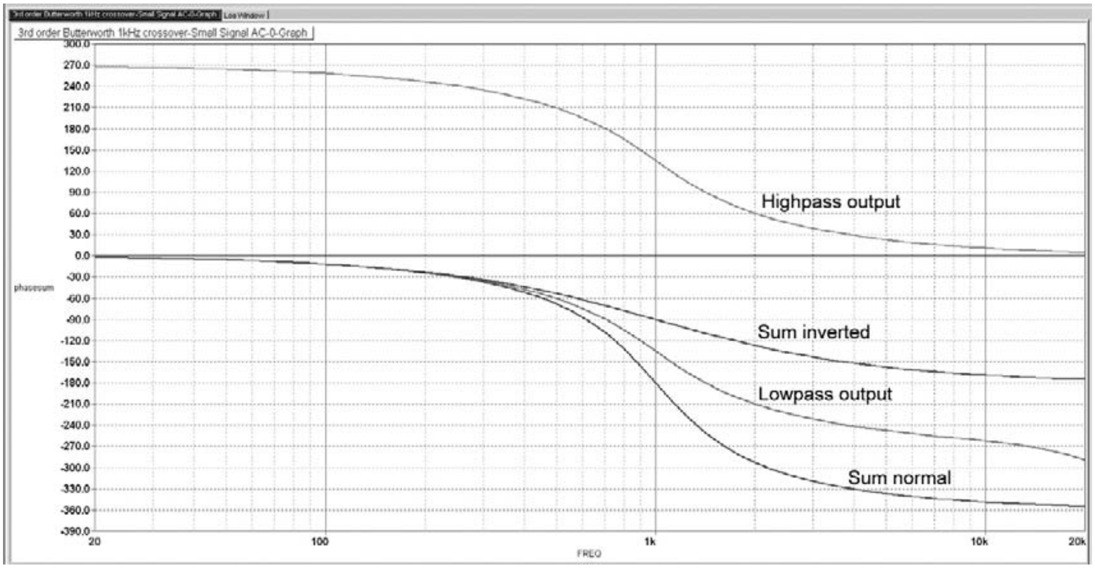


Figure 4.28: The phase response of a third-order Butterworth crossover, for normal and inverted connections.

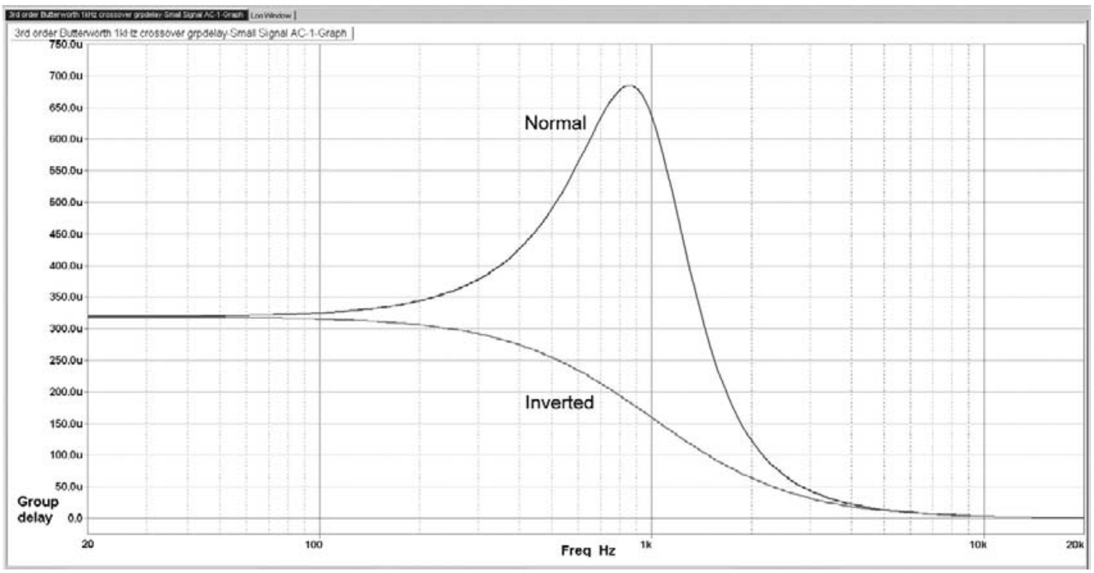


Figure 4.29: The group delay response of a third-order Butterworth crossover. The normal connection has a big peak in delay, but with one output inverted there is just a roll-off.

4.7.2 Third-Order Linkwitz–Riley Crossover

The response of the third-order Linkwitz–Riley crossover alignment is shown in Figure 4.30. The summed amplitude response has a dip of exactly −3 dB at the crossover frequency, and each filter output is at −6 dB. In this case, reversing the phase of one output does not convert it into a hump, but gives exactly the same response with the same dip, because of the differing phase relationships. As we have seen, a hump in the response can be flattened by using a frequency offset that lowers the cutoff frequency of the lowpass filter, and raises the cutoff frequency of the highpass filter. Clearly a dip cannot be dealt with like that, as pulling apart the crossover frequencies, as we have done before, is just going to make the dip deeper.

Instead, if we take the phase-reversed case and push the two cutoff frequencies together, by using an offset ratio of 0.872 times, so the HP cutoff is now 0.872 kHz and the LP cutoff is 1.15 kHz, we get the maximally flat result with only ±0.33 dB of ripple. See Figures 4.31 and 4.32. The crossover point is now at –4.5 dB. This procedure only works with one phase reversed. If applied to the in-phase case we just dig a deeper hole for ourselves; the dip becoming −4.4 dB deep at the bottom.

There is no point in inverting one output as regards the amplitude response (assuming no offset is applied), because it is unchanged. We do, however, get a much better phase

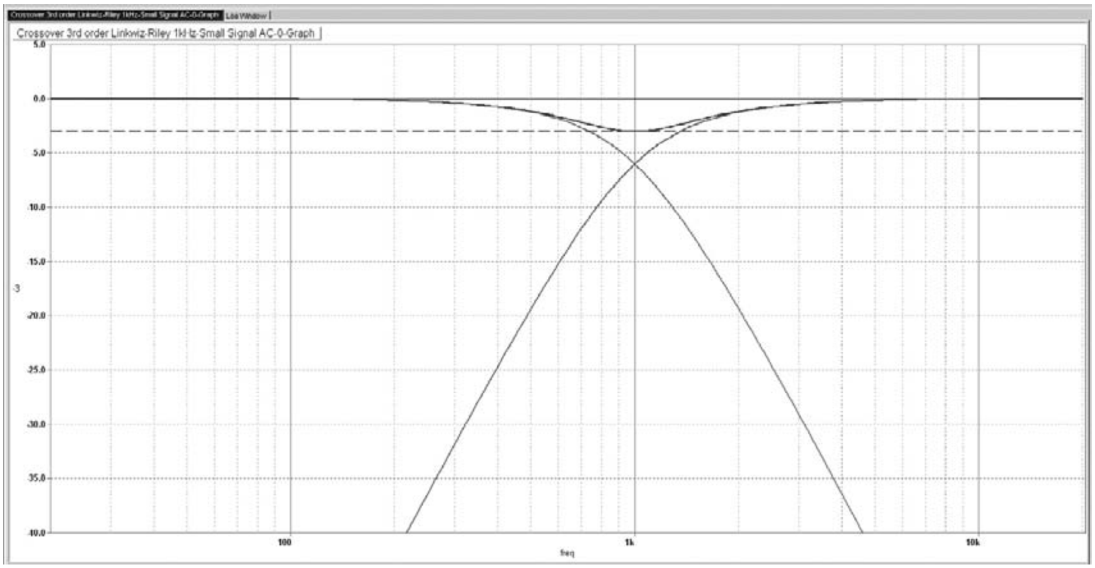


Figure 4.30: The frequency response of a third-order Linkwitz-Riley crossover. The in-phase summation of the two filter outputs gives a −3 dB dip. The dashed line is at −3 dB.

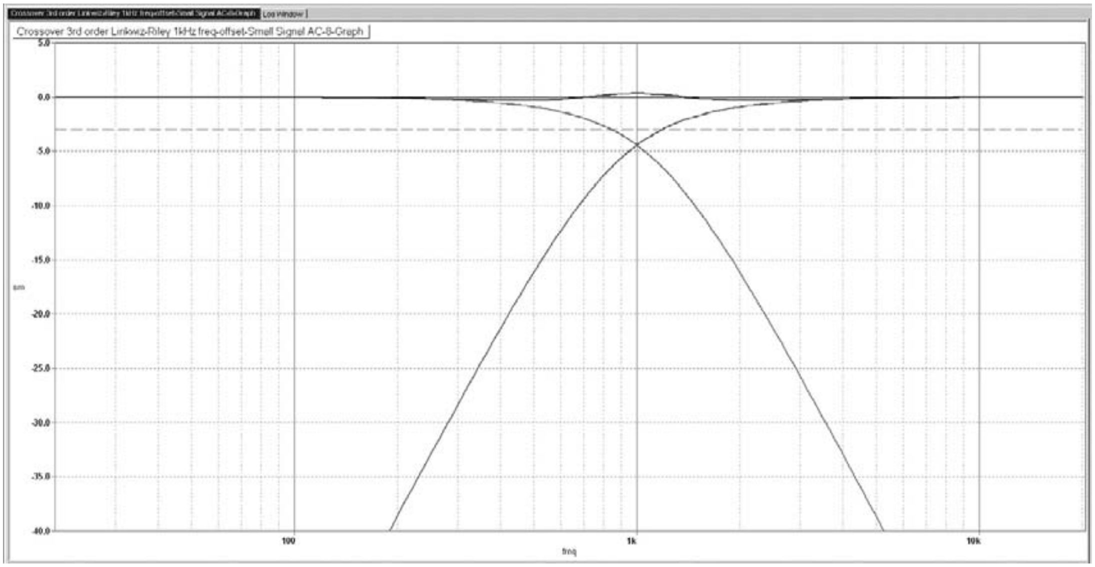


Figure 4.31: The frequency response of a third-order Linkwitz–Riley crossover, with frequency offset of 0.872 times. The reversed-phase summation of the two filter outputs now gives a ripple of only ±0.33 dB.

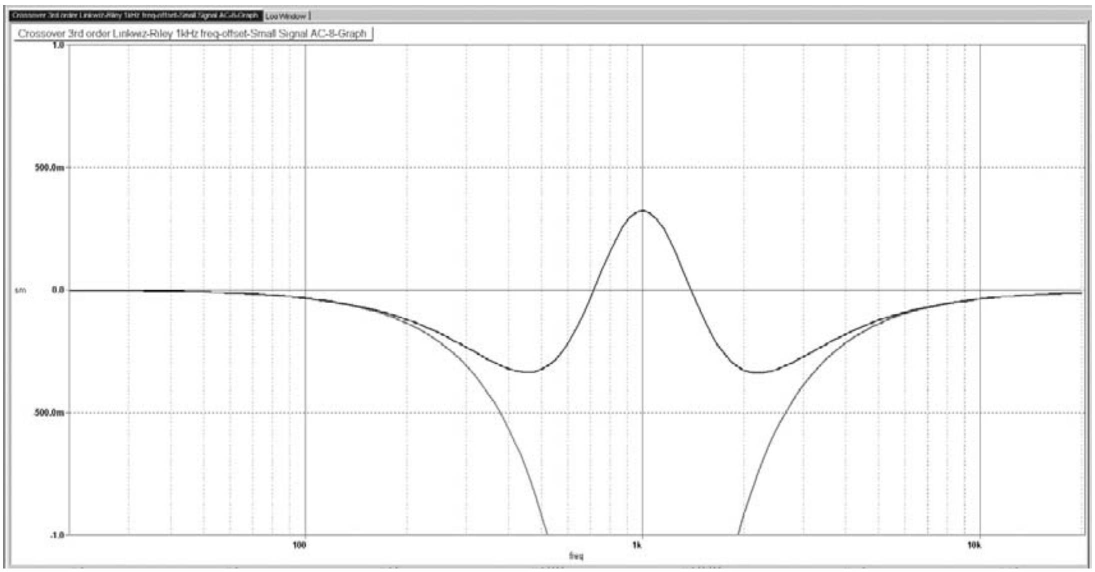


Figure 4.32: Zooming in on the frequency response of a third-order Linkwitz–Riley crossover, with frequency offset of 0.872 times and reversed-phase summation. Ripple is only ±0.33 dB.

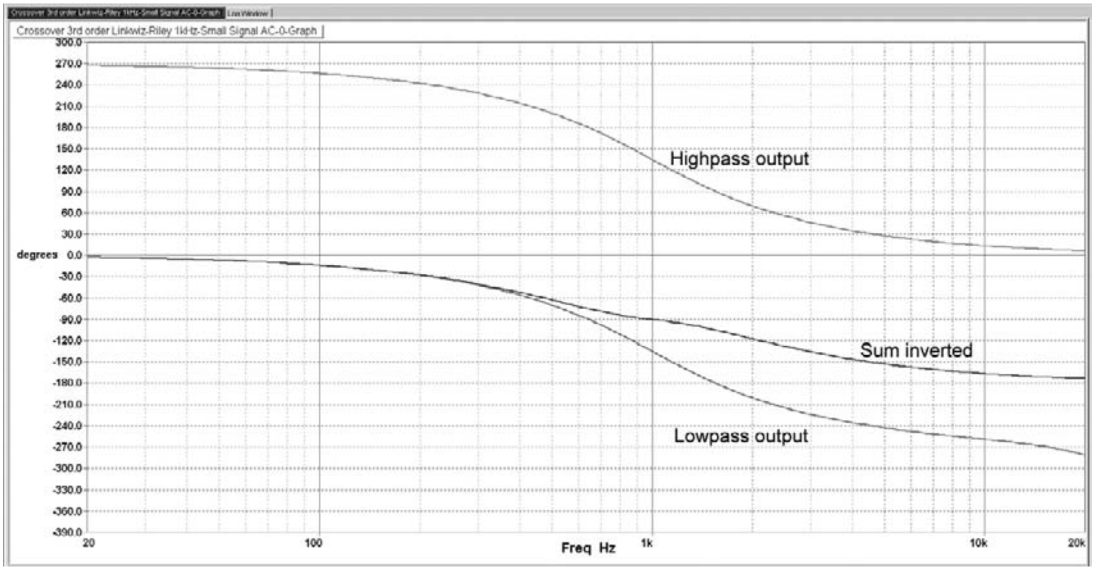


Figure 4.33: The phase response of a third-order Linkwitz–Riley crossover, for inverted connection only.

response, as for the third-order Butterworth. A second-order allpass phase response is converted to a first-order allpass phase response. The latter only is shown in Figure 4.33.

You will notice that the phase response shown in Figure 4.33 looks very much the same as for the Butterworth version in Figure 4.28 above. There seems little point in using valuable space by repeating diagrams that are very similar, and so phase responses are only shown for selected crossovers.

The third-order Linkwitz–Riley (without frequency offset) has a power response with a −3 dB dip at crossover; it is not a CPC type.

The third-order Linkwitz–Riley crossover is neither linearphase nor minimumphase. The group delay has a peak just below the crossover frequency.

4.7.3 Third-Order Bessel Crossover

If a Bessel filter characteristic is used for a third-order crossover, the in-phase summation, as in Figure 4.34, shows a deep dip that bottoms out at −13.5 dB. This looks distinctly unpromising.

Phase-inverting one of the outputs gives a +3.0 dB hump, as in Figure 4.35, which is very reminiscent of what we got from the second-order Butterworth crossover with one output phase-reversed, though in this case the hump is rather narrower because we are using

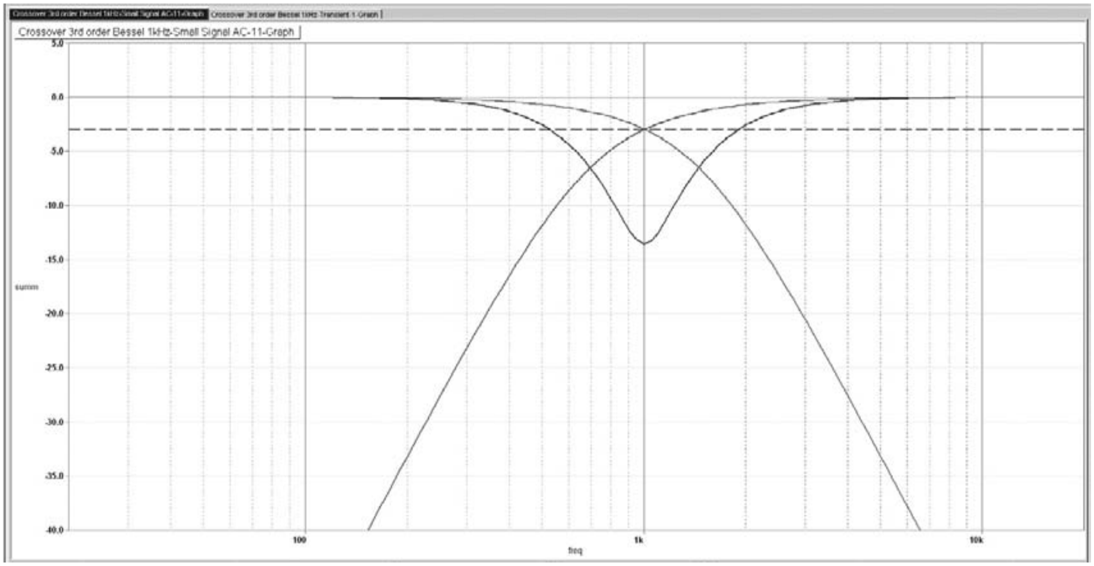


Figure 4.34: The frequency response of a third-order Bessel crossover resulting from in-phase summation and giving a nasty dip. The dashed line is at −3 dB.

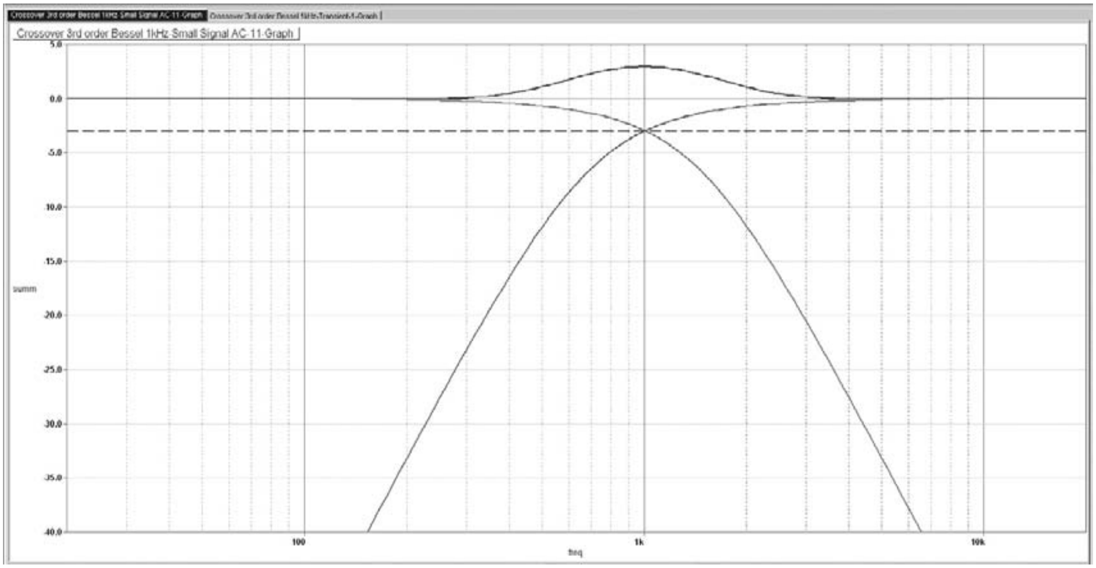


Figure 4.35: The frequency response of a third-order Bessel crossover, resulting from summation with one of the filter outputs phase-reversed. The dashed line is at −3 dB.

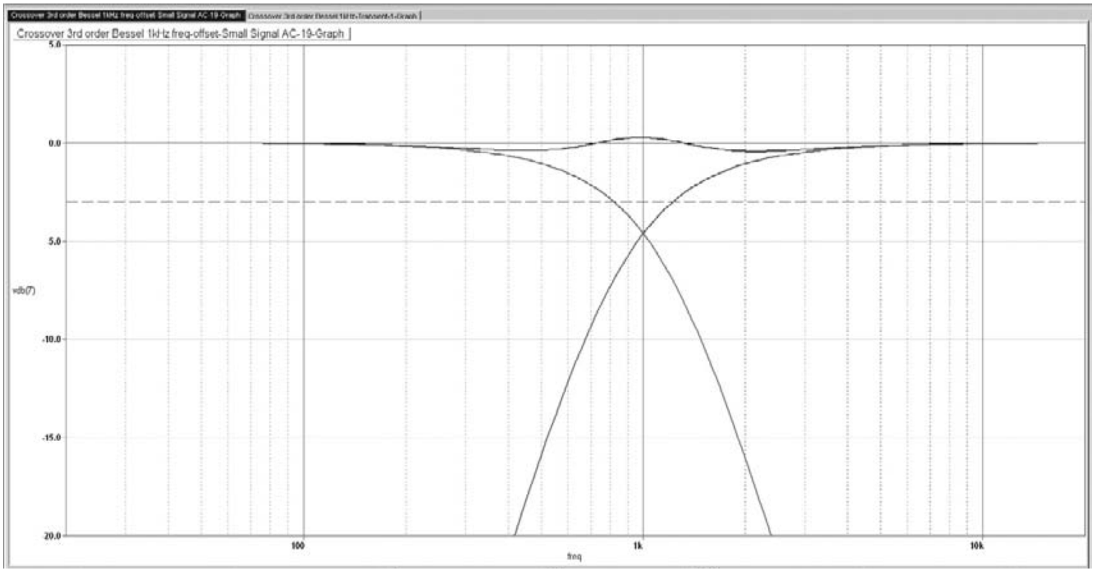


Figure 4.36: The frequency response of third-order Bessel crossover, summation with one output phase-reversed and both cutoff frequencies offset by 1.22x for maximal flatness. Ripple is only ±0.35 dB. Note vertical scale now has –20 dB at bottom.

third-order filters. This immediately suggests that the hump could be dealt with as before, by frequency offsetting the filters and so spreading their cutoff frequencies apart.

In this case offsetting each filter cutoff frequency by a ratio of 1.22, so that the highpass cutoff is at 1.22 kHz while the lowpass cutoff is at 1/1.22 = 0.820 kHz, gives the maximally flat sum with a peak and two dips each at 0.35 dB away from the 0 dB line; see Figure 4.36. The crossover point is now at –4.6 dB instead of –3 dB.

4.7.4 Third-Order 1.0 dB-Chebyshev Crossover

Using 1.0 dB-Chebyshev filters gives us the unpleasant peak at +4 dB, seen in Figure 4.37, with 1 dB dips on either side. This does not look like a good starting point for a flat crossover. You can see that the crossover point is now at −1 dB, because the cutoff frequency of a third-order Chebyshev filter is defined as the point where the rolloff response has fallen again to the level of the 1 dB dip after rising briefly back to 0 dB.

If one output is inverted we get instead a gentle dip, as in Figure 4.38, which looks more promising as a subject for frequency offsetting. As we saw with the third-order Linkwitz–Riley crossover, to tackle a dip in the combined response, it is necessary to push the curves together rather than pull them apart, and so the offset ratio will be less than 1.0.

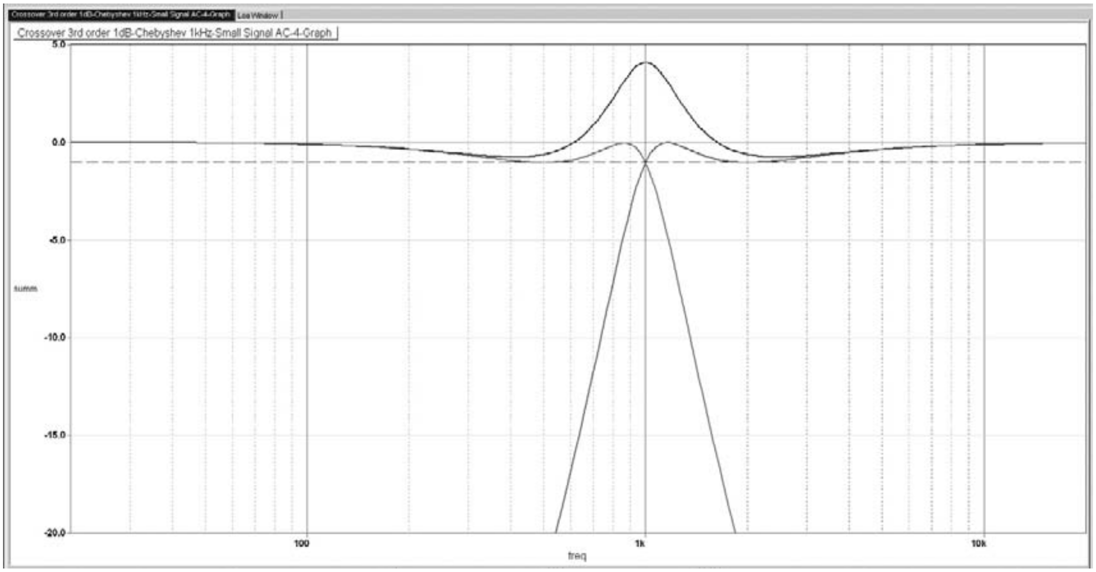


Figure 4.37: The frequency response of a third-order 1.0 dB-Chebyshev crossover, with in-phase summation; a +4 dB peak with dips either side. The dashed line is now at −1 dB.

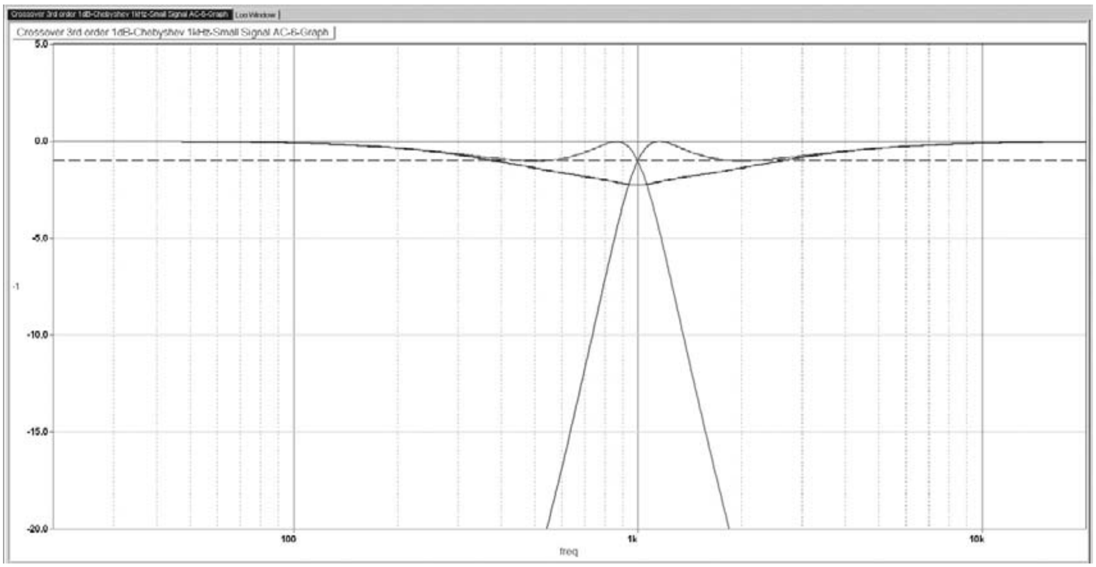


Figure 4.38: The frequency response of a third order 1.0 dB-Chebyshev crossover, resulting from summation with one of the filter outputs phase-reversed. The dashed line is at −1 dB.

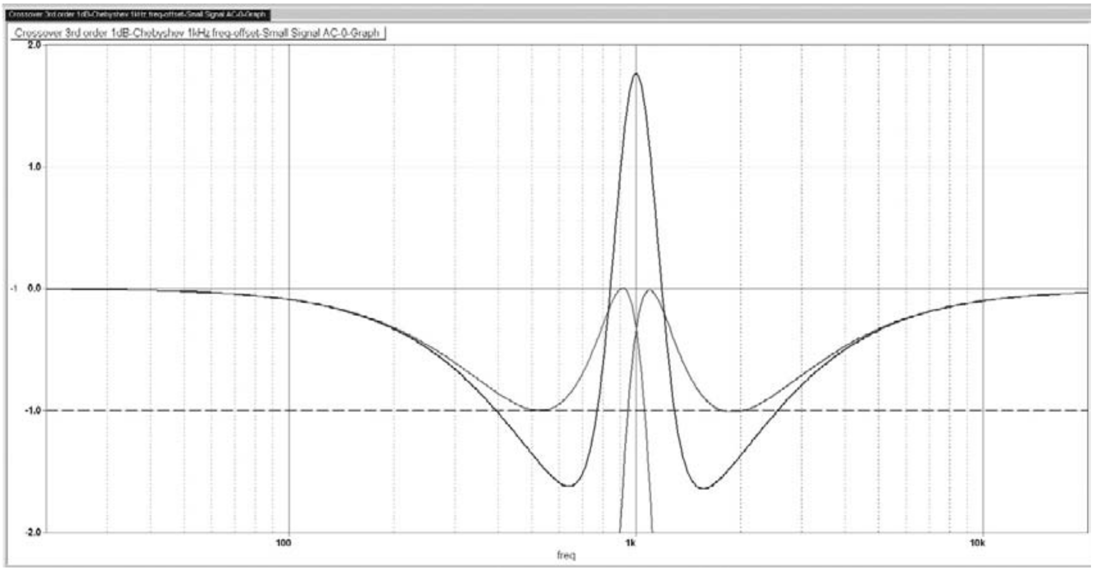


Figure 4.39: Zooming in on the frequency response of a third-order 1.0 dB-Chebyshev crossover, with frequency offset of 0.946 times and reversed-phase summation. Ripple is ±1.6 dB, vertical scale ±2 dB.

An offset ratio of 0.946 (making the highpass cutoff 0.946 kHz and the lowpass cutoff 1/0.946 = 1.057 kHz), gives the maximally flat response shown in Figure 4.39. The response deviations are ±1.6 dB, and once more the Chebyshev filter does not look like a promising start for a crossover design.

4.8 Fourth-Order Crossovers

The fourth-order crossovers give still greater separation between the drive units than the first-, second-, or third-order types. The 24 dB/octave slopes minimise the chance of exciting out-of-band drive unit irregularities like the tweeter or mid-range resonance will be excited. Modulation distortion will be further reduced, though probably not to a great extent. The sensitivity to driver time-delay misalignments is further lessened due to the narrower crossover region. The fourth-order Butterworth crossover does not gives a flat summed amplitude response but the fourth-order Linkwitz–Riley famously does. For the Butterworth and Linkwitz–Riley versions, inverting one output gives a useless response with a deep notch at the crossover frequency.

Since fourth-order filters are used, their extra phase-shift gives outputs that are 360° apart, which is the same as being in-phase and so eliminates lobing errors and tilting of the vertical coverage pattern in the crossover region; this is a major advantage.

Fourth-order crossovers have further reduced sensitivity to driver time misalignments because of their steeper 24 dB/octave slopes. Vance Dickason records [2] that a fourth-order Butterworth crossover and a 2-inch time-misalignment give an amplitude response error of about 1 dB, smaller than that of any other crossover examined so far.

Despite their advantages, fourth-order filters are rarely used in passive crossovers, because the greater number of expensive inductors increases the losses due to their resistance, and increasing the inductor wire gauge to reduce these losses puts the cost up even more. Fourth-order crossovers also require a greater number of big capacitors.

I was inspired by Vance Dickason [2] to investigate some more exotic filters as possible candidates for crossovers; the linear-phase filter, the Gaussian filter, and the Legendre filter. It has to be said that on examination none look very promising. Other unusual filters such as transitional and synchronous types are looked at in Chapter 7.

4.8.1 Fourth-Order Butterworth Crossover

The fourth-order Butterworth sums to give a +3.0 dB hump at the crossover frequency, as in Figure 4.40, while the summation with one phase reversed gives a deep crevasse as in Figure 4.41; this is the exact opposite of the behaviour of the second-order Butterworth. A frequency offset of 1.128 times reduces the hump to a maximally flat ripple of ±0.47 dB, as seen in Figure 4.42.

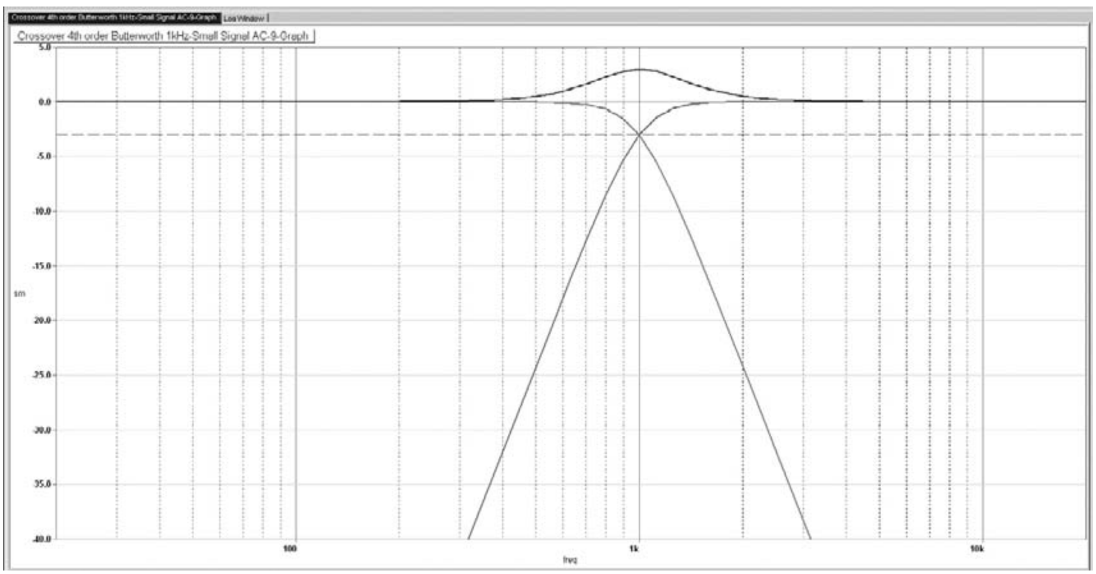


Figure 4.40: The frequency response of a fourth-order Butterworth crossover, resulting from in-phase summation. The dashed line is at −3 dB.

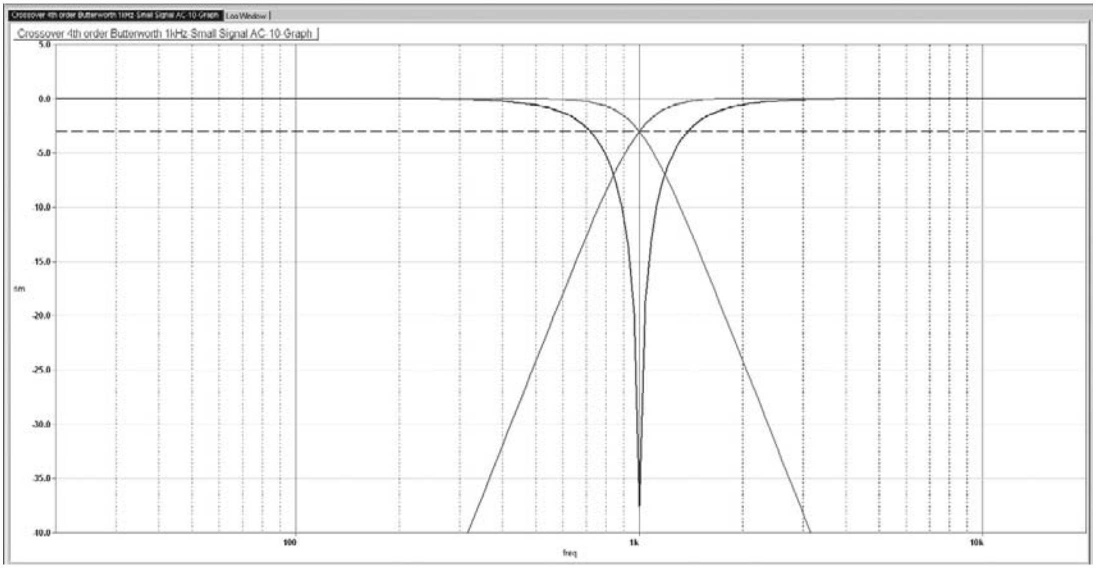


Figure 4.41: The frequency response of a fourth-order Butterworth crossover with reverse-phase summation. The dashed line is at −3 dB.

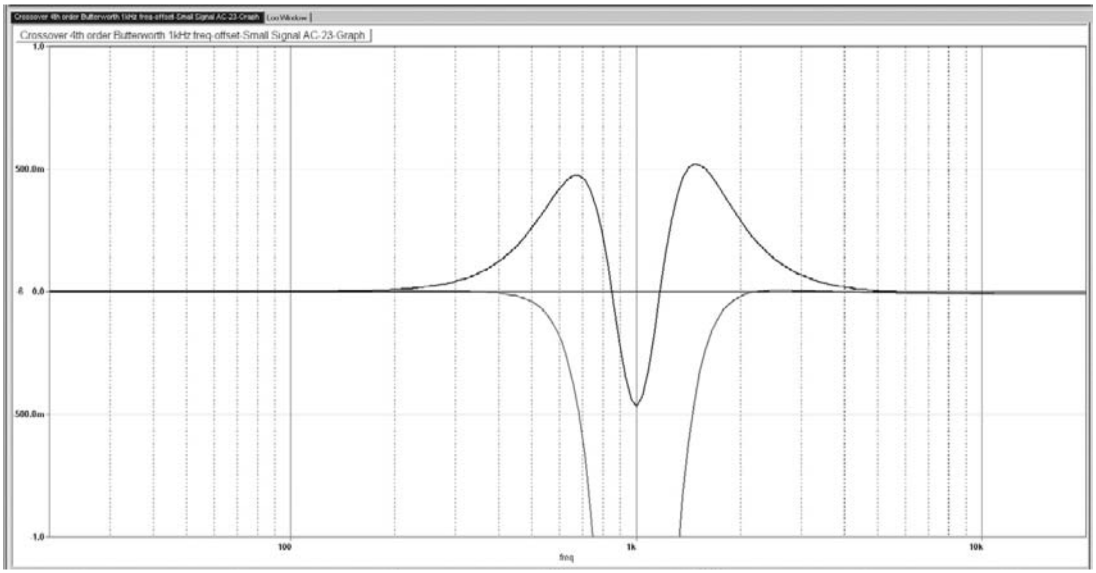


Figure 4.42: Zooming in on the frequency response of a fourth-order Butterworth crossover with in-phase summation, and an optimal frequency-offset ratio of 1.128 times. The error is ±0.47 dB.

The fourth-order Butterworth crossover produces a flat power response and so it is a CPC crossover. The outputs are 360° apart in phase at all times. This is equivalent to being in-phase and so there is no tilting of the vertical coverage pattern in the crossover region. The summed group delay has a significant peak just below the crossover frequency.

Given that the best possible amplitude response flatness obtainable by frequency offset is ±0.47 dB, there seems no reason to use the fourth-order Butterworth in preference to the fourth-order Linkwitz-Riley crossover. The phase and group delay plots are therefore not shown.

4.8.2 Fourth-Order Linkwitz–Riley Crossover

The fourth-order Linkwitz–Riley is considered by many the best crossover alignment of the lot. The in-phase response sums to completely flat, making it an APC type; see Figure 4.43. (The reversed-phase response in Figure 4.44 has a deep crevasse at the crossover frequency and is of no value.) The outputs are 360° apart in phase at all times so there is no lobe-tilting in the crossover region. The crossover point is at −6 dB. There is a −3 dB dip in the power response so it is not a CPC crossover.

One of the beauties of this type of crossover is the ease of its design. It is normally made by cascading two second-order Butterworth filters with identical cutoff frequencies, and identical Qs of 0.7071 (1/√2). Thus, it is sometimes called a "squared Butterworth" filter, or, less logically, a "−6 dB Butterworth" filter.

Figure 4.45 shows the −3 dB dip in the power response. The two outputs are at −6 dB at the crossover point, and are regarded as uncorrelated, so they RMS-sum to give −3 dB.

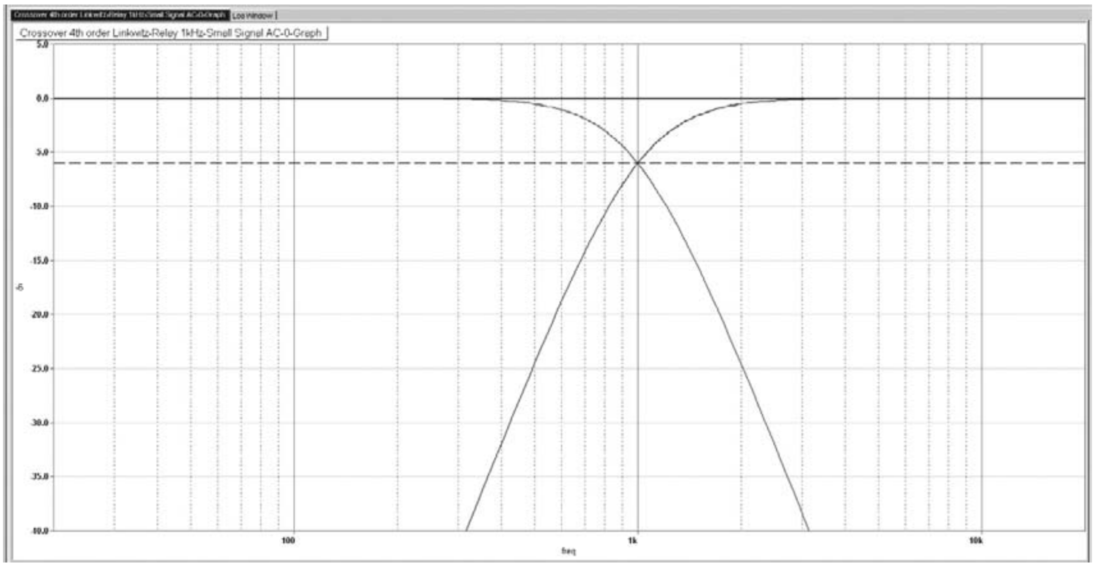


Figure 4.43: The frequency response of a fourth-order Linkwitz–Riley crossover, resulting from in-phase summation. The dashed line is at −6 dB.

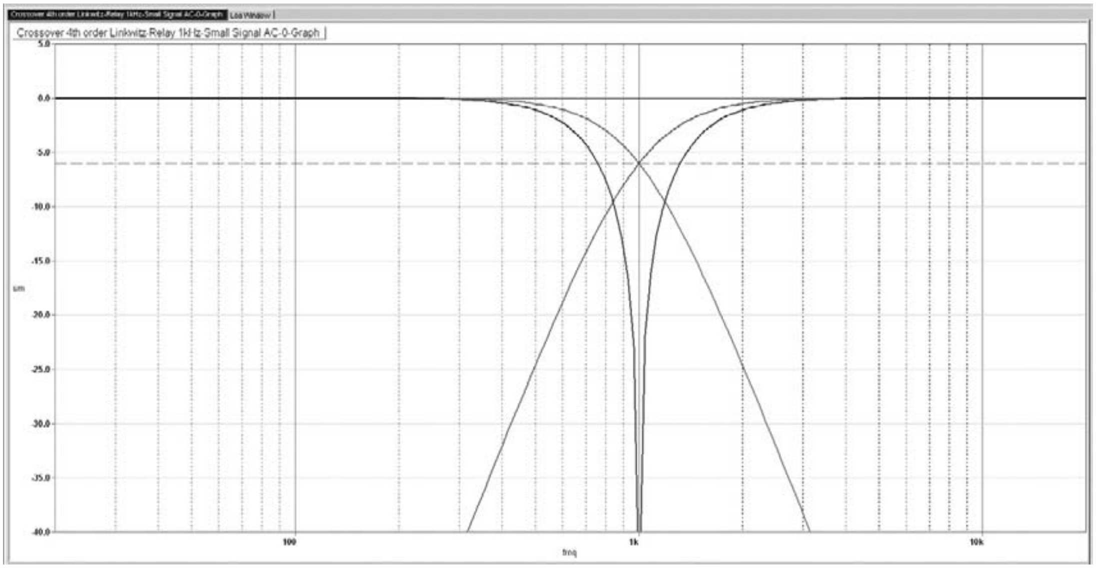


Figure 4.44: The frequency response of a fourth-order Linkwitz–Riley crossover, resulting from phase-inverted summation. The dashed line is at −6 dB.

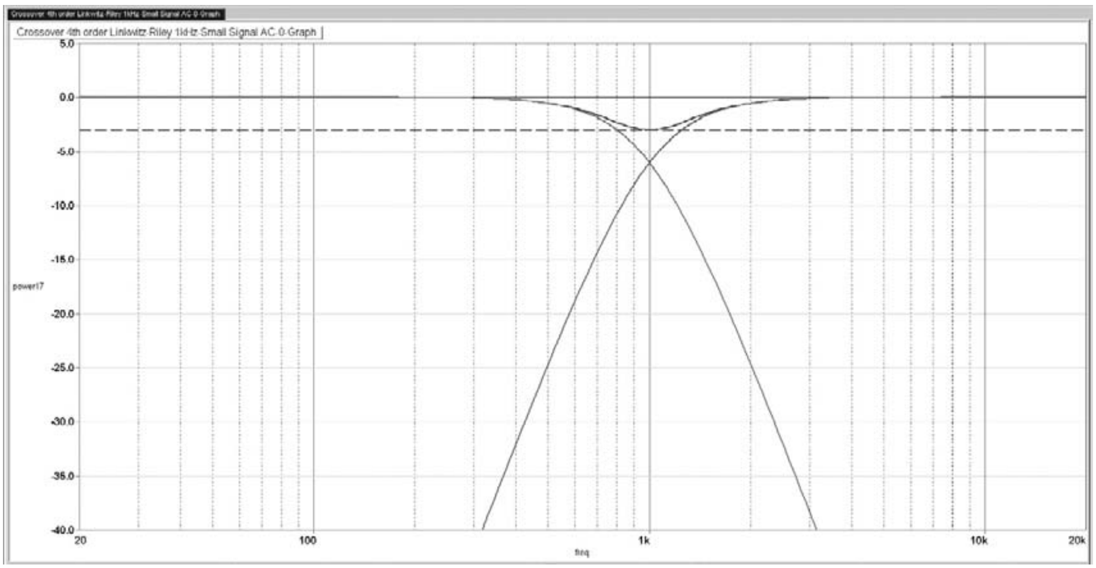


Figure 4.45: The power response of a fourth-order Linkwitz–Riley crossover, resulting from in-phase summation. The dashed line is at −3 dB.

While this is not ideal, it is a modest dip, and the use of fourth-order filters means it is not wide, so the effect on the reverberant energy in a listening space will be correspondingly small.

Figure 4.46 illustrates how the outputs are 360° apart in phase at all times. This is equivalent to being in-phase and so eliminates tilting of the vertical coverage lobes in the crossover region. The summed phase swings through 360° across the audio band and thus emulates a second-order allpass filter.

The summed group delay in Figure 4.47 has a flat LF region at 450 usec, which is a longer delay than any of the lower-order crossovers we have looked at; however it is much less than the 3.2 msec audibility threshold at 500 Hz, as quoted in Chapter 3. The group delay shows a moderate peak of 540 usec just below the crossover frequency.

Passive versions of the fourth-order Linkwitz–Riley are relatively uncommon because of the power losses and the number of components in a passive fourth-order crossover, but one example of a loudspeaker using the technology was the KEF Model 105 [5] of which the first version was released in 1977.

The fourth-order Linkwitz–Riley crossover is widely considered to be the best. It sums to a flat amplitude response, and its power response has a dip of limited width and only

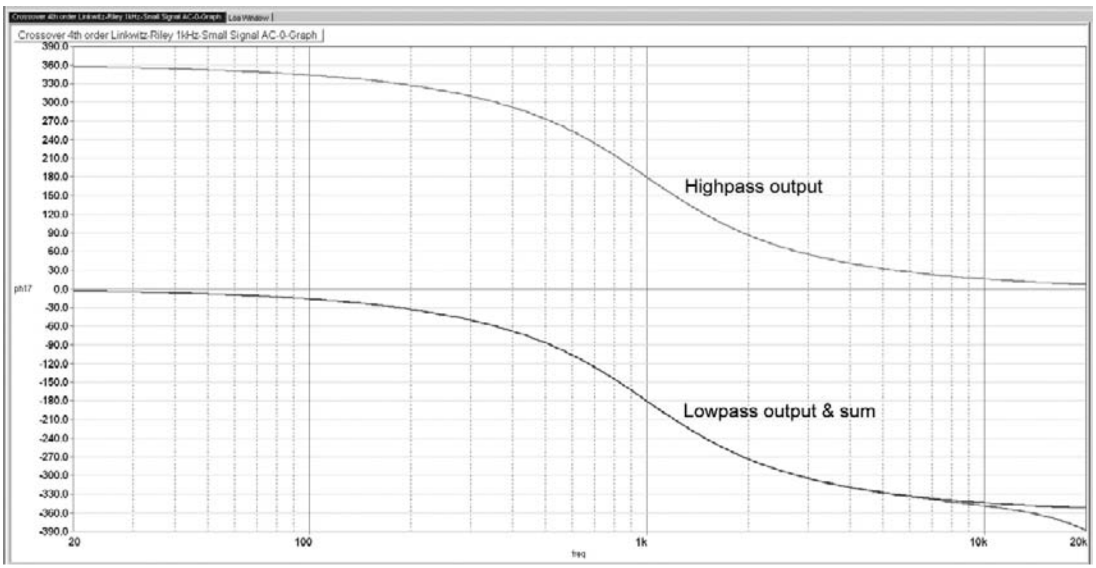


Figure 4.46: The phase response of a fourth-order Linkwitz–Riley crossover, in-phase connection. There is a constant 360° phase difference between the two outputs. (The summed phase trace is on top of the lowpass output trace).

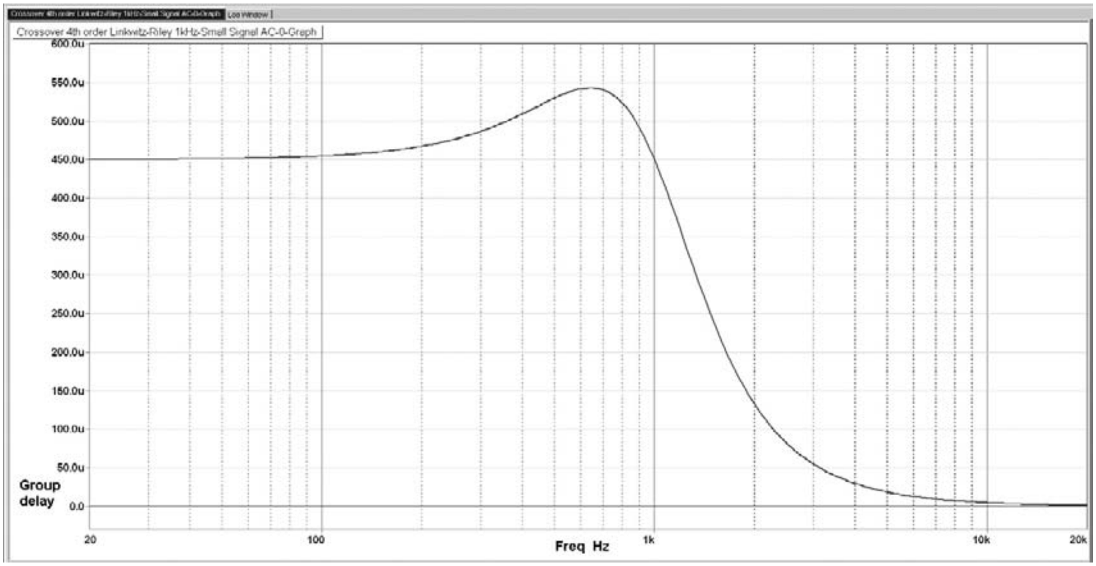


Figure 4.47: The group delay response of a fourth-order Linkwitz–Riley crossover, in-phase connection.

−3 dB deep. There is no lobe tilting and the 24 dB/octave slopes are considered adequate for the vast majority of drive units. On the debit side, the phase response is that of a second-order allpass filter rather than a first-order, and the group delay is relatively long at 450 usec, and has a peak.

4.8.3 Fourth-Order Bessel Crossover

The fourth-order Bessel has its crossover point at −3 dB. It sums to give a −2.6 dB dip at the crossover frequency, as in Figure 4.48, while the summation with one phase reversed gives a +2 dB hump and two −1 dB dips, as in Figure 4.49. The phase-reversed case does not look like a suitable case for frequency-offset treatment as pulling the cutoff frequencies apart will pull down the hump but deepen the dips, while pushing them together will pull up the dips but make the hump worse.

It looks more promising for the in-phase case, but pushing the cutoffs together for this crossover actually deepens the dip due to the phase-shifts involved. Pulling them apart by an increasing amount makes the dip more shallow to begin with, but before the response begins to straddle the 0 dB line, it undergoes a fairly complicated set of changes with two new dips appearing either side of the central one, and they deepen as the offset ratio increases.

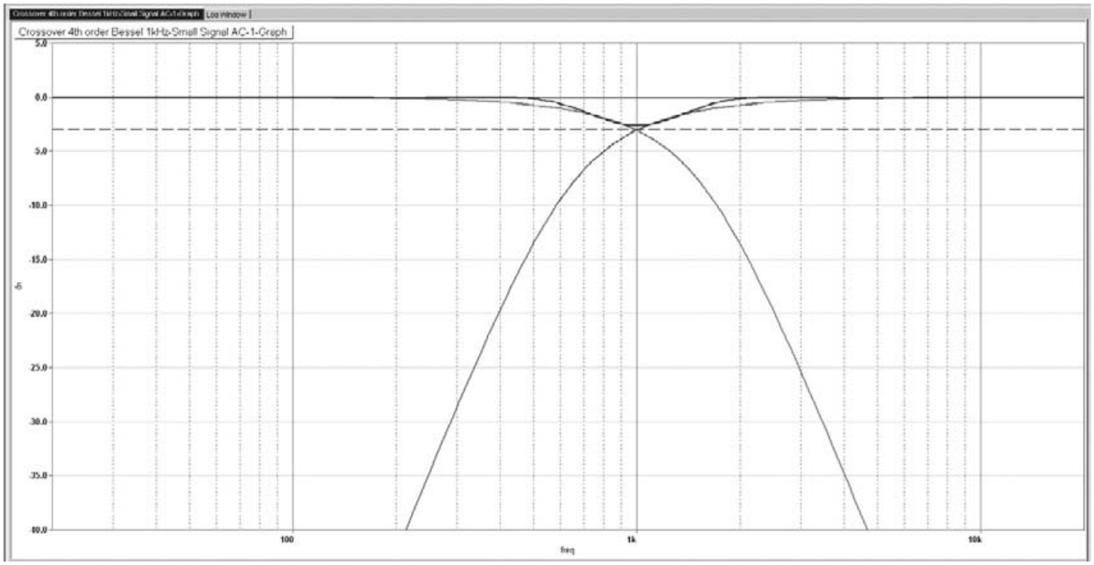


Figure 4.48: The frequency response of a fourth-order Bessel crossover, resulting from in-phase summation. The dashed line is at −3 dB.

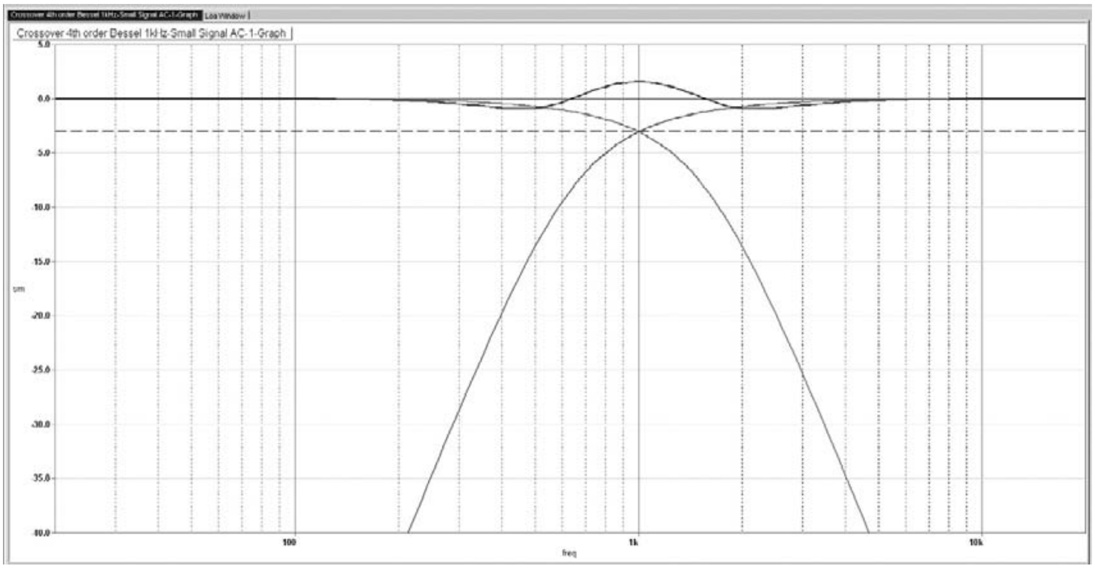


Figure 4.49: The frequency response of a fourth-order Bessel crossover, with reversed-phase summation. The dashed line is at −3 dB.

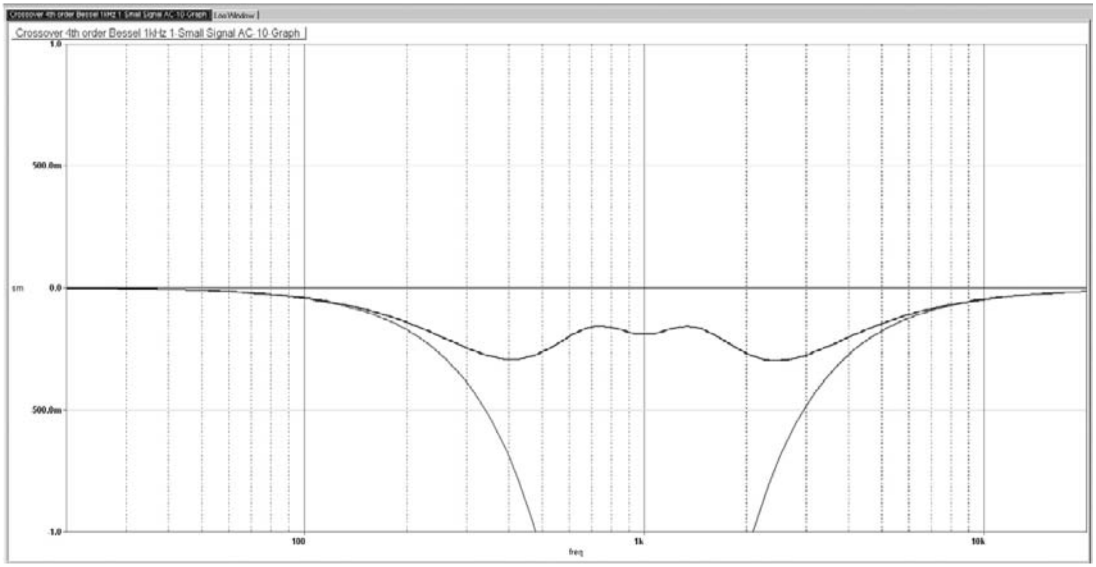


Figure 4.50: Zooming in on the frequency response of a fourth-order Bessel crossover, with an offset ratio of 1.229 times and in-phase summation. Vertical scale is ± 1 dB.

In this situation it is hard to say what constitutes the maximally flat solution, but a promising candidate is shown in Figure 4.50, where a frequency offset of ratio of 1.229 times gives only −0.18 dB at the crossover frequency, with dips either side that are −0.3 dB deep.

The summed group delay response has a moderate peak just below the crossover frequency.

4.8.4 Fourth-Order 1.0 dB-Chebyshev Crossover

Using the 1.0 dB-Chebyshev characteristic for each filter gives an in-phase summed response with a central peak at the crossover frequency of +2.22 dB, with the dips on either side going down to +0.15 dB, as in Figure 4.51. These dips are above 0 dB because a fourth-order Chebyshev filter has its passband ripples above the 0 dB line. All even-order Chebyshev filters have response peaks above 0 dB, while all odd-order Chebyshev filters have dips below it.

The summed response with one output phase-inverted, seen in Figure 4.52, looks very much the same, except that the central peak is taller at +3.65 dB. Neither summed response looks as though it could be significantly flattened by the use of a frequency offset, and I have not attempted it.

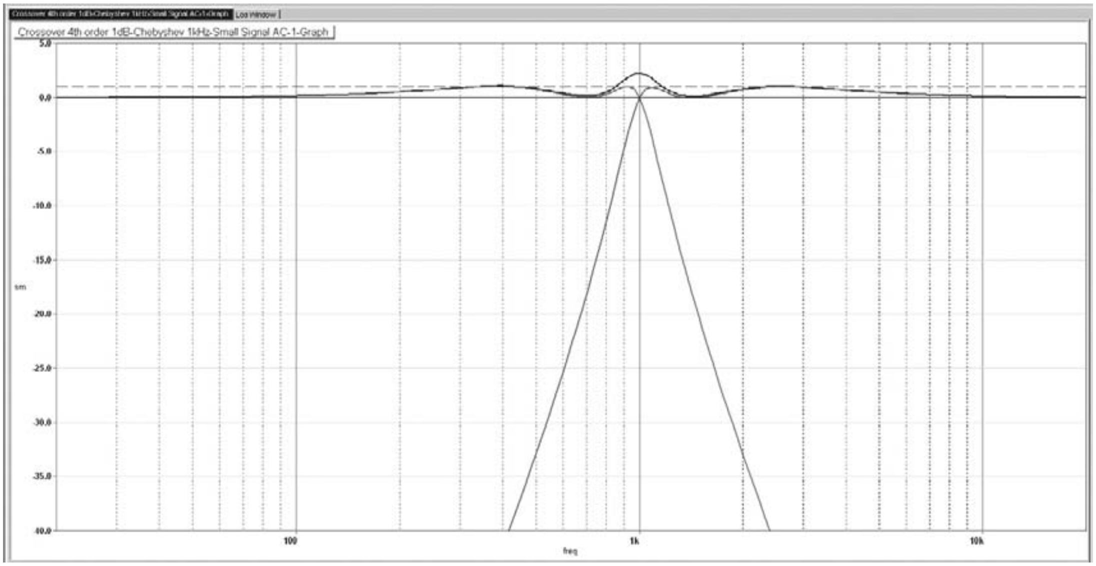


Figure 4.51: The frequency response of a fourth-order Chebyshev crossover, resulting from in-phase summation. The dashed line is at –3 dB.

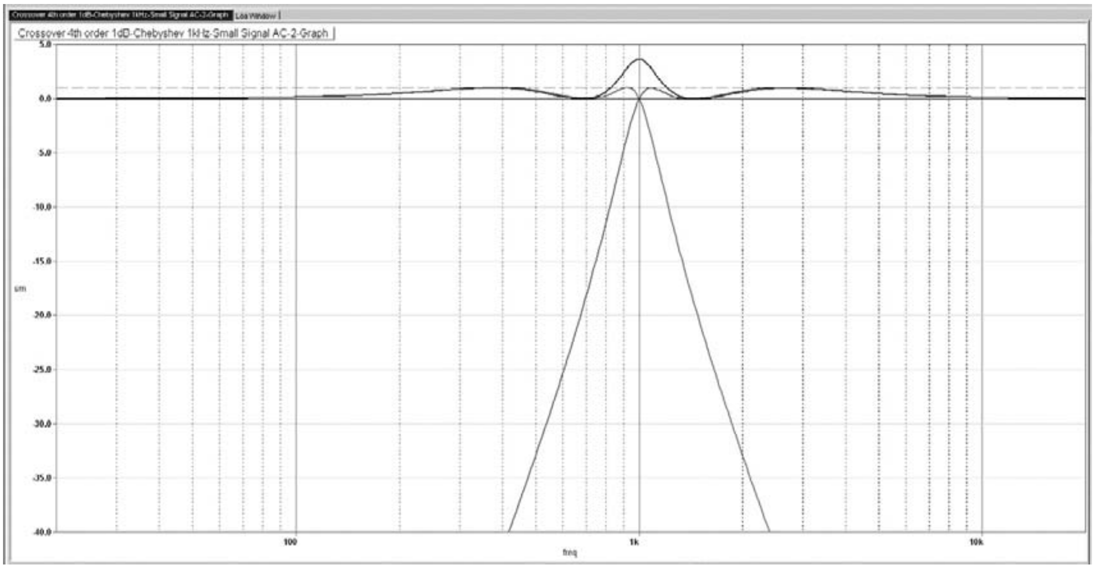


Figure 4.52: The frequency response of a fourth-order Chebyshev crossover, resulting from reversed-phase summation. The dashed line is at –3 dB.

4.8.5 Fourth-Order Linear-Phase Crossover

There is not a great deal of information out there about the design of linear-phase filters, and some of what there is appears to be contradictory. The crossover here uses linear-phase filters as defined by Linear-X Systems Filtershop [6], which for a fourth-order linear-phase filter consists of two cascaded second-order stages; the first has a cutoff frequency of 1.334 and a Q of 1.316, while the second has a cutoff frequency of 0.7496 and a Q of 0.607. With this filter structure the crossover point is at −4.5 dB. The summed response with the outputs in-phase has a gentle +1.2 dB hump at crossover, dipping very slightly below the 0 dB line on either side; see Figure 4.53.

The summation with one output phase-reversed has a deep hole in it but is notable because it has an unusual flat portion at −8.7 dB around the crossover frequency, as in Figure 4.54. It is in fact very flat indeed, to within 0.01 dB across the visibly flat part. While this is strange and rather interesting, the deep dip does not look like a good starting point for the frequency-offset process.

Applying frequency offset to the in-phase case of Figure 4.42, we find that an offset ratio of 1.0845 times gives a maximally flat response with deviation of ±0.66 dB around the 0 dB line; see Figure 4.55. This is poor compared with other crossovers.

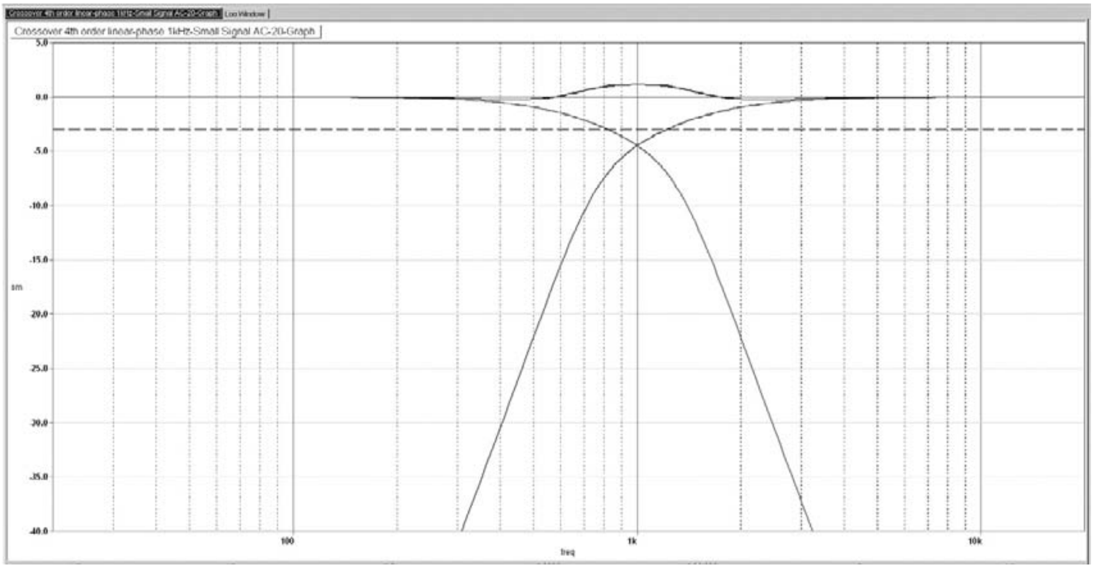


Figure 4.53: The frequency response of a fourth-order linear-phase crossover, resulting from in-phase summation. The dashed line is at −3 dB.

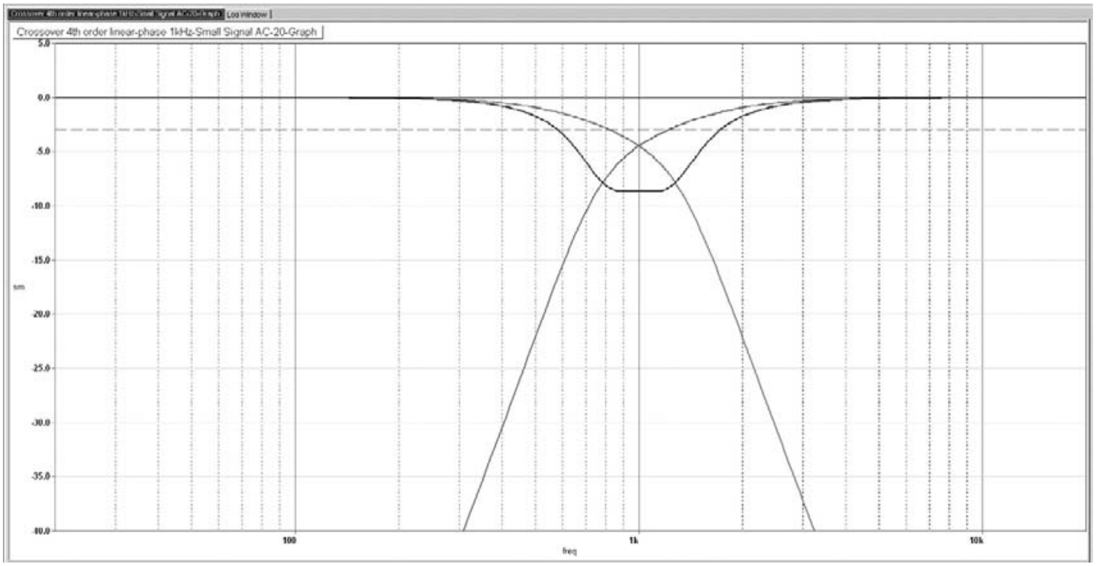


Figure 4.54: The frequency response of a fourth-order linear-phase crossover, resulting from reversed-phase summation. The dashed line is at −3 dB.

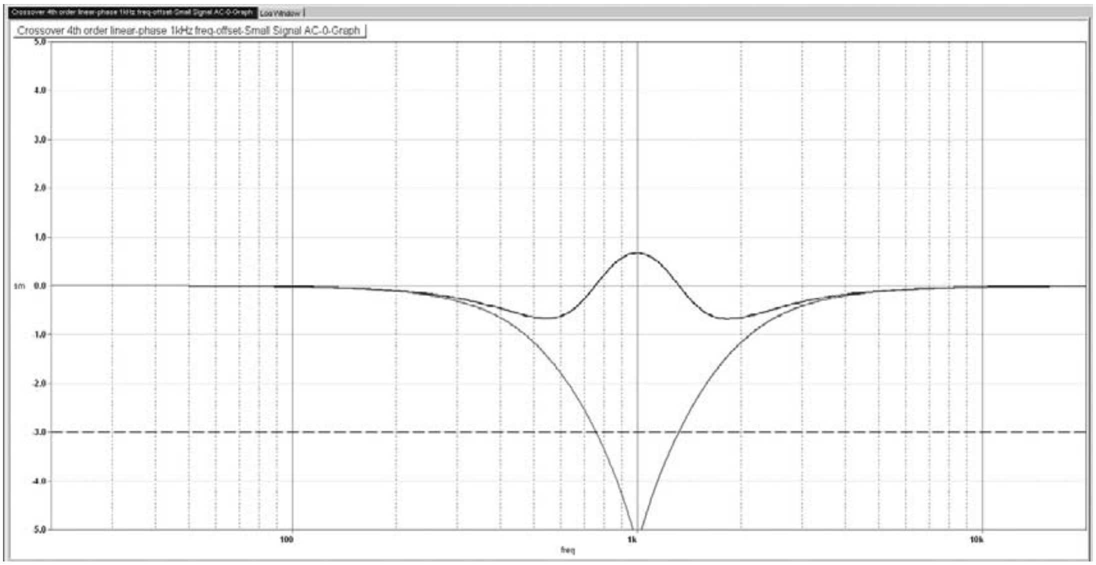


Figure 4.55: The frequency response of a fourth-order linear-phase crossover with a frequency offset of 1.0845 times, and in-phase summation, showing errors of ± 0.66 dB. The dashed line is at −3 dB.

Despite its promising name, the linear-phase filter crossover cannot be made to sum flatter than ±0.66 dB, which makes it a very doubtful choice.

4.8.6 Fourth-Order Gaussian Crossover

As explained in Chapter 7, the Gaussian filter characteristic is based on time-domain considerations, being designed for no overshoot on a step function input while keeping rise and fall times as fast as possible. This response is closely connected to the fact that the Gaussian filter has the minimum possible group delay. Gaussian filters come in various kinds identified by a dB suffix, such as "Gaussian-6 dB" and "Gaussian-12 dB," though the differences in amplitude response are very small. The amplitude response is very similar indeed to that of a Bessel filter.

Designing the lowpass and highpass filters for a cutoff of 1 kHz gives a crossover point at −2.8 dB, as in Figure 4.56. The reversed-phase connection gives the crevasse in Figure 4.57, which will not be considered further.

Looking at Figure 4.56, it is clear that the filter cutoff frequencies need to be pushed together to get a flatter summed response. The depth of the dip is minimised at −1.05 dB by a frequency-offset ratio of 0.8045 times; thus the highpass filter cutoff becomes 0.8045 kHz and the lowpass filter cutoff becomes 1/0.8045 = 1.243 kHz, giving the response in Figure 4.58. Altering the ratio either up or down gives a deeper dip.

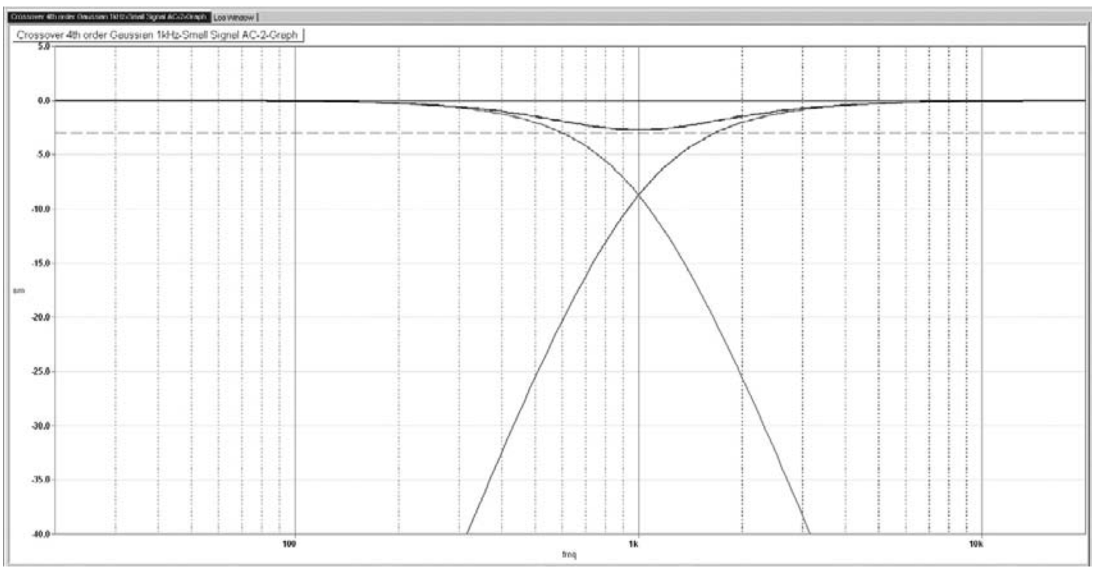


Figure 4.56: The frequency response of a fourth-order Gaussian crossover, resulting from in-phase summation. The dashed line is at −3 dB.

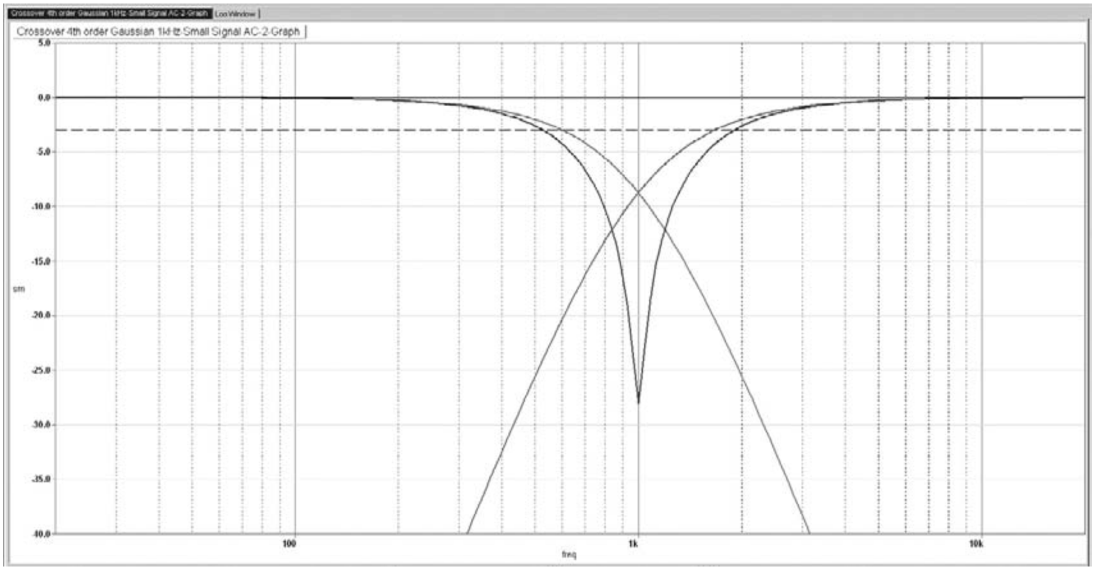


Figure 4.57: The frequency response of a fourth-order Gaussian crossover, resulting from reversed-phase summation. The dashed line is at −3 dB.

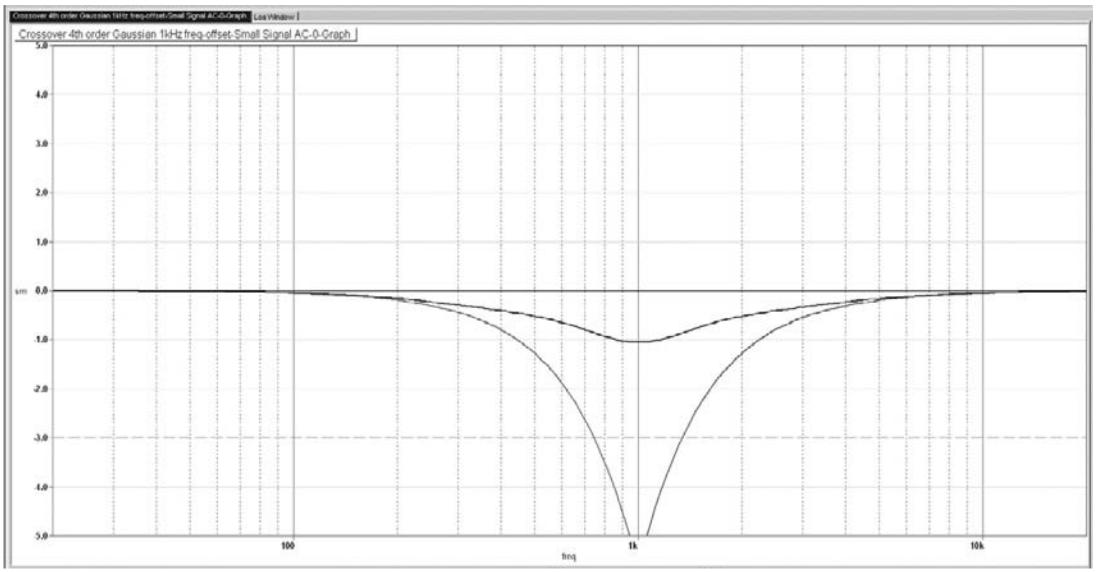


Figure 4.58: The frequency response of a fourth-order Gaussian crossover with a frequency offset of 0.8045 times to give minimal depth of dip. In-phase summation. The dashed line is at −3 dB.

Once again, there is not a great deal of information to be had on Gaussian filters, and some of what there is appears to be contradictory. The crossover here uses Gaussian filters as defined by Linear-X Systems Filtershop [6], which for the fourth-order consist of two cascaded second-order stages; the first has a cutoff frequency of 0.9930 and a Q of 0.636, while the second has a cutoff frequency of 1.0594 and a Q of 0.548. It is very possible that if you use a different design methodology from me—see Chapter 7 for more details—you will get different cutoff frequencies. This is irritating but not a cause for despair as adjusting the frequency-offset ratio to suit is straightforward.

However, the Gaussian alignment appears to have no great advantages. Its rolloffs are slow; if that suits your design intentions the Bessel crossover has a better maximal flatness. (The Bessel has two dips −0.3 dB deep as opposed to the Gaussian −1.05 dB single dip.)

The summed group delay response has a moderate peak just below the crossover frequency.

4.8.7 Fourth-Order Legendre Crossover

The Legendre filter (sometimes called the Legendre–Papoulis filter) is optimised for the greatest possible slope at the passband edge without passband ripples—in other words it is a monotonic filter, with a response that always goes downwards. It gives a faster rolloff than the Butterworth characteristic, but the drawback is that the passband is not maximally flat as is the Butterworth; instead it begins to slope gently down until the rapid roll-off begins. This can be seen in Figure 4.59, which also shows that in-phase summation gives a

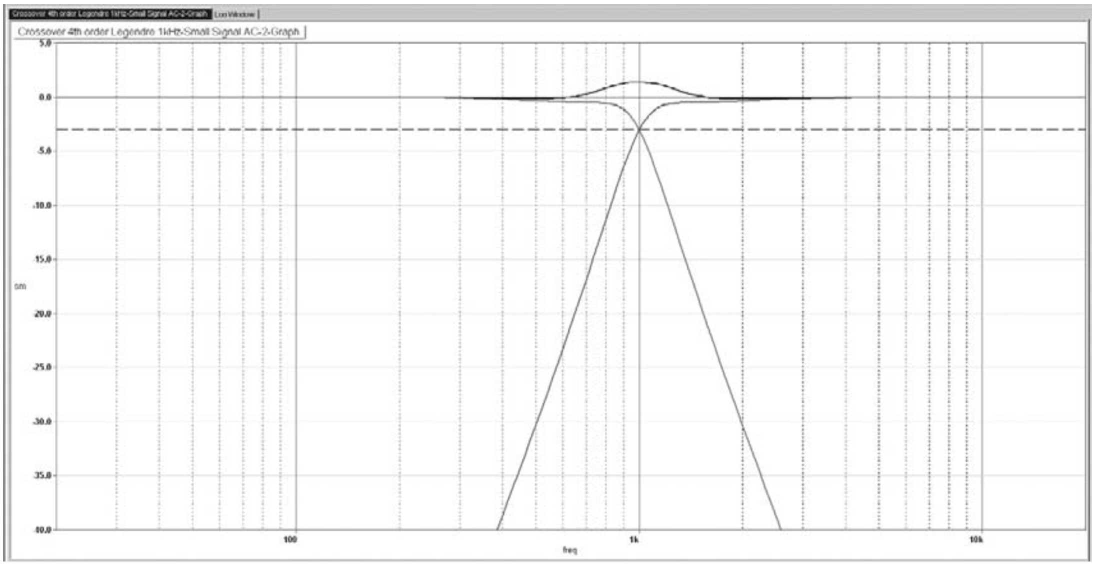


Figure 4.59: The frequency response of a fourth-order Legendre crossover, resulting from in-phase summation. The dashed line is at −3 dB.

narrow hump reaching +1.4 dB at the crossover point. It may not be visible in the plot, but the summed response actually dips below the 0 dB line by 0.1 dB on either side of the hump.

A fourth-order Legendre filter is normally constructed by cascading a second-order stage with a Q of 2.10 with another second-order stage of Q = 0.597. A Q of 2.10 is quite high for a standard Sallen & Key filter and requires a capacitor ratio of 17.6 to achieve it; this indicates that component sensitivity will be higher than usual and care will be needed to get an accurate response.

The summation with one phase reversed has a shallow dip of −2.0 dB around the crossover point, as shown in Figure 4.60. It is not too clear which of these can be best flattened by the use of a frequency offset, so I tried both.

The rapid rolloff of the Legendre filters means that quite small amounts of frequency offset have large effects. An offset ratio of only 1.029 times gives the maximally flat response for the in-phase condition, illustrated in Figure 4.61. The twin peaks are +0.3 dB high and the central dip is −0.39 dB deep.

Applying a frequency offset to the reversed-phase case gives an intriguing shape to the curve but larger errors, the −2 dB dip reaching its minimum depth of −1.05 dB with an offset ratio of 1.041 times, as shown in Figure 4.62. Clearly the in-phase option is better,

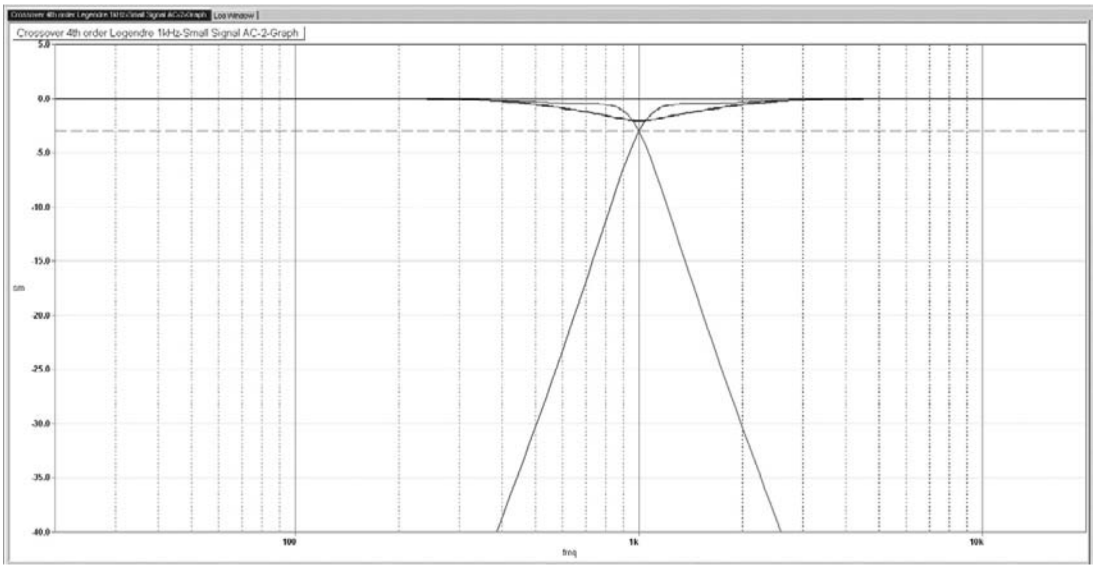


Figure 4.60: The frequency response of a fourth-order Legendre crossover, resulting from reversed-phase summation. The dashed line is at −3 dB.

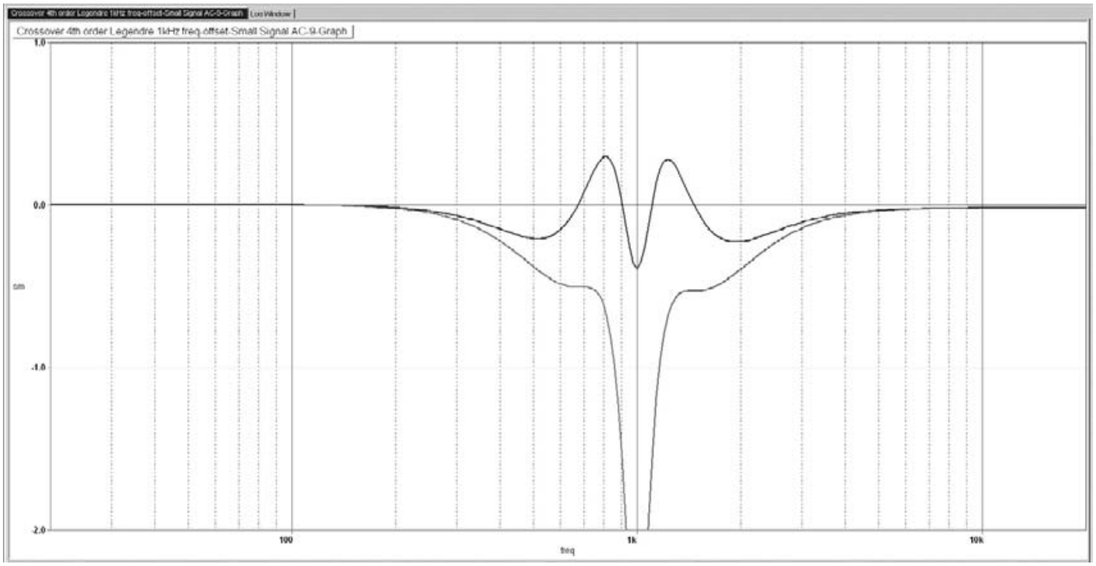


Figure 4.61: The frequency response of a fourth-order Legendre crossover with a frequency offset of 1.029 times, and in-phase summation. The dashed line is at –3 dB.

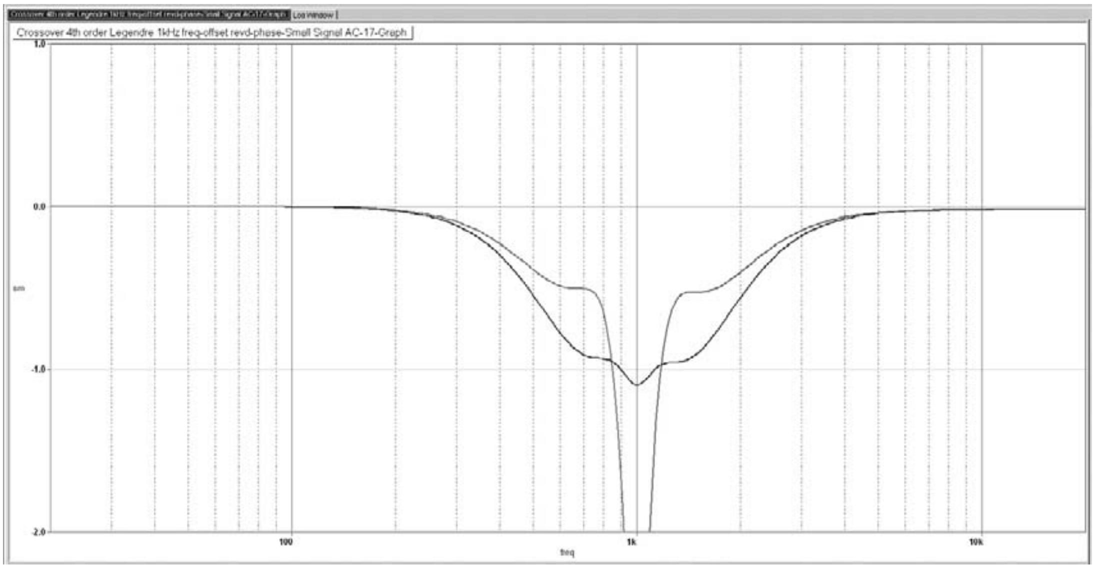


Figure 4.62: The frequency response of a fourth-order Legendre crossover with a frequency offset of 1.041 times, and reversed-phase summation. The dashed line is at –3 dB.

unless you are trying to keep abreast of the times, and perhaps compensating for a 1 dB peak in a drive unit response.

The group delay of the summed output has a significant peak just below the crossover frequency.

4.9 Higher-Order Crossovers

As we have seen, fourth-order crossovers provide a wide range of possible alignments, and good slopes of 24 dB/octave, which are generally considered to be adequate for maintaining separation between drive units. The fourth-order Linkwitz–Riley, with the absolute flatness of its summed response and its delightfully straightforward filter design, is considered by many the best crossover alignment known at present. The question nonetheless arises if crossovers of fifth, sixth or higher order could give any further advantages. There seems to be no consensus as to whether the extra steepness of the slopes gives significant benefits, and there is an issue about the perceptibility of the group delay. According to Siegfried Linkwitz, "Crossover filters of higher order than LR4 are probably not useful, because of an increasing peak in group delay around the crossover frequency."

Nonetheless, crossovers of up to the eighth order (with 48 dB/octave slopes) are sometimes used, mainly for sound reinforcement applications. These are usually of the Linkwitz–Riley type. According to Dennis Bohn of Rane Corporation [7], going from a fourth-order Linkwitz–Riley to an eighth-order Linkwitz–Riley halves the effective width of the crossover region from 1.5 octaves to 0.75 octaves, and gives more linear drive unit operation and greater driver protection, and hence better power handling. He makes the point that the system designer must have a very good knowledge of the drive unit characteristics to use such steep slopes effectively.

Eighth-order crossover filters raise questions about component sensitivity. Expensive precision capacitors (and possibly resistors) are likely to be required. Eighth-order crossovers are probably best realised by DSP techniques.

At least one company makes fifth-order passive crossovers. CDT Audio [8] manufactures crossovers for automotive audio that are stated to effectively give fifth-order slopes by the use of elliptical filters.

4.10 Determining Frequency Offsets

The frequency offsets required for maximal flatness are given above for the common (and some not-so-common) crossover alignments. However, if you are bravely striking out on your own with an unconventional filter not covered here, you may find it useful to be able to determine your own frequency offsets for maximal response flatness.

There are several ways of doing this. It would be possible to do it purely mathematically, working from the complex filter responses and manipulating the cutoff frequencies by some sort of optimisation technique. However, the rather heavyweight mathematics involved in that approach does not appeal to all of us. An alternative simulation method which requires a bit of work but no turn-the-paper-sideways algebraic manipulation is as follows:

1. Simulate the lowpass and highpass filters, designed for the same cutoff frequency, (typically 1.0 kHz) and driven from the same signal source. Check that they are giving the response shapes and cutoff frequencies that you expect.
2. Sum the two filter outputs. In most simulators this can be done simply by plotting the arithmetical sum, so there is no need to add an electrical summing stage to the simulation. The reverse-phase connection can be checked by plotting the difference of the filter outputs instead of the sum.
3. If the summed response is not ruler flat, determine how it needs to be altered. Very roughly, if there is a peak around the crossover frequency you will need to move the filter cutoff frequencies apart (offset ratio greater than 1), and if there is a dip you need to move them together (offset ratio less than 1).
4. Now there is a bit of work; move one of the cutoff frequencies, by changing the component values in the relevant filter, and see how the response changes. Normally five or six attempts will get the response as flat as you are going to get it; in many cases this means equal deviations above and below the 0 dB line. Changing the cutoff frequencies is much easier if you have designed the filters as Sallen & Key types and scaled the components so that they are equal. For example, a fourth-order highpass Sallen & Key filter commonly consists of two cascaded second-order stages. These will have two equal capacitors in each stage, and by manipulating the resistor values it can be arranged that all four capacitors have the same value. The cutoff frequency can then be changed by at worst typing in four equal values; if your simulator has parameter facilities you should be able to set thing up so typing in a single value changes all four capacitors.
5. You will note that in the previous step we performed the cut-and-try on one filter only, to save effort. This means that the crossover frequency will move away from the original design value, but ignore this as you concentrate on the shape of the summed response. When you have the best response you can get, one cutoff frequency will be unchanged at 1.0 kHz, but the other is altered to say 1.4 kHz. To get the crossover point back where it should be, the 1.4 times ratio needs to be equally distributed between the two filter cutoff frequencies. This is done by taking the square root of the ratio, so the 1.4 kHz cutoff becomes 1.183 kHz, and the 1 kHz cutoff becomes 1/1.183 = 0.845 kHz. If you have got the calculations right, the crossover frequency will now be back at 1 kHz. It is wise to check that point with a final simulation.

Several examples of this process are given in the descriptions of the crossover alignments above.

Table 4.1 includes a summary of the frequency offsets required for maximal flatness with the crossover alignments dealt with in this chapter.

4.11 Summary of Crossover Properties

Table 4.1: Summarises the Properties of the Crossovers Examined So Far, Including the Frequency Offsets Required for Maximal Flatness. (n/d = not determined)

Order	Crossover Type	Phase	APC?	CPC?	Lobe Error?	Freq-Offset x	Deviation +/− dB
1	First-order	In-phase	Linear	YES	YES	None	Flat
1	First-order	Reversed	YES	YES	YES	None	Flat
2	Butterworth	Reversed	NO	YES	NO	1.30	0.45
2	Linkwitz–Riley	In-phase	YES	NO	NO	None	Flat
2	Bessel	Reversed	NO	NO	NO	1.45	0.07
2	1.0 dB-Chebyshev	Reversed	NO	NO	NO	1.53	1.60
3	Butterworth	Either	YES	YES	YES	None	Flat
3	Linkwitz–Riley	Reversed	NO	NO	YES	0.872	0.33
3	Bessel	Reversed	NO	NO	YES	1.22	0.35
2	1.0 dB-Chebyshev	Reversed	NO	NO	n/d	0.946	1.60
4	Butterworth	In-phase	NO	YES	NO	1.128	0.47
4	Linkwitz–Riley	In-phase	YES	NO	NO	None	Flat
4	Bessel	In-phase	NO	NO	n/d	1.229	0.30
4	1.0 dB-Chebyshev	In-phase	NO	NO	n/d	n/d	n/d
4	linear-phase	In-phase	NO	NO	n/d	1.084	0.66
4	Gaussian	In-phase	NO	NO	n/d	0.804	1.05
4	Legendre	In-phase	NO	NO	n/d	1.029	0.39

4.12 Filler-Driver Crossovers

The filler-driver crossover concept was introduced by Erik Baekgaard of Bang & Olufsen in 1977 [9]. It uses an extra drive unit in order to obtain a linear-phase response. This is one of those not-too-common situations where wading through the complex algebra that describes the crossover response is instructive. If you take the equation for the response of a second-order Butterworth crossover, and then compare it with the corresponding mathematical description of a linear-phase crossover, there is a missing term in the former. The idea is to add an extra drive unit, called a "filler-driver" that will supply that missing term, which is equivalent to a bandpass filter of low Q centred on the crossover frequency.

The second-order Butterworth alignment is normally used with one output inverted to get somewhere near a flat response (without frequency offset it gives a broad +3 dB hump at crossover). If the outputs are in-phase then there is a deep notch at the crossover frequency. This notch is filled in precisely by the filler drive unit, as shown in Figure 4.63, where the summed response is exactly flat, and additionally we get a flat phase response at 0° and a flat group delay response. If we assume the filler drive unit has a flat response then it must be fed via a bandpass filter that is −3 dB down an octave away from the centre frequency on each side; this corresponds to a Q of 0.6667. Ways of realising the required filter appear

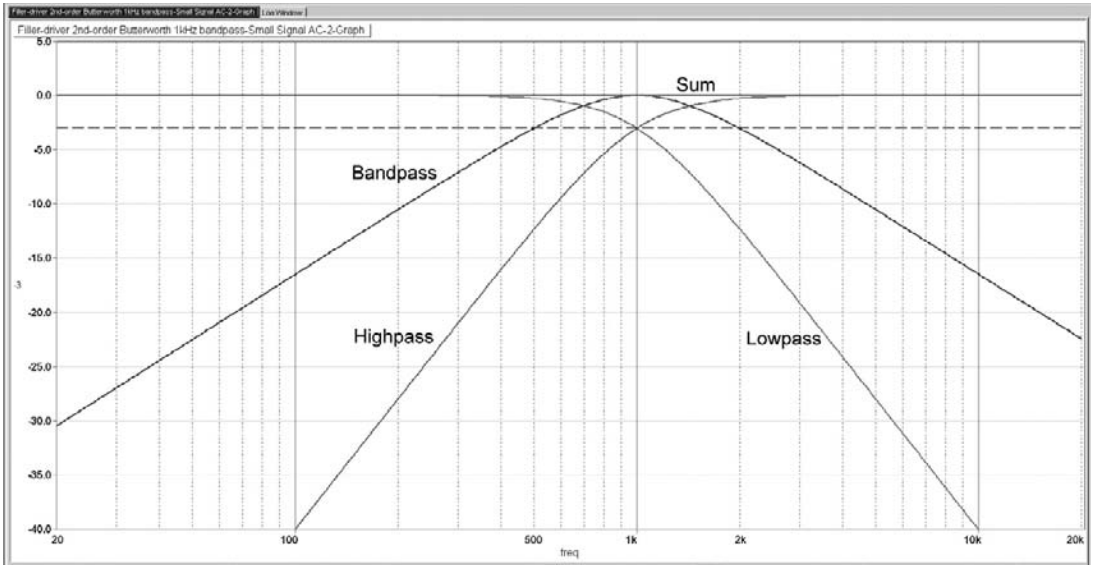


Figure 4.63: The flat amplitude response of a second-order Butterworth crossover (in-phase) with an added bandpass filler-driver.

to be strangely absent from most discussions of the filler-driver technique, so I present here a straightforward method.

Since the required Q of 0.6667 is less than 0.707, it cannot be obtained with a standard MFB (Rauch) filter. The Deliyannis version is used instead (see Chapter 10 for the design procedure); the passband gain cannot be set independently, and we find we get a loss of 1.125 times (−1.03 dB), which must be made up. MFB filters also give a phase-inversion that must be undone as the filler contribution must be in-phase; if its phase is reversed the summed response is still dead flat but it is allpass rather than linear phase, with a peak in the group delay. The A3 stage in Figure 4.64 neatly performs both corrections. Note that the value of R5 eerily echoes the resistor values in the other filters. Component tolerance errors will cause ripples in the amplitude, phase, and group delay responses.

Since the bandpass filter is a simple second-order affair, the slopes at either side of the crossover frequency are only 6 dB/octave, which places quite severe demands on the filler drive unit. The filler-driver concept can also be used with fourth-order Linkwitz–Riley crossovers [10], but the filler slopes remain at 6 dB/octave.

While the filler-driver concept is unquestionably ingenious, it has not caught on. The filler-driver takes up room on the baffle and makes it harder to get the two main drivers close together, worsening time-alignment problems. It significantly complicates an active crossover audio system, as we need an extra set of filters and an extra pair of power

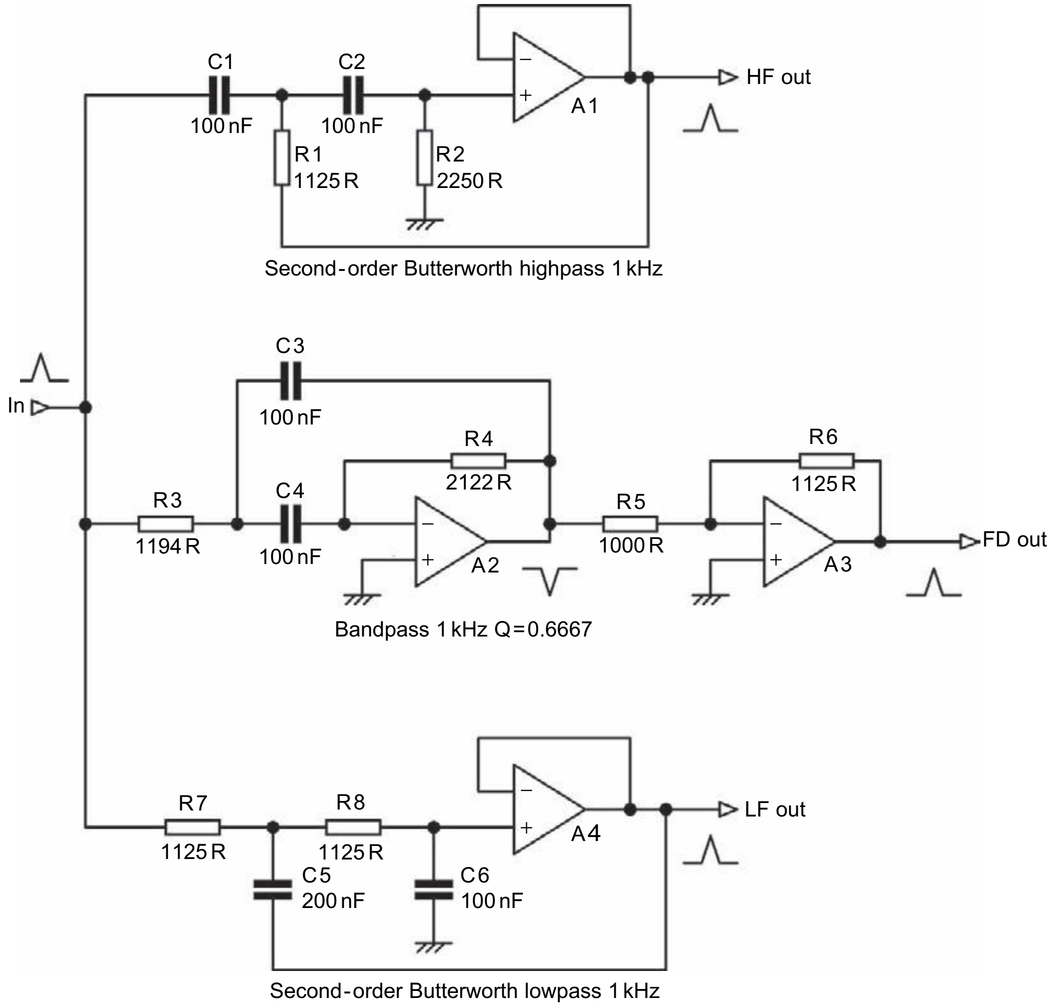


Figure 4.64: The schematic of a second-order Butterworth crossover (in-phase) with added bandpass filler-driver output.

amplifiers, not to mention the two extra drive units (assuming stereo). These will not be cheap because of the wide bandwidth of operation required, and cost is added to the loudspeaker enclosures because of the extra cut-outs, terminals, etc. that are required.

4.13 The Duelund Crossover

This idea, originated by Steen Duelund [11], is also usually regarded as a filler-driver crossover concept. It is even harder to explain without recourse to complex algebra, but the basic plan is to use two cascaded bandpass filters to generate a MID output, rather than simply filling in a notch. This output has 12 dB/octave slopes on each side rather than the 6 dB/octave slopes of the Baekgaard filler-driver. The demands on the driver concerned, however, are still quite severe. There is much more information on the Duelund crossover at [12] and [13].

4.14 Crossover Topology

For the sake of simplicity, all the crossovers we have looked at in this chapter are two-way; they divide the audio spectrum into two bands only. For the purposes of illustration, a 1 kHz crossover frequency was used throughout. However, most active crossover systems will use three- or four-way crossovers to split the audio spectrum into three or four bands, to reduce the demands on the drive units and generally get the full benefits of active crossover technology. This introduces some extra complications.

It would appear to be very simple to make a three-way crossover by combining two two-way crossovers. We will use our usual example crossover specification with an LF/MID crossover frequency of 400 Hz and a MID/HF crossover frequency of 3 kHz. There are two highpass filters and two lowpass filters, and the standard way of connecting them is shown in Figure 4.65.

It has however been pointed out by several people, amongst them Siegfried Linkwitz in Linear Audio Volume 0 [13] and in [14], that this crossover topology is defective. If all three outputs are summed, they do not sum to exactly flat. This is best illustrated when two crossover frequencies are close together. Consider the fourth-order Linkwitz–Riley

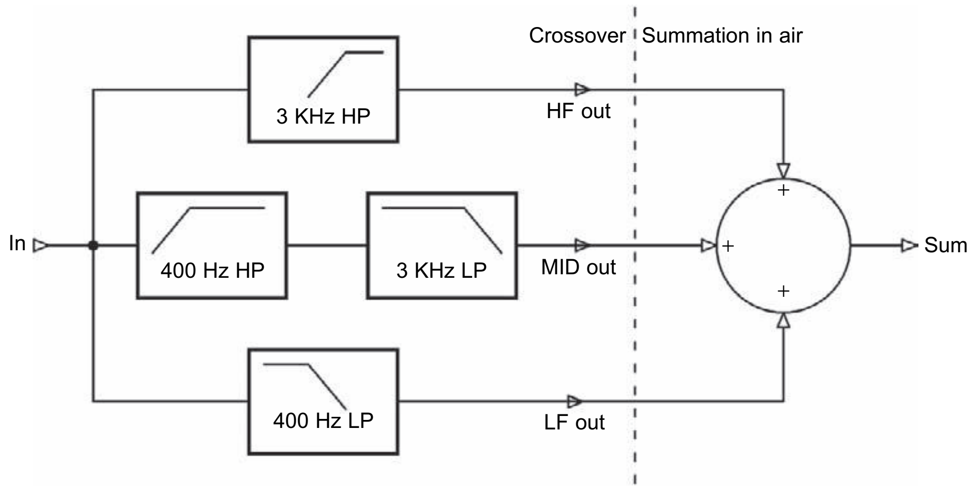


Figure 4.65: A three-way crossover made by combining a 400 Hz two-way crossover with a 3 kHz two-way crossover. In-line topology.

crossover in Figure 4.66, where the MID crosses over to the LF at 200 Hz, but this crosses over again to a sub-woofer at 50 Hz; the rest of the crossover is left out to keep things simple. This is what we might call an in-line crossover, because the filters follow one after the other; the outputs from it are shown in Figure 4.67, where you will note that the LF output never gets up to 0 dB as the crossover frequencies are only two octaves apart. As a check we sum the subwoofer output SUB with the LF output of the 50 Hz highpass filter

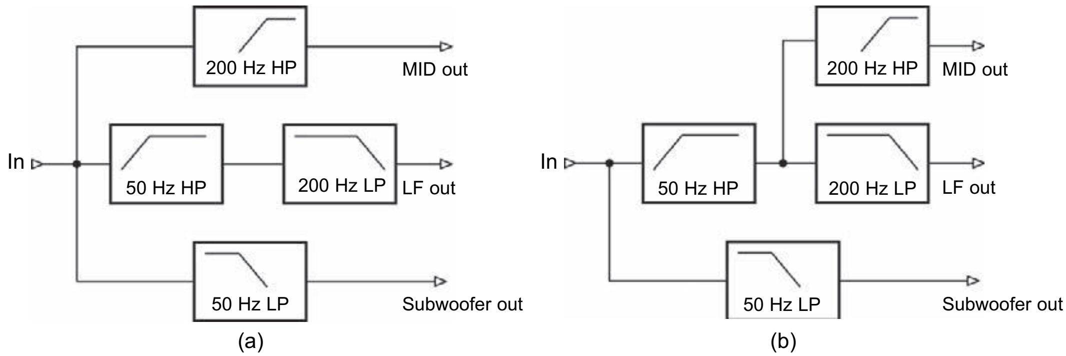


Figure 4.66: Three-way crossovers with frequencies at 50 Hz and 200 Hz: (a) in-line topology; (b) branching topology.

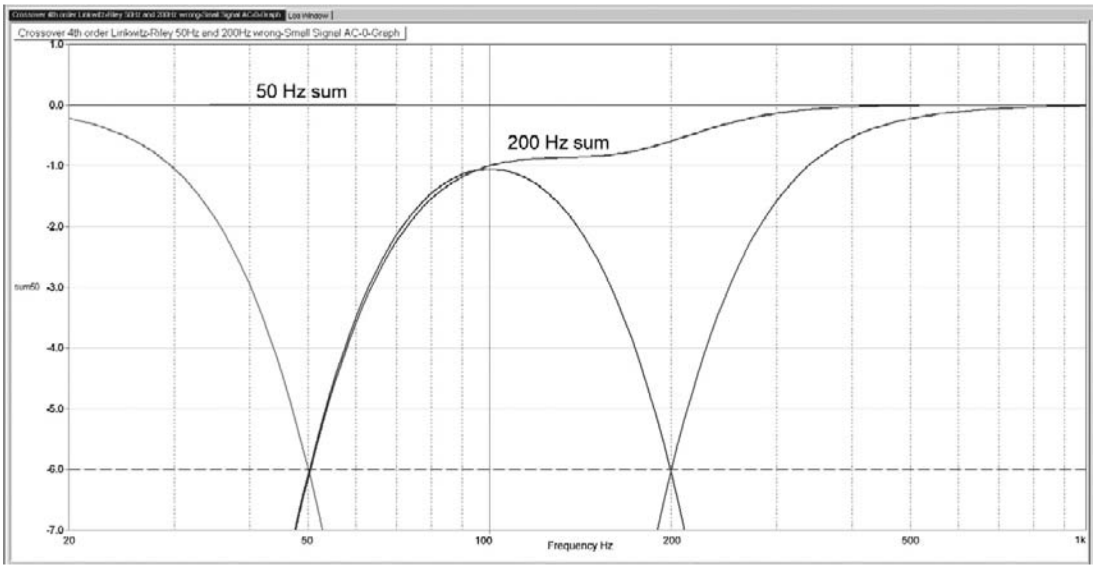


Figure 4.67: Summed LF and MID outputs with in-line filter topology as in Figure 4.66. The 200 Hz sum response is bent out of shape.

(before it has been through the 200 Hz lowpass filter) and we get the dead flat line labelled "50 Hz sum," which proves that the 50 Hz Linkwitz–Riley filters are doing their stuff.

Next, we sum the MID and LF outputs; we expect an output that will rolloff around 50 Hz, as we have not included the subwoofer output. What we get is labelled "200 Hz sum" in Figure 4.67 and I think you can see at once that it has a warped look which indicates something is wrong. The problem is that the LF signal has passed through the 50 Hz highpass filter, while the MID signal has not. 50 Hz is only two octaves away from 200 Hz, and while the amplitude response of the 50 Hz highpass filter is only down by 0.53 dB at 100 Hz and 0.03 dB at 200 Hz, the phase-shift is still 40° short of its ultimate value of 360°, and this spoils the MID-LF summation.

The cure for this problem is to rearrange our in-line crossover into a branching topology, as in Figure 4.66. Now both signals going to the 200 Hz filters have passed through the 50 Hz highpass filter, and there is no extra relative phase-shift to mess things up. The result is seen in Figure 4.68 where the "200 Hz sum" now has the response we expect. The error with the in-line topology reaches a maximum of –0.74 dB at 150 Hz.

That demonstrates the phase-shift effect. But what happens when we sum all three outputs, as we will in real life? Figure 4.69 gives the unwelcome answer that something is now amiss with the summation of the SUB and LF outputs; the summed level being 0.8 dB low

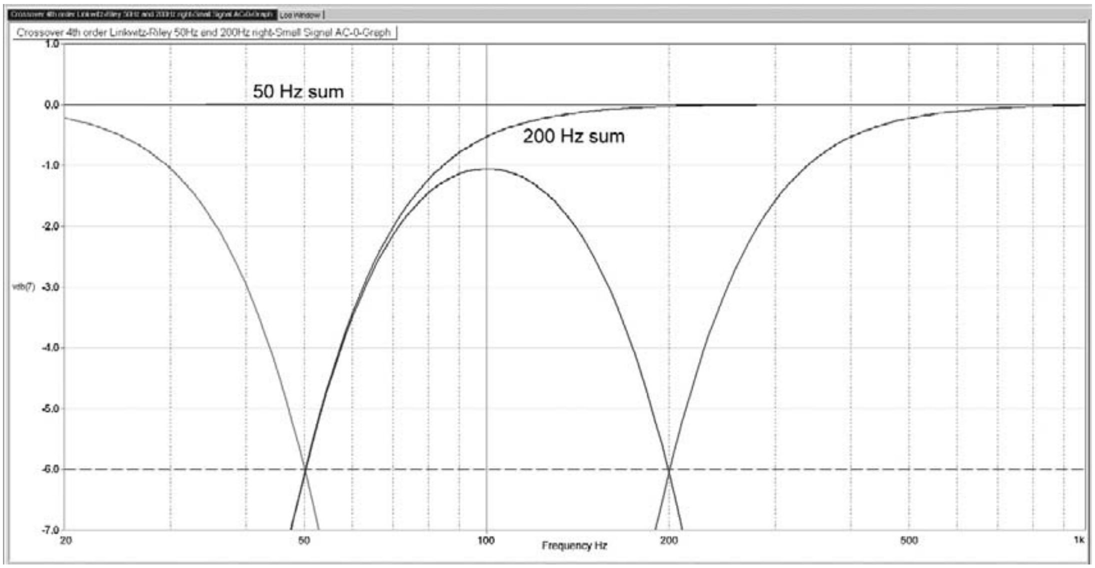


Figure 4.68: Summed LF and MID outputs with branching filter topology as in Figure 4.66. The 200 Hz sum response is now as expected.

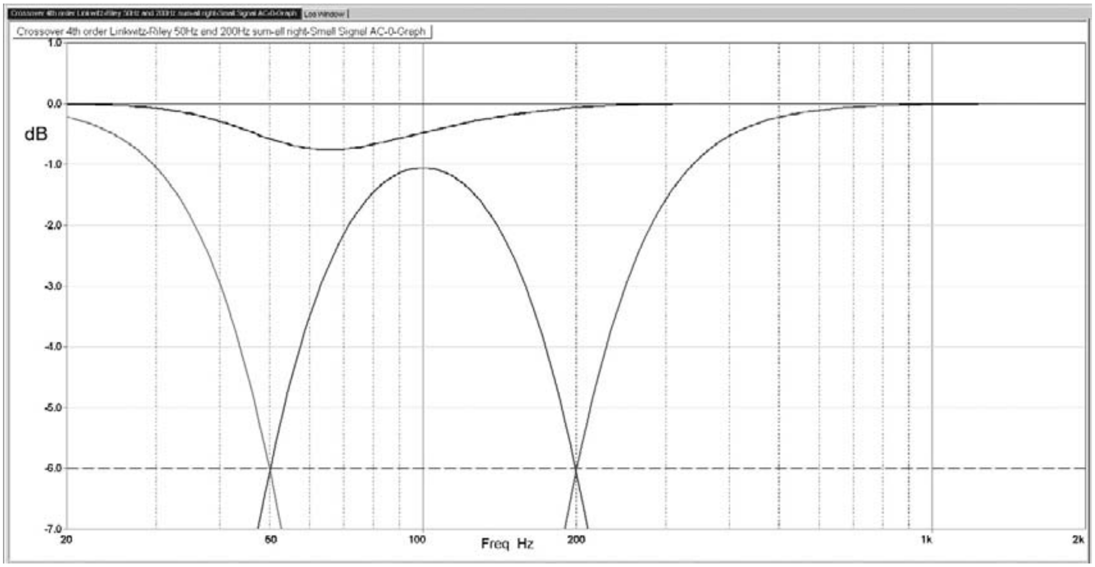


Figure 4.69: Summed SUB, LF, and MID outputs with branching filter topology as in Figure 4.66b.

around 65 Hz. The problem is similar to the one we have just solved; the LF output has been phase-shifted by going through the 200 Hz lowpass filter, but the SUB output has not. Going back to the in-line topology makes it worse; the sag in the summed response is now slightly deeper and much wider, extending from 40 Hz to 200 Hz.

Before we panic, we must recognise that bringing in a subwoofer output as a demonstration deliberately put the crossover frequencies as close together as is plausible, and most crossovers will have a much greater frequency spacing. If we go back to our example crossover frequencies of 400 Hz and 3 kHz, using the branching topology as in Figure 4.70, and sum all three outputs, then we get the much smaller error of a −0.20 dB dip around 530 Hz. If we try the alternative branching topology in Figure 4.70, then there is still a −0.20 dB dip, but now around 2.2 kHz, as in Figure 4.71. The in-line topology gives the same depth of dip but it is once more much wider, extending between the crossover frequencies.

Clearly the branching topology is superior, but we do seem to have uncovered an inherent problem with three-way crossovers. The MID output must go through two filters, and one of the other outputs can be put through two filters by branching; but no matter which of the two ways in Figure 4.70 you choose, the remaining output of the three can only go through one filter. There will always be a phase-shift problem, though fortunately an error of 0.20 dB is going to be negligible compared with driver irregularities.

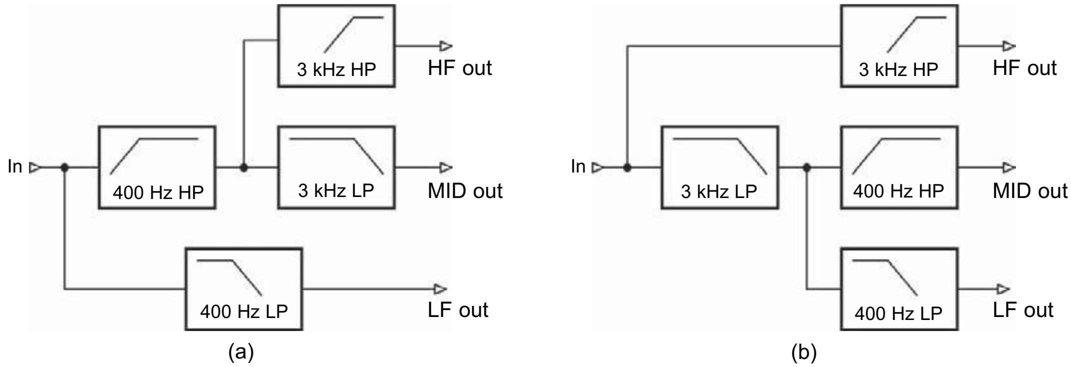


Figure 4.70: The three-way crossover of Figure 4.65 converted to the branching filter topology in two different ways.

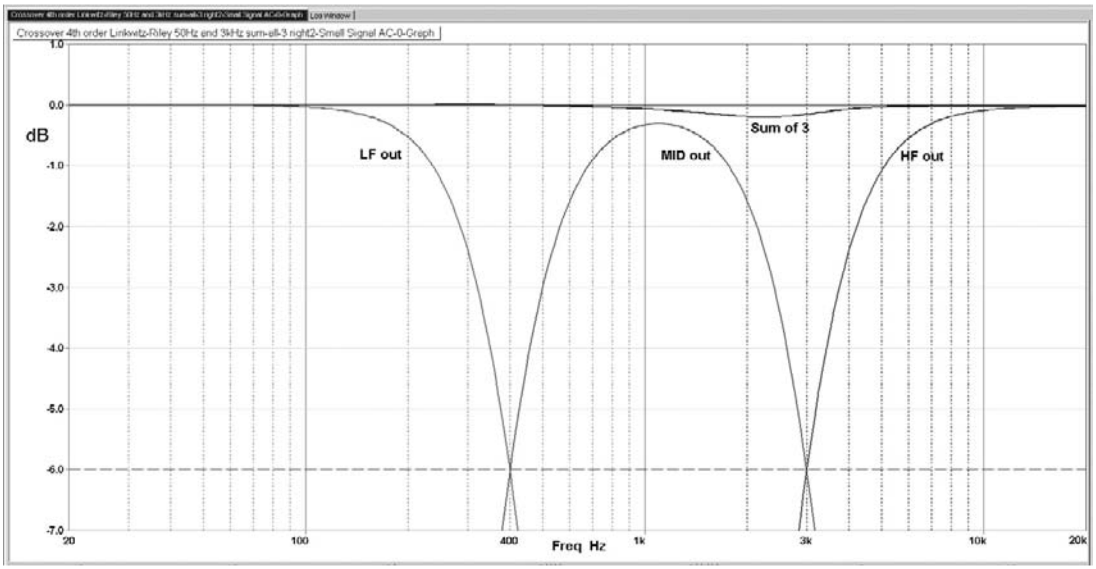


Figure 4.71: The summed LF, MID, and HF outputs, with branching crossover topology as in Figure 4.70, show a −0.2 dB dip around 2.2 kHz.

Earlier in this chapter we made extensive use of offsetting the filter frequencies to minimise response deviations. We can do that here. Applying an offset of 0.961 times to the 3 kHz highpass filter and moving its cutoff down to 2.88 kHz, converts the –0.20 dB dip into an innocuous ±0.11 dB ripple in the summed response, as in Figure 4.72. I think we can live with that.

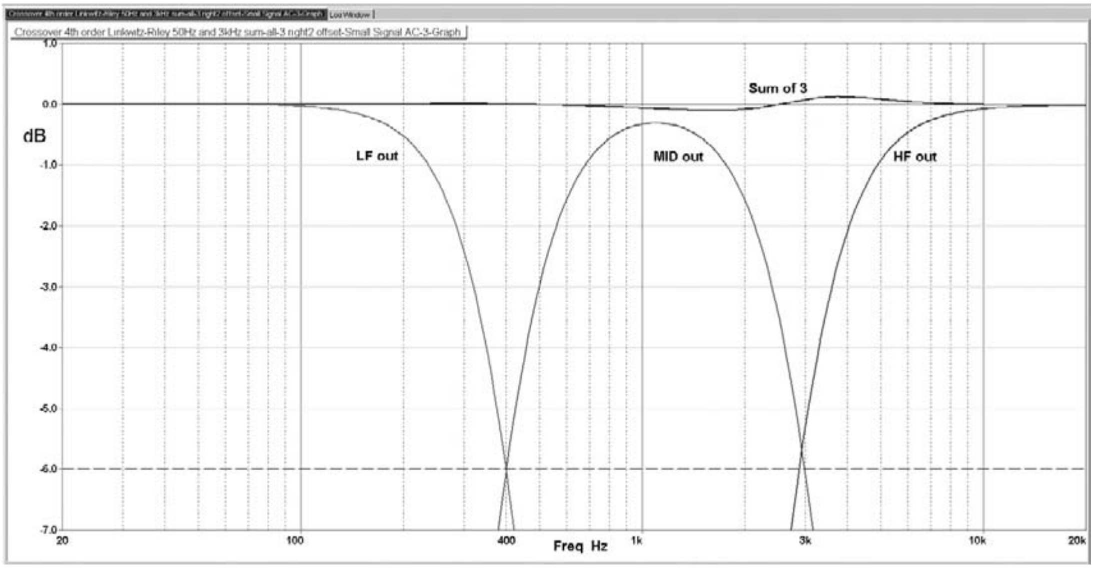


Figure 4.72: The summed LF, MID, and HF outputs as in Figure 4.71, but with the HF 3 kHz highpass filter cutoff frequency offset by 0.961 times to 2.88 kHz.

4.15 Crossover Conclusions

This has been a long chapter, and we have looked at a lot of different crossover types, with yet more to be found in Chapter 5 on notch crossovers and Chapter 6 on subtractive crossovers. To briefly summarise the best ones:

First-order crossovers have tempting properties but put great demands on the drive units. The second-order Linkwitz–Riley crossover sums flat but does not have adequate slopes. The third-order Butterworth crossover sums flat but shows lobing errors. The third-order Linkwitz–Riley crossover does not sum flat. The fourth-order Linkwitz–Riley crossover sums flat with no lobing errors and good slopes. Of the types not yet examined, I draw your attention to the very clever NTM crossover in Chapter 6.

References

[1] R.H. Small, Constant voltage crossover network design, JAES 19 (1) (1971) 12–19.
[2] V. Dickason, The Loudspeaker Design Cookbook, seventh ed., Audio Amateur Press, Peterborough, NH, 2006, ISBN: 1882580478.
[3] KEF, http://www.kef.com/GB/Loudspeakers/QSeries2010/Floorstanding, 2010 (accessed 12.10).
[4] R. Miller, A Bessel filter crossover, and its relation to others, Rane Note 147, Rane Corporation, West Mukilteo, WA, USA, 2006.
[5] KEF, KEF 105 description: http://www.kef.com/history/en/1970/model105.asp, 2010 (accessed 12.10).

[6] C. Strahm, Linear-X Systems Filtershop, http://www.linearx.com/products/software/FilterShop/Fshop_01.htm, 2010 (accessed 12.10).
[7] D. Bohm, Linkwitz-Riley crossovers: a primer, Rane Note 160, Rane Corporation, West Mukilteo, WA, USA, 2005.
[8] cdtAudio, http://www.cdtaudio.com/crossovers_08.htm, 2010 (accessed 12.10).
[9] E. Baekgaard, A novel approach to linear phase loudspeakers using passive crossover networks, JAES 25 (5) (1977) 284–284 (filler-driver).
[10] A. Megann, http://www.soton.ac.uk/~apm3/diyaudio/Filler_drivers.html, 1999 (Second and fourth-order crossovers with filler-drivers), 2010 (accessed 12.10).
[11] S. Duelund, http://www.duelundaudio.com/Articles_by_Steen_Duelund.asp, 2010 (accessed 12.10).
[12] J. Kreskovsky, http://www.musicanddesign.com/Duelund_and_Beyond.html, 2010 (accessed 12.10).
[13] S. Linkwitz, http://www.linkwitzlab.com/crossovers.htm (Duelund crossover).
[14] S. Linkwitz, STEREO—from live to recorded and reproduced, Linear Audio vol. 0, p. 109. Published by Jan Didden 2010. http://www.linearaudio.nl/ (crossover topology), 2010 (accessed 12.10).

This page intentionally left blank

CHAPTER 5

Notch Crossovers

The crossovers examined in Chapter 4 are all-pole crossovers, which means that the filters used are relatively simple lowpass and highpass types, though of varying order and filter characteristic (Butterworth, Bessel, etc.). It is also possible to contrive crossovers that have notches (or to get mathematical, zeros) built into the rolloff, typically giving a much steeper filter slope, to begin with at least, than an all-pole crossover of a practical order. This can be very useful when drive units that are otherwise acceptable misbehave badly when taken just outside their intended operating range. Neville Thiele [1] gives the example of a horn loudspeaker being used near its cutoff frequency. He also cites the case of a mono subwoofer, where its contribution must be rolled-off as quickly as possible out of band to prevent it contaminating the stereo localisation cues from the main speakers.

5.1 Elliptical Filter Crossovers

A crossover design based on the use of elliptical filters was published by Bill Hardman in Electronics World in 1999 [2] and this has been the basis for much discussion on the subject. The amplitude response of the published design is shown in Figure 5.1. The original version had a crossover point at 1.5 kHz, but I have modified the filter frequencies for 1 kHz instead, to match all the other crossover response examples in this book. Nothing else that might affect the response has been changed.

The two deep notches symmetrically placed on either side of the crossover point are at 529 Hz and 1.84 kHz, less than an octave away from the crossover point. It is obvious that their presence greatly increases the initial roll-off slope, making it more effective than a fourth-order Linkwitz–Riley alignment from this point of view.

It is, however, vital that the outputs of a crossover should sum to as nearly flat as possible, and looking at the spiky nature of Figure 5.1, you are probably thinking that this is going to be very difficult to arrange. In fact, it can be done relatively easily. Figure 5.2 shows the in-phase summation; the central hump is only –0.13 dB down, while the dips on either side are about –0.9 dB deep. The phase-reversed summation has a deep notch at the crossover frequency and is of no value.

Looking at Figure 5.2, one cannot help feeling it might be able to improve the flatness of the summation by tweaking the filter frequencies. However, applying a frequency offset is in fact not very helpful, because the centre rises much more than the dips do; for example,

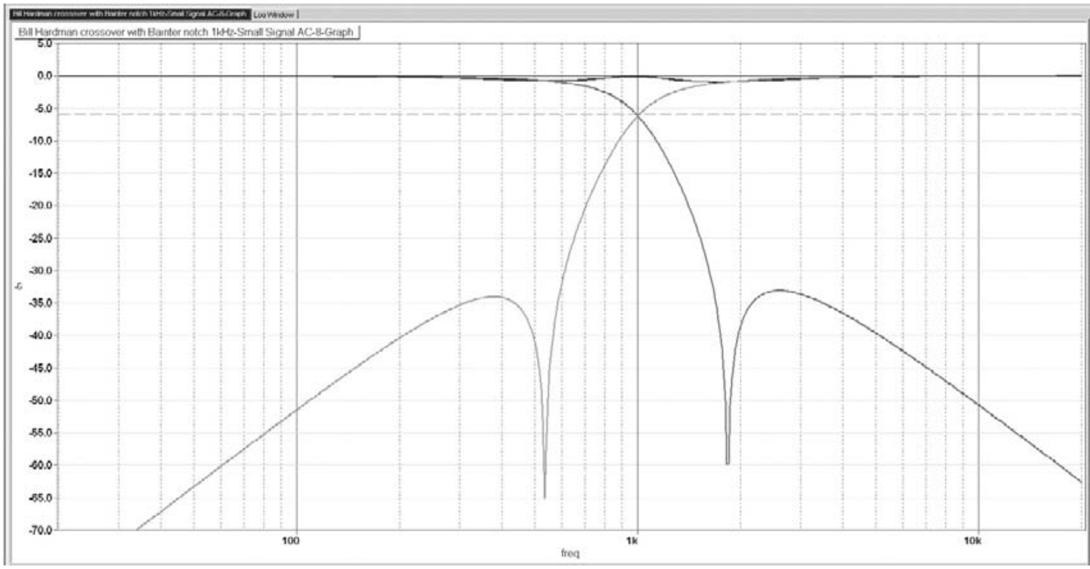


Figure 5.1: Amplitude response of the Hardman elliptical crossover.

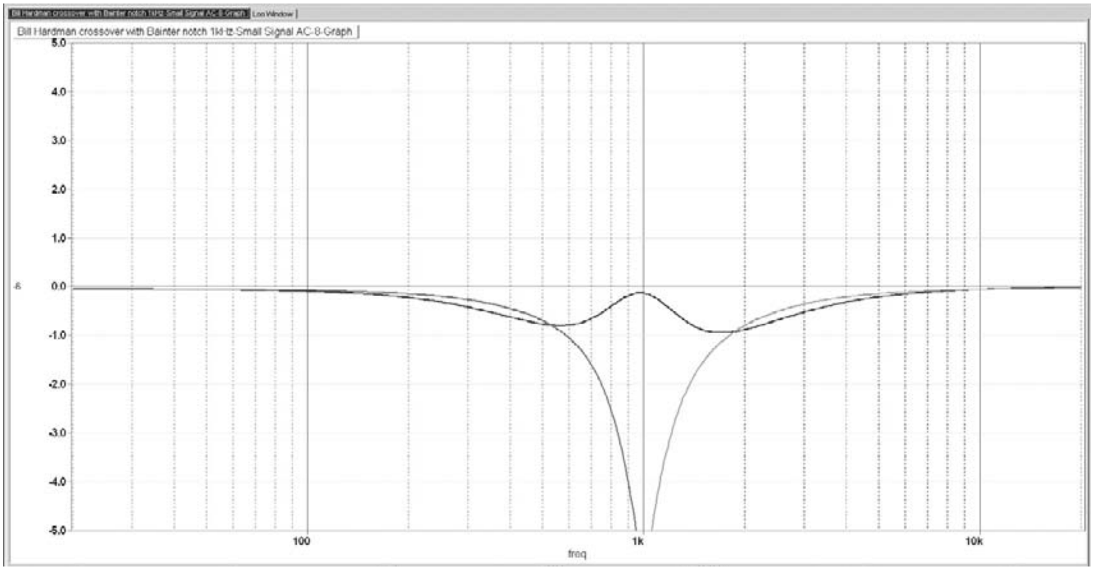


Figure 5.2: Zooming in on the in-phase summation of the Hardman elliptical crossover.

an offset ratio of 1.045 gives a central peak of +0.4 dB with dips of –0.7 dB, which is not much improvement, if any, on the original Hardman alignment.

The elliptical filters used in this crossover are the usual combinations of notch filters and all-pole filters. Figure 5.3 shows that the lowpass path is composed of a lowpass notch and

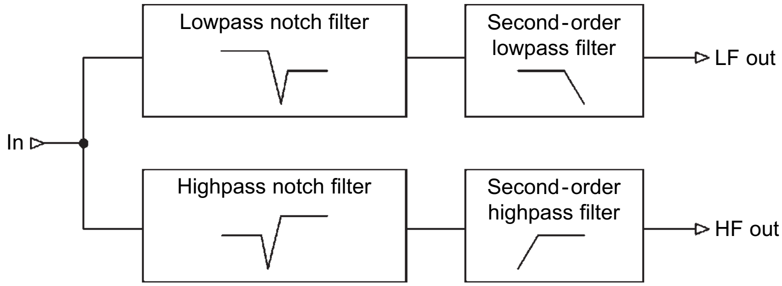


Figure 5.3: Block diagram of the Hardman elliptical crossover architecture.

a second-order lowpass Sallen & Key filter; the latter has equal component values but a non-standard Q which places it somewhere between a Linkwitz–Riley and a Bessel filter characteristic. The highpass path is composed of a highpass notch and a second-order highpass Sallen & Key filter with the same non-standard value of Q.

Figure 5.3 underlines the difference between a lowpass notch and a highpass notch. These are not your more familiar symmetrical notches that go up to 0 dB on either side of the central crevasse. A lowpass notch response starts out at 0 dB at low frequencies, drops down into the crevasse, and then comes back to level out at a lower gain, say –10 dB; the lowpass notch can be seen in both Figures 5.3 and 5.4. Conversely, a highpass notch has a response that is 0 dB at high frequencies, drops down, and then comes back up to say –10 dB at low frequencies.

Figure 5.4 shows how it works for the lowpass path. The notch filter provides a lowpass notch; as frequency increases there is some peaking just before the rolloff into the crevasse. This is cancelled out by the low-Q, relatively slow rolloff from the all-pole lowpass filter, giving a fast rolloff around the crossover frequency, and also gives a notch deepened by the increasing attenuation of the lowpass filter. The ultimate rolloff slope, well above the notch frequency, is only 12 dB/octave because it comes entirely from the second-order lowpass filter.

The original Hardman crossover used Bainter filters to create the notches followed by second-order Sallen & Key filters. Bainter filters are popular for making elliptical filters of the type shown here because they give good deep notches where the depth does not depend on the matching of passive components, but only on the open-loop gain of the opamps. You can see lovely deep notches in Figures 5.1 and 5.4; it is in a sense a pity that the notch depth is not in any way critical to get a good summed response. So long as a distinct notch is visible the effect on the summed response is negligible.

A design for a Hardman crossover is shown in Figure 5.5. While the essential filter characteristics are closely based on those in the original article [2], the crossover frequency, as noted, has been altered from 1.5 kHz to 1.0 kHz, and the general impedance level of the circuitry reduced by a factor of approximately twenty, to reduce noise. Any further

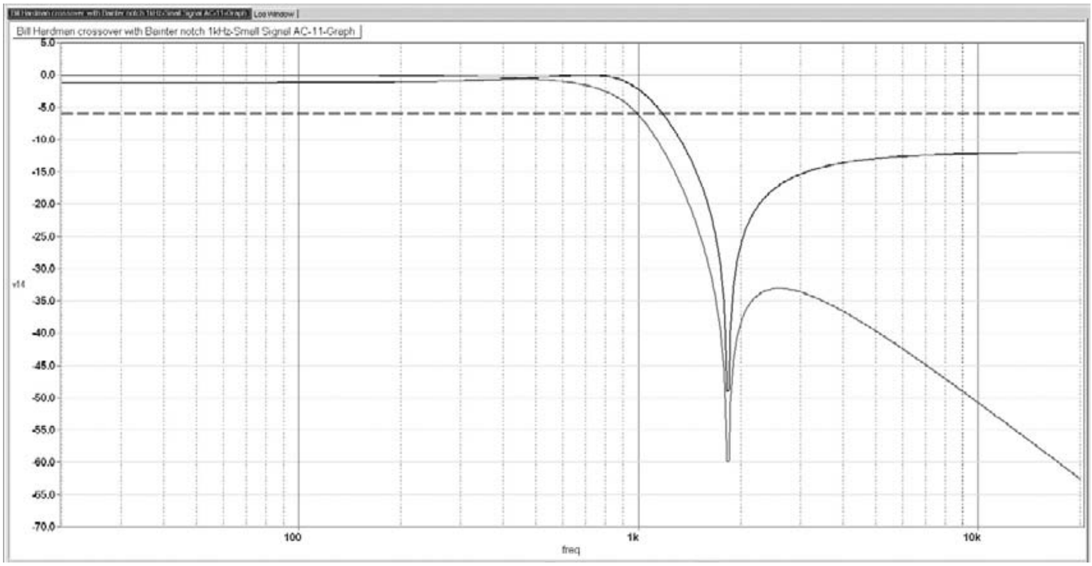


Figure 5.4: How the Hardman elliptical lowpass filter works. The upper plot is the output of the lowpass notch filter only, the second plot adds in the effect of the lowpass filter.

impedance reduction would risk overloading the opamps (assumed to be 5532s) with a consequent increase in distortion. As with many other circuits in this book, it has been designed to use standard capacitor values, with the resistor values coming out as whatever they do. It is much cheaper to get an exact value by combining resistors rather than by combining capacitors. The simplest way to scale the circuit to implement other crossover frequencies is to change all the capacitors in the same ratio.

Figure 5.5 shows that the two Bainter notch filters for lowpass and highpass are very similar, with only three components differing in value. This sort of convenient behaviour is what makes the Bainter filter so popular. The relationship between R3 and R4 in the lowpass filter, and between R13 and R14 in the highpass filter, determine the kind of notch produced. If R4 is greater than R3, you get a lowpass notch. If R3 = R4, you get the standard symmetrical notch. If R3 is greater than R4, you get a highpass notch. The value of R6 (or R16) sets the notch Q.

There are, however, some aspects of this circuit that are less convenient. The lowpass notch response does not actually start out at 0 dB at low frequencies; instead it has a gain of +10.6 there. The response then drops into the crevasse, and comes back up to 0 dB. There is then the gain of +1.3 dB from the lowpass filter, giving a total of 11.9 dB of passband gain, which may not fit well into the gain/headroom scheme planned for the crossover. The highpass notch response has a passband gain of 0 dB at high frequencies, and after the notch comes back up to –10.8 dB. With the final highpass filter added the passband gain is +1.2 dB.

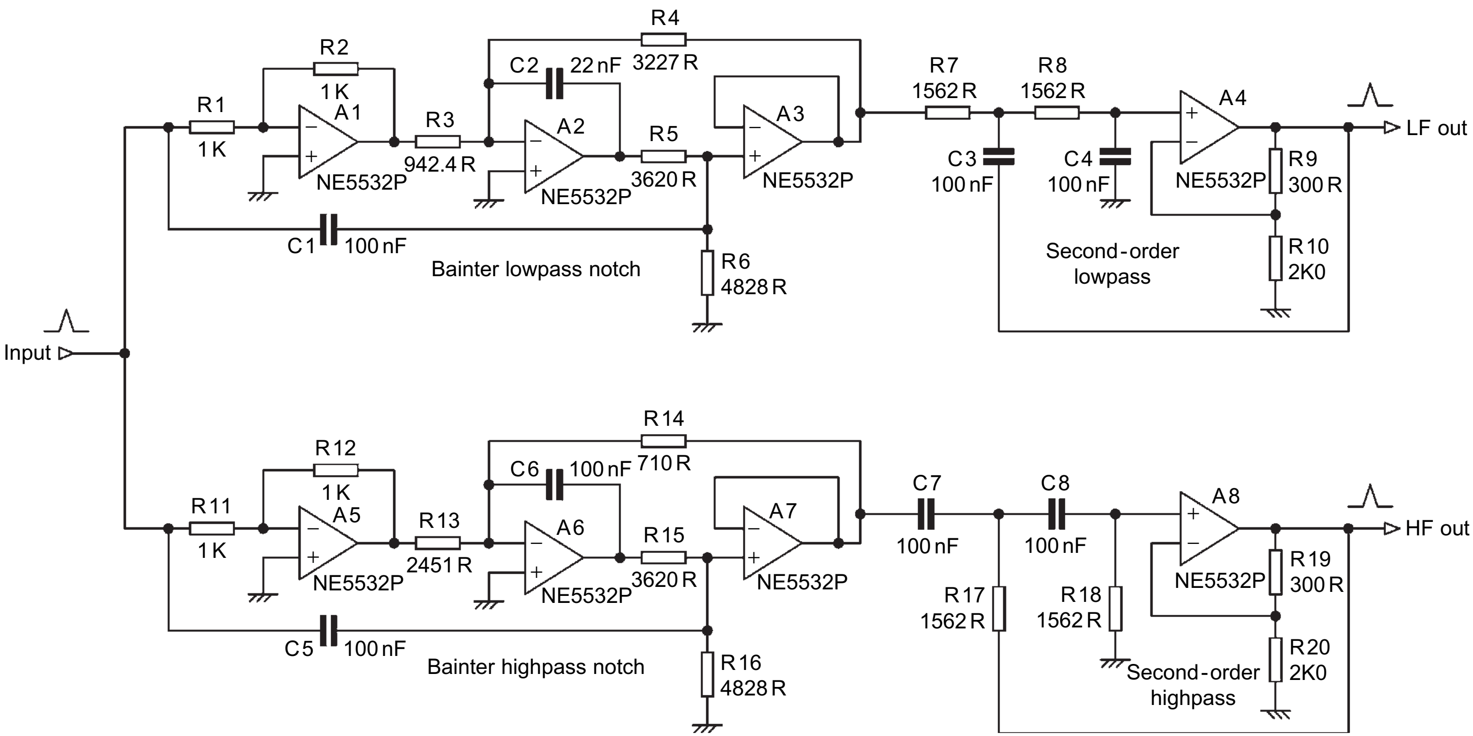


Figure 5.5: Schematic diagram of a Hardman elliptical crossover with a crossover frequency of 1 kHz. The outputs are at different levels—see text for details.

Therefore we have a level difference of 11.9 – 1.2 = 10.7 dB between the two outputs, and this will have to be accommodated somewhere in the crossover system design. There is more information on the Bainter filter and other notch filters in Chapter 9.

A most interesting paper on the use of Chebyshev filters (ripples in the passband), inverse-Chebyshev (notches in the stopband), and elliptical filters (ripples in the passband *and* notches in the stopband) with the added feature of a variable crossover frequency, was published in the JAES by Regalia et al. in 1987 [3]. It is well worth studying.

5.2 Neville Thiele Method™ (NTM) Crossovers

One of the better-known notch crossovers is the Neville Thiele Method™ crossover, introduced by Neville Thiele in an AES paper in 2000 [1]. This does not appear to consist of elliptical filters as such (as far as my knowledge of elliptical filters goes, anyway), but a rather more subtle arrangement that sums to unity much more accurately than the Hardman crossover we have just looked at. One of the few examinations of this technique that has been published is that by Rod Elliot [4].

I should say at once that the Neville Thiele Method™ or NTM is a proprietary technology addressed by US Patent 6,854,005 and assigned to Techstream Pty Ltd, Victoria, AU, and that if you plan to use it for anything other than a private project you might want to talk to them about licensing issues. The information given here is published by permission and is derived solely from the public-domain references [1] and [5], which I have to say are not an easy read. I have never seen a schematic of a manufactured crossover using this technology, nor have I ever deconstructed any related hardware.

Using references [1] and [4], it appears that the lowpass path of a sixth-order NTM crossover filter (eighth-order versions are also possible) consists of a bridged-T lowpass notch filter, followed by a second-order lowpass filter with a Q of about 1.6; this is followed in turn by two first-order filters. This structure, together with the matching highpass path, is shown in Figure 5.6.

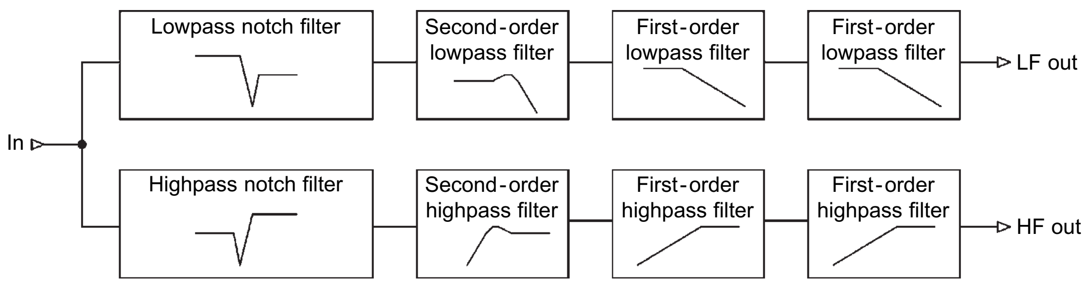


Figure 5.6: Block diagram of an NTM crossover.

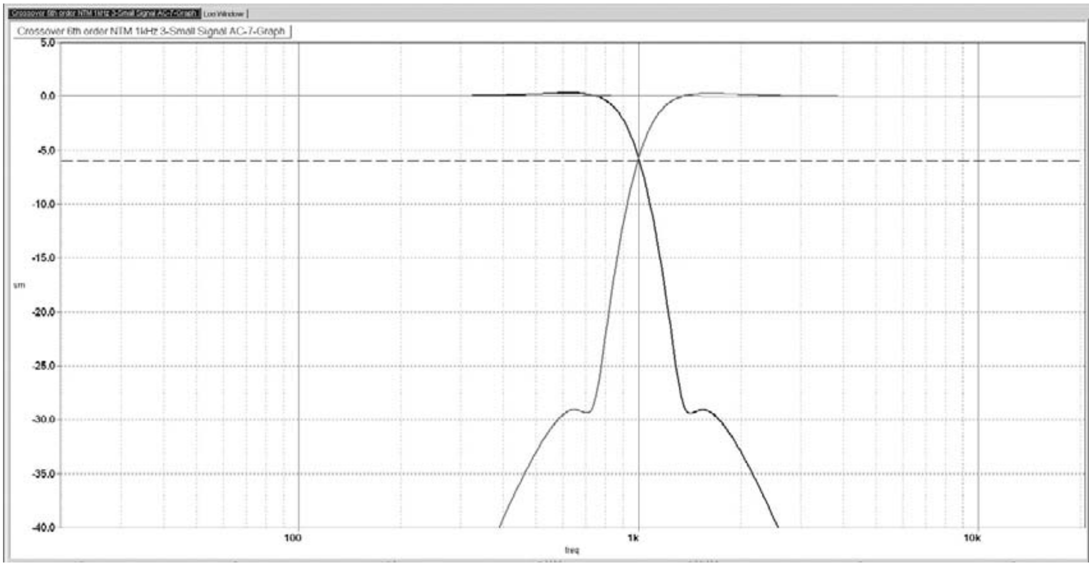


Figure 5.7: The NTM crossover. The dotted line is –6 dB.

The result of this rather complicated-looking block diagram is shown in Figure 5.7. Each filter output has a fast rolloff after the crossover point, terminating in a shallow notch; the response then comes back up a bit but then settles down to an ultimate 24 dB/octave rolloff. The filter responses may not look very promising for summation, but in fact they do add up to an almost perfectly flat response when the phase of one output is reversed. If summed in-phase there is a central crevasse about 12 dB deep, of no use to anyone.

Figure 5.8 shows a much closer view of the summed response. The bump below the crossover frequency peaks at +0.056 dB, while the flat error in level above the crossover frequency is at –0.031 dB. Since these very small errors are asymmetrical about the crossover point, it seems more likely that they are due to opamp limitations or similar causes, rather than anything inherent in the crossover. They are negligible compared with transducer tolerances, or with the summation errors of crossovers that approximate flatness by using frequency offsetting; the performance is much better than that of the Hardman elliptical crossover, which has flatness errors of 0.9 dB.

Figure 5.9 attempts to show how the NTM crossover works; the response of each filter stage in the lowpass path is shown separately. (The responses of the two first-order filters have been combined into a single response for the two when cascaded.) The bridged-T filter creates a lowpass response that comes back up to –5 dB after it has been down in the notch. You will note that the notch is much shallower than that of the more complex Bainter filter

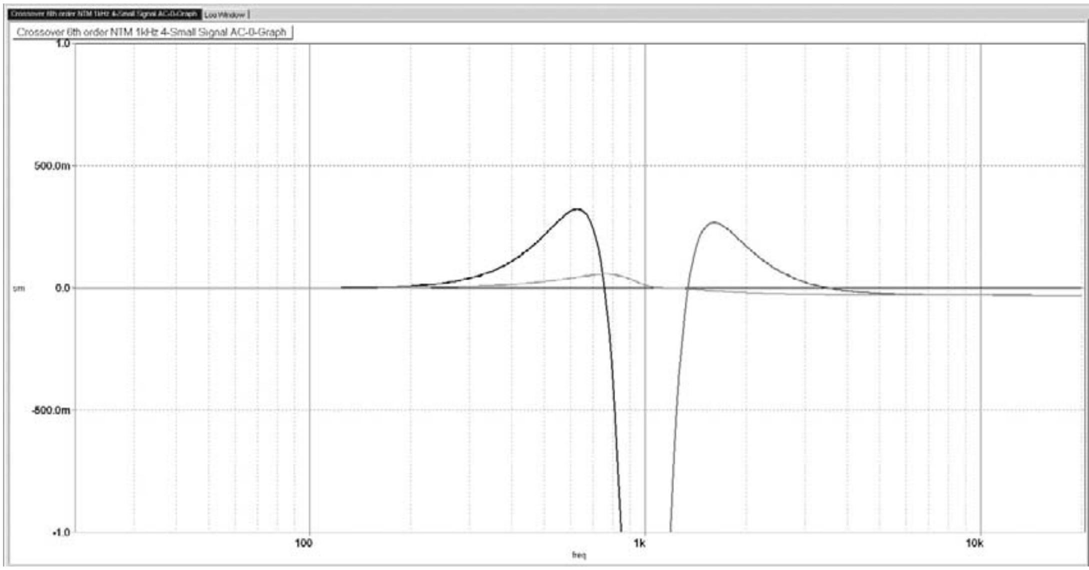


Figure 5.8: Zooming in on the summation of the NTM crossover, showing very small errors. The vertical scale is ±1 dB.

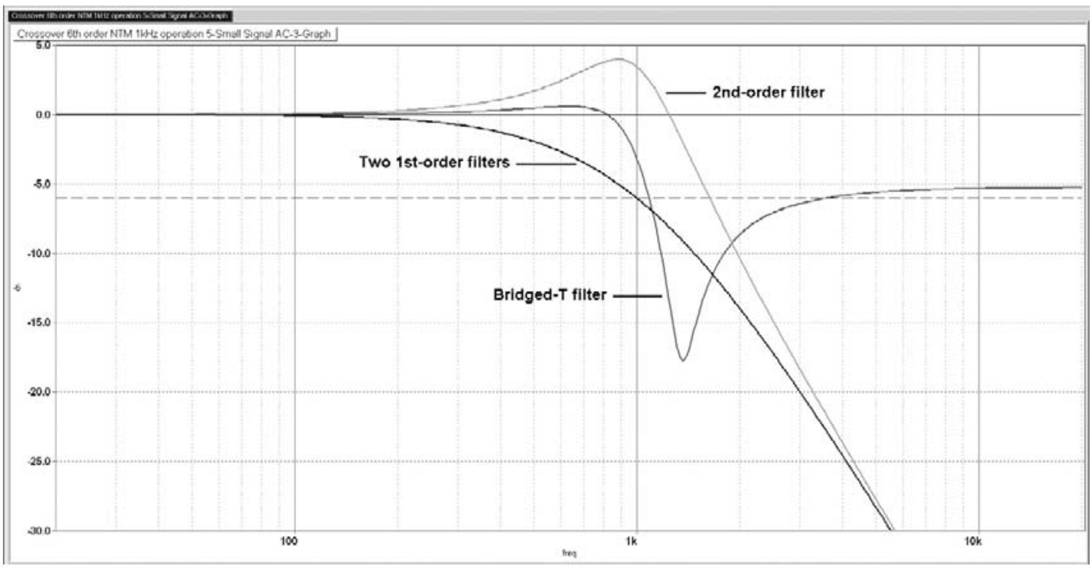


Figure 5.9: The operation of a lowpass of an NTM crossover filter, which combines the lowpass notch, a peaking second-order lowpass filter, and two cascaded first-order lowpass filters. (The last is shown as one plot here.)

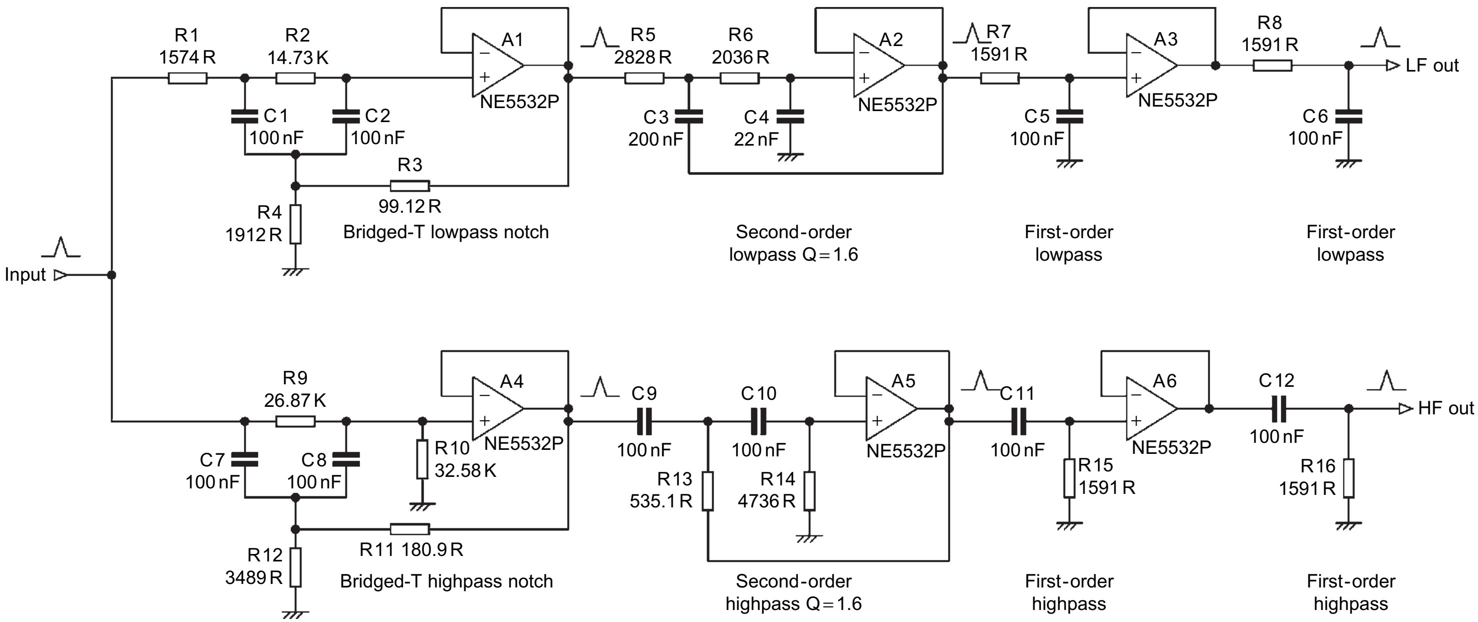


Figure 5.10: Schematic of my NTM implementation with a crossover frequency of 1 kHz.

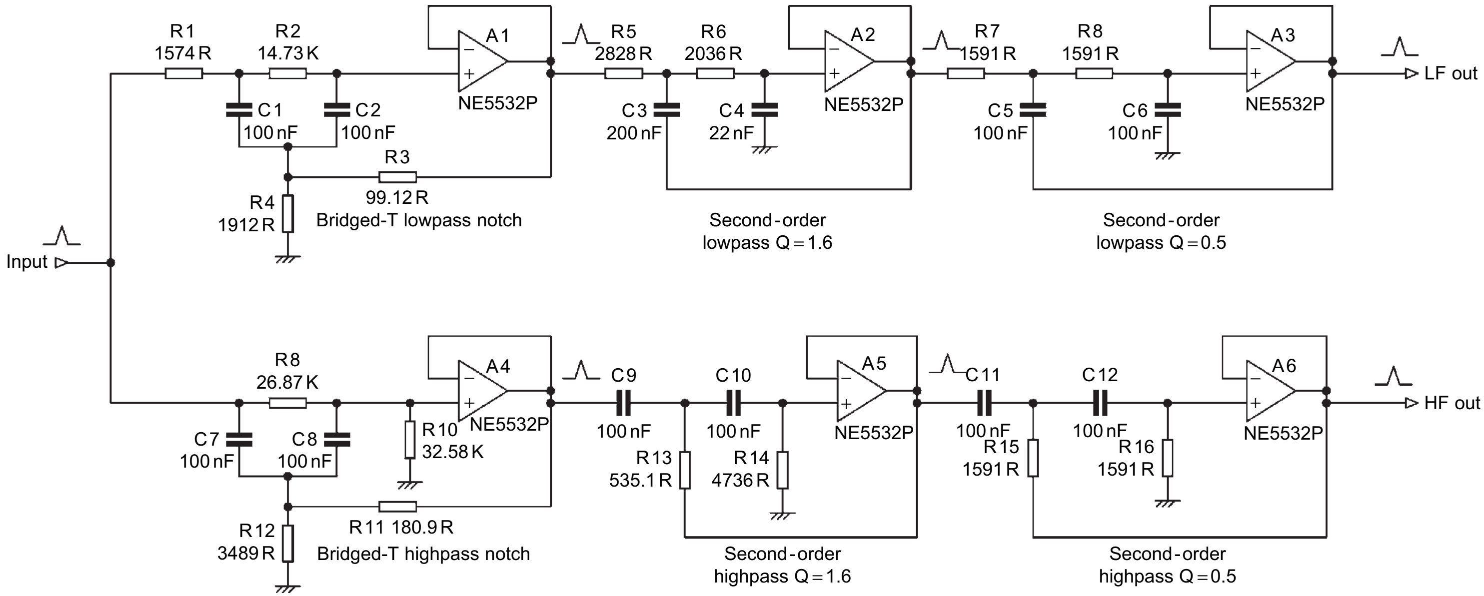


Figure 5.11: Replacing the two first-order filters with an equivalent second-order Sallen & Key stage. This saves an opamp section as no output buffer is required.

used for the Hardman crossover, but this has only a very small effect on the summed response, and the simplicity of the bridged-T circuit is welcome. You can see how the response is –3.0 dB at 1 kHz.

The next stage in the crossover is a second-order lowpass filter with a cutoff frequency of 1.0 kHz and a Q of approximately 1.6. This is quite a high Q for a second-order filter and gives considerable peaking of the response before the ultimate 12 dB/octave rolloff.

The two first-order filters both have a cutoff frequency of 1.0 kHz, and so their combined response is down –6 dB at 1.0 kHz, and also has an ultimate slope of 12 dB/octave. The action of these two stages together could be described as that of a synchronous filter, as described in Chapter 7. If implemented as RC networks they must be separated by a suitable unity-gain buffer to give the correct response.

Figure 5.10 shows the schematic of my version of an NTM crossover based wholly on the information given in [1] and [4].

An interesting point is that the two first-order filters can be combined into a single Sallen & Key second-order stage with a Q of 0.5. As Figure 5.11 shows, the component values used are exactly the same, and you might wonder if anything is gained by doing this. The answer is a resounding yes; Figure 5.10 is incomplete in that it does not show output buffers after the final first-order filters. These will be required to drive whatever equalisation or output networks follow the crossover, because the final RC network must not be loaded if it is to give the correct response. Using a Sallen & Key second-order stage makes this output buffer unnecessary and so saves an opamp section. This modification is shown in Figure 5.11.

References

[1] N. Thiele, Loudspeaker crossovers with notched responses, JAES 48 (9) (2000) 784.
[2] B. Hardman, Precise Active Crossover, Electronics World, London, 1999, p. 652.
[3] Regalia, Fujii, Mitra, Sanjit, Neuvo, Active RC crossover networks with adjustable characteristics, JAES 35 (1/2) (1987) 24–30.
[4] R. Elliot, http://sound.westhost.com/articles/ntm-xover.htm, 2010 (accessed 12.10).
[5] N. Thiele, Crossover Filter System and Method, US Patent 6,854,005, 2005 (Assigned to Techstream Pty Ltd, Victoria, AU).

CHAPTER 6

Subtractive Crossovers

6.1 Subtractive Crossovers

What you might call the standard crossover architecture has a lowpass filter to generate the LF signal for the bass drive unit, and a corresponding highpass filter to create the HF signal for the tweeter/midrange driver. (Obviously that only describes a two-way crossover and a three-way crossover has more filters.) That is not, however, the only way to do it, and one of the alternatives is the subtractive crossover. In this you just have one filter, say the lowpass which gives the LF signal, and you make the HF signal by subtracting the LF signal from the original input, so that HF = 1 − LF. This process is illustrated in Figure 6.1, which shows a first-order subtractive crossover. Higher-order subtractive crossovers can be implemented by replacing the first-order filter with a higher-order version; Figure 6.3 shows a second-order subtractive crossover. There is no necessity for the one filter used to be the lowpass type. It is equally possible to use a highpass filter, and generate the LF signal by subtracting the highpass filter output from the original input. Subtractive crossovers are also called derived crossovers as one output is derived from the other instead of being filtered independently, and constant-voltage crossovers because of the way that their outputs sum to reconstruct the original waveform. The constant-voltage crossover, first properly described by Dick Small [1], one of the great pioneers of scientific loudspeaker design, is a subtractive crossover.

The attraction of the subtractive crossover is that it promises a perfect response—surely the subtraction process, followed by the summation that takes place in the air in front of the speaker, must give a ruler-flat combined response? Well, the answer is that it does—but not in a way that is generally useful. Exactly why this is so will be described shortly.

Subtractive crossovers also promise—and indeed deliver—perfect waveform reconstruction. It is described in Chapter 4 how a first-order conventional crossover can do this reconstruction, but second-order and higher conventional crossovers cannot, because of their phase-shift behaviour. Nonetheless, this ability is not as advantageous as it appears, because it is of little value when compared with the serious compromises inherent in the simple subtractive crossover.

The subtractive process also offers perfect matching between the crossover characteristics of the HF and LF paths. Since there is only one filter, there cannot be a mismatch between filters because of component tolerances. Instead, we have the requirement that the

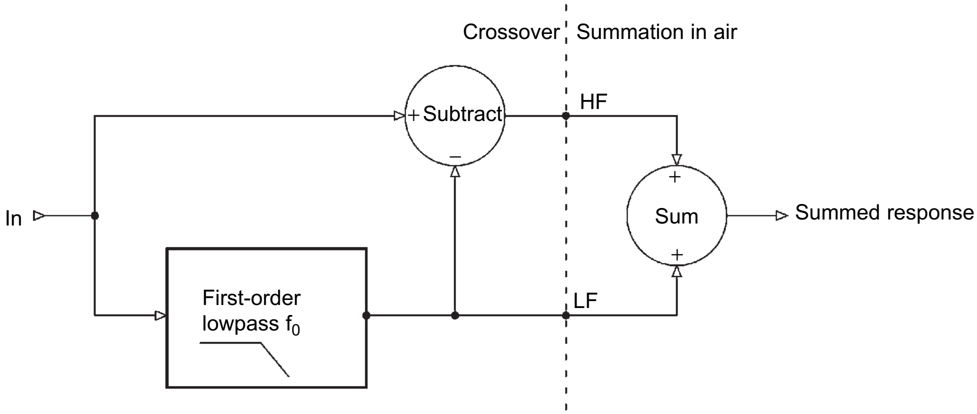


Figure 6.1: First-order subtractive crossover. The lowpass filter output is subtracted from the incoming signal to get the HF output.

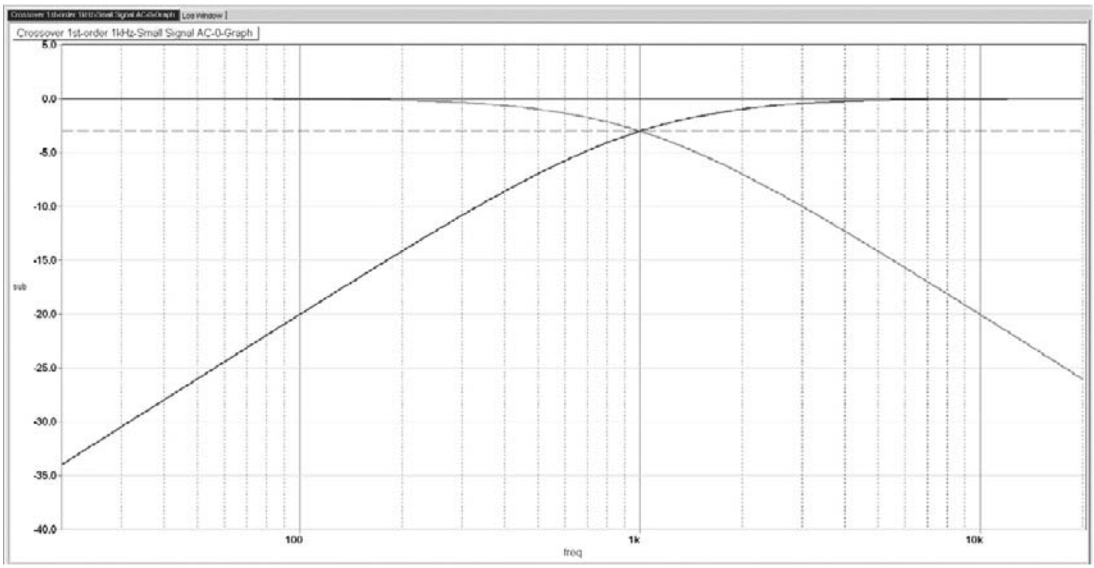


Figure 6.2: Frequency response of first-order crossover; both filter output plus their sum (straight line at 0 dB) dashed line is at −3 dB.

subtraction must be done accurately, so that the HF signal really is 1 − HF rather 1 − 0.99 HF. This is straightforward to arrange because the subtractor will be subject only to resistor tolerances, and accurate resistors are much cheaper than accurate capacitors.

An intriguing aspect of the subtractive crossover is the prospect of saving some serious money on filter capacitors. Since only one filter is used, the number of expensive precision capacitors is halved. Certainly we have to pay for the subtractor circuit, but that is cheap by comparison.

Relatively few designs for subtractive crossovers have been published; one example that deserves examination was put forward by Christhof Heinzerling in Electronics World in 2000 [2]. This design includes what looks to me like rather ambitious biquad equalisation to extend the bass response of the LF unit, and a boost/cut control for very low frequencies based on the Baxandall circuit. While I don't agree with every statement made in the article, it is well worth tracking down and perusing.

Probably the best-known subtractive crossover is a design incorporating a time delay, based on the work of Lipshitz and Vanderkooy [3], and published by Harry Baggen, in Elektor in 1987 [4]. The purpose of adding a time delay in one of the paths of a subtractive crossover is to make the filter slopes equal (without it the output derived by subtraction always has an inadequate 6 dB/octave slope) and obtain a linear-phase crossover. A linear-phase crossover has a combined output phase-shifted by an amount proportional to frequency; in other words it introduces a pure time delay only, while other crossovers have phase-shifts that change non-linearly with frequency and act like allpass filters. The snag is that the time delay has to be very accutrately matched to the filter to get the steeper crossover slopes.

All these issues are looked at in detail below.

6.1.1 First-Order Subtractive Crossovers

A first-order subtractive crossover is shown in Figure 6.1. The circular summing element to the right is not part of the crossover, but simply shows how the acoustic outputs from the HF and LF drive units sum together as air pressure in front of the loudspeaker box. It signifies the purely mathematical process of addition and it does not affect the operation of the crossover itself in any way. The circular subtracting element is of course part of the crossover, but likewise it performs a pure subtraction and nothing else.

Figure 6.2 shows the response. The LF lowpass filter output is as expected, with its −3 dB point at 1 kHz. The HF output obtained by subtraction is exactly the same as would be obtained from a first-order highpass filter with a 1 kHz −3 dB cutoff frequency. The summed response is a flat line at 0 dB, exactly as for the conventional first-order crossover. The phase behaviour is also identical, and it reconstructs waveforms in the same way.

The main advantage is that there is only one filter, so there is no problem with filter matching. The economic advantages are small because we have saved only one capacitor, and the subtractor circuit may cost more than that. As always, the unsolvable problem with first-order crossovers is that the 6 dB/octave slopes are just not adequate for directing frequencies to the appropriate drive unit.

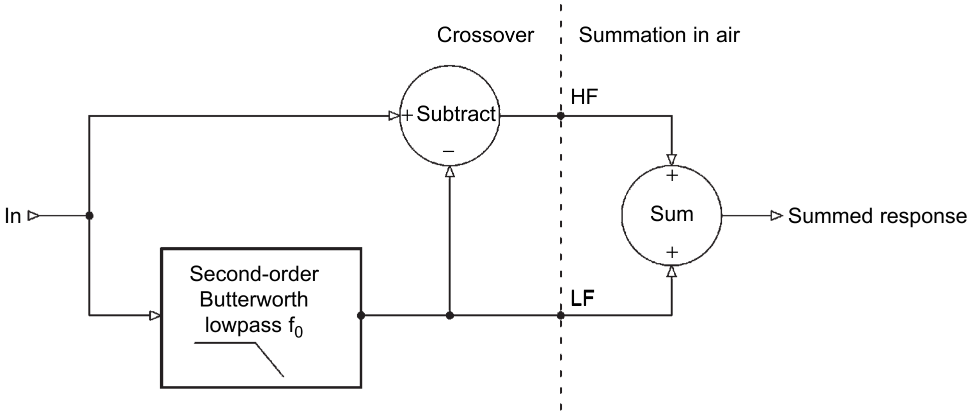


Figure 6.3: Second-order subtractive crossover. The Butterworth lowpass filter output is subtracted from the incoming signal to get the HF output.

6.1.2 Second-Order Butterworth Subtractive Crossovers

We have just seen how the first-order subtractive crossover gives some interesting results, though these are in fact the same as you get with a conventional two-filter crossover, and the advantages of the subtractive method are limited. A second-order crossover that can reconstruct a waveform (as a subtractive crossover can) sounds, however, like it might be more interesting. Figure 6.3 shows a second-order subtractive crossover, which you will note is exactly the same as the first-order version, except that the first-order filter has been swapped for a second-order filter.

If we choose a second-order Butterworth for our filter, we get the response of Figure 6.4, which was probably not what you were expecting. The lowpass output is a maximally flat Butterworth response, as it must be, but the derived HF output is very different. Firstly, its slope is only 6 dB/octave, while the LF output has the expected 12 dB/octave slope. Worse still, the HF output has a peak +2.4 dB high just above the crossover frequency; while the final summed response may be flat, that peak represents a lot of extra energy being delivered to the midrange/tweeter at a frequency below the crossover point, which is not good place for it, probably leading to excessive coil excursions and increased distortion. It is the phase-shift of the lowpass filter that causes the 6 dB/octave slope and the response peaking.

What we have built here is an asymmetrical crossover—one with unequal slopes. The 6 dB/octave slope of the HF output is not adequate for frequency separation of normal drive units. You will note that the method I have used here derives the HF output from the LF

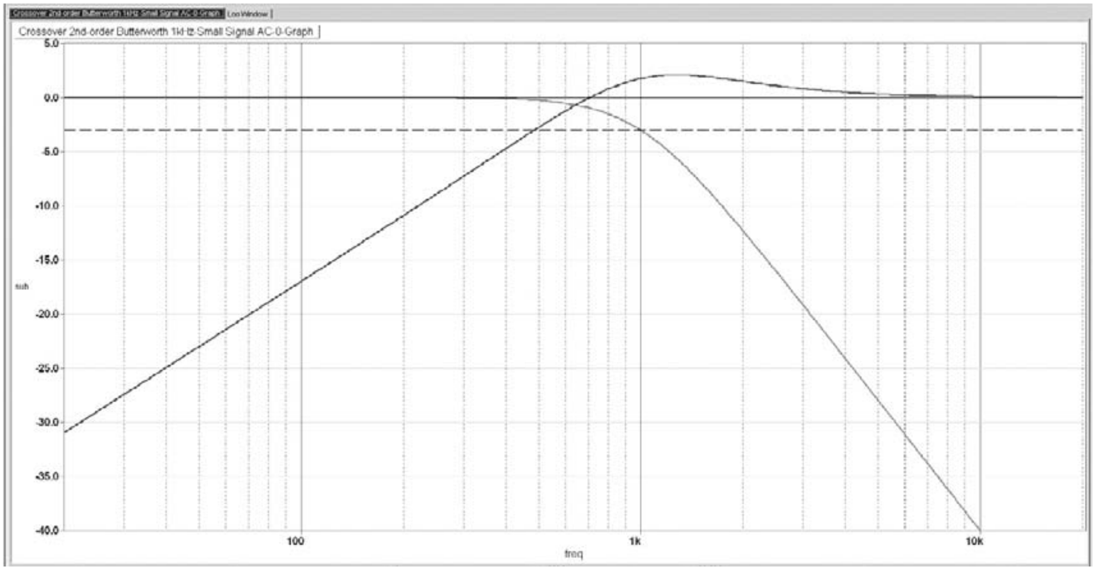


Figure 6.4: Frequency responses of second-order Butterworth subtractive crossover; both filter outputs plus their sum (straight line at 0 dB) dashed line is at −3 dB.

output, and there is a good reason for that. The derived output always has a slope of 6 dB/octave, as you will shortly see, and deriving the HF output makes it very clear that its feeble slope extends well into the bass end of the plot, so an excessive amount of LF energy is going to get into the midrange/tweeter, with a high risk of damage. If you really wanted to use a subtractive crossover of the simple sort described here, you would derive the LF output from the highpass filter output. That would give a decent slope to the highpass output, and the 6 dB/octave lowpass slope would then be failing to keep HF out of the LF drive unit. This will probably lead to some unfortunate response irregularities as the LF drive unit will be working outside its intended frequency range, but it is not likely to be damaged.

6.1.3 Third-Order Butterworth Subtractive Crossovers

A third-order subtractive crossover can be made in just the same way by replacing the second-order lowpass filter with a third-order one, and carrying out the same subtraction. If we plug in a third-order Butterworth, we find that the results are no better—in fact they are rather worse; see Figure 6.5. The crossover is still asymmetrical, for despite the use of a third-order filter instead of a second-order one, the HF output still only has a slope of 6 dB/octave. The unwelcome peak in the response is still there; now it is at slightly below the

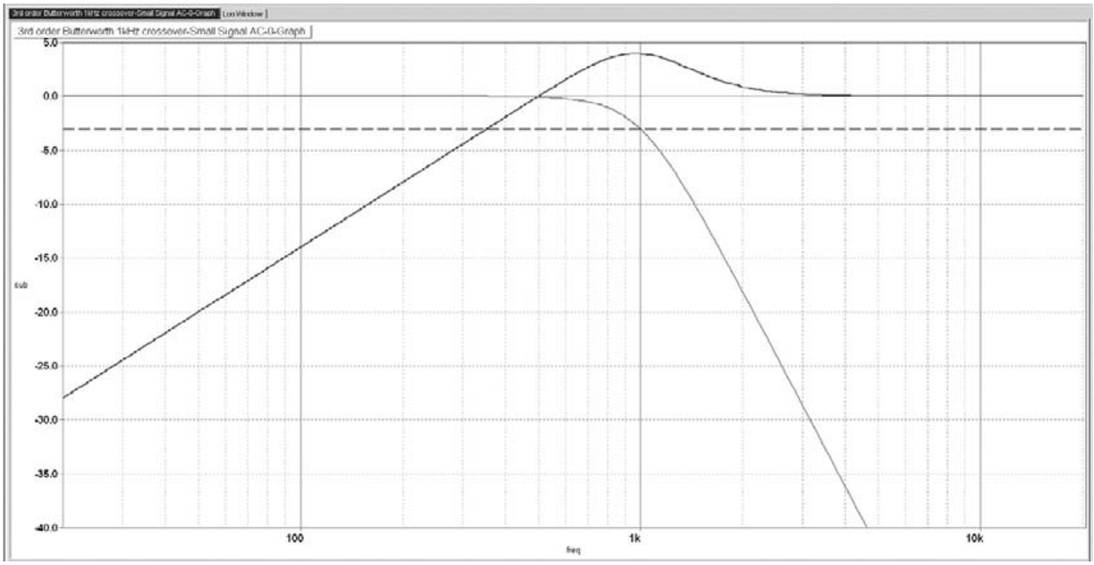


Figure 6.5: Frequency response of third-order Butterworth subtractive crossover; both filter outputs plus their sum (straight line at 0 dB) dashed line is at −3 dB.

crossover frequency and it has grown to +4.0 dB in height. Third-order filters are clearly not the answer.

6.1.4 Fourth-Order Butterworth Subtractive Crossovers

If we try a fourth-order Butterworth as the lowpass filter, the results are much the same; see Figure 6.6. The LF output is the direct output of the lowpass filter, and so is what we expect, rolling off at a satisfactory 24 dB/octave. The HF output slope stays stubbornly at 6 dB/octave, and the peak moves down a little in frequency and grows in height to +5.2 dB. The crossover is still asymmetric- in fact it is more asymmetric, with the LF slope now being four times that of the HF slope.

You may be thinking at this point that we are making a crass mistake by using Butterworth filters, and some other filter characteristic like Bessel or Chebyshev would give better results. The most popular fourth-order crossover is the Linkwitz-Riley alignment (equivalent to two cascaded second-order Butterworth filters) so let’s see if using that for the lowpass filter makes a revolutionary difference.

Not at all. As Figure 6.7 shows, the crossover is still highly asymmetrical because the HF output still has that useless 6 dB/octave slope. The height of the peak is slightly less at +4.3 dB, but that’s precious little help.

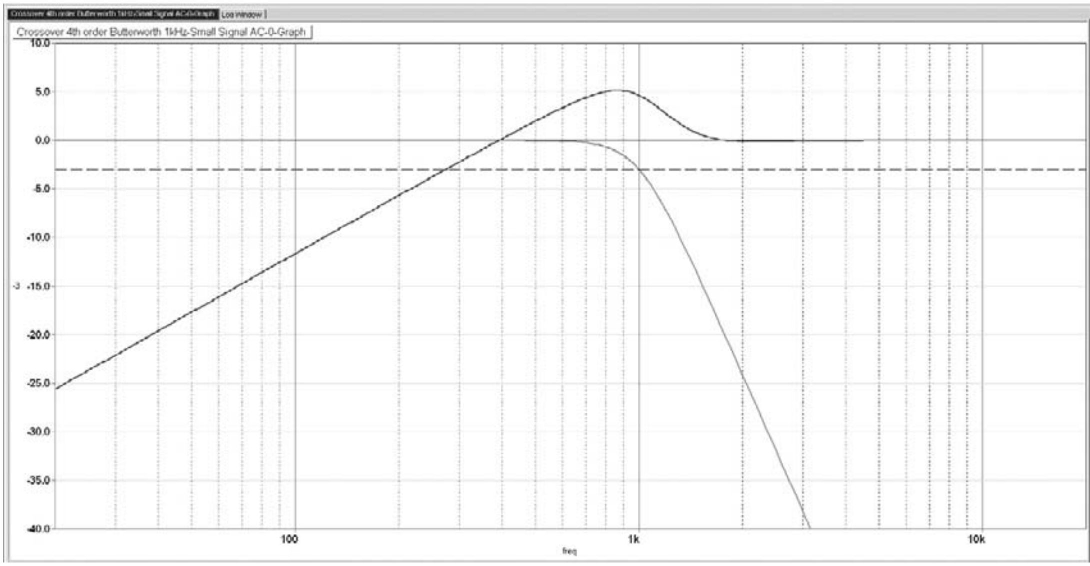


Figure 6.6: Frequency response of fourth-order Butterworth subtractive crossover; both filter outputs plus their sum (straight line at 0 dB) dashed line is at −3 dB.

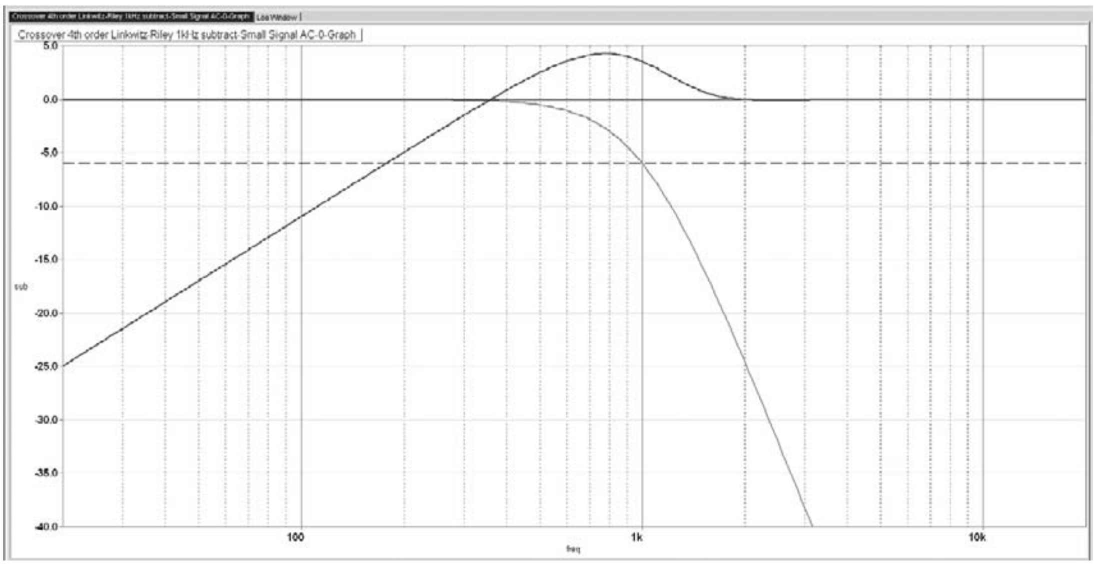


Figure 6.7: Frequency response of fourth-order Linkwitz-Riley subtractive crossover; both filter outputs plus their sum (straight line at 0 dB) dashed line is at −6 dB.

6.2 Subtractive Crossovers with Time Delays

In 1983 Lipshitz and Vanderkooy [3] proposed that linear-phase crossover networks could be produced by a subtractive method, the key idea being that a time delay inserted in the unfiltered path would compensate for the phase-shift in the lowpass filter and allow crossovers to be designed with symmetrical slopes of useful steepness. The basic arrangement is shown in Figure 6.8.

To the best of my knowledge, the only practical design of this sort of crossover that has been published was by Harry Baggen, in a famous article in Elektor in 1987 [4]. It was a three-way crossover based on fourth-order Linkwitz–Riley filters. Since the highpass outputs were derived by subtraction, using the time-delay concept, only two Linkwitz–Riley filters were required, to some extent making up for the extra cost of the subtractors and the second-order allpass filters used to create the delays. This crossover may be over twenty years old, but its conceptual significance is such that it is still being actively discussed today.

The block diagram of this crossover is shown in Figure 6.9. The crossover frequencies were nominally 500 Hz and 5 kHz, but the actual frequencies calculated from the original component values are 512 Hz and 5.12 kHz. The 512 Hz fourth-order Linkwitz–Riley lowpass filter gives the LF output; while its phase-shift is compensated for in the lower path by the delay filter t1. A highpass signal is derived from it by Subtractor 1. The circuit section including the 5.12 kHz lowpass filter, a delay block t2 and Subtractor 2, is as shown in Figure 6.8, and derives the HF output. The signal from the 5.12 kHz lowpass filter then has the signal from the 512 Hz lowpass filter subtracted from it to create the MID output; note that another t2 delay block is inserted into this path to allow for the phase-shift in the 5.12 kHz lowpass filter.

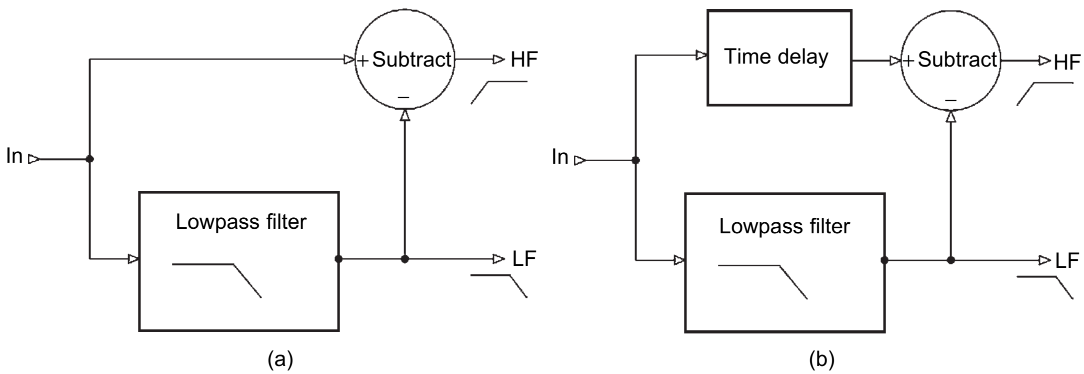


Figure 6.8: Basic subtractive crossover is at (a) adding a time delay in the unfiltered path (b) allows symmetrical-slope crossover outputs to be derived.

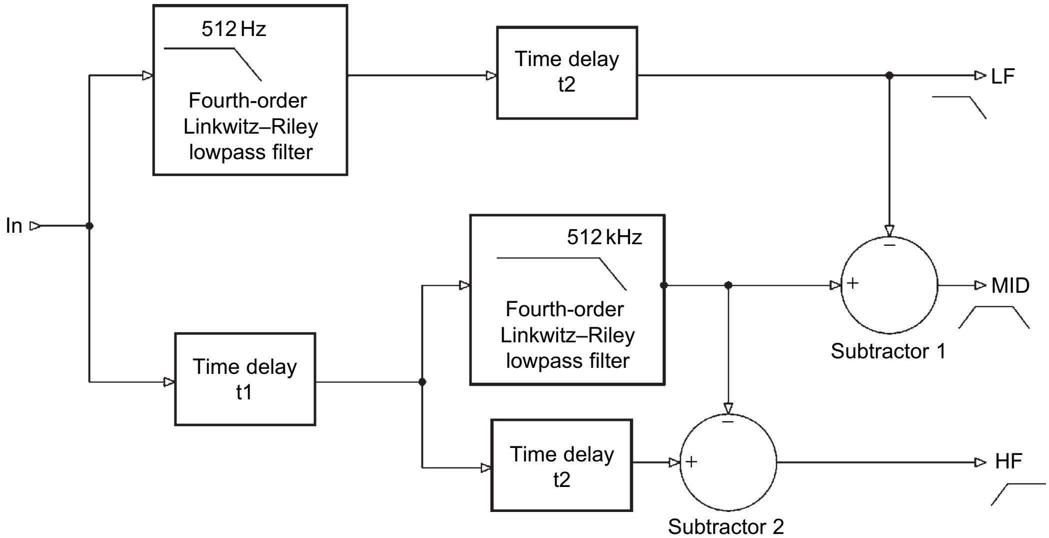


Figure 6.9: Block diagram of Elektor subtractive three-way crossover with time delays.

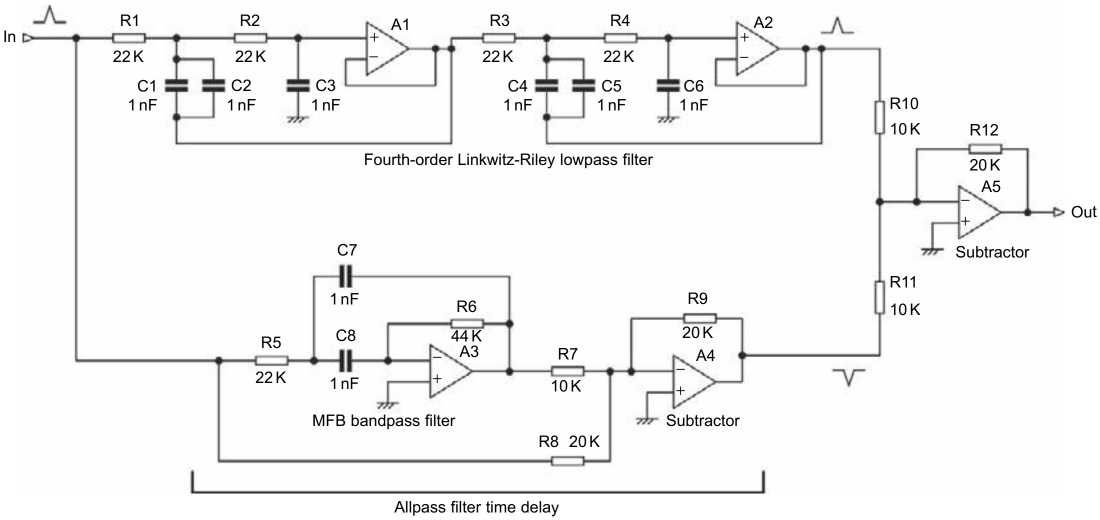


Figure 6.10: Schematic of the MID/HF section of the Elektor subtractive three-way crossover with the original component values.

Since the three-way nature of the crossover makes it quite complex, I though it best to examine the time-delay principle by looking at only one section of it. The MID/HF subtractive crossover circuitry is shown in Figure 6.10, with the original component values. The fourth-order Linkwitz-Riley lowpass filter is a standard configuration made up of two cascaded Butterworth second-order filters A1, A2. The time delay t2 in the other path is realised by a second-order allpass filter, made up of a multiple-feedback bandpass filter A3 and the shunt-feedback stage A4. This implements the 1-2BP second-order allpass configuration, where the signal is fed to a second-order bandpass filter, multiplied by two, and then subtracted from the original signal. It is not what you might call intuitively obvious, but this process gives a flat amplitude response and a second-order allpass phase response. Since the MFB bandpass filter phase-inverts, the subtraction can be performed by simple summation using A4. The MFB bandpass filter has unity gain at its resonance peak, so R7 needs to be half the value of R8 to implement the scaling by two. The operation and characteristics of this configuration are much more fully described in Chapter 10 on time-domain filtering.

We now have two signals, one lowpass filtered and one time delayed, and the former must be subtracted from the latter to derive the highpass output. This can again be done by a simple summing stage, in this case A5, because the delayed signal has been phase-inverted by A4 so summing is equivalent to subtraction. The alert reader—and I trust there is no other sort here—will have noticed that the phase of the signals going to the subtractor A5 in Figure 6.10 is the opposite of those shown in Figure 6.9; this is because in the complete crossover the signal entering the MID/HF crossover circuitry has already been phase-inverted by the delay circuitry t1.

You are possibly thinking that the impedance levels at which this circuitry operates are rather higher than recommended in this book, and you are quite right. When the Elektor crossover was published in 1987, the 5532 opamp was still expensive, and so the crossover used TL072s. These opamps have much a much inferior load-driving capability, with even light loading degrading their distortion performance, so low-impedance design was not practicable.

Figure 6.11 shows the two outputs, with nice symmetrical 24 dB/octave slopes, crossing over at −6 dB very close to 5 kHz. However, for this plot the vertical scale has been extended down to −80 dB, and you can see that something goes wrong at about −60 dB, with the derived HF output 24 dB/octave slope quite suddenly reverting to a shallow 6 dB/octave. It is highly unlikely that a 6 dB/octave slope at such a low level could cause any drive-unit problems, but alarm bells ring in the distance because this is a simulation, and one of the most dangerous traps in simulation is that it enables you to come up with an apparently sound circuit that actually depends critically on component values being exactly correct. Further investigation is therefore called for…

I suspected that the abrupt shallowing of the slope was due to the delay not being exactly matched to the lowpass-filter characteristics, and to test this hypothesis I increased the allpass delay by about 2% by changing R5 to 22.5 kΩ and R6 to 45 kΩ. This raised the level at which the derived highpass output slope became shallower quite dramatically to −25 dB, as shown in Figure 6.12.

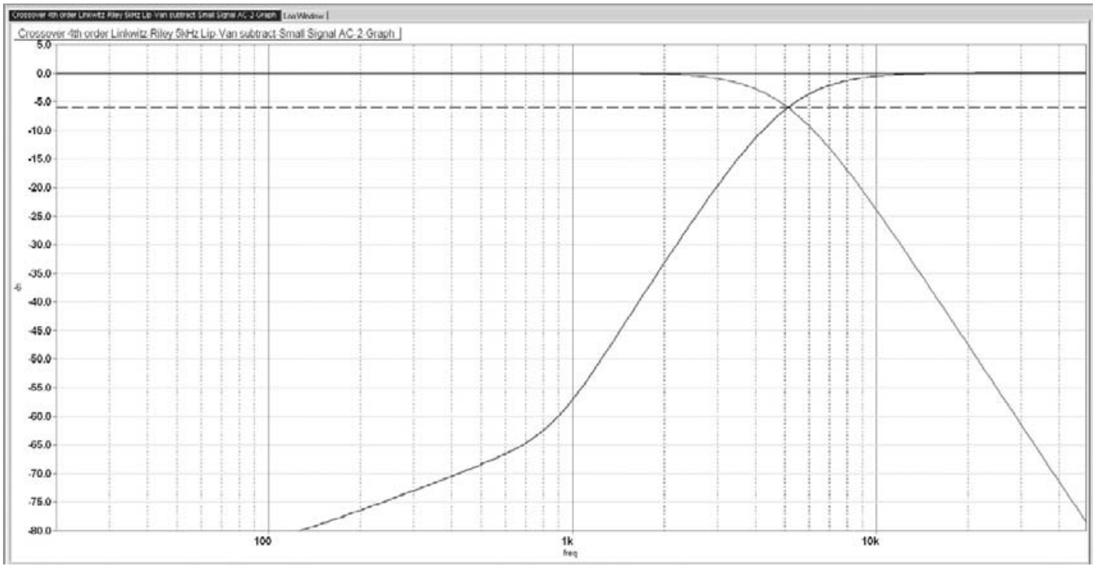


Figure 6.11: The MID/HF crossover is only symmetrical down to −60 dB. Dashed line is at −6 dB.

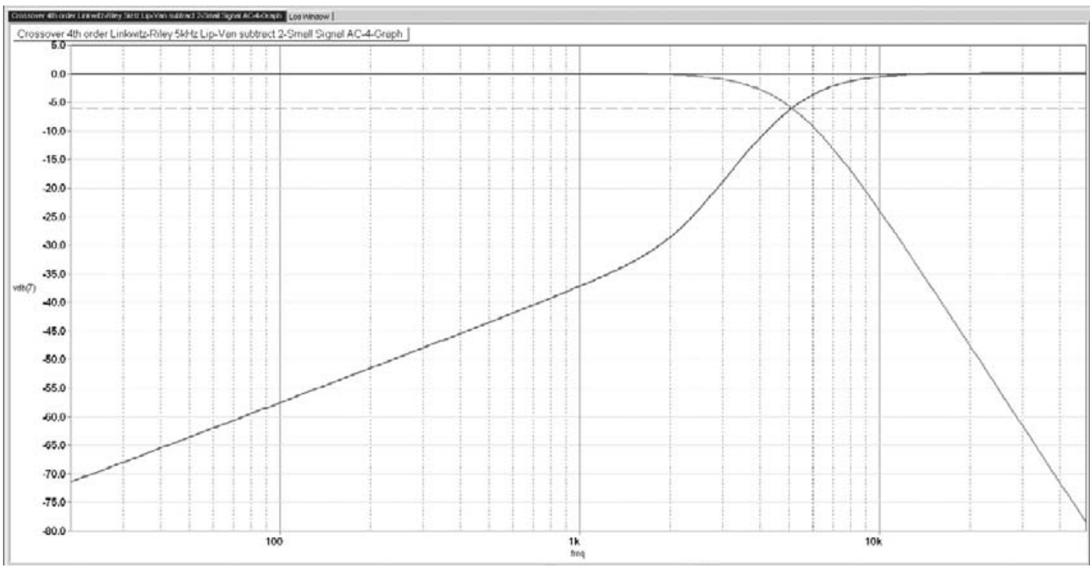


Figure 6.12: With a 2% time error the subtractive crossover is only symmetrical down to −25 dB. Dashed line is at −6 dB.

In reality you would probably find that the delay errors were larger, as they also depend on C7 and C8 in the allpass filter, and these may not be more accurate than ±5%. If all other components are completely accurate, setting C7 and C8 so they are both 5% high causes the 24 dB/octave slope of the derived signal to become 6 dB/octave at only −15 dB, which is certainly going to interfere with proper crossover operation. Setting both 5% low gives the same result. Various other twiddlings and tweakings of C7 and C8 have similar effects on the slope, which always reverts to −6 dB/octave. This is obviously not a statistically rigorous analysis of the likely errors in the delay filter, but things are not looking promising.

I then turned to the lowpass filter, introducing assorted 5% errors into the four capacitors there. It did not come as a total shock to find once more that the derived signal slope was severely compromised.

Since the whole process depends on subtraction, it seemed very likely that the components defining the accuracy of the subtraction would be critical, and so it proved. For accurate subtraction R10 has to be equal to R11; errors in R12 can only result in a gain error. Increasing R11 by just 1% to 22.2 kΩ gives the disconcerting result seen in Figure 6.13, where a notch has appeared at 1.6 kHz; above the notch the slope is increased from 3 kHz on down, while at lower frequencies the response has become flat at just below −40 dB. Changing R11 by 1% in the other direction to 21.8 kΩ, gives a similar levelling-out just below −40 dB but the notch is absent. This is a really discouraging result, stemming from a limit-of-tolerance (1%) change to one single component. Any process depending on subtraction will have problems when the output is required to be low, because of the way that errors are magnified when the difference between two large quantities is taken. This is bad enough when just two components affect the result, but here we have one relatively complex circuit trying to cancel out a property of another relatively complex circuit. There is a lot that can go wrong.

These results strongly suggest that while a linear-phase crossover may be a very desirable goal—though not everyone would agree with on that—this sort of subtractive crossover needs to be constructed with an impractical degree of precision to work properly. Rod Elliot has come to the same conclusion [5].

6.3 Performing the Subtraction

It is not hard to come up with a circuit that performs a subtraction. Any balanced line input stage does this; it takes the in-phase (hot) input and subtracts the out-of-phase (cold) input from it to remove common-mode signals occurring in the input cable or elsewhere. This process is described in detail in Chapter 16, where a variety of balanced-input stages are described. Here we only need a simple subtraction without variable gain or any other

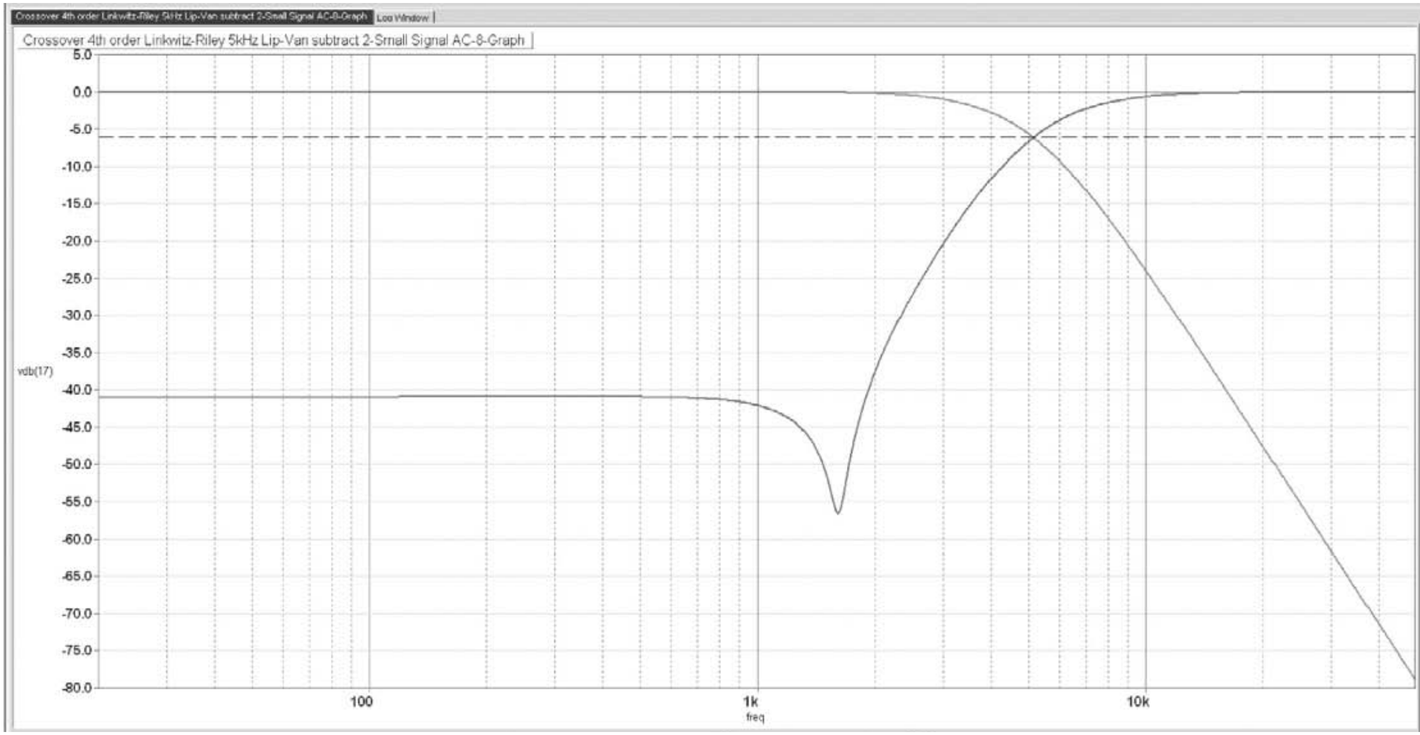


Figure 6.13: Altering the subtraction resistor R11 by 1% gives the derived highpass signal a notch and a flat section in its response. Dashed line is at −6 dB.

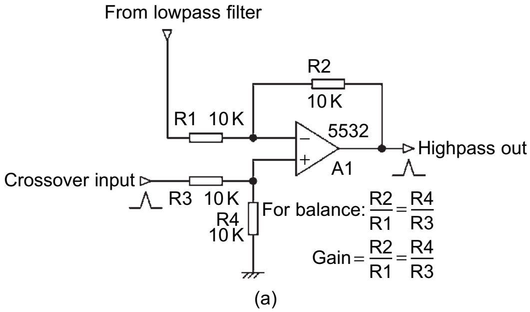


Figure 6.14: Subtractor circuits: the version in (a) is 6.6 dB noisier than that at (b), which has lower resistor values.

complicating features. We do, however, need an accurate subtraction; in the world of balanced line inputs, this is described as a high common-mode rejection ratio.

The standard balanced-input/subtractor is shown in Figure 6.14. It is very often built with four 10 kΩ resistors as a compromise between noise performance and achieving reasonably high input impedances, and the measured noise output using a NE5532 opamp is −105.1 dBu. However, in an active crossover application we can assume that both inputs will be driven from opamp outputs (almost certainly from more NE5532s) and so the resistor values can be reduced drastically. This lowers the Johnson noise from the resistors and also means that the opamp current noise is flowing through less resistance and so creating less voltage noise. The effect of the opamp voltage noise is unchanged.

Reducing the four resistors from 10 kΩ to 820 Ω, as in Figure 6.14, reduces the noise output from −105.1 dBu to −111.7 dBu, a helpful improvement of 6.6 dB in performance for no cost at all. The value of 820 Ω is chosen as it can be driven by another NE5532 without significantly impairing its distortion performance. This is a good example of low-impedance design.

If one of the inputs to be subtracted is phase-inverted, then simply summing the two inputs together is equivalent to a subtraction. This is the method used in the Elektor crossover described earlier in this chapter.

References

[1] R.H. Small, Constant voltage crossover network design, JAES 19 (1) (1971) 12–19.
[2] C. Heinzerling, Adaptable Active Speaker System (subtractive), Electronics World, London, 2000, p. 105.
[3] S.P. Lipshitz, J. Vanderkooy, A family of linear-phase crossover networks of high slope derived by time delay, JAES 31 (1983) 2.
[4] H. Baggen, Active phase-linear crossover network, Elektor (1987) 61.
[5] R. Elliot, Private communication (2010).

CHAPTER 7

Lowpass & Highpass Filter Characteristics

Active filters are the building blocks of active crossovers. Filter design is an enormous subject, and it is of course quite impossible to cover even its audio aspects in a single chapter. There are many excellent and comprehensive textbooks on filters [1–3], and there would be no point in trying to create another one here. This chapter shows how to put together the blocks that make up a filter, but the design of each block to component-value level is covered in Chapter 8.

Filter design is at the root highly mathematical, and it is no accident that popular filter characteristics such as Bessel and Chebyshev are named after mathematicians, who usually did not live long enough to see their mathematics applied to practical filters. A notable exception is the Butterworth characteristic, probably the most popular and useful characteristic of all; Stephen Butterworth was a British engineer.

Here, however, I am going to avoid the complexities of pole and zero placement etc., and concentrate on practical filter designs that can be adapted for use at different frequencies by a simple process of scaling component values. Most filter textbooks give complicated equations for calculating the amplitude and phase response at any desired frequency. This gets very cumbersome if you are dealing with a lot of different filters. I have assumed access to a SPICE simulator, which gives all the information you could possible want, much more quickly and efficiently than wrestling with calculations manually or in a spreadsheet. Free simulator packages can be downloaded.

Filters are either passive or active. Passive or LCR filters use resistors, inductors, and capacitors only. Active filters use resistors, capacitors, and gain elements such as opamps; active filter technology is usually adopted with the specific intent of avoiding inductors and their well-known limitations. Nevertheless, there are some applications where LCR filters are essential, such as the removal of high frequencies from the output of Class-D power amplifiers, which would otherwise upset audio test gear. The answer, as described by Bruce Hofer [4] is a passive LCR roofing filter.

There is no reason why you could not make a line-level passive crossover using only inductors, capacitors, and resistors, but it is difficult to think of any reason why you'd want to.

Likewise, you could design a crossover made up of active filters that used only inductors and resistors (i.e., no capacitors) but this would also come under the heading of "perverse electronics."

7.1 Active Filters

Active filters do not normally use inductors as such, though configurations such as gyrators that explicitly model the action of an inductor are sometimes used. The active element need not be an opamp; the Sallen and Key configuration requires only a voltage follower, which in some cases can be a simple BJT emitter-follower. Opamps are usual nowadays, however. This chapter deals only with lowpass and highpass filters; bandpass, notch and allpass filters are dealt with in later chapters.

7.2 Lowpass Filters

Lowpass filters have obvious uses in crossover design, keeping the MID and HF material out of the LF path, and keeping HF material out of the MID path. They are also used to explicitly define the upper limit of the audio bandwidth in a system, say at 50 kHz. This is much better than letting it happen by the casual accumulation of a lot of first-order roll-offs in succeeding stages. This bandwidth definition is not a duplication of input RF filtering which, as described in Chapter 16, must be passive and positioned before the incoming signals encounter any electronics at all, as active circuitry can demodulate RF. Lowpass filters are used in sound reinforcement systems to protect power amplifiers and loudspeakers against ultrasonic oscillation from outside sources or in the system itself. Lowpass filters are defined by their order, Q, cutoff frequency, and characteristic, e.g., Butterworth.

7.3 Highpass Filters

Highpass filters are used in active crossovers for keeping the LF out of the MID and HF paths, and keeping LF and MID material out of the HF path. They are also commonly used to explicitly define the lower limit of the audio bandwidth in a system, say at 20 Hz. In this case relying on the happenstance accumulation of first-order LF roll-offs is distinctly dangerous because loudspeakers are vulnerable to damage from unduly low frequencies. Highpass filters are defined by their order, Q, cutoff frequency, and characteristic.

7.4 Bandpass Filters

Bandpass filters as such are not much used in crossover design, except perhaps occasionally as part of an equalisation scheme, but they are of great utility as a basis from which to make notch filters and allpass filters by subtractive methods.

The Q required rarely exceeds a value of 5, which can be implemented with relatively simple active filters, such as the multiple-feedback type. Higher Q's or independent control of all the resonance parameters require the use of the more complex biquad or state-variable filters. Bandpass and notch filters are said to be "tuneable" if their centre frequency can be altered relative easily, say by changing only one component value. Bandpass filters are dealt with in detail in Chapter 9. They are defined by their Q and centre frequency.

7.5 Notch Filters

Notch filters are used in some active crossovers, for example [5]. They are an integral part of elliptic filters, inserting zeros (notches or nulls) in the stop-band. When someone says "notch filter" we naturally think of the most common version—the symmetrical notch which goes back up to 0 dB on either side of the central crevasse. However, there are also lowpass notch and highpass notch filters, and these are highly relevant to crossover design. Notch filters are dealt with in detail in Chapter 9. They are defined by their Q and centre frequency.

7.6 Allpass Filters

Allpass filters are so-called because they have a flat frequency response, and so pass all frequencies equally. Their point is that they have a phase-shift that *does* vary with frequency, and this is often used for delay correction in active crossovers. You may occasionally see a reference to an all-stop filter, which has infinite rejection at all frequencies; this is a filter designer's joke, sort of. Allpass filters are dealt with in detail in Chapter 10.

7.7 The Order of a Filter

Strictly speaking, the order of a filter is the highest power of frequency that occurs in the complex algebraic equation that describes its behaviour. Since that is the sort of complication we are trying to avoid, we have to here rely on somewhat less formal definitions. For example, the ultimate roll-off slope of a lowpass or highpass filter is 6 dB/octave times the order of the filter. A first-order filter (which is just a single RC time-constant) rolls off at 6 dB/octave, a second-order filter at 12 dB/octave, and so on. For most kinds of filter the number of capacitors is equal to the order of the filter.

A bandpass filter must be at least second-order; it can also be fourth-order, which represents two bandpass filters cascaded, perhaps with their centre frequencies offset to give a flattish section rather than a sharp peak at the frequencies of interest; in radio receivers this is referred to as staggered tuning. It is impossible to have a first-order bandpass or notch (band-reject) response.

7.8 Filter Cutoff Frequencies and Characteristic Frequencies

The cutoff frequency of a lowpass or highpass filter is a measure of where in the spectrum its roll-off starts. In many cases it is defined as the frequency at which the gain is down by −3 dB (= 1/√2). The word "cutoff" is perhaps unfortunate because it seems to imply a frequency response that drops suddenly, like falling off a cliff. So-called "brickwall" filters with very fast roll-offs do exist, but the filters used in active crossover design are rarely higher than fourth-order and the start of the roll-off is actually quite gentle. However, nobody's going to change the word now.

Filter types like the Butterworth and Bessel characteristics have their cutoff frequencies defined at the −3 dB point, whereas the Linkwitz-Riley filter may be defined at the −3 dB or the −6 dB point. To an extent the choice of attenuation is arbitrary; it would be quite possible to design these filters with cutoff defined at −4.5 dB or whatever, but there would be no point at all in doing so. For the Butterworth filter in particular, using −3 dB makes the mathematics very simple.

Other filters like the Chebyshev type, which has amplitude ripples in the passband, have their cutoff defined as the point where the gain passes through 0 dB for the last time as the frequency increases (in the lowpass case) and the steady roll-off begins.

The characteristic frequency of a lowpass or highpass filter is not at all the same thing as its cutoff frequency, and the two terms should not be confused. Taking lowpass filters as an example, the amplitude response off all second-order filters will eventually become a 12 dB/octave straight line heading downwards. If you extend that line upwards and to the left it will eventually cut the 0 dB gain line, and the point where it cuts it is the characteristic frequency. You can design a Bessel filter and a Chebshev filter so that their responses converge on the same 12 dB/octave line and so have the same characteristic frequency, but they will have different cutoff frequencies.

Characteristic frequencies are not widely used in crossover design but they are useful when dealing with state variable filters; this is discussed in an excellent article by Ramkumar Ramaswamy [6]. This book uses cutoff frequencies exclusively.

7.9 First-Order Filters

A first-order filter is just a single RC time-constant, with an ultimate roll-off of 6 dB/octave. Some writers seem to regard first-order filters as having a fixed Q of 0.5, though quite how this might be helpful is unclear. This is absolutely NOT the same thing as a second-order filter with a Q of 0.5. There is no such thing as a Butterworth first-order filter or a Bessel first-order filter; just a first-order filter. Most filter design software does not issue a warning if you ask for a "Butterworth first-order filter"—it simply serves up a first-order filter with no further comment. The cutoff frequency for a first-order filter is always that frequency at which the output has fallen to −3 dB ($1/\sqrt{2}$) of the passband gain.

A filter made up only of cascaded first-order stages with the same cutoff frequency is called a synchronous filter; it has a very slow roll-off and is in general not very useful. It is, however, described later in this chapter.

7.10 Second-Order and Higher-Order Filters

Second-order bandpass responses are basically all the same, being completely defined by centre frequency, Q, and gain. Second-order and higher highpass and lowpass filters are also defined by cutoff frequency, Q, and gain but come in many different types or characteristics, one of which is selected as a compromise between the need for a rapid roll-off, flatness in the passband, a clean transient response, and desirable group-delay characteristics. The Butterworth (maximally flat) characteristic is the most popular for many applications, not least because it is refreshingly simple to design. Filters with passband ripples in their amplitude response, such as the Chebyshev or the elliptical types, have not found favour for in-band filtering, such as in active crossovers, but were once widely used for applications like ninth-order anti-aliasing filters in front of ADCs, and similar reconstruction filters after DACs; the passband ripples being accepted as inescapable if the filters were to be kept to an acceptable level of complexity. Such filters have mercifully been made obsolete by oversampling, pushing the unwanted frequencies a long way up the spectrum, so post-DAC filters are nowadays usually simple second- or third-order Butterworth types.

The Bessel filter characteristic gives a maximally flat group delay (maximally linear phase response) across the passband, and so better preserves the waveform of filtered signals, but it has a much slower roll-off in amplitude response than the Butterworth.

The cutoff frequency for second- and higher-order filter is to a certain extent a matter of definition. The cutoff frequency of a Butterworth filter is always taken as the −3 dB frequency because this fits in very neatly with the delightfully straightforward design equations. Bessel filters normally use the same definition, but a Linkwitz-Riley filter might be specified as either −3 dB or −6 dB, the latter being useful because when a lowpass-highpass pair of these filters is used to make a Linkwitz-Riley crossover, the filters are −6 dB at the crossover frequency. The cutoff frequency definitions for Chebyshev filters are more complicated, depending on whether the filter is odd-order or even-order; this is dealt with later in this chapter.

7.11 Filter Characteristics

There are many different types of recognised filter response, each with their own special properties. Their study is complicated by the fact that some of them go by several different names, and you need to keep a close eye on the terminology that an author is using. The well-known filter types are shown in Table 7.1 below. Some lesser known types are shown in Table 7.2.

We will look at some of these filter types, in each case first examining the simplest possible instances, the second-order versions; there are no first-order versions because a first-order filter is just a first-order filter, in other words a simple single time-constant. All the filters in this section are lowpass filters with a cutoff frequency of 1 kHz. In every case there is an equivalent highpass filter.

For the simpler filters—the Butterworth, Linkwitz-Riley, and the Bessel—we will look at the amplitude response, the phase-shift response, the response to a step input, and the group-delay curve; we then go on to compare the response of the higher-order versions. Chebyshev filters have a more complicated response and some explanation of this is required before we get into the same details as for the earlier filters.

The filters used to generate the information are shown in Figure 7.1. They are all unity-gain Sallen and Key configurations, differing only in Q. Full details of how

Table 7.1: The Popular Filter Types

Common Name	Other Names Used	Main Features
Butterworth	Maximally flat	Maximally flat amplitude, easy design
Linkwitz-Riley	Butterworth-squared, Butterworth-6 dB	Two cascaded make 4th order filter summing to allpass
Bessel	Thomson, Bessel-Thomson	Maximally flat group-delay, slower roll-off than Butterworth
Chebyshev	Chebyshev Type-I	Passband ripple, faster roll-off than Butterworth
Inverse Chebyshev	Chebyshev Type-II	Stopband has notches, faster roll-off
Elliptical	Cauer, Zolotarev, complete-Chebyshev	Passband ripple & stop band notches, fastest roll-off

Table 7.2: Lesser-Known Filter Types

Common Name	Other Names Used	Main Features
Transitional		Compromise between any two filter characteristics
Linear-Phase	Butterworth-Thomson	Compromise between Butterworth and Bessel
Gaussian		Optimised for step risetime without overshoot
Legendre	Legendre-Papoulis	Passband monotonic but not maximally flat
Synchronous		Very slow roll-off
Ultraspherical		Generalisation of other filter types
Halpern		Similar to Legendre

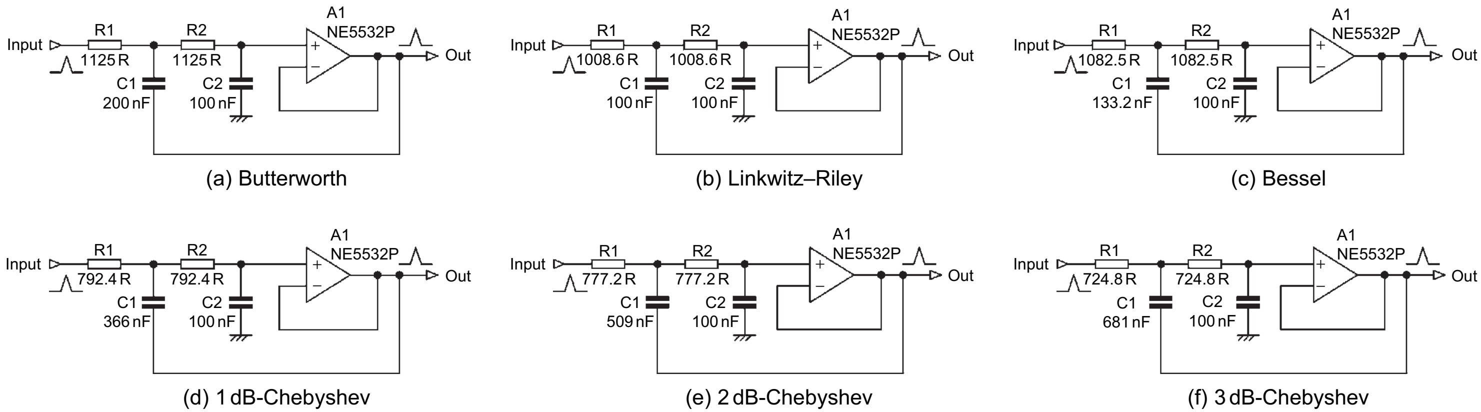


Figure 7.1: Second-order Butterworth, Linkwitz–Riley, Bessel, and Chebyshev lowpass filters. Cutoff frequency is 1.00 kHz.

to design Sallen and Key stages for a desired cutoff frequency and Q are given in Chapter 8.

7.11.1 Butterworth Filters

The Butterworth filter characteristic is one of the most popular because it has a maximally flat frequency response.

The cutoff frequency is defined as the –3-dB point. Butterworth filter design is relatively easy in its high-order versions (third and above) because every stage making up the complete filter has the same cutoff frequency—only the Qs vary. The second-order Butterworth filter has a Q of $1\sqrt{2}$ (= 0.7071). Some books use the term "damping factor" a (= 1/Q) rather than Q, but here we will stick with Q. Pay attention, 007.

The Butterworth response was introduced by Stephen Butterworth (1885–1958) in 1930 [7]. It is remarkable that not only was Butterworth a physicist engineer rather than a mathematician, but in this case the filter is actually named after the person that put it into use.

Figure 7.2 shows the amplitude response, with a sharper "knee" than other filters such as the Bessel. The ultimate slope is 12 dB/octave, determined by the fact that this is a second-order filter. The phase-shift is shown in Figure 7.3; it is 90° at the cutoff frequency.

As Figure 7.4 demonstrates, the Butterworth response gives a modest degree of overshoot when it is faced with a step input. Maximal flatness in the frequency response does *not* mean no overshoot, and does *not* mean a flat group delay (Figure 7.5). For that you need a Bessel filter.

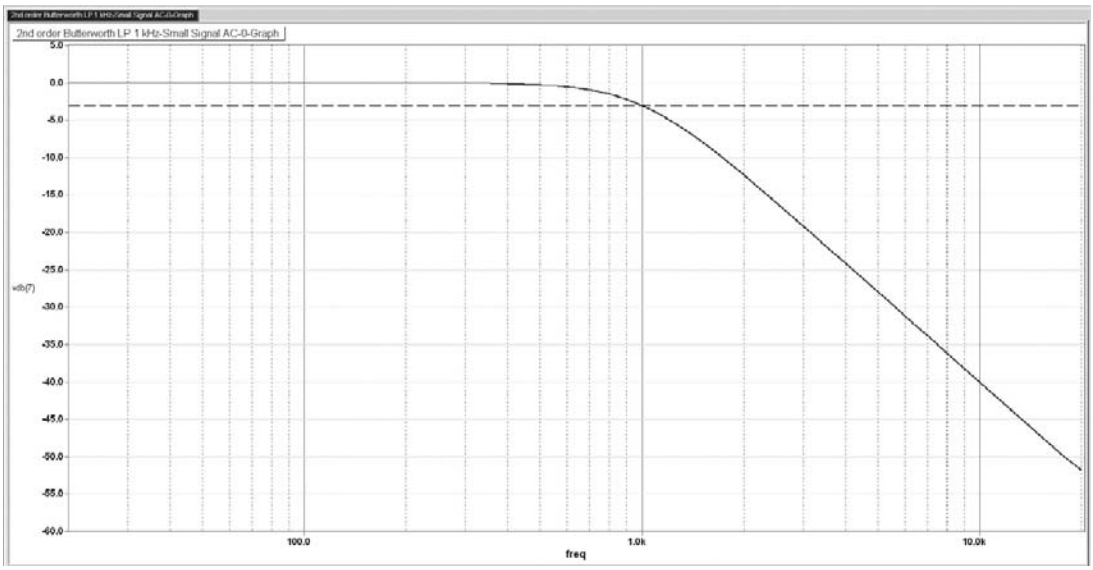


Figure 7.2: Amplitude response of second-order Butterworth lowpass filter (Q = 1√2 = 0.7071). Cutoff frequency is 1.00 kHz. Attenuation at 10 kHz is −40 dB.

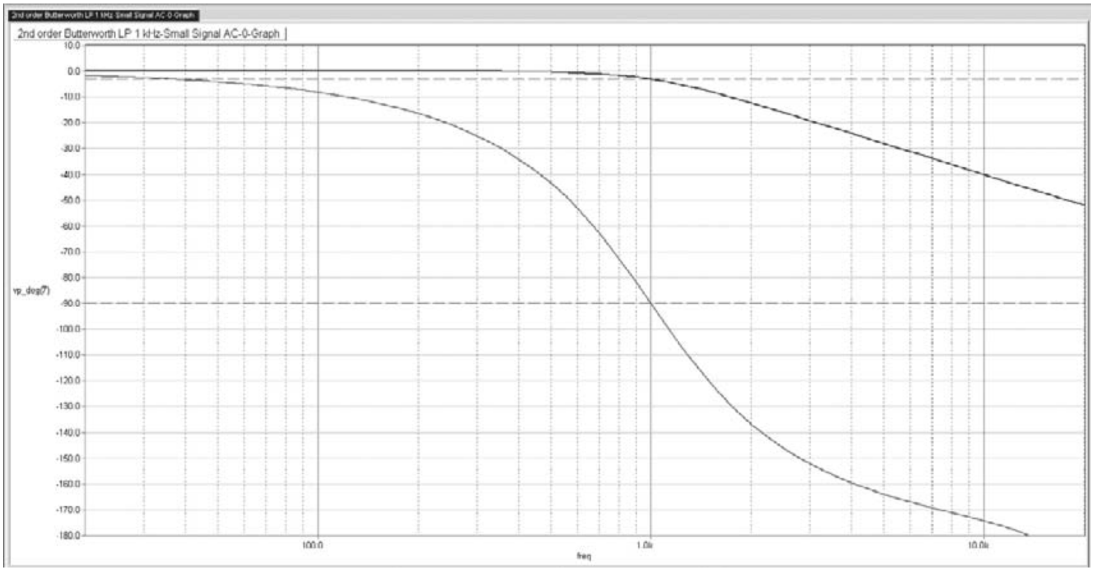


Figure 7.3: Phase response of second-order Butterworth lowpass filter. Cutoff frequency is 1.00 kHz. The phase-shift is 90° at 1.00 kHz. Upper trace is amplitude response.

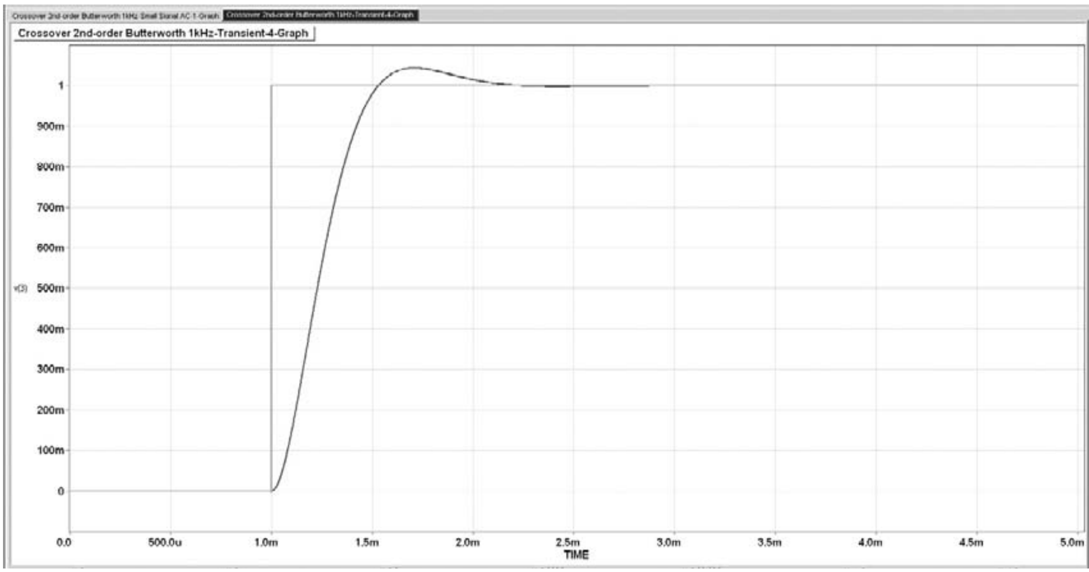


Figure 7.4: Step response of second-order Butterworth lowpass filter. Cutoff frequency is 1.00 kHz. There is significant overshoot of about 4%.

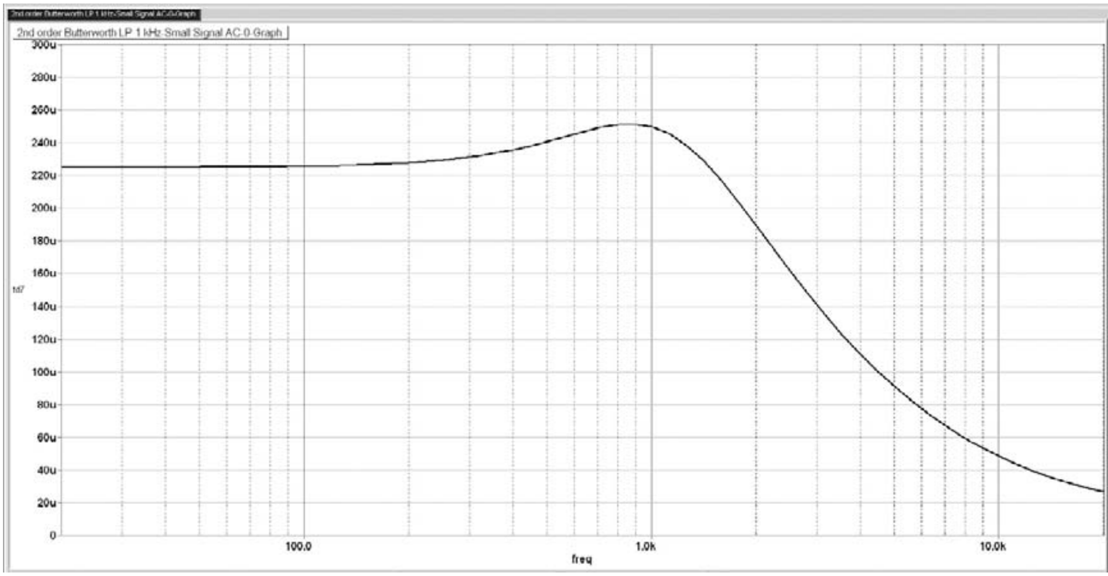


Figure 7.5: Group delay of second-order Butterworth lowpass filter. Cutoff frequency is 1.00 kHz. The delay peaks by 15% near the cutoff frequency.

7.11.2 Linkwitz–Riley Filters

The lowest Q normally encountered in a second-order filter is 0.50, which is used for second-order Linkwitz–Riley (LR-2) crossovers. This is less than the Q of 0.58 used for second-order Bessel filters. Note that 0.5 is the square of the Butterworth Q which is 0.7071 (1/√2). They are sometimes called Butterworth-squared filters because the fourth-order version is often implemented by cascading two identical second-order Butterworth stages. They are also called Butterworth-6 dB filters because there is a −6 dB level at the cutoff frequency of the two cascaded filters.

Designing for 1.00 kHz, using the equations given later, you will actually get −6.0 dB at 1.00 kHz; the −3 dB point is at 633.8 Hz. The phase-shift is 90° at 1.00 kHz. Use a frequency scaling factor of 1.578, that is, design for 1.578 kHz and you will get −3 dB at 1.00 kHz. The phase-shift is then 90° at 1.578 kHz.

The cutoff frequency is defined as the –3-dB point.

Figure 7.6 demonstrates the slower roll-off of the Linkwitz–Riley compared with the Butterworth filter. The attenuation at 10 kHz is −32.5 dB instead of −40 dB for the Butterworth. The ultimate slope is still 12 dB/octave, as this is determined by the fact that this is a second-order filter, and is not affected by the Q chosen. The phase response is shown in Figure 7.7. There is no step-response overshoot (Figure 7.8) and no group-delay peak (Figure 7.9).

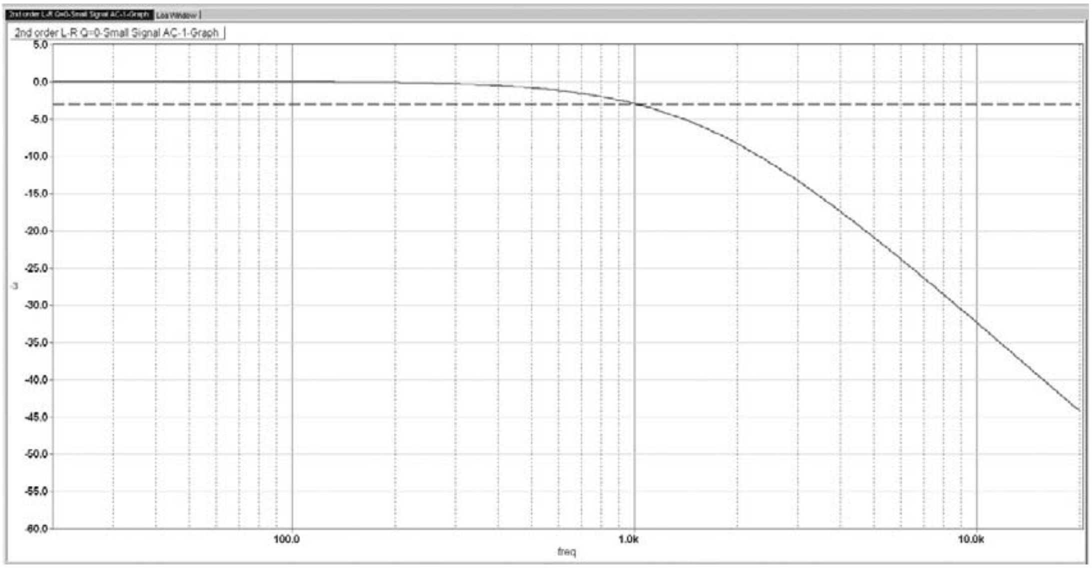


Figure 7.6: Amplitude response of second-order Linkwitz–Riley lowpass filter. (Q = 0.50) −3 dB at 1 kHz. Attenuation at 10 kHz is −32.5 dB.

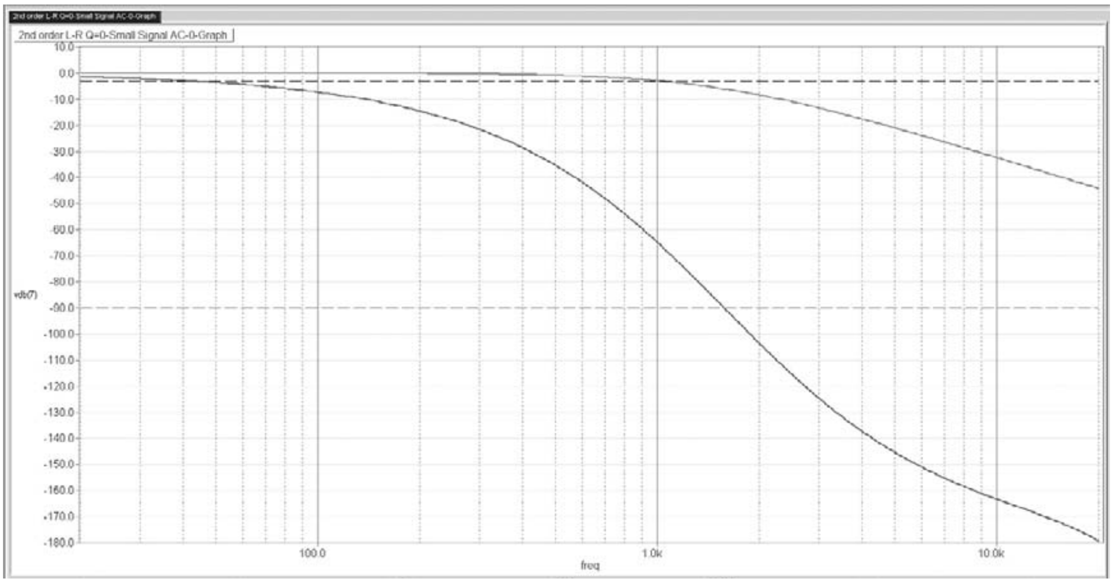


Figure 7.7: Phase response of second-order Linkwitz–Riley lowpass filter. −3 dB at 1.00 kHz. The phase-shift is 90° at 1.578 kHz. Upper trace is amplitude response.

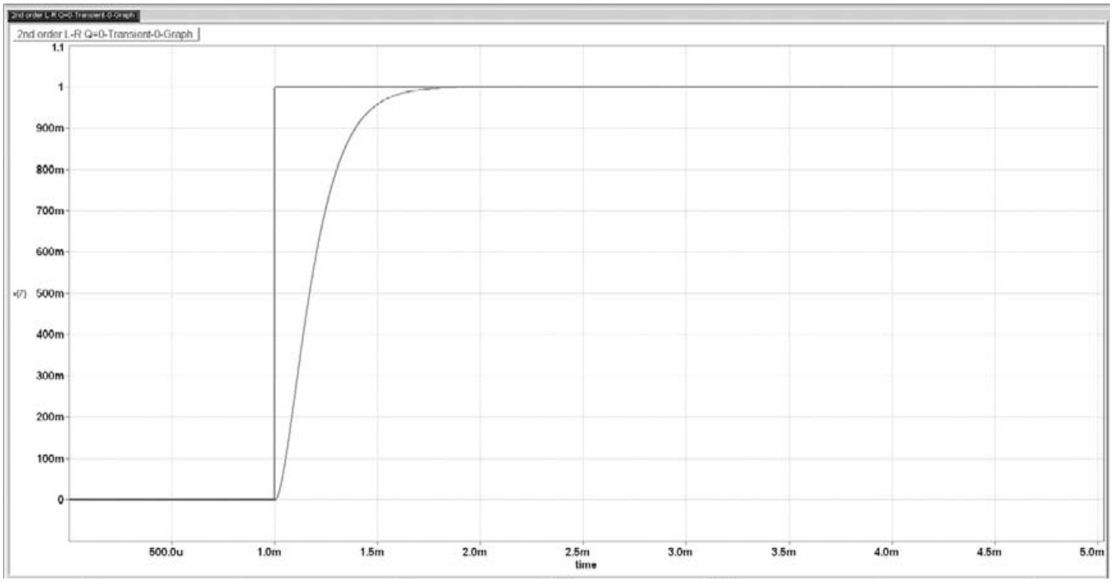


Figure 7.8: Step response of second-order Linkwitz–Riley lowpass filter. −3 dB at 1.00 kHz. There is no overshoot.

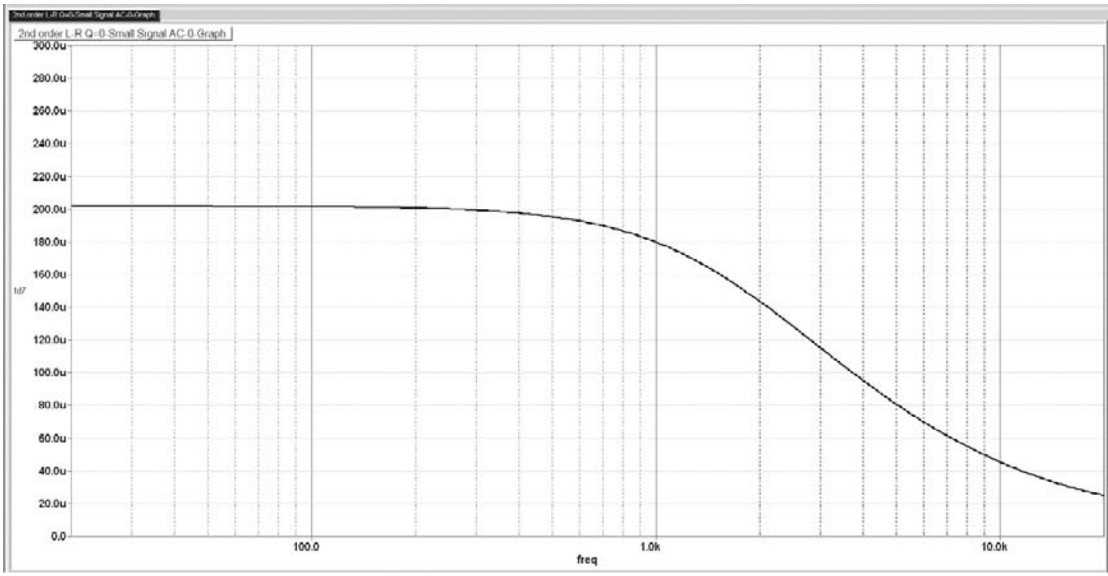


Figure 7.9: Group delay of second-order Linkwitz–Riley lowpass filter. −3 dB at 1.00 kHz. There is no peaking in the delay.

7.11.3 Bessel Filters

As is common in the world of filters, Bessel filters as such were not invented by Bessel at all. Friedrich Wilhelm Bessel 1784–1846) was a German mathematician, astronomer, and was long dead by the time that anyone thought of applying his mathematics to electrical filtering. He systematised the Bessel functions, which, to keep the level of confusion up, were actually discovered by Daniel Bernoulli (1700–1782).

The first man to put Bessel functions to work in filters was W.E. Thomson [8], and that is why Bessel filters are sometimes called Thomson filters or Bessel–Thomson filters. According to Ray Miller [9], Thomson was actually anticipated by Kiyasu [10] working in Japan in 1943, but given the date it is not surprising that communications with the West were somewhat compromised. It is important to remember that the term Bessel–Thomson does *not* refer to a hybrid between Bessel and Thomson filters, because they are the same thing. This is in contrast to Butterworth–Thomson transitional filters, which *are* hybrids between Butterworth and Thomson (i.e., Bessel) filters.

The Bessel filter gives the closest approach to constant group delay; in other words the group delay curve is maximally flat. The downside is that the amplitude response roll-off is slow- actually very slow compared with a Butterworth filter. The cutoff frequency is defined as the –3-dB point.

If you design for 1.00 kHz using the equations, you will actually get −4.9 dB at 1.00 kHz; the −3 dB point is at 777 Hz. There is however 90° phase-shift at 1.00 kHz. Use a frequency scaling factor of 1.2736, that is, design for 1.2736 kHz and you will get −3 dB at 1.00 kHz.

The Bessel responses are shown in Figures 7.10–7.13. The attenuation at 10 kHz is −36 dB, which is worse than the Butterworth (−40 dB), but better than the Linkwitz-Riley (−32.5 dB).

Note that the flat part of the group delay curve extends further up in frequency than for the Linkwitz–Riley filter, and yet the group delay does not peak like the Butterworth filter. If you want to combine maximally flat group delay with filtering, the Bessel filter is mathematically the best answer possible. If you want to create group delay *without* filtering, then you need an allpass filter, which gives delay with a flat frequency response. See Chapter 10 on time-domain filtering.

7.11.4 Chebyshev Filters

Pafnuty Lvovich Chebyshev (1821–1894) was a Russian mathematician. His name can be alternatively transliterated as Chebychev, Chebyshov, Tchebycheff, or Tschebyscheff.

There is only one kind of Bessel response, and only one kind of Butterworth response, but there are an infinite number of Chebyshev responses, depending on the amount of passband ripple you

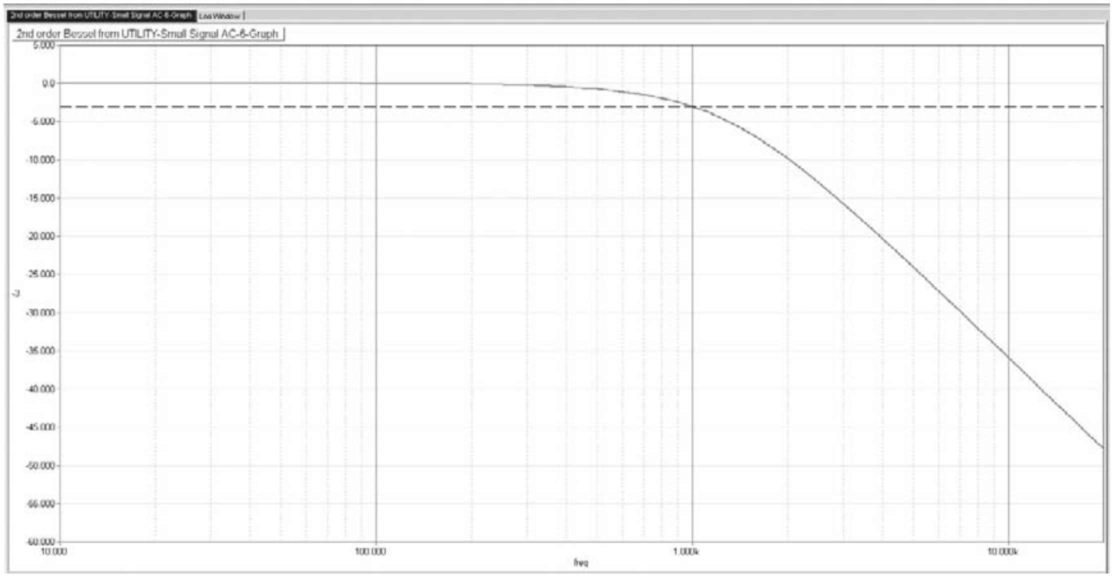


Figure 7.10: Amplitude response of second-order Bessel lowpass filter (Q = 0.578). Cutoff frequency is 1 kHz. Attenuation at 10 kHz is −36 dB.

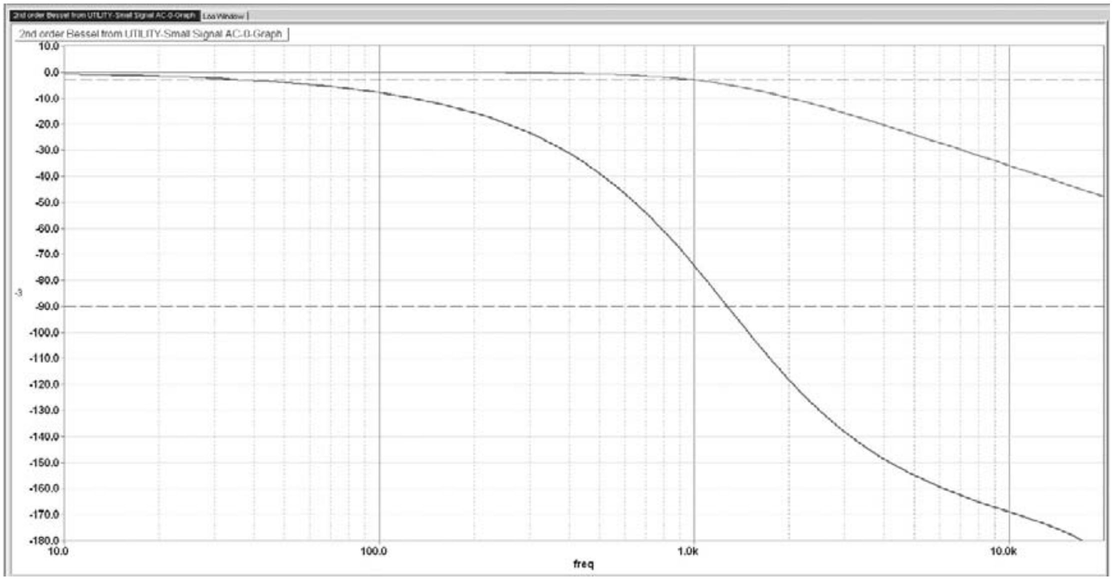


Figure 7.11: Phase response of second-order Bessel lowpass filter. −3 dB at 1.00 kHz. The phase-shift is 90° at 1.3 kHz. Upper trace is amplitude response.

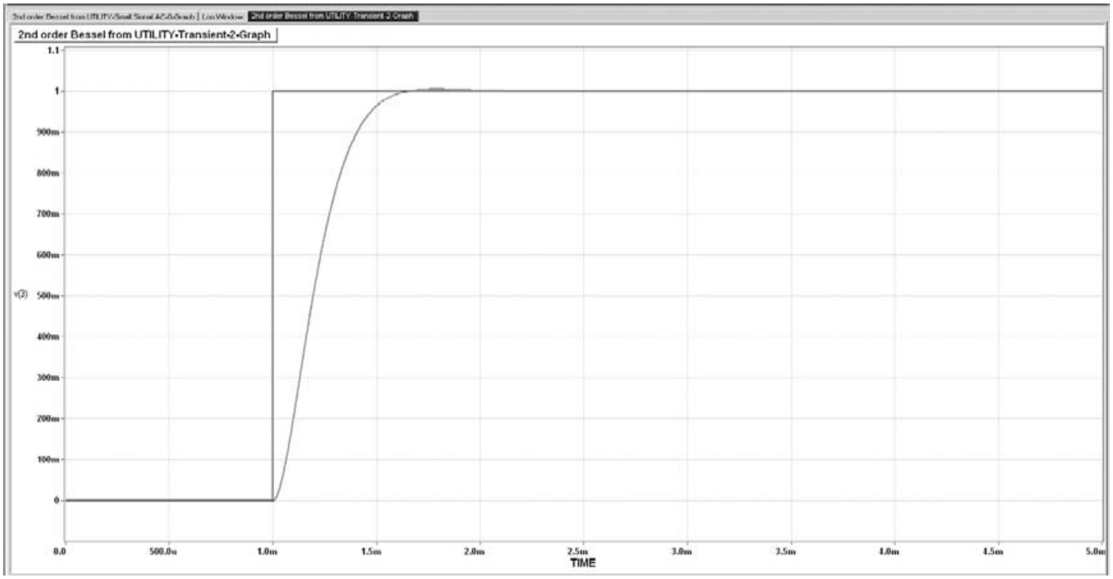


Figure 7.12: Step response of second-order Bessel lowpass filter. −3 dB at 1.00 kHz. There is no overshoot.

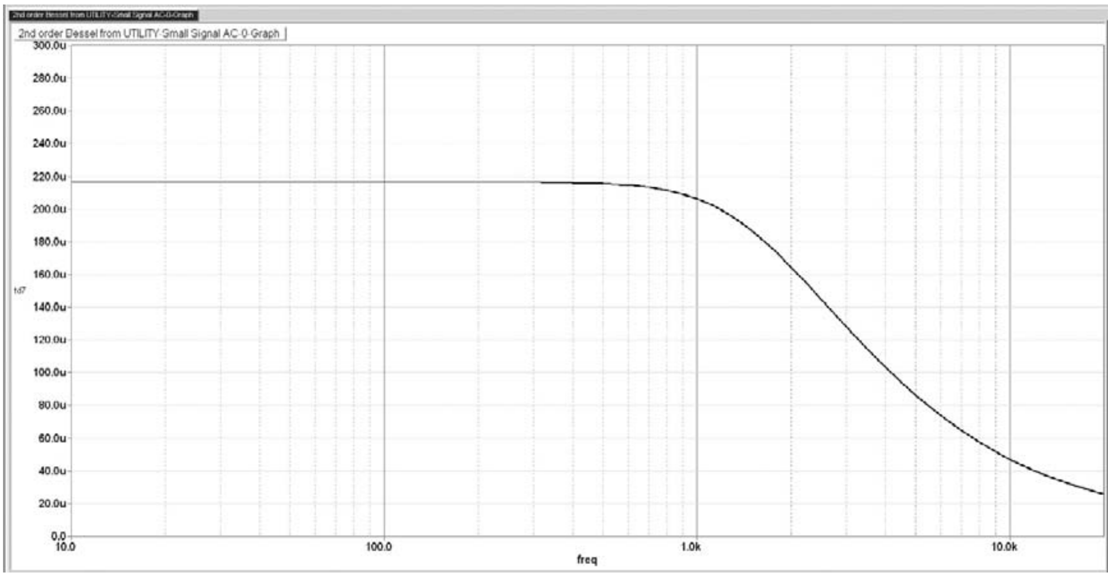


Figure 7.13: Group delay of second-order Bessel lowpass filter. −3 dB at 1.00 kHz. There is no peaking and it is maximally flat.

are prepared to put up with in order to get a faster roll-off. As the ripple increases, the roll-off becomes sharper. The passband ripple might range from 0.5 dB to 3 dB in practical use, with all values between possible. There is, however, nothing whatever to stop you from designing a Chebyshev filter with 0.1, 0.01, or even 0.0001 dB passband ripple, though there is very little point in doing it, as the result is in practice indistinguishable from a Butterworth characteristic; in fact with zero ripple it *is* a Butterworth characteristic. Filter design programs like Filtershop will not let you attempt Chebyshev filters with more than 3 dB of passband ripple.

The Chebyshev filter accepts a non-flat passband as the price for a faster roll-off. The transient response of a Chebyshev filter to a pulse input shows more overshoot and ringing than a Butterworth filter.

It should be said that the famous "passband ripple" does not look much like ripple for lower-order Chebyshev filters; in the second-order case it simply means that the response peaks gently above the 0 dB line by the amount selected, and then falls away as usual; see the 1 dB-Chebyshev response in Figure 7.14, which peaks by the desired +1 dB at 750 Hz, and then passes through 0 dB at the cutoff frequency.

The fourth-order 1 dB-Chebyshev filter has two +1 dB peaks before roll-off, and this is starting to look more like a real "ripple"; the response also passes through 0 dB at the cutoff frequency. For all Chebyshev filters, even-order versions have peaks above the 0 dB line, and a response going through 0 dB at cutoff. Odd-order versions have dips below it.

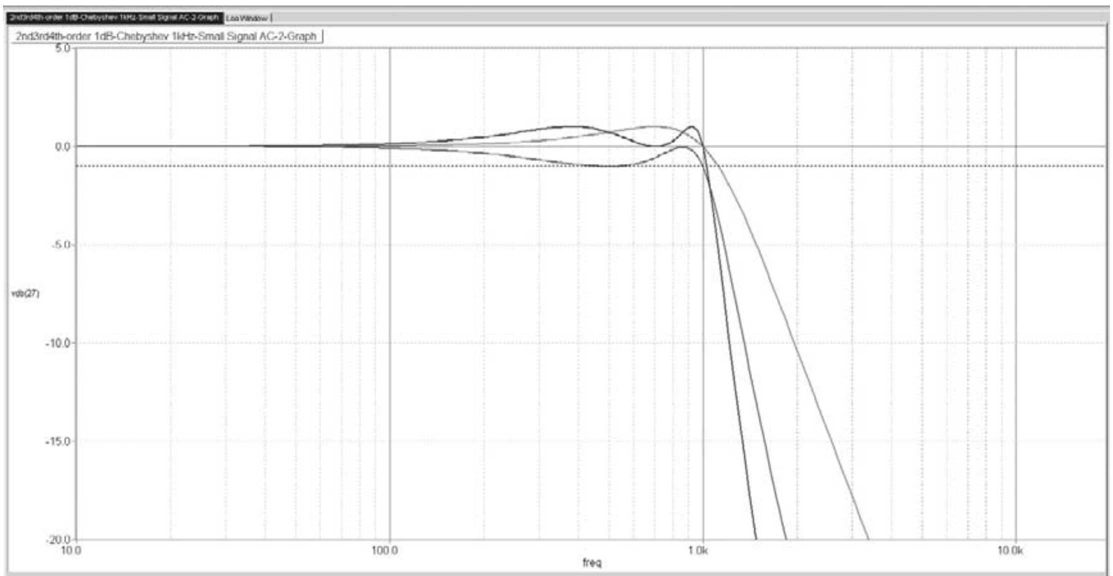


Figure 7.14: Amplitude responses of second-, third-, fourth-order 1 dB-Chebyshev lowpass filters, showing the 1 dB ripples in the passband. Cutoff frequency is 1 kHz. Dotted line at −1 dB.

For higher-order filters, each deviation from and return to the 0 dB line requires a second-order stage in the filter.

Figure 7.15. shows how the response of an odd-order (third, in this case) 1 dB-Chebyshev filter dips 1 dB below the 0 dB line, then goes back up to exactly 0 dB again before falling to pass through the −1 dB line at the cutoff frequency. In the same way a 2 dB-Chebyshev odd-order filter dips 2 dB below the 0 dB line, then goes back up to 0 dB before falling to go through the −2 dB line at the cutoff frequency, and the same applies to a 3 dB-Chebyshev odd-order filter.

It is important to realise that in a Chebyshev filter the frequencies at which the passband ripples occur cannot be controlled independently. Once you have specified the passband ripple and the cutoff frequency of the filter, that's it—you have no more control over the response. This means that attempts to use the ripples of a Chebyshev crossover filter to equalise humps or dips in a drive unit response are unlikely to be successful. It is far better to keep the crossover and equalisation functions in separate circuit blocks.

For a given filter order, a steeper cutoff can be achieved by allowing more passband ripple, as shown by the fourth-order Chebyshev response in Figure 7.16, which has been designed for 1 dB, 2 dB, and 3 dB of passband ripple. The improvement is, however, not exactly stunning. Looking at the response at 3 kHz, the 1-dB case is down −33 dB, the 2-dB case is

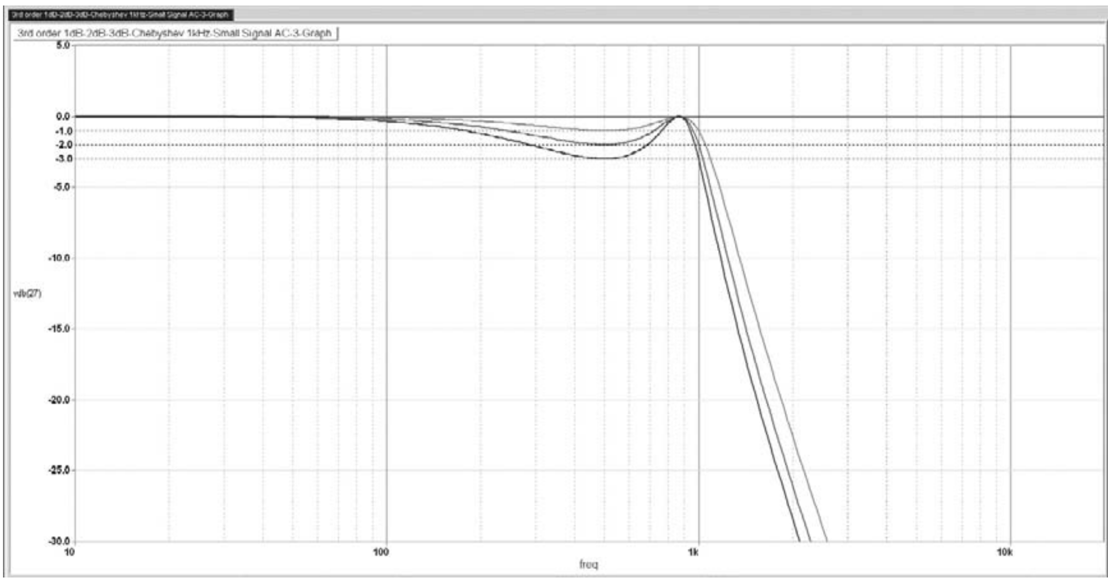


Figure 7.15: Amplitude responses of third-order Chebyshev lowpass filter designed for 1 dB, 2 dB, and 3 dB of passband ripple. The curves go through −1 dB, −2 dB, and −3 dB at the cutoff frequency of 1 kHz.

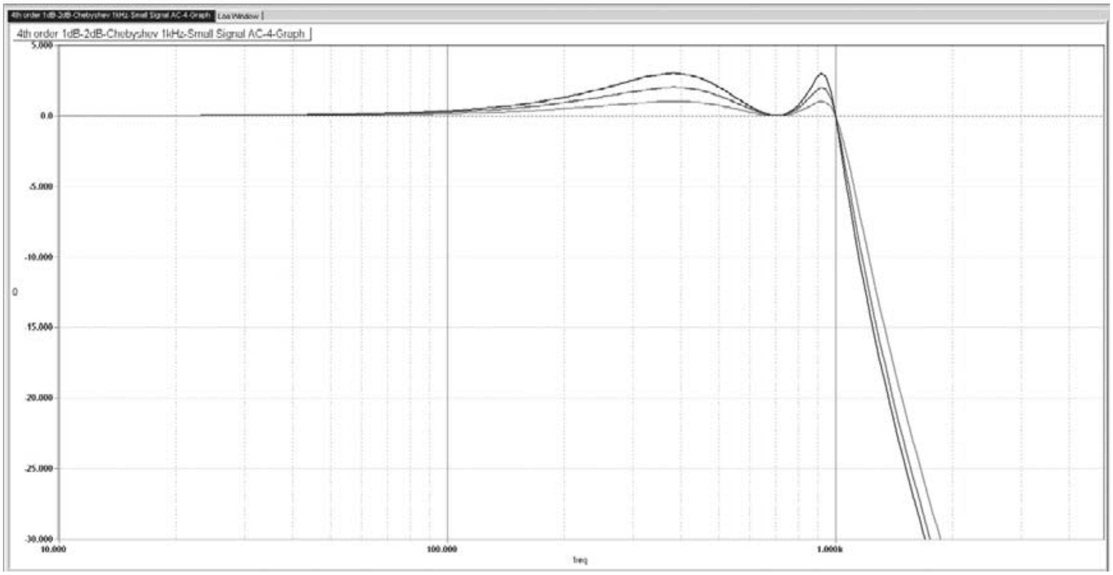


Figure 7.16: Amplitude responses of fourth-order Chebyshev lowpass filter designed for 1 dB, 2 dB, and 3 dB of passband ripple. All the curves go through 0 dB at the cutoff frequency of 1 kHz.

down −35 dB, and the 3-dB case is down −37 dB. In general, if you want a steeper roll-off you go for a higher-order filter.

7.11.5 1 dB-Chebyshev Lowpass Filter

The second-order 1 dB-Chebyschev lowpass filter has a 1 dB peak in the passband at 750 Hz. The amplitude response is shown in Figure 7.17; note that the attenuation at 10 kHz is −39 dB, which is actually slightly worse than the −40 dB of the Butterworth filter. The other responses are shown in Figures 7.18–7.20.

7.11.6 3 dB-Chebyshev Lowpass Filter

This has a 3 dB peak in the passband at 710 Hz, slightly lower than for the 1 dB-Chebyshev. The amplitude response is shown in Figure 7.21; note that the attenuation at 10 kHz is −43 dB, which is somewhat better than the −40 dB of the Butterworth filter. The other responses are shown in Figures 7.22–7.24.

7.12 Higher-Order Filters

In many cases second-order filters do not give a fast enough roll-off. As we have seen, going from a second-order Butterworth filter to a second-order 3-dB Chebyshev filter does

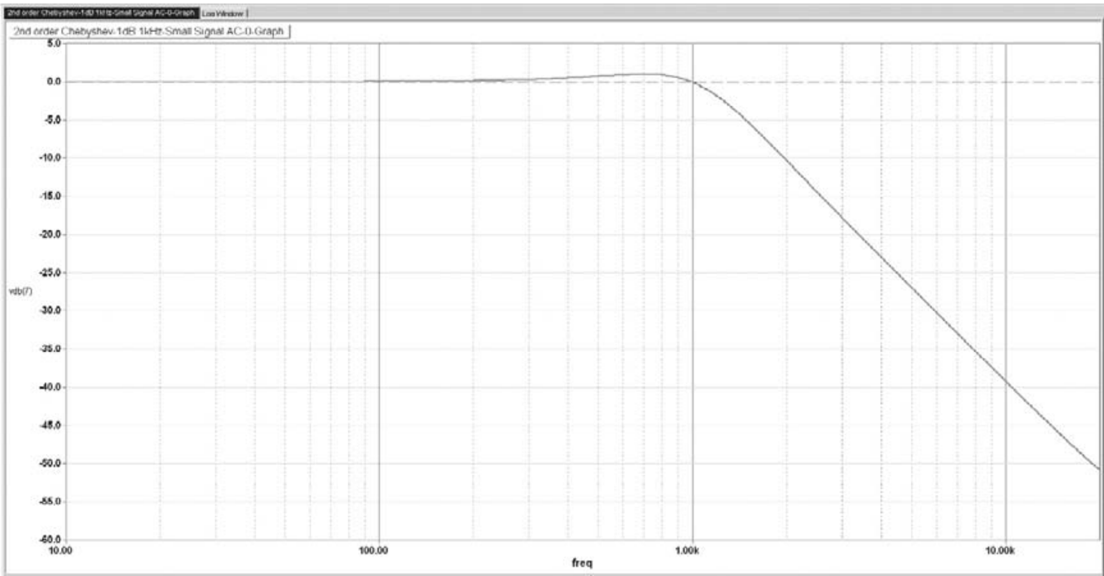


Figure 7.17: Amplitude response of a second-order 1 dB-Chebyshev lowpass filter (Q = 0.9565). Cutoff frequency is 1 kHz. Attenuation at 10 kHz is −39 dB.

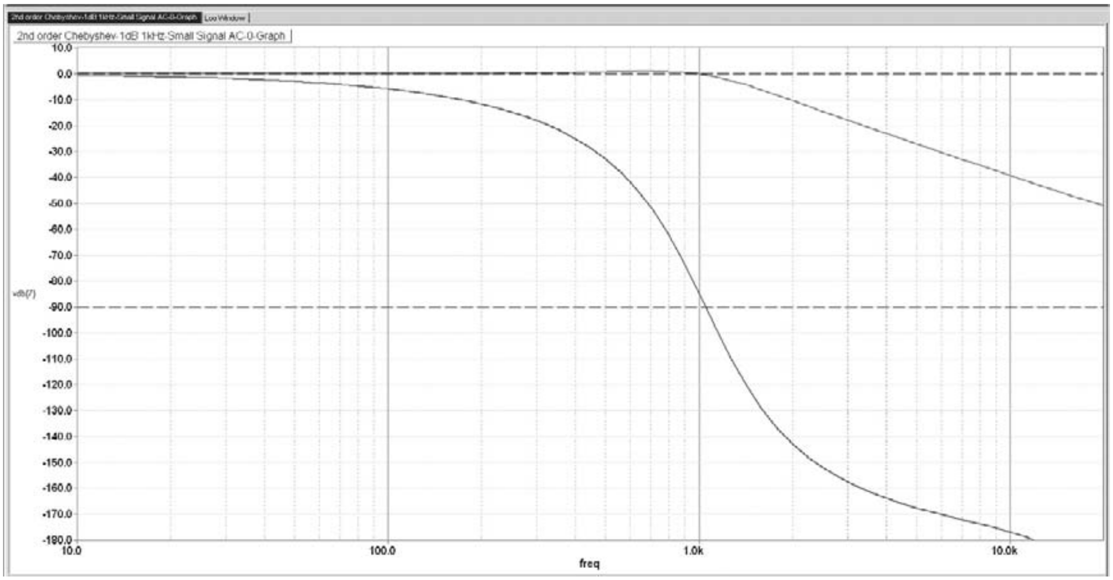


Figure 7.18: Phase response of a second-order 1 dB-Chebyshev lowpass filter. 0 dB at 1.00 kHz. The phase-shift is 90° at 1.1 kHz. Upper trace is amplitude response.

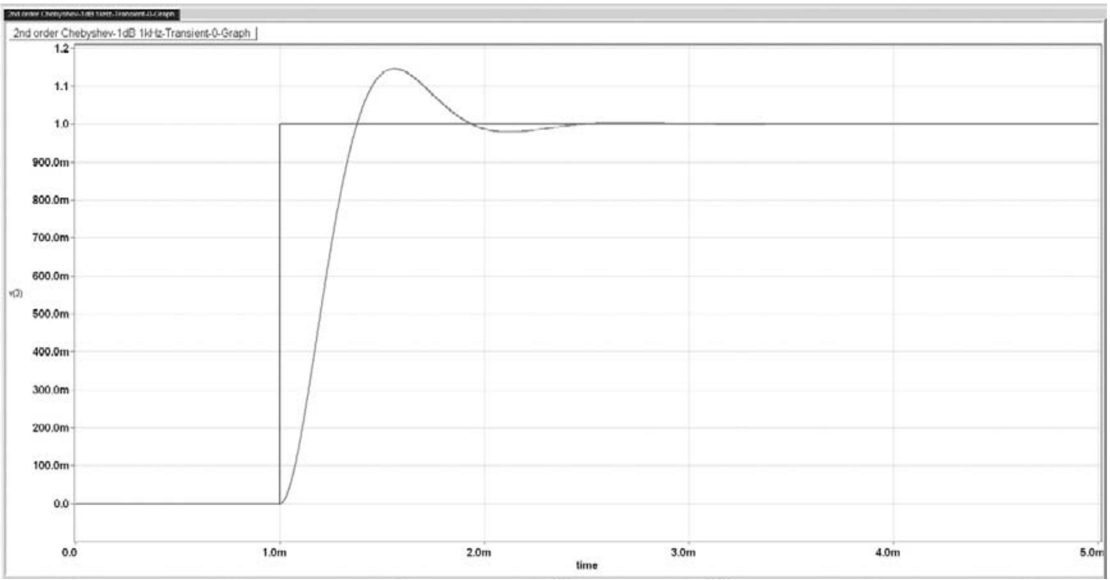


Figure 7.19: Step response of a second-order 1 dB-Chebyshev lowpass filter. 0 dB at 1.00 kHz. There is significant overshoot of 14%, followed by undershoot of 2%.

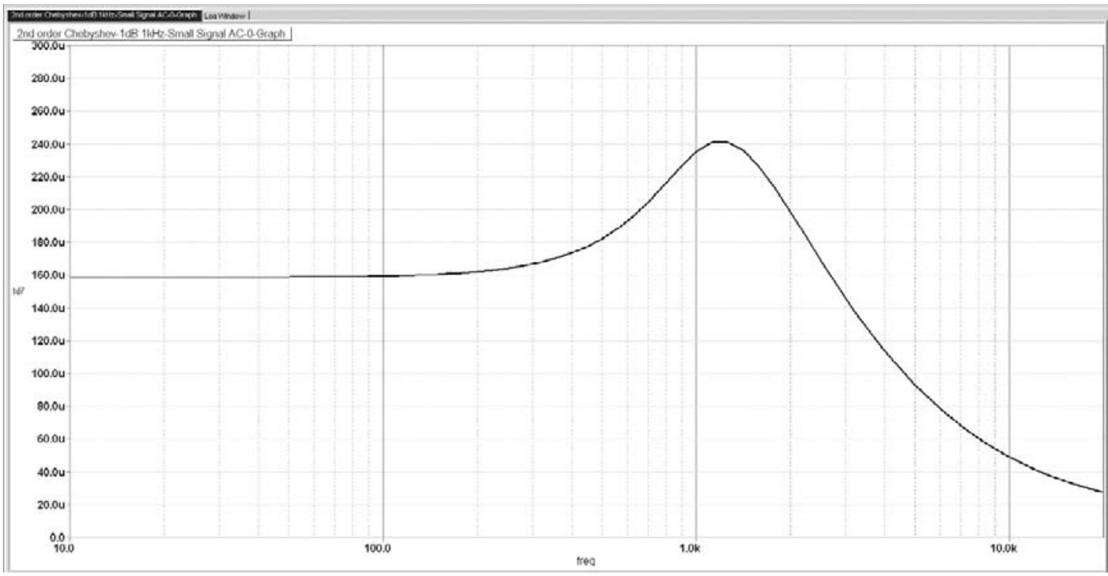


Figure 7.20: Group delay of a second-order 1 dB-Chebyshev lowpass filter. 0 dB at 1.00 kHz. There is now considerable peaking in the delay, reaching 50%.

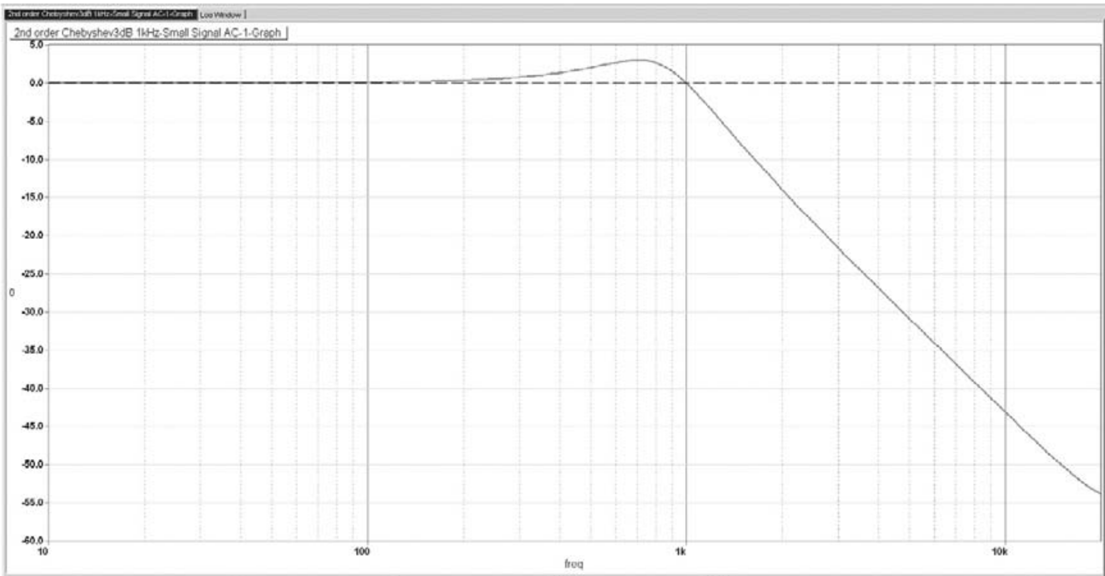


Figure 7.21: Amplitude response of a second-order 3 dB-Chebyshev lowpass filter (Q = 1.305). Cutoff frequency is 1 kHz. Attenuation at 10 kHz is −43 dB.

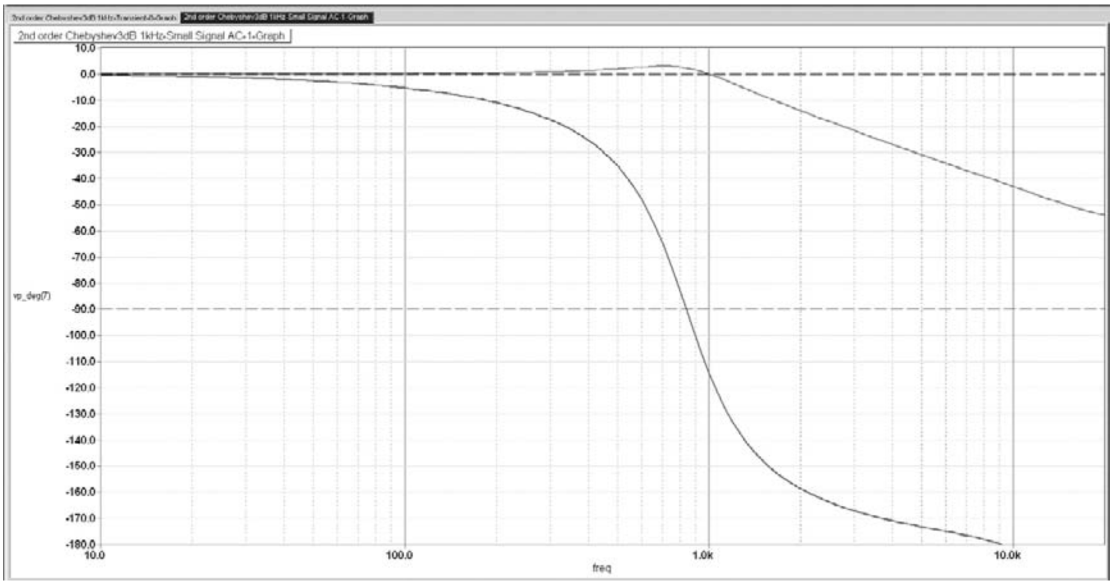


Figure 7.22: Phase response of a second-order 3 dB-Chebyshev lowpass filter. 0 dB at 1.00 kHz. The phase-shift is 90° at 0.84 kHz. Upper trace is amplitude response.

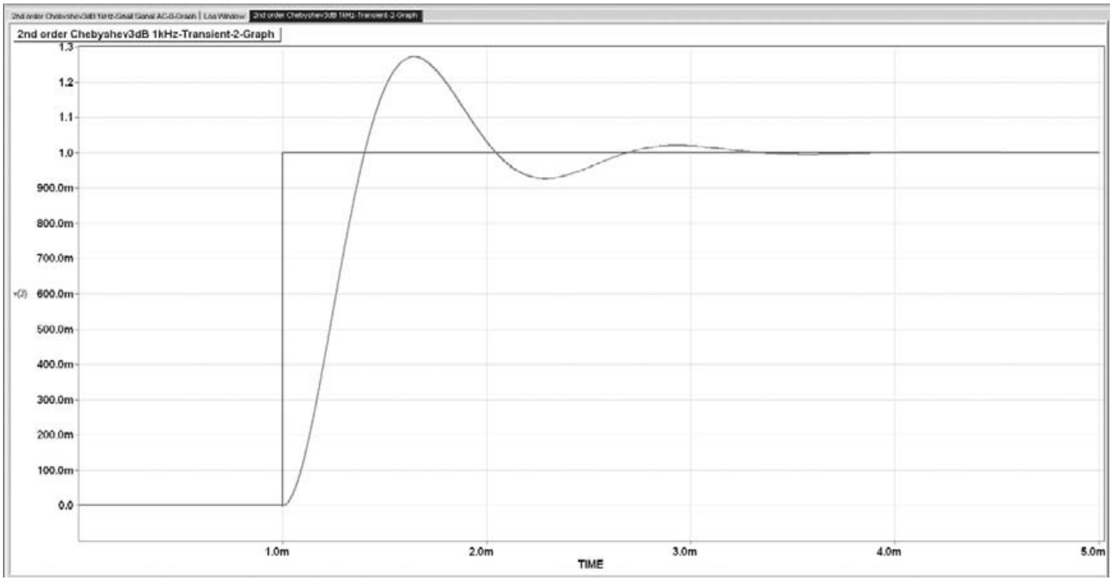


Figure 7.23: Step response of a second-order 3 dB-Chebyshev lowpass filter. 0 dB at 1.00 kHz. There is a serious overshoot of 28%, followed by an undershoot of 8% and continuing damped oscillation.

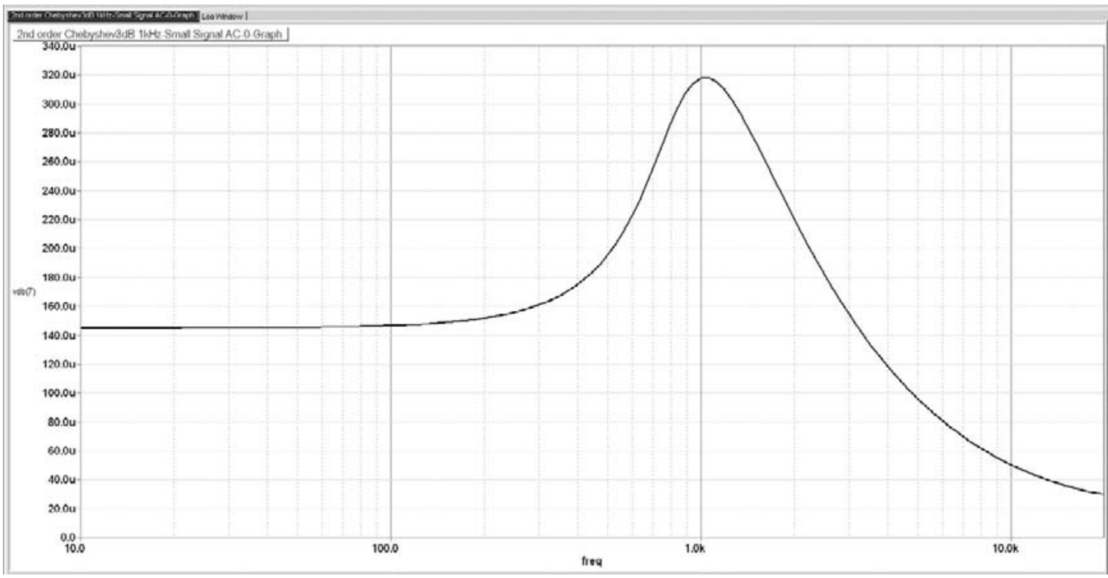


Figure 7.24: Group delay of a second-order 3 dB-Chebyshev lowpass filter. 0 dB at 1.00 kHz. The peaking is now huge at 120%; note scale change to fit it in.

not help much; in a 1 kHz filter we gain only 3 dB more attenuation at the expense of a 3 dB peak in the passband, which is not very helpful in most circumstances.

If you want a faster roll-off as well as a maximally flat passband, then the answer is to go to a higher order of Butterworth filter. Likewise, if you have a 1 dB-Chebyshev filter that gives a passband ripple you either want or can put up with, but if you need a steeper roll-off, then a higher order filter is the way to go.

The mathematical equations describing the response of a second-order filter contains the square of frequency, and cannot be simplified, but it is a mathematical fact (which I devoutly hope you will take on trust, because this is not the place for me to prove it) that any third-order equation of this sort (technically, a polynomial) can be broken down into a second-order equation multiplied by a first-order equation, which represents a second-order filter cascaded with a first-order filter. Furthermore, it is equally a fact that any complex equation describing a high-order filter, be it fourth, fifth, or higher, can be factorised, into a combination of second order and (sometimes) first order equations multiplied together. This leads directly to making a high-order filter by cascading first- and second-order filters. Thus a fourth-order filter can be made by cascading two second-order filters, and a fifth-order filter can be made by cascading two second-order filters and a first-order filter, and so on.

It is not compulsory to break things down as much as possible, into second and first-order stages; it would be quite possible to make a fifth-order filter by cascading a third-order stage with a second-order stage. However, using only second- and first-order stages is almost always the method adopted because although it may use more amplifiers, it is the best way to achieve minimum component sensitivities; in other words the circuits are more tolerant of component value tolerances. There is much more on component sensitivity in the next chapter.

One thing you cannot do is make a fifth-order filter by cascading five first-order stages; nor can you do it by cascading one second-order stage with three first-order stages. The rules are that an odd-order filter will be composed of a number of second-order stages and one first-order stage, while an even-order filter will be made up solely of second-order stages.

Putting together higher-order filters of the all-pole type (that term will be explained shortly) is actually quite straightforward if you have the information you need. It is provided here.

The order of filter stages is important. Filter textbooks usually put the stages in order of increasing Q, so that large signals will be attenuated by earlier stages and so are less likely to cause clipping in the high-Q stages. This policy probably derives from telecommunications practice, where clipping is much more of a problem than a small increase in the noise floor. However, for high-order lowpass filters at least, if noise performance is more important, then the order should be reversed, so that the low-gain low-Q stages attenuate the higher noise from the earlier high-Q stages.

7.12.1 Butterworth Filters Up to Eighth-Order

Figure 7.25 shows how to make Butterworth filters up to the eighth order. All you need to know is the cutoff frequency and the Q of each second-order stage, and the cutoff frequency of the first order stages. This information is given in Table 7.3. You are not very likely to need seventh- or eight-order filters in a crossover design, but if you do, then the cutoff frequencies and Qs are given at the end of Table 7.3; both seventh- and eighth-order filters require four stages. The stages are arranged in the conventional way, in order of increasing Q, though as explained in the previous section this may not always be the best order for crossover work. The parameters are given to four decimal places, though not all of this accuracy will be needed for practical circuit design.

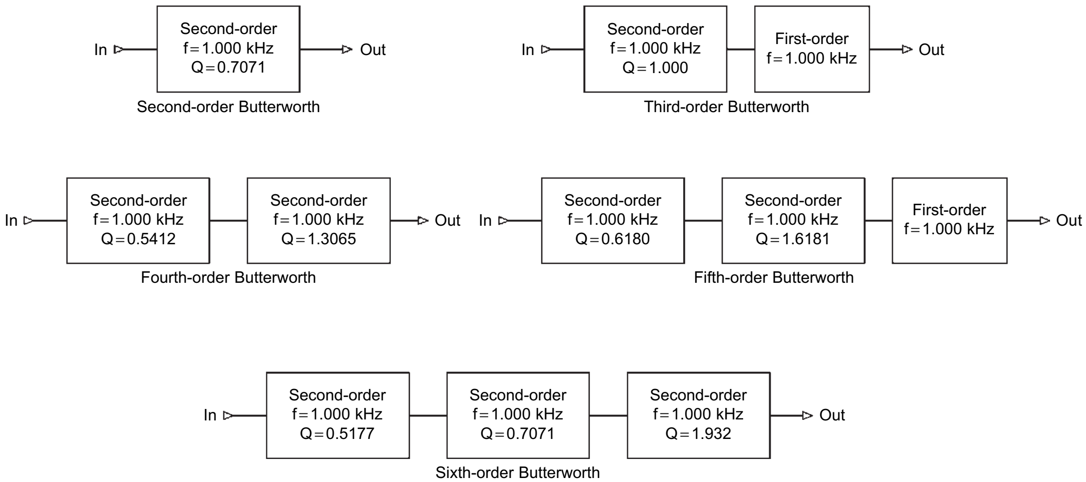


Figure 7.25: Butterworth lowpass filters up to sixth order made from second- and first-order stages cascaded; all stages have the same cutoff frequency of 1.00 kHz.

Table 7.3: Frequencies and Q's for Butterworth Filters up to Eighth Order. Stages Are Arranged in Order of Q, with the First-Order Section at the End for the Odd Order Filters

Order	Freq 1	Q 1	Freq 2	Q 2	Freq 3	Q 3	Freq 4	Q 4
2	1.0000	0.7071						
3	1.0000	1.0000	1.0000	n/a				
4	1.0000	0.5412	1.0000	1.3065				
5	1.0000	0.6180	1.0000	1.6181	1.0000	n/a		
6	1.0000	0.5177	1.0000	1.7071	1.0000	1.9320		
7	1.0000	0.5549	1.0000	0.8019	1.0000	2.2472	1.0000	n/a
8	1.0000	0.5098	1.0000	0.6013	1.0000	0.8999	1.0000	2.5628

Butterworth filter design is relatively easy because every stage has the same cutoff frequency—only the Q's vary. To design a second-order Butterworth filter that is −3 dB at 1 kHz, you use the standard second-order filter design equations (see Chapter 8) to make a single stage with a cutoff frequency of 1 kHz and a Q of 0.7071.

For a third-order Butterworth filter you consult Table 7.3, which shows you need a second-order stage with a cutoff frequency of 1 kHz and a Q of 1.000, followed by a first-order stage that also has a cutoff frequency of 1 kHz.

A fourth-order Butterworth filter has a second-order stage with a cutoff frequency of 1 kHz and a Q of 0.5412, followed by another second-order stage with a cutoff frequency of 1 kHz and a Q of 1.3065; The Q's are always different. The process is the same for higher-order filters. The amplitude responses for second-, third-, and fourth-order Butterworth filters are compared in Figure 7.26.

Table 7.3 shows that the higher the order of the filter, the wider the ranges of Q's used in it. It is a handy property of Butterworth filters that the maximum Q's needed are relatively low, and therefore do not require very precise components to achieve acceptable accuracy.

The next few diagrams (Figures 7.27 to 7.29) demonstrate just how these high-order filters work. In Figure 7.27 the upper trace is the output of the second order stage; because its Q is 1.0 rather than 0.7071 it peaks by +1 dB just before the roll-off. When this is combined with the slow roll-off of the first-order stage, the combined response is held up by the peak to give

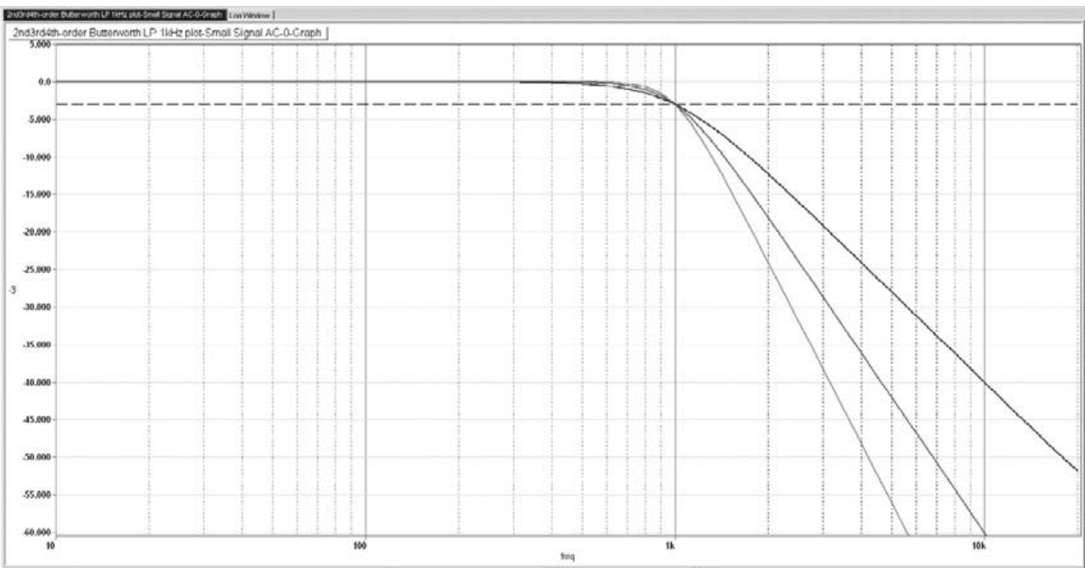


Figure 7.26: Amplitude responses of second-, third-, and fourth-order Butterworth lowpass filters. Cutoff frequency is 1.00 kHz.

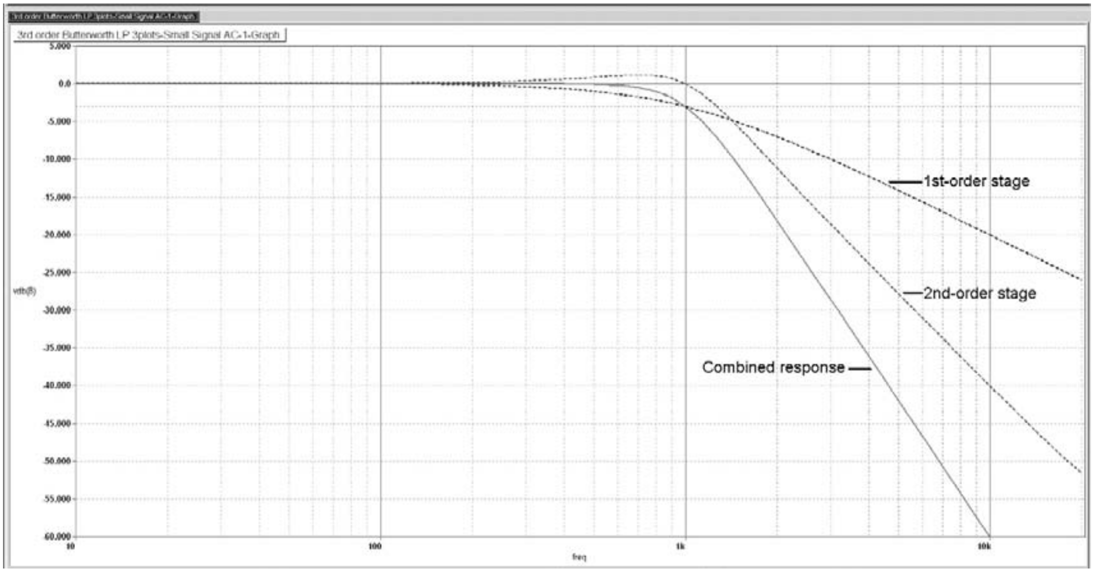


Figure 7.27: Amplitude response of second-order stage, first-order stage, and the final output for a third-order Butterworth lowpass filter. Cutoff frequency is 1 kHz.

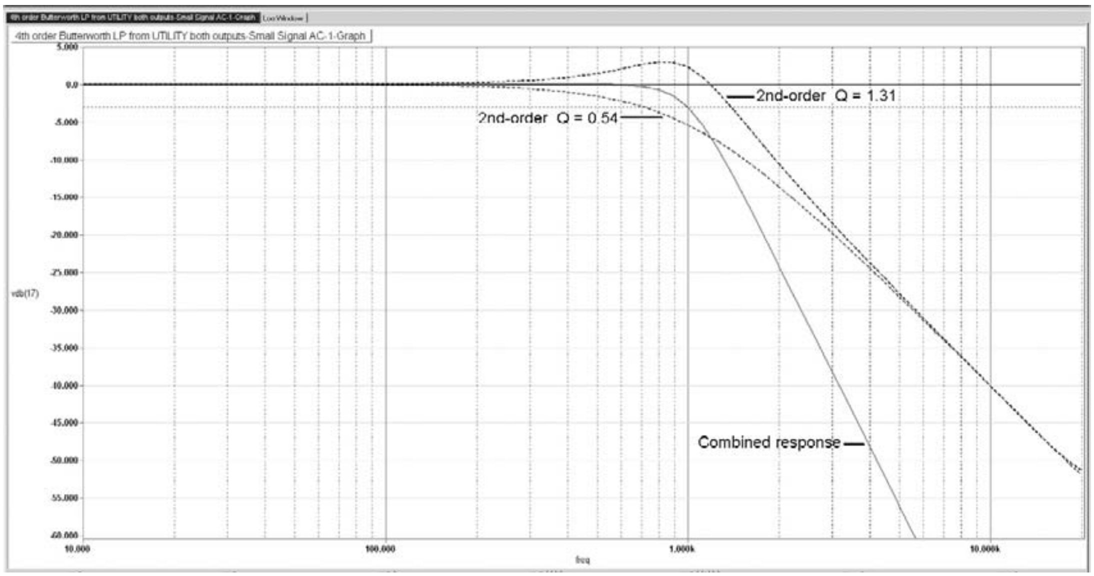


Figure 7.28: Amplitude response of the two second-order stages in a fourth-order Butterworth lowpass filter. Cutoff frequency is 1 kHz.

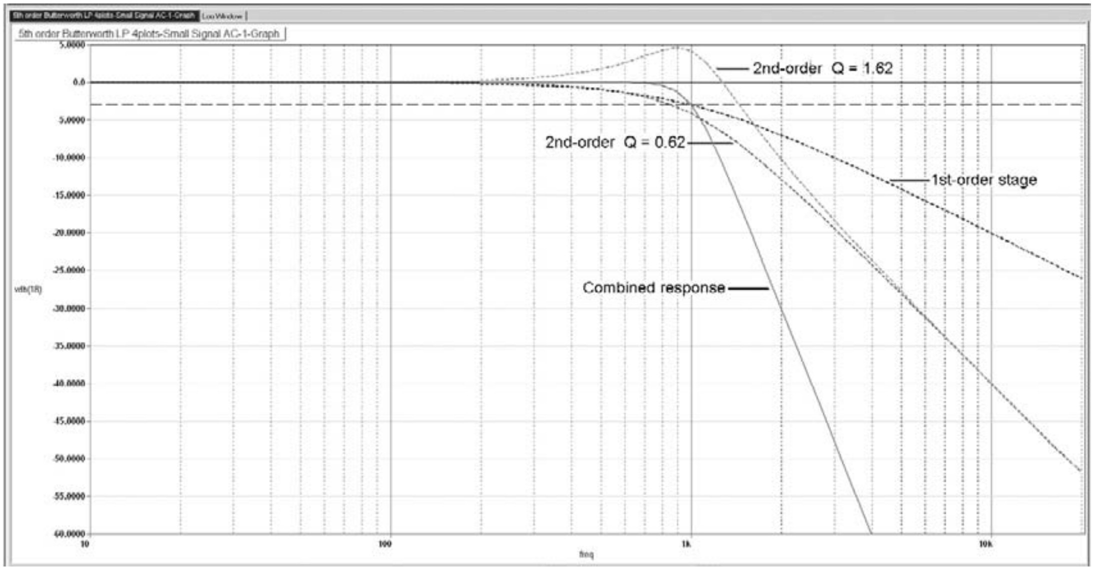


Figure 7.29: Amplitude response of the three stages in a fifth-order Butterworth lowpass filter. Cutoff frequency is 1 kHz.

a maximally flat response until it falls suddenly around the cutoff frequency. This naturally only works if the first-order cutoff frequency exactly matches the second-order cutoff frequency; properly mismatching these will give either an unduly slow roll-off or unwanted peaking in the passband. The Q of the second-order stage also needs to be just right to get the requisite amount of peaking. The higher the filter order, the more precise this matching needs to be, and the greater the demands on component accuracy.

Because we have a second-order stage cascaded with a first-order stage, the ultimate roll-off rate is 18 dB/octave.

The fourth-order filter is made up of two second-order stages cascaded. One of these has a relatively low Q of 0.54 while the other has a relatively high Q of 1.31. Figure 7.28 shows that the latter gives a sharper peak in the passband than the second-order stage did in the third-order filter, and this sharper peak is cancelled out by the other second-order stage, which rolls off faster than a first-order stage. Because we have two second-order stages cascaded, the ultimate roll-off rate is 24 dB/octave.

The fifth order Butterworth filter of Figure 7.29 is more complicated, being made up of three stages. As with the fourth-order filter, there are two second-order stages, one with high Q and the other with low Q. The peaking in the high Q stage is higher, but interacts with the low Q stage and the first-order stage to once more give a maximally flat passband.

7.12.2 Linkwitz–Riley Filters Up to Eighth-Order

Linkwitz–Riley filters, with Butterworth filters, are the only ones that have all stages set to the same cutoff frequency, as shown in Table 7.4. Unlike Butterworth filters, only a few different values of Q are used. Using the table with the stage cutoff frequencies shown will give an amplitude response −6 dB down at the cutoff frequency; this is the "natural" cutoff attenuation for a Linkwitz–Riley filter. If you want to use the more familiar −3 dB criterion for cutoff, then you will need to change all the stage frequencies in Table 7.4 to 1.543.

A comparison of second-, third-, and fourth-order order filters is shown in Figure 7.30.

Table 7.4: Frequencies and Qs for Linkwitz–Riley Filters up to Eighth Order. Stages Are Arranged in Order of Increasing Q, Odd-Order Filters Have the First-Order Section at the End with No Q Shown

Order	Freq 1	Q 1	Freq 2	Q 2	Freq 3	Q 3	Freq 4	Q 4
2	1.0000	0.5000						
3	1.0000	0.7071	1.0000	n/a				
4	1.0000	0.7071	1.0000	0.7071				
5	1.0000	0.7071	1.0000	1.0000	1.0000	n/a		
6	1.0000	0.5000	1.0000	1.0000	1.0000	1.0000		
7	1.0000	0.5412	1.0000	1.0000	1.0000	1.3066	1.0000	n/a
8	1.0000	0.5412	1.0000	0.5412	1.0000	1.3066	1.0000	1.3066

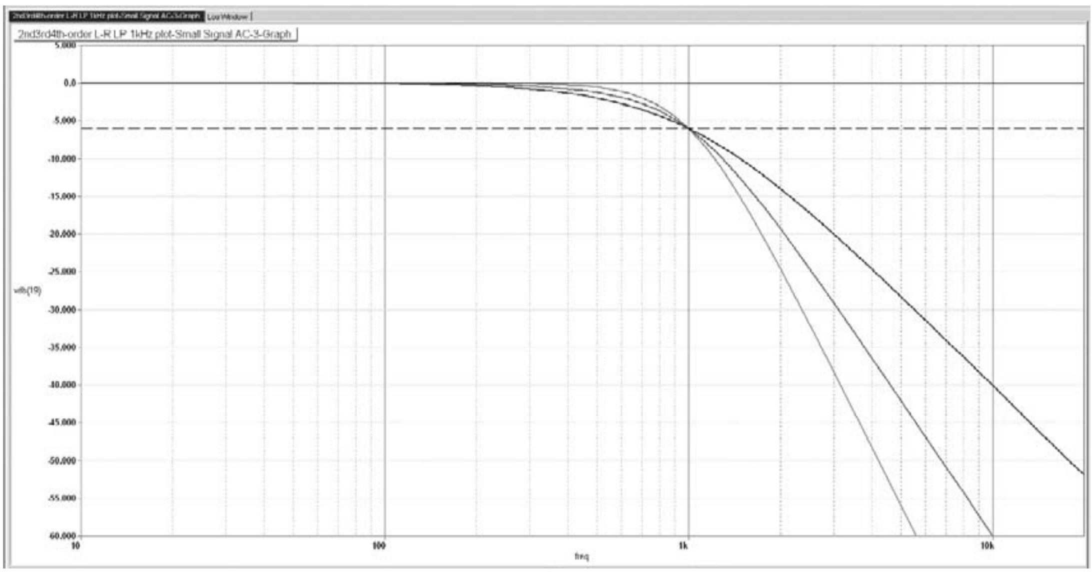


Figure 7.30: Amplitude response of second-, third-, and fourth-order Linkwitz–Riley lowpass filters. Note that 1 kHz cutoff is at −6 dB.

You will note that the fourth-order filter is made up of two cascaded Butterworth filters with Q = 1/√2 (= 0.7071) and this is the most common arrangement for fourth-order Linkwitz–Riley crossovers. For Linkwitz–Riley filters above fourth order, higher Q values must be used.

7.12.3 Bessel Filters Up to Eighth-Order

Bessel filters have different cutoff frequencies as well as different Q's for each stage in the higher-order filters. The cutoff frequencies here are based on the amplitude response being 3 dB down at the required frequency. Thus, to design a second-order lowpass Bessel filter that is −3 dB at 1 kHz, you use the standard second-order filter design equations in Chapter 8 to make a stage with a cutoff frequency of 1.27 kHz and a Q of 0.5773.

A comparison of second-, third-, and fourth-order order filters is shown in Figure 7.31.

Table 7.5 shows the cutoff frequencies and Q's for each stage. These are no longer all the same, as they were for Butterworth and Linkwitz–Riley filters so a bit more care is required. The frequencies are normalised on 1.000, so if you want a third-order Bessel lowpass filter with a cutoff of 1.00 kHz, then you design a second-order stage with a cutoff frequency of 1.452 kHz and a Q of 0.691, and cascade it with a first-order section having a cutoff frequency of 1.327 kHz. A vital point is that if you want a highpass Bessel filter with a cutoff of 1.00 kHz for the other half of a crossover, then you must use the reciprocal of

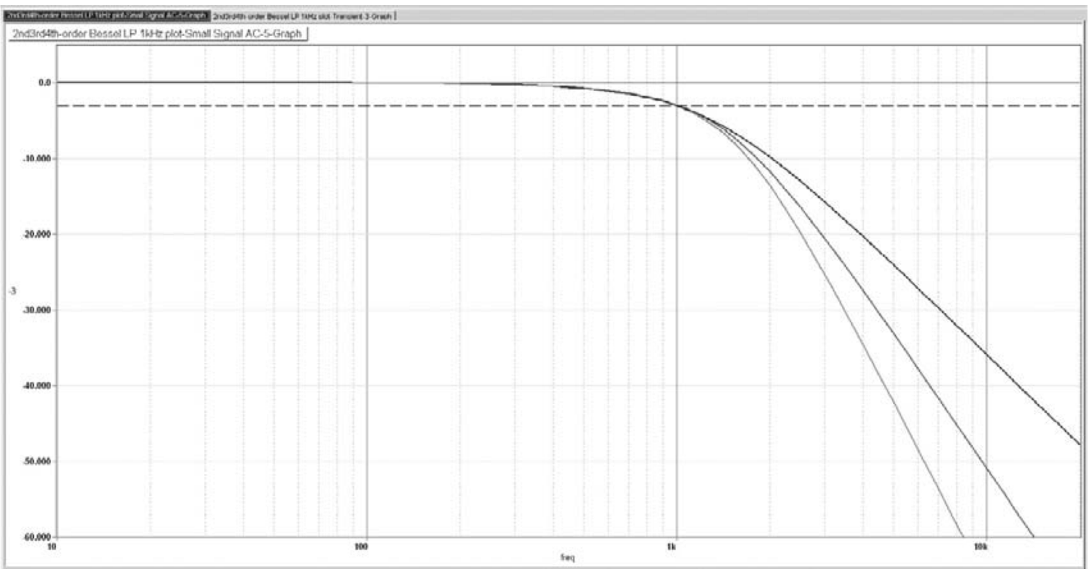


Figure 7.31: Amplitude response of second-, third-, and fourth-order Bessel lowpass filters. Cutoff frequency 1 kHz. Compare Butterworth filters in Figure 7.26.

Table 7.5: Frequencies and Qs for Bessel Lowpass Filters up to Eighth Order. For Highpass Filters Use the Reciprocal of the Frequency. Stages Are Arranged in Order of Increasing Q, Odd-Order Filters Have the First-Order Section at the End with No Q Shown

Order	Freq 1	Q 1	Freq 2	Q 2	Freq 3	Q 3	Freq 4	Q 4
2	1.2736	0.5773						
3	1.4524	0.6910	1.3270	n/a				
4	1.4192	0.5219	1.5912	0.8055				
5	1.5611	0.5635	1.7607	0.9165	1.5069	n/a		
6	1.6060	0.5103	1.6913	0.6112	1.9071	1.0234		
7	1.7174	0.5324	1.8235	0.6608	2.0507	1.1262	1.6853	n/a
8	1.7837	0.5060	1.8376	0.5596	1.9591	0.7109	2.1953	1.2258

the frequency in each case, because a highpass filter is a lowpass filter mirrored along the frequency axis, if you see what I mean. Thus, a third-order highpass Bessel filter is made up of a second-order stage with a cutoff frequency of 1/1.452 kHz = 688.5 Hz and a Q of 0.691, cascaded with a first-order section having a cutoff frequency of 1/1.327 kHz = 753.6 Hz. The lowpass and highpass third-order filters are both −3 dB at 1 kHz.

7.12.4 Chebyshev Filters Up to Eighth-Order

Like Bessel filters, Chebyshev filters have different cutoff frequencies as well as different Q's for each stage in the higher-order filters. Once again, if you want a highpass Chebyshev filter, then you must use the reciprocal of the frequency in each case.

The stage frequencies and Q's for Chebyshev filters with passband ripple of 0.5, 1, 2, and 3 dB are shown in Tables 7.6 to 7.9. You can see that as the passband ripple increases (and the steepness of roll-off also increases), higher Q's are required. For the high-order filters, these high Q's are a serious problem as they require great component precision to implement them with the required accuracy. The usual Sallen & Key and Multiple-FeedBack filter configurations are not up to the job and more sophisticated circuitry is necessary. It is fair to say that filters like these nowadays are avoided like the plague; if an eighth-order 3 dB-Chebyshev filter is the answer, then you might want to look at changing the question. If such sharp filtering is really required (and that situation is not very likely in crossover design), then it will probably be cheaper to convert to the digital domain and use DSP techniques, which can provide stable and accurate filtering of pretty much any kind you can imagine.

Table 7.6: Frequencies and Qs for 0.5 dB-Chebyshev Lowpass Filters up to Eighth Order. For Highpass Filters Use the Reciprocal of the Frequency. Stages Are Arranged in Order of Increasing Q, Odd-Order Filters Have the First-Order Section at the End with No Q Shown

Order	Freq 1	Q 1	Freq 2	Q 2	Freq 3	Q 3	Freq 4	Q 4
2	1.2313	0.8637						
3	1.0689	1.7062	0.6265	n/a				
4	0.5970	0.7051	1.0313	2.9406				
5	0.6905	1.1778	1.0177	4.5450	0.3623	n/a		
6	0.3962	0.6836	0.7681	1.8104	1.0114	6.5128		
7	0.5039	1.0916	0.8227	2.5755	1.0080	8.8418	0.2562	n/a
8	0.2967	0.6766	0.5989	1.6107	0.8610	3.4657	1.0059	11.5308

Table 7.7: Frequencies and Qs for 1 dB-Chebyshev Lowpass Filters up to Eighth Order. For Highpass Filters Use the Reciprocal of the Frequency. Stages Are Arranged in Order of Increasing Q, Odd-Order Filters Have the First-Order Section at the End with No Q Shown

Order	Freq 1	Q 1	Freq 2	Q 2	Freq 3	Q 3	Freq 4	Q 4
2	1.0500	0.9565						
3	0.9971	2.0176	0.4942	n/a				
4	0.5286	0.7845	0.9932	3.5600				
5	0.6552	1.3988	0.9941	5.5538	0.2895	n/a		
6	0.3532	0.7608	0.7468	2.1977	0.9953	8.0012		
7	0.4800	1.2967	0.8084	3.1554	0.9963	10.9010	0.2054	n/a
8	0.2651	0.7530	0.5838	1.9564	0.8506	4.2661	0.9971	14.2445

Table 7.8: Frequencies and Qs for 2 dB-Chebyshev Lowpass Filters up to Eighth Order. For Highpass Filters Use the Reciprocal of the Frequency. Stages Are Arranged in Order of Increasing Q, Odd-Order Filters Have the First-Order Section at the End with No Q Shown

Order	Freq 1	Q 1	Freq 2	Q 2	Freq 3	Q 3	Freq 4	Q 4
2	0.9072	1.1286						
3	0.9413	2.5516	0.3689	n/a				
4	0.4707	0.9294	0.9637	4.5939				
5	0.6270	1.7751	0.9758	7.2323	0.2183	n/a		
6	0.3161	0.9016	0.7300	2.8443	0.9828	10.4616		
7	0.4609	1.6464	0.7971	4.1151	0.9872	14.2802	0.1553	n/a
8	0.2377	0.8924	0.5719	2.5327	0.8425	5.5835	0.9901	18.6873

Table 7.9: Frequencies and Qs for 3 dB-Chebyshev Lowpass Filters up to Eighth Order. For Highpass Filters Use the Reciprocal of the Frequency. Stages Are Arranged in Order of Increasing Q, Odd-Order Filters Have the First-Order Section at the End with No Q Shown

Order	Freq 1	Q 1	Freq 2	Q 2	Freq 3	Q 3	Freq 4	Q 4
2	0.8414	1.3049						
3	0.9160	3.0678	0.2986	n/a				
4	0.4426	1.0765	0.9503	5.5770				
5	0.6140	2.1380	0.9675	8.8111	0.1775	n/a		
6	0.2980	1.0441	0.7224	3.4597	0.9771	12.7899		
7	0.4519	1.9821	0.7920	5.0193	0.9831	17.4929	0.1265	n/a
8	0.2243	1.0337	0.5665	3.0789	0.8388	6.8251	0.9870	22.8704

Comparisons of second-, third-, and fourth-order order Chebyshev filters of various kinds can be seen above in Figures 7.14 to 7.16.

7.13 More Complex Filters—Adding Zeros

All the filter types we have looked at so far can be made by plugging together second-order and first-order lowpass or highpass stages in cascade. This is true no matter how high the filter order. They are technically known as "all-pole filters" which basically means that they combine different sorts of roll-off, but once it is properly begun the response *stays* rolled-off; it does not come back up again. There are no deep notches in the amplitude response; once the roll-off is established it just keeps on going down. However, filters with notches in their stopband that plunge to the infinite depths (in theory, at least) have their uses, and their study makes up a large proportion of filter theory.

When a faster roll-off than Butterworth is required, without the passband amplitude ripples of the Chebyshev, one possibility is the Inverse Chebyshev, which adds a notch in the response just outside the passband. Notches have steep and accelerating slopes as you approach the actual notch frequency, so the roll-off is much accelerated.

Elliptical (Cauer) filters permit ripple in the passband *and* have notches in the response just outside the passband, and offer even steeper roll-off slopes. They are also very economical on hardware. Suppose you need a serious lowpass filter that must roll-off from −0.5 dB to −66 dB in a single octave (not very likely in crossover design, but stay with me). This would need a 13th-order Butterworth filter, or an 8th-order Chebyshev, but a 5th-order Cauer filter can do the job, with much greater economy in components and also in power, because less opamps are required.

In filter design the notch frequencies are known as "zeros" because they are the frequencies at which the complex equations describing the filter response give a value of zero—in other words infinite attenuation. Real filters do not have infinitely deep notches, as the depth usually depends on component tolerances and opamp gain-bandwidths.

The design procedures for these filters are not at all straightforward, and I am simply going to show some design examples. These can have their component values scaled in the usual away to obtain different cutoff frequencies.

It is of course always possible to add notches to a filter response by, for example, cascading a Butterworth filter with a notch filter placed suitably in the stopband. This is, however, not as efficient as using Inverse Chebyshev or Cauer filters (though it is conceptually much simpler), because in the latter the notch is properly integrated with the filter response and so better passband flatness and sharper roll-offs are obtained.

7.13.1 Inverse Chebyshev Filters (Chebyshev Type II)

The Inverse Chebyshev filter, also known as the Chebyshev Type II filter, does not have amplitude ripples in the passband; instead it has notches (zeros) in the stopband. Like the Chebyshev filter, it is directed towards getting a faster roll-off than a Butterworth filter while meeting other conditions; it offers a maximally flat passband, a moderate group delay, and an equi-ripple stop band. The cutoff frequency is usually defined to be at the −3 dB level, though other definitions can be used. The Inverse Chebyshev filter uses zeros so it is not an all-pole filter. These filters are complicated to design, and it would not be a good use of space to try and plod through the procedures here. Instead, I am presenting a finished design which can be easily scaled for different frequencies.

Figure 7.32 shows a simple fifth-order Inverse Chebyshev filter with a cutoff frequency of 1 kHz. It is made up of two lowpass notch filters followed by a first-order lowpass filter. The most familiar notch filter is the symmetrical sort, where the gains on either side of the notch are the same; there are, however, also lowpass notch filters where as the frequency

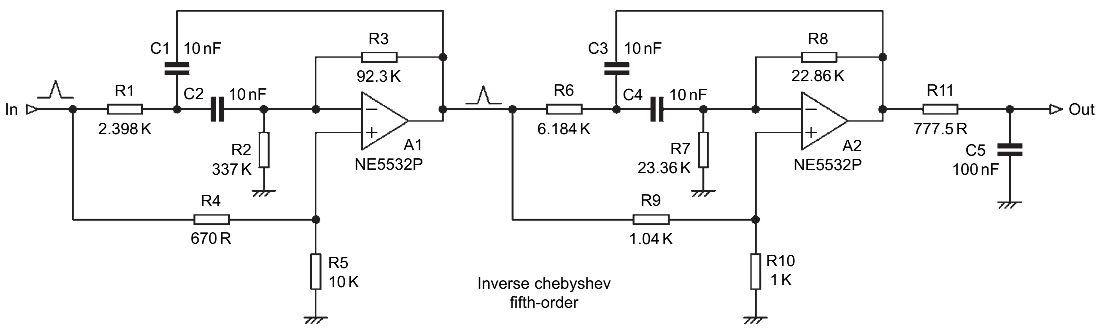


Figure 7.32: A fifth-order Inverse Chebyshev lowpass filter. Made up of two lowpass notch filters followed by a first-order lowpass. Cutoff frequency 1 kHz.

increases, the response dives down into the notch, but comes up again to level out at a lower gain, and highpass notch filters where the gain levels out to be lower on the low-frequency side of the notch. There is more on this in Chapter 9 on bandpass and notch filters.

Here the two lowpass notch filters are of the Deliyannis-Friend [11], [12] type, which consists basically of a Multiple-FeedBack (MFB) lowpass filter with an extra signal path via the non-inverting input that generates the notch by cancellation. It is not exactly obvious, but these filter stages are non-inverting in the passband. The notch depth is critically dependent on the accuracy of the ratio set by R4, R5. The complete filter has an overall gain in the passband of +1.3 dB. The filter structure is based on an example given by Van Valkenburg [13].

The amplitude response in Figure 7.33 shows how the first notch in the stopband, at 1.22 kHz, is very narrow and makes the roll-off very steep indeed. However, the response naturally also starts to come back up rapidly, and this is suppressed by the second notch (at 1.88 kHz), which has a lower Q and is therefore broader, and buys time, so to speak, for the final first-order filter to start bringing in a useful amount of attenuation. Each time the response comes back up it reaches −17 dB; this is what is meant by an equi-ripple stop band. Variations on this type of filter with non-equal stop band ripples may not be officially Inverse Chebyshev filters, but they can be useful in specific cases.

To scale this circuit for different frequencies you can alter the capacitors, keeping C1 = C2 and C3 = C4 and the ratios of C1, C3, and C5 the same, or you can change resistors R1, R2, R3,

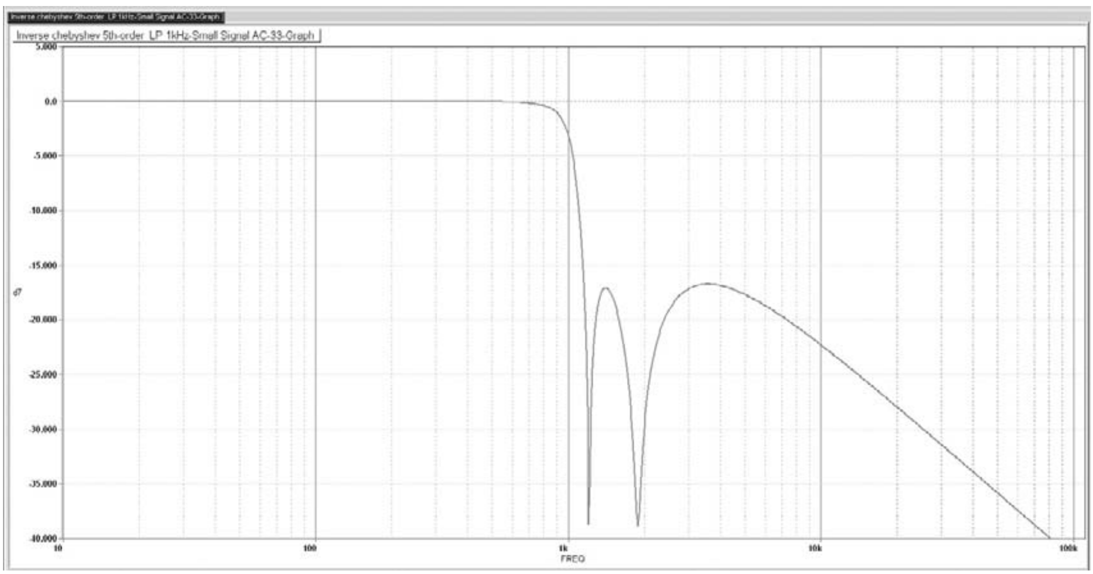


Figure 7.33: The amplitude response of a fifth-order Inverse Chebyshev lowpass filter. Note how the passband is maximally flat before the very steep roll-off. Cutoff frequency 1 kHz.

and R6, R7, R8, and R11 keeping all their ratios the same. You can, of course, change both capacitors and resistors to get appropriate circuit impedances. R4 and R5 can be altered but their ratio must remain the same, and likewise with R9 and R10; remember that the accuracy of this ratio controls notch depth. R11 and C5 can be altered as you wish, quite separately from any other alterations to the circuit, so long as their time-constant is unchanged.

More details on the design of Inverse Chebyshev filters can be found in Van Valkenburg [14].

7.13.2 Elliptical Filters (Cauer Filters)

Elliptical filters (also called Cauer filters) are basically a combination of Chebyshev and Inverse Chebyshev filters; amplitude ripples in the passband are accepted as the price of a faster roll-off, and there are also one or more notches (zeros) in the stopband to steepen the roll-off rate. They have a sharp cut off, high group delay, and the greatest stop band attenuation. They are sometimes called complete-Chebychev filters or Zolotarev filters.

As for the Chebyshev filter, the definition of an elliptic filter cutoff frequency depends on the passband ripple amplitude. In most filter design software any value of attenuation can be defined as the cutoff point.

The design of elliptical filters is not simple, and even the authors of filter textbooks that are a morass of foot-long complex equations are inclined to say things like "…it is rather involved…" and recommend you use published tables to derive component values. Regrettably, the use of these tables is in itself rather hard going, so here I am just going to give one example of how these filters are put together. This can be easily scaled for different frequencies. There is another example in Chapter 5 on notch crossovers.

The amount of passband ripple in an elliptic filter is sometimes quoted as a "reflection coefficient" percentage ρ, which as you might imagine is a hangover from transmission-line theory, and not in my opinion a very helpful way of putting it.

Wilhelm Cauer (June 24, 1900–April 22, 1945) was a German mathematician and scientist. He is most noted for his work on the analysis and synthesis of electrical filters [15] and his work marked the beginning of the field of network synthesis. He was shot dead in his garden in Berlin-Marienfelde in Berlin by Soviet soldiers during the capture of the city in 1945.

Elliptical filters are commonly implemented by combining a notch filter with an all-pole filter such as a Butterworth type. The notches used are not in general symmetrical notches that go up to 0 dB either side of the central crevasse—they are usually lowpass or highpass notch filters. A lowpass notch response starts out at 0 dB at low frequencies, plunges into the crevasse, and then comes up again to flatten out at a lower level, often −10 dB. When combined with an all-pole lowpass filter this gives much better high-frequency attenuation than a symmetrical notch. Conversely, a highpass notch has a response that is 0 dB at high frequencies, but comes back up to around −10 dB at low frequencies. There is more on lowpass and highpass notches in Chapter 9 on bandpass and notch filters.

Figure 7.34 shows a simple third-order elliptical filter with a cutoff frequency of 1 kHz and a reflection coefficient of 20%. The passband ripple is therefore very small at 0.2 dB, and there is only one notch in the stopband. The filter structure is based on an example given by Williams and Taylor [16]. The first stage around A1 is a lowpass notch filter, made up of a twin-T notch filter with its Q (notch sharpness) enhanced by positive feedback through C1, the amount being fixed by R5, R6 which set the closed-loop gain of A1. C4 is added to make a lowpass notch rather than a symmetrical notch. The output of this second-order stage is the upper trace in Figure 7.35, and you can see it has been arranged to peak gently just before the roll-off. When this is combined with the first-order lowpass filter R7, C5, the final response is the third-order lower trace in Figure 7.35. The 0.2 dB ripple in the passband is just visible. You will observe that as the frequency increases, once the drama of the notch is over the ultimate rollo-ff slope is only 6 dB/octave, because the lowpass notch response is now flat and so only the final first-order filter is contributing to the roll-off. This is the price you pay for implementing a fast roll-off with a filter that is only third-order. Fourth-order filters that have an ultimate roll-off slope of 12 dB/octave are described in Chapter 5 on notch crossovers.

To scale this circuit for different frequencies you can alter all the capacitors, keeping their ratios to each other the same, or the resistors R1–4 and R7. You can of course change both to get appropriate circuit impedances. R5 and R6 can be altered but their ratio must remain the same.

We have compared the roll-off of the previous filters by looking at the attenuation at 10 kHz, a decade above the 1 kHz cutoff frequency. That is less helpful here because of the way the

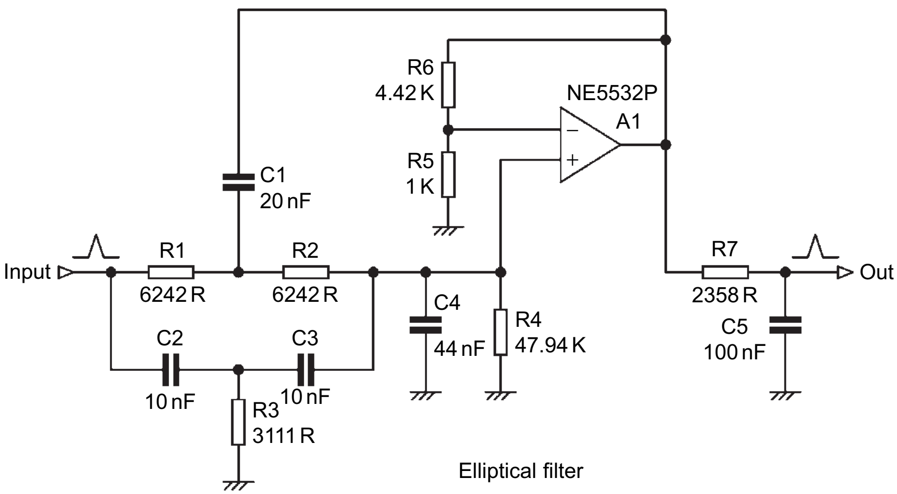


Figure 7.34: A third-order elliptical lowpass filter. Made up of a lowpass notch filter followed by a first-order lowpass. Cutoff frequency 1 kHz.

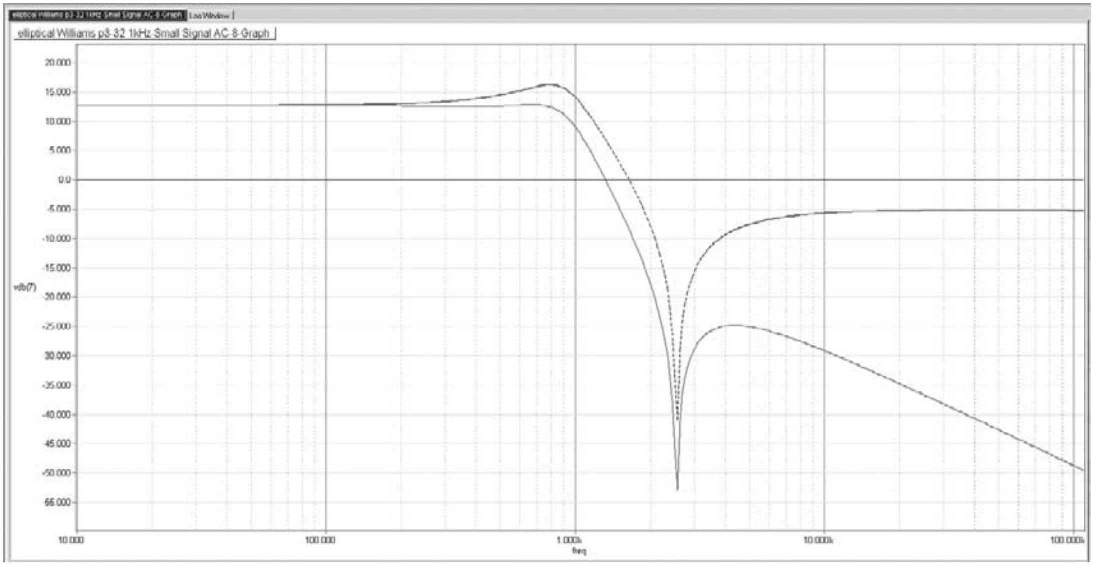


Figure 7.35: The amplitude response of the third-order elliptical lowpass filter. The lower trace is the final output, and the upper trace is the signal from the first stage. Cutoff frequency 1 kHz.

amplitude response comes back up at high frequencies; the response at 10 kHz is −42 dB, but in my simulation the attenuation at the bottom of the notch (at 2.6 kHz) was about −65 dB. The passband gain is +12.7 dB, because of the positive feedback applied to the notch network, and in many cases this will be less than convenient. Elliptical filters can sometimes be very useful for crossover use, for if we have an otherwise good drive unit with some nasty behaviour just outside its intended frequency range, the notch can be dropped right on top of it.

This elliptical filter and the Inverse Chebyshev filter are the only ones in this chapter that are not all-pole filters. More details on the design of elliptical filters can be found in Williams and Taylor [17], and Van Valkenburg [18].

7.14 Some Lesser-Known Filter Characteristics

The filter types described above are the classic types. Their characteristics depend on particular mathematical relationships such as Bessel functions or Chebyshev polynomials. There is however no rule that you must restrict yourself to using the "official" filters. There are in fact an infinite number of filter characteristics that can be used if necessary. As an example, look at a simple second-order filter. It only has two parameters that define its behaviour—the cutoff frequency and the Q. It is the Q that determines the filter type; we have just seen that a Q of 0.50 gives a Linkwitz–Riley filter, a Q of 0.578 gives a Bessel filter, a Q of 0.707 gives a Butterworth filter, a Q of 0.956 gives a 1 dB-Chebyshev filter, and so on; we have noted that the cutoff frequencies most be scaled appropriately for a

Table 7.10: Frequencies and Qs for Recognised Second-Order Filter Types

Filter Type	Freq	Q
Linkwitz-Riley	1.578	0.5000
Bessel	1.2736	0.5773
Butterworth	1.0000	0.7071
Chebyshev 0.5 dB	1.2313	0.8637
Chebyshev 1.0 dB	1.0500	0.9565
Chebyshev 2.0 dB	0.9072	1.1286
Chebyshev 3.0 dB	0.8414	1.3049

given filter type. There is however nothing whatever to stop you using a Q of 0.55, 0.66, or whatever you feel you need.

The recognised kinds of second-order filter are summarised in Table 7.10. Things become more complicated with filters of third-order and above, where more parameters are needed to define the filter.

7.14.1 Transitional Filters

When people talk about "transitional filters" they are most often referring to filters that are a compromise between Bessel and Butterworth characteristics [19]. Filtershop offers transitional filters, but it describes them as a Gaussian passband spliced together with a Chebyshev stop band. There are an infinite number of versions of such a filter, depending on the attenuation level at which the splicing occurs; Filtershop offers 3 dB, 6 dB, and 12 dB options. All the filters described below are all-pole filters.

7.14.2 Linear-Phase Filters

As we have seen, Bessel filters have a maximally flat group delay, avoiding the delay peak that you get with a Butterworth filter. Discussions on filters always remark that the Bessel alignment has a slower roll-off, but often fail to emphasise that it is a *much* slower roll-off. If the requirement for a maximally flat delay is relaxed to allow equi-ripple group delay of a specified amount, then the amplitude roll-off can be much faster. The equi-ripple group delay characteristic is more efficient in that the group delay remains flat (within the set limits) further into the stop band.

This is very much the same sort of compromise as in the Chebyshev filter, where the rate of amplitude roll-off is increased by tolerating a certain amount of ripple in the passband amplitude response. Linear-phase filters are all-pole filters.

Filters of this kind are frequently referred to as "Linear Phase" filters, but are sometimes (and more accurately) called Butterworth–Thomson filters. For some reason, they never seem to be

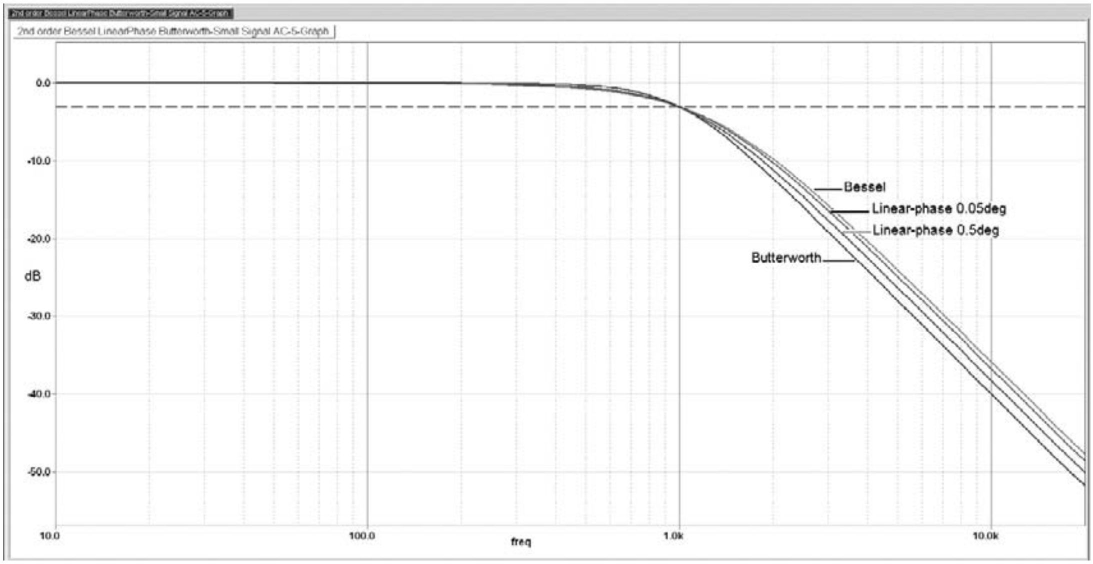


Figure 7.36: The amplitude response of Bessel, Linear-phase 0.05 degree, Linear-phase 0.5 degree, Butterworth lowpass filters. Cutoff frequency is 1 kHz in all cases.

called Butterworth–Bessel filters, though this means exactly the same thing. The parameter m describes the move from Thomson to Butterworth, with m = 0 meaning pure Butterworth, and m = 1 meaning pure Bessel. Any intermediate value of m yields a valid transitional filter. Linear-phase filters can alternatively be characterised by an angle parameter, so you can have a Linear-phase 0.05 degree filter, or a Linear-phase 0.5 degree filter. This refers to the amount of deviation in the filter phase characteristics caused by allowing ripples in the group delay curve.

Figure 7.36 shows the amplitude responses of Bessel, Linear-phase 0.05 degree, Linear-phase 0.5 degree, Butterworth lowpass filters. This makes it very clear how linear-phase filters provide intermediate solutions between the Bessel and Butterworth characteristics.

The frequencies and Q's for linear-phase filters up to the 8th order are shown in Table 7.11.

7.14.3 Gaussian Filters

A Gaussian filter is essentially is essentially a time-domain filter, optimised give no overshoot to a step function input while minimizing the rise and fall time; this response is closely connected to the fact that the Gaussian filter has the minimum possible group delay. Gaussian filters come in various kinds identified by a dB suffix, such as "Gaussian-6 dB" and "Gaussian-12 dB," which indicate the attenuation level at which the roll-off steepens, though the amplitude response differences between them are very small. The amplitude response is also very similar indeed to that of a Bessel filter, as demonstrated in Figure 7.37.

Table 7.11: Frequencies and Q's for Linear-Phase Filters Up to Eighth Order. Stages Are Arranged in Order of Increasing Q, Odd-Order Filters Have the First-Order Section at the End with No Q Shown

Order	Freq 1	Q 1	Freq 2	Q 2	Freq 3	Q 3	Freq 4	Q 4
2	1.0000	0.6304						
3	1.2622	0.9370	0.7923	n/a				
4	1.3340	1.3161	0.7496	0.6074				
5	1.6566	1.7545	1.0067	0.8679	0.5997	n/a		
6	1.6091	2.1870	1.0741	1.1804	0.5786	0.6077		
7	1.9162	2.6679	1.3704	1.5426	0.8066	0.8639	0.4721	n/a
8	1.7962	3.1146	1.3538	1.8914	0.8801	1.1660	0.4673	0.6088

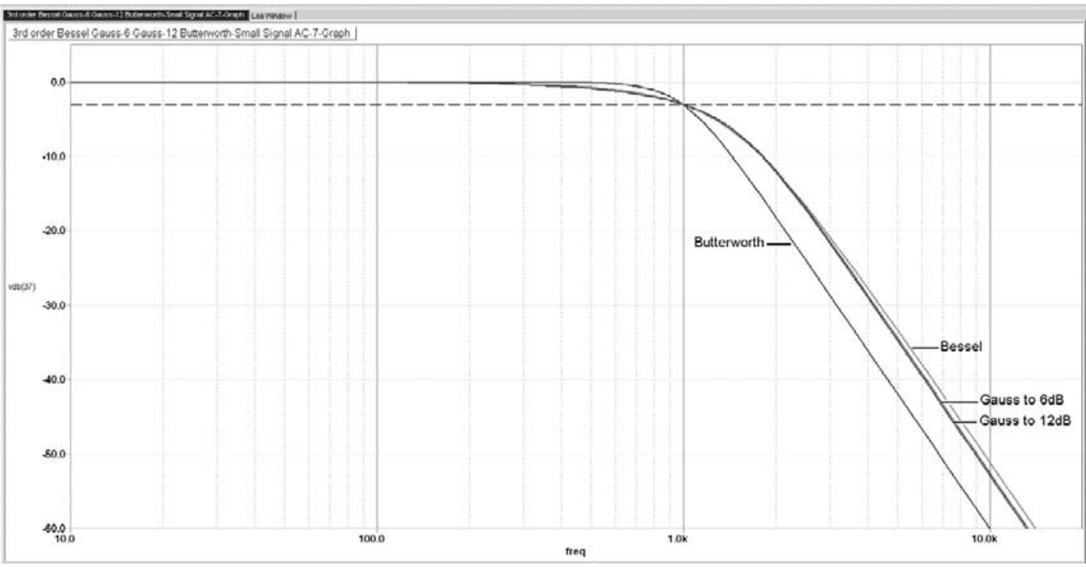


Figure 7.37: The amplitude response of third-order Bessel, Gaussian-6 dB, Gaussian-12 dB, and Butterworth lowpass filters. Cutoff frequency is 1 kHz in all cases.

Gaussian filters are named after Johann Carl Friedrich Gauss (1777–1855), a German mathematician and scientist, because their mathematical derivation stems from the same basic equations used to derive the Gaussian distribution in statistics.

Figure 7.38 compares the step response of third-order Gaussian-12 dB and Butterworth lowpass filters. The Gaussian output rises faster than the Butterworth but has no overshoot at all. The Bessel step response is similar to that of the Gaussian. Gaussian filters are all-pole filters.

The frequencies and Q's for Gaussian filters up to the 8th order are shown in Table 7.12.

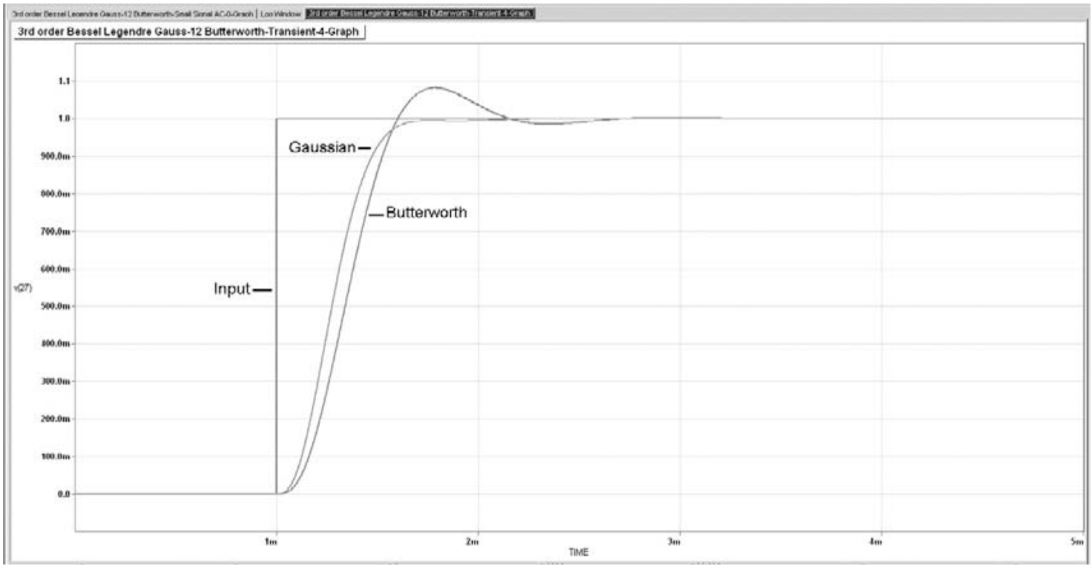


Figure 7.38: The step response of third-order Gaussian-12 dB, and Butterworth lowpass filters. Cutoff frequency is 1 kHz.

Table 7.12: Frequencies and Q's for Gaussian Filters Up to Eighth Order. Stages Are Arranged in Order of Increasing Q, Odd-Order Filters Have the First-Order Section at the End with No Q Shown

Order	Freq 1	Q 1	Freq 2	Q 2	Freq 3	Q 3	Freq 4	Q 4
2	0.9170	0.6013						
3	0.9923	0.5653	0.9452	n/a				
4	0.9930	0.6362	1.0594	0.5475				
5	1.0427	0.6000	1.1192	0.5370	1.0218	n/a		
6	1.0580	0.6538	1.0906	0.5783	1.1728	0.5302		
7	1.0958	0.6212	1.1358	0.5639	1.2215	0.5254	1.0838	n/a
8	1.1134	0.6644	1.1333	0.5994	1.1782	0.5537	1.2662	0.5219

7.14.4 Legendre–Papoulis Filters

The Legendre–Papoulis filter is a monotonic all-pole filter; in other words the response is always downwards. It is optimised for the greatest slope at the passband edge, given this condition. It gives faster attenuation than the Butterworth characteristic, but the snag is that the passband is not maximally flat as is the Butterworth; instead it slopes gently until the rapid roll-off begins. This is the crucial difference. Legendre–Papoulis filters can be useful in applications that need a steep cutoff at the passband edge but cannot tolerate passband ripples, or in cases where a Chebyshev I filter produces too much group delay at the passband edge.

It is essentially a compromise between the Butterworth and the Chebyshev filters; it has the maximum possible roll-off rate for a given filter order while maintaining a monotonic amplitude response. Legendre–Papoulis filters are often simply called Legendre filters.

A second-order Legendre–Papoulis response is exactly the same as the Butterworth response. With a second-order filter the only parameter you have to play with is Q, and the highest Q that can be used without causing peaking (in which case the response would no longer be monotonic) is the familiar $1/\sqrt{2}$ which gives the Butterworth response.

Legendre filters are only different from Butterworth filters for third order and above. The frequencies and Q's for Legendre-Papoulis filters up to the 8th order are shown in Table 7.13.

The amplitude responses of third-order Legendre–Papoulis and Butterworth filters are compared in Figure 7.39. You will note that the differences are not very great. Compared with the Butterworth, the Legendre–Papoulis sags by 0.5 dB around 500 Hz, which may or may not be negligible depending on your application. Once the roll-off has begun, the Legendre–Papoulis is usefully if not dramatically superior; at 2 kHz it gives 3.7 dB more attenuation, while at 3 kHz it gives 4.3 dB more. After that the difference is effectively constant at 4.4 dB as the two curves must ultimately run parallel at −18 dB/octave, both filters being third order. Legendre–Papoulis filters are all-pole filters.

The Legendre–Papoulis filter was proposed by Athanasios Papoulis in 1958 [20]. It is also sometimes known as an "Optimum L" or just "Optimum" filter. The filter design is based on Legendre polynomials; their French inventor, Adrien-Marie Legendre (1752–1833) was

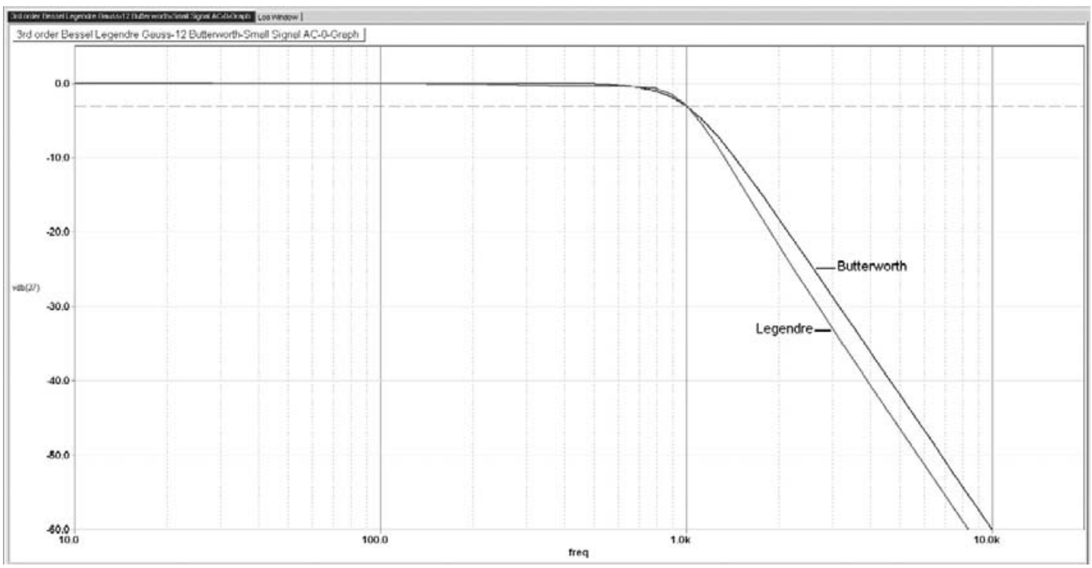


Figure 7.39: Amplitude response of third-order Legendre and Butterworth filters. Cutoff frequencies are 1 kHz.

yet another mathematician who did not live to see his mathematics applied to electrical filters.

The fact that the response is monotonic, but not maximally flat, means that Legendre–Papoulis filters are of little use in many hi-fi applications. For example, a subsonic filter for a phono input needs to be maximally flat, as a slow early roll-off will cause increased RIAA errors in the low-frequency part of the audio range. However, it is possible that Legendre–Papoulis filters may be of use in crossover design.

The step response of the Legendre–Papoulis filter is similar to that of the Butterworth. Figure 7.40 shows that the response is slightly slower, with about the same amount

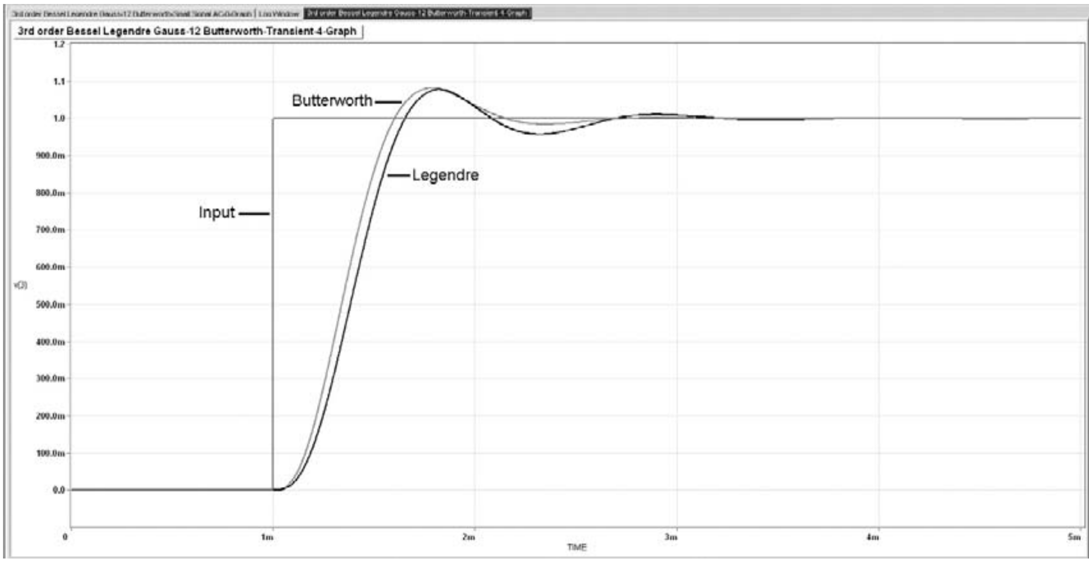


Figure 7.40: Step response of third-order Legendre and Butterworth filters. Cutoff frequencies are 1 kHz.

Table 7.13: Frequencies and Q's for Legendre–Papoulis Filters Up to Eighth Order. Stages Are Arranged in Order of Increasing Q, Odd-Order Filters Have the First-Order Section at the End with No Q Shown

Order	Freq 1	Q 1	Freq 2	Q 2	Freq 3	Q 3	Freq 4	Q 4
2	1.000	0.707						
3	0.9647	1.3974	0.6200	n/a				
4	0.9734	2.1008	0.6563	0.5969				
5	0.9802	3.1912	0.7050	0.9082	0.4680	n/a		
6	0.9846	4.2740	0.7634	1.2355	0.5002	0.570		
7	0.9881	5.7310	0.8137	1.7135	0.5531	0.7919	0.3821	n/a
8	0.9903	7.1826	0.8473	2.1807	0.6187	1.0303	0.4093	0.5573

of overshoot, but after that the undershoot is significantly greater. The Legendre–Papoulis is essentially a frequency-domain filter.

7.14.5 Synchronous Filters

Synchronous filters consist of a number of identical first-order stages in cascade. Since there are no second-order stages the response is inevitably monotonic. The term "synchronous" refers to the fact that all the stages have an identical cutoff frequencies, and therefore identical time-responses; it has nothing to do with system clocks or digital data transmission. You cannot produce a synchronous filter simply stringing together a chain of R's and C's; each first-order RC stage must be isolated from the next by a buffer stage to prevent loading and interaction effects. Figure 7.41 shows second-, third-, and fourth-order synchronous filters made in this way. All are designed to be −3 dB at 1 kHz overall, so the stage frequencies for the higher-order filters are increased so that the overall response passes through this point. For example, the fourth-order filter has four stages with cutoff (–3 dB) frequencies of 2.303 kHz, each of which give −0.75 dB at 1 kHz; when the four stages are used together the attenuation is therefore −3 dB at 1 kHz. The cutoff frequencies required for second-, third-, and fourth-order synchronous filters are given in Table 7.14.

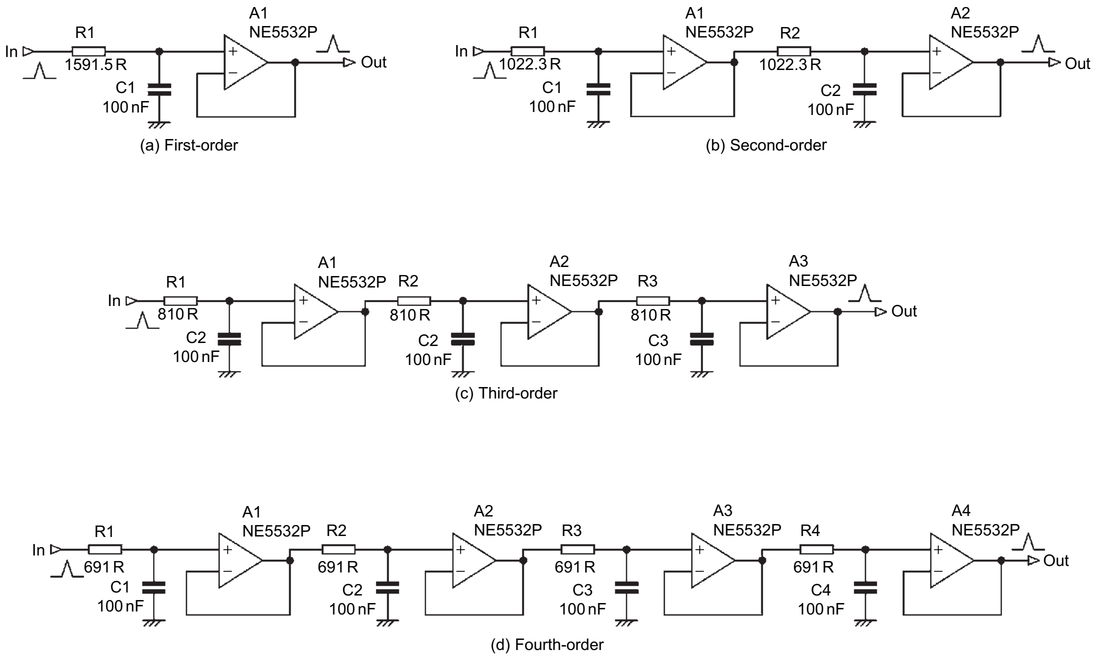


Figure 7.41: First-, second-, third-, and fourth-order synchronous lowpass filters constructed from repeated first-order filter sections, all designed to be −3 dB at 1 kHz.

Table 7.14: Cutoff Frequencies for Synchronous Filters That Give Cutoff at 1.000 When Stages Cascaded

Order	Cutoff Frequency
1	1.0000
2	1.5568
3	1.9649
4	2.3033

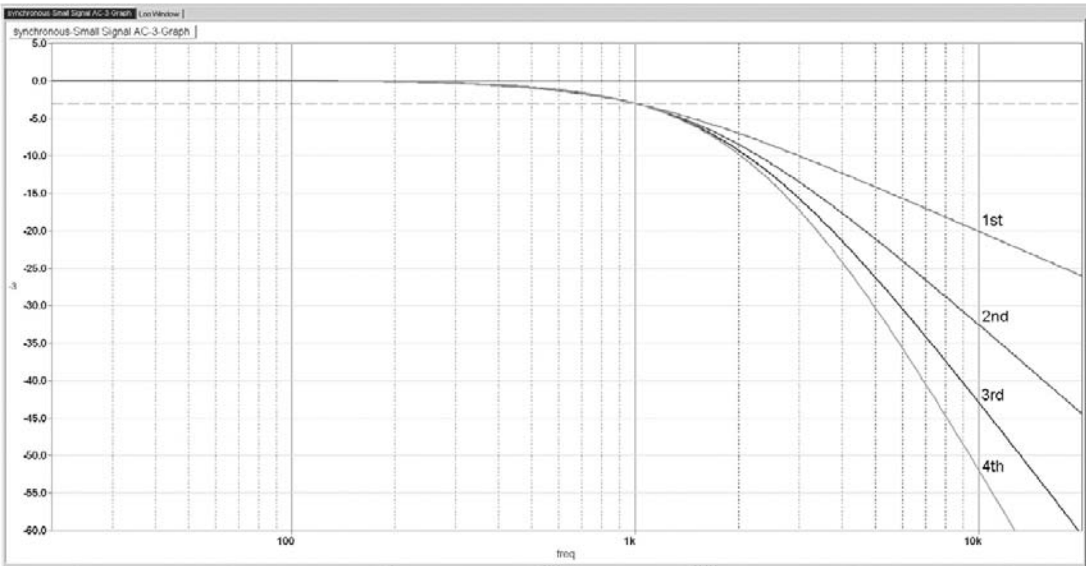


Figure 7.42: Amplitude response of second-, third-, and fourth-order synchronous lowpass filters, all designed to be −3 dB at 1 kHz.

Figure 7.42 shows the amplitude response, which has an even slower roll-off than the Bessel filter, because of a complete lack of internal peaking. The response differences at frequencies below the −3 dB point are very small. It does not at present appear very likely that synchronous filters will be useful in crossover filters as such of their very slow rate of roll-off, but is a branch of filter technology to be aware of because they may be appropriate for specific equalisation of drive unit roll-offs.

Looking at Figure 7.41, you can see that making high-order synchronous filters by stringing together first-order stages uses a lot of opamps. However, a second-order stage with a Q of 0.5 is equivalent to two cascaded first-order stages, and the circuits shown in Figure 7.43 save one opamp section for a second-order filter, and two opamp sections for both third-order filter and fourth-order synchronous filters. A second-order stage with a Q of 0.5 is

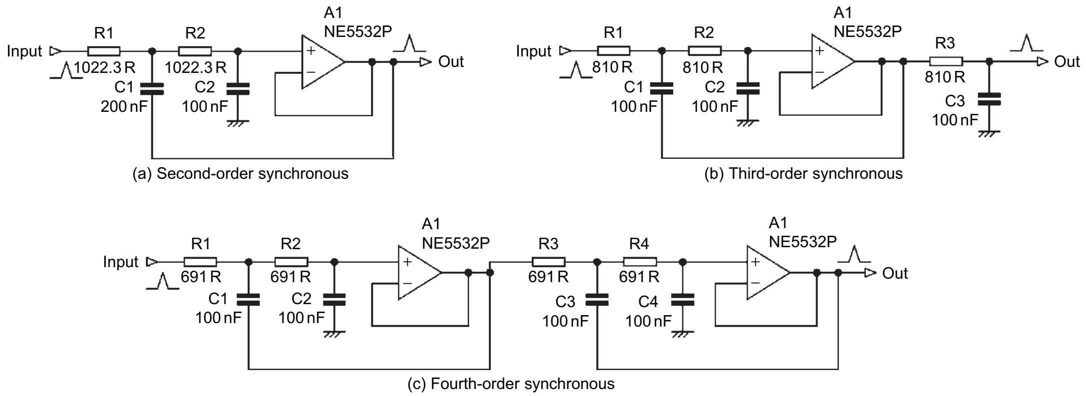


Figure 7.43: A more efficient way to make higher-order synchronous lowpass filters using second-order filters with Q = 0.5. All are −3 dB at 1 kHz.

also equivalent to a second-order Linkwitz–Riley filter, but this does not hold for third-order and higher filters.

It is not possible to replace three cascaded first-order stages with a second-order stage of any Q because the ultimate slope is 12 dB/octave instead of 18 dB/octave. It might be possible to replace them with a third-order filter of suitably low Q but the parts saving would be less.

7.15 Other Filter Characteristics

We have just looked at a number of obscure filter types; this however by no means exhausts the recognised filter characteristics that exist.

Ultraspherical filters [21] are based on ultraspherical polynomials. How, you may inquire, can something be more spherical than spherical? To answer that in a suitably short space is quite beyond my powers, but rest assured that the whole business is solidly rooted in higher mathematics. Ultraspherical polynomials are sometimes called Gegenbauer polynomials, after their inventor, and inhabit a fairly elevated area of mathematics. They are generalisations of Legendre polynomials and Chebyshev polynomials, and are special cases of Jacobi polynomials. The response of an ultraspherical filter is set by four parameters; the filter order and the cutoff frequency as usual, plus ε which controls the attenuation at the cutoff frequency, and α which controls the shape of the amplitude response. Ultraspherical filters are a class of filters rather than a distinct type, and they include Butterworth, Legendre, and Chebyshev filters as special cases; for example $\alpha = \infty$ gives a Butterworth filter, while $\alpha = -0.5$ gives a Chebyshev.

There are also inverse ultraspherical filters, in the same way that there are inverse Chebyshev filters [22].

Another rare bird is the Halpern filter [23]. This is related to the Legendre–Papoulis filter in that it rolls off faster than a Butterworth but slower than a Chebyshev. Likewise, the Halpern filter has a monotonic passband without being maximally flat. The peak group delay near the cutoff is significantly lower than that of a Chebyshev filter but not as good as Butterworth or Legendre–Papoulis.

I don't pretend that the information in this final section is going to be digestible to anyone without a quite advanced knowledge of mathematics, but do not fret. There is at present nothing to suggest that these filter types have anything to offer crossover design that is not already covered by the better-known filter types dealt with in detail above, so there seems no point in going further into the rather heavy mathematics required.

References

[1] Williams, Taylor, Electronic Filter Design Handbook, fourth ed., McGraw-Hill, USA, 2006, ISBN 0-07-147171-5.
[2] van Valkenburg, Analog Filter Design, Holt-Saunders International Editions, 1982, ISBN 4-8338-0091-3.
[3] H.M. Berlin, Design of Active Filters with Experiments, Blacksburg, VA, 1978, p. 85.
[4] B. Hofer, Switch-mode amplifier performance measurements, Electronics World, USA, 2005, p. 30.
[5] H. Bill, Precise active crossover, Electronics World, USA, 2009, p. 652.
[6] R. Ramaswamy, A universal continuous-time active filter, Linear Audio, USA, vol. 1, Publ. Jan Didden, 2011.
[7] S. Butterworth, On the theory of filter amplifiers, Exp. Wireless Wireless Engineer 7 (1930) 536–541.
[8] W.E. Thomson, Delay networks having maximally flat frequency characteristics, Proc. IEEE (Part 3) 96 (1949) 487–490.
[9] R. Miller, A Bessel filter crossover, and its relation to others, Rane Note 147, Rane Corporation, USA, 2006.
[10] Z. Kiyasu, On a design method of delay networks, J. Inst. Electron. Commun. Eng. 26 (1943) 598–610.
[11] T. Delyannis, High-Q factor circuit with reduced sensitivity, Electron. Lett. 4 (26) (1968) 577–579.
[12] J. Friend, A single operational-amplifier biquadratic filter section, IEEE ISCT Digest Technical Papers, 1970, p. 189.
[13] van Valkenburg, Analog Filter Design, Holt-Saunders International Editions, USA, 1982, p. 375, ISBN 4-8338-0091-3.
[14] van Valkenburg, Analog Filter Design, As [12], p. 365.
[15] W. Cauer, Theorie der linearen Wechselstromschaltungen, vol. I, Akad. Verlags-Gesellschaft Becker und Erler, Leipzig, 1941, (in German) (Theory of Linear AC Circuits, Vol 1).
[16] Williams, Taylor, Electronic Filter Design Handbook, fourth ed., McGraw-Hill, USA, 2006, p. 2.76, ISBN 0-07-147171-5.
[17] Williams, Taylor, Electronic Filter Design Handbook, As [15].
[18] van Valkenburg, Analog Filter Design, As [12], p. 385.
[19] L. Paarmann, Design & Analysis of Analog Filters—A Signal Processing Perspective, Kluwer Academic Publishers, 2001, ISBN: 0792373731, p. 234, 235 (Transitional filters).
[20] A. Papoulis, Optimum filters with monotonic response, Proc. IRE 46, (1958) 606–609.
[21] L. Paarmann, Design & Analysis of Analog Filters—A Signal Processing Perspective, As [19], p. 245.
[22] C.A. Corral, Inverse ultraspherical filters, J. Analog Integr. Circuits Signal Process. 42 (2) (2005) 147–159. Springer, Netherlands, ISSN 0925-1030.
[23] L. Paarmann, Design & Analysis of Analog Filters—A Signal Processing Perspective, As [19] p. 255.

This page intentionally left blank

CHAPTER 8

Designing Lowpass and Highpass Filters

8.1 Designing Real Filters

In Chapter 7 we saw how just about any filter you can dream up can be implemented by cascading first- and second-order filter blocks. In this chapter we look at the best ways to turn those theoretical blocks into reality. There are many ways to do this, but for the moderate Q values normally required, the two simplest and the best methods are the well-known Sallen and Key configuration and the multiple-feedback (MFB) configuration.

All the filters in this chapter have been designed with a cutoff frequency of 1 kHz, which allows quick comparisons and quick scaling of the examples to get the filter frequency you want. The use of the term "cutoff" might be taken to imply that the filter has a response that drops steeply; this is in general not the case, and even the most gently sloping filter has a "cutoff" frequency. For the Butterworth filter the cutoff frequency is the point at which the response has fallen by 3 dB, that is, to $1/\sqrt{2}$ of its passband value. Other filters such as the Chebyshev have different definitions of cutoff. The term "transition frequency" means the same thing.

In the examples shown here, the capacitor values have been chosen to put the resistor values mostly in the range 700 Ω – 1 kΩ, in order to minimise Johnson noise without putting excessive loading on opamps that will increase distortion. Most examples are of the lowpass version.

8.2 Component Sensitivity

The accuracy of the response of a filter obviously depends on the precision of the components with which it is made. What is less obvious is that in some cases the response is affected very greatly by small changes in certain components. This sort of "sensitivity" has nothing whatever to do with the sensitivity of an equipment input, as in "Sensitivity: 500 mV rms for full output." It is unfortunate that two different concepts have ended up with the same name, but no one is going to change it now.

As described in Chapter 7, high-order filters are usually made up from cascaded second-order sections, plus a first-order section for the odd-order versions. These sections usually have different Qs, and the higher the Q, the greater the component sensitivity, and the more the stage amplifies component errors. Precision components are expensive, and so there is a strong incentive to use the lowest Qs possible for a filter design.

While cascaded second-order sections are the norm, identical filters can be made up from higher-order sections, so that a sixth-order filter could be redesigned as two cascaded third-order sections instead of three second-order sections. This is rarely if ever done, and as you might imagine there are excellent reasons for this. A third-order section has greater component sensitivity, and any saving made in the number of opamps is likely to be less than the extra cost of more precise components to get a filter of the same accuracy as the original. It is also possible to make a fourth-order filter in a single stage, but the component sensitivity is predictably even worse.

Component sensitivity may appear to have a somewhat academic flavour about it, but in fact, keeping it as low as possible is crucial to making an economic design, as the cost of components, especially capacitors, increases very steeply with increased accuracy. The alternative is some sort of cut-and-try in production, which is also going to be costly and time-consuming, and will require accurate measuring equipment.

Component sensitivity is expressed as a ratio. There is a different figure for every component and how it affects every parameter of the stage. Thus, every component in a second-order filter has two sensitivities; one for its effect on cutoff frequency and one for its effect on Q. Thus if a certain resistor R in a filter has a sensitivity with respect to cutoff frequency of −0.5, that means that a 1% increase in the component will cause a −0.5% reduction in the cutoff frequency. This is the usual frequency sensitivity in second-order Sallen and Key filters.

Component sensitivities are dealt with in the relevant section for each type of filter described.

8.3 First-Order Lowpass and Highpass Filters

First-order filters are the simplest possible; they are completely defined by one parameter; the cutoff frequency.

The usual first-order filter is just a resistor and capacitor. Figure 8.1a shows the normal (non-inverting) lowpass version in all its beautiful simplicity. To give the calculated first-order response it must be driven from a very low impedance and see a very high impedance looking into the next stage. It is frequently driven from an opamp output, which provides the low driving impedance, but often requires a unity-gain stage to buffer its output, as shown.

Sometimes it is necessary to invert the phase of the audio to correct an inversion in a previous stage. This can be handily combined with a shunt-feedback first-order lowpass filter as shown in Figure 8.1b. This stage inherently provides a low output impedance so there are no loading effects from circuitry downstream. The signal level can be adjusted either up or down by giving R1 and R2 the required ratio.

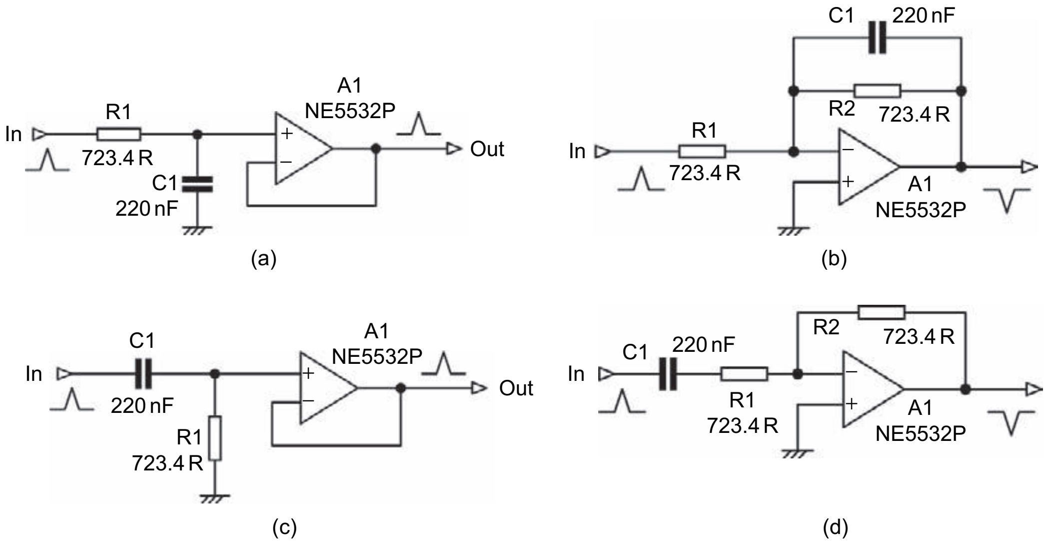


Figure 8.1: First-order filters: (a) non-inverting lowpass; (b) inverting lowpass; (c) non-inverting highpass; (d) inverting highpass. Cutoff frequency (–3 dB) is 1 kHz in all cases.

Here are the design and analysis equations for the first-order filters. The design equations give the component values required for given cutoff frequency; the analysis equations give the cutoff frequency when the existing component values are plugged in. Analysis equations are useful for diagnosing why a filter is not doing what you planned.

For both versions the design equation is:

$$R = \frac{1}{2\pi f_0 C} \text{ (choose C)}, \tag{8.1}$$

and the analysis equation is:

$$f_0 = \frac{1}{2\pi RC} \tag{8.2}$$

For the inverting lowpass version you must use R2 and C1 in the equations. The passband gain A = R2/R1.

Highpass versions of these filters are shown in Figure 8.1c and 5.1d. The design and analysis equations are the same, but for the inverting highpass you must use R1 and C1 in the equations. In the case of Figure 8.1d, it may be advisable to put a small (say, 100 pF) capacitor across R2 to ensure HF stability.

There are no issues with component sensitivities in a first-order filter; the cutoff frequency is inversely proportional to the product of R and C so the sensitivity for either component is always 1.0. In other words, a change of 1% in either component will give a 1% change in the cutoff frequency.

8.4 Second-Order Filters

Second-order filters are much more versatile. They are defined by two parameters; the cutoff frequency and the Q, and as described in Chapter 7, higher-order filters are usually made by cascading second-order stages with carefully-chosen cutoff frequencies and the Qs. Odd-order filters require a first-order stage as well.

There are many ways to make a second-order filter. The simplest and most popular are the well-known Sallen and Key configuration, and the Multiple FeedBack (MFB) filters. The latter are better-known in their bandpass form, but can be configured for either lowpass or highpass operation. MFB filters give a phase-inversion so can only be used in pairs (e.g., in a Linkwitz-Riley 4th-order configuration) or in conjunction with another stage that re-inverts the signal to get it back in phase.

These two filter types are not suitable for producing high Q characteristics, as for Q's above about 3, such as might be needed in a high-order Chebyshev filter, they begin to show excessive component sensitivities, and also may start to be affected by the finite gain and bandwidth of the opamps involved. For higher Qs more complex multi-opamp circuit configurations are used that do not suffer from these problems. Having said that, high Q filters are not normally required in crossover design.

8.5 Sallen & Key Second-Order Filters

The Sallen and Key filter configuration was introduced by R.P. Sallen and E. L. Key of the MIT Lincoln Laboratory in 1955 [1]. It became popular as it is relatively easy to design and the only active element required is a unity-gain buffer, so in pre-opamp days it could be effectively implemented with a simple emitter-follower, and before that with a cathode-follower.

An example of a second-order low-pass Sallen & Key filter is shown in Figure 8.2. Its operation is very simple; at low frequencies the capacitors are effectively open-circuit so it acts as a simple voltage-follower. At higher frequencies C2 begins to shunt the signal to ground, and this reduces the output of the follower, causing the "bootstrapping" of C1 to be less effective, and so causing it to also shunt signal away. Hence there are two roll-off mechanisms working at once and we get the familiar 12 dB/octave filter slope.

There are two basic ways to control the Q of a second-order lowpass Sallen and Key stage. In the first, the two capacitors are made unequal, and the greater the ratio between them the higher the Q, and a unity-gain buffer is used. In the second method, the two capacitors are made equal (which will usually be cheaper as well as more convenient) and Q is fixed by setting the gain of the amplifier to a value greater than one. The first method usually gives superior performance, as filter gain is often not wanted in a crossover, but may be more costly. The Q of a Sallen & Key filter can also be controlled by using non-equal resistors in the lowpass case, or non-equal capacitors in the highpass case, but this is unusual.

There are likewise two methods of Q-control for a second-order highpass Sallen and Key stage, but in this case it is the two resistors that are being set in a ratio, or held equal while the amplifier gain is altered.

While the basic Sallen & Key configuration is a simple and easy to design, it is actually extremely versatile, as you will appreciate from the many different versions in this chapter. It is even possible to design bandpass Sallen & Key filters, though this chapter is confined to lowpass and highpass versions.

The lowpass version of the Sallen & Key configuration in its many variants is examined here first, followed by the highpass version.

8.6 Sallen & Key Lowpass Filter Components

A fundamental difficulty with lowpass Sallen & Key filters is that the resistor values are usually all the same, but in general the capacitors are awkward values. This is the wrong way round from the designer's point of view; using two resistors in parallel or series to get the exact required value— or at any rate close enough to it— is cheap. Capacitors however are relatively expensive, and paralleling them to get a given value is therefore a relatively costly process. Putting capacitors in series to get the right value is not sensible because you will have to use bigger and more expensive capacitors. A 110 nF capacitance could be obtained by putting two 220 nF parts in series, but the alternative of 100 nF plus 10 nF in parallel is going to be half the cost or less, and also occupy less PCB area. The relatively precise capacitors required for accurate crossover filters do not come cheap, especially if the polypropylene type is chosen to prevent capacitor distortion; they are almost certainly the most expensive components on the PCB, and in a sophisticated crossover there may be a lot of them. For this reason I have gone to some trouble to present lowpass filter configurations that keep as many capacitors as possible the same value; the more of one value you buy, the cheaper they are.

Highpass filters do not have this problem. The capacitors are usually all the same value, and it is the resistors that come in awkward values, and paralleling them is cheap and takes up little PCB space.

8.7 Sallen & Key Second-Order Lowpass: Unity Gain

Figure 8.2 shows a very familiar circuit—a second-order lowpass Sallen & Key filter with a cutoff (−3 dB) frequency of 1 kHz and a Q of 0.707, giving a maximally flat Butterworth response. The pleasingly simple design equations for cutoff (−3 dB) frequency f_0 and Q are given below. Other recognised second-order responses can be designed by taking the Q values from Table 8.1.

Table 8.1: Second-Order Sallen and Key Unity-Gain Lowpass Qs and Capacitor Ratios for Various Filter Types

Type	FSF	Q	C1/C2 Ratio
Linkwitz–Riley	1.578	0.500	1.000
Bessel	1.274	0.578	1.336
Linear-Phase 0.05deg	1.210	0.600	1.440
Linear-Phase 0.5deg	1.107	0.640	1.638
Butterworth	1.000	0.707	2.000
0.5 dB-Chebyshev	1.231	0.864	2.986
1.0 dB-Chebyshev	1.050	0.956	3.663
2.0 dB-Chebyshev	0.907	1.129	5.098
3.0 dB-Chebyshev	0.841	1.305	6.812

In this chapter the component values are exact, just as they came out of the calculations, with no consideration given to preferred values or other component availability factors; these are dealt with later in Chapter 12. However it is worth pointing out now that in Figure 8.2, C1 would have to be made up of two 100 nF capacitors in parallel.

The main difference in the circuit that you will notice from textbook filters is that the resistor values are rather low and the capacitor values correspondingly high. This is an example of low-impedance design, where low resistor values minimise Johnson noise and reduce the effect of the opamp current noise and common-mode distortion. The measured noise output is −117.4 dBu. This is after correction by subtracting the test gear noise floor.

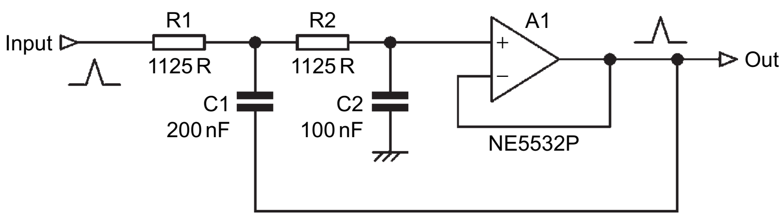


Figure 8.2: The classic second-order low-pass unity-gain Sallen & Key filter. Cutoff frequency is 1 kHz and Q = 0.7071 for a Butterworth response.

Here are the design and analysis equations for the general unity-gain Sallen & Key lowpass filter; in other words they cover all characteristics, Butterworth, Bessel, etc. The design equations give the component values required for given cutoff frequency and Q; the analysis equations give the cutoff frequency and Q when the existing component values are fed in.

Design equations: Begin by choosing a value for C2.

$$R = \frac{1}{2Q(2\pi f_0)C_2}, \tag{8.3}$$

$$C_1 = 4Q^2C_2 \tag{8.4}$$

Analysis equations:

$$f_0 = \frac{1}{2\pi R\sqrt{C_1C_2}}, \tag{8.5}$$

$$Q = \frac{1}{2}\sqrt{\frac{C_1}{C_2}} \tag{8.6}$$

Table 8.1 gives the Qs and capacitor ratios required for the well-known second-order filter characteristics.

It is important to remember that a Q of 1 does *not* give the maximally flat Butterworth response; the value required is $1/\sqrt{2}$, that is, 0.7071. The lowest Q you are likely to encounter are Sallen & Key filters with a Q of 0.5 sometimes used in second-order Linkwitz–Riley crossovers, but these are not favoured because the 12 dB/octave roll-off of the highpass filter is not steep enough to reduce the excursion of a driver when a flat frequency response is obtained, nor to attenuate drive-unit response irregularities [2].

The input impedance of this circuit is of importance because if it loads the previous stage excessively then the distortion of that stage will suffer. The input impedance is high at low frequencies where the series impedance of the shunt capacitors is high. It then falls with increasing frequency, reaching 2.26 kΩ at the 1 kHz cutoff frequency if we use the component values in Figure 8.2. Above this frequency it falls further, finally levelling out around 4 kHz at 1.125 kΩ, the value of R1. This is because at high frequencies there is no significant signal at the opamp non-inverting input, so C1 is effectively connected to ground at one end; its impedance is now very low so the input impedance of the filter looks as if R1 was connected directly to ground at its inner terminal.

Another consideration when contemplating opamp loading and distortion is the amount of current that the filter opamp must deliver into the capacitor C1. This is significant. If the circuit of Figure 8.2 is driven with 10 Vrms, the current through C1 is very low at LF, and then peaks gently at 1.9 mA rms at 650 Hz. The C1 current then reverses direction as frequency increases and C1 is required to absorb most of the input current, reaching −8.9 mA at 20 kHz. This sounds alarming, but at these frequencies the filter output is very low, (about 32 dB down) so the effect on the opamp is not as bad as it sounds.

Sallen & Key lowpass filters have a lurking problem. When they are implemented with opamps, the response does not carry on falling forever at the filter slope—instead it reaches a minimum and starts to come back up at 6 dB/octave. This is because the filter action relies on C1 seeing a low impedance to ground, and the impedance of the opamp output rises with frequency due to falling open-loop gain and hence falling negative feedback. When the circuit of Figure 8.2 is built using a TL072, the maximum attenuation is −57 dB at 21 kHz, rising again and flattening out at −15 dB at 5 MHz. More capable opamps such as the 5532 give much better results, though the effect is still present. While the rising response can be countered by adding a suitable first-order RC filter, (preferably in front of the filter to reduce possible intermodulation) Sallen & Key filters should not be used to reject frequencies well above the audio band; a lowpass version of the multiple-feedback filter is preferred.

8.8 Sallen & Key Second-Order Lowpass Unity-Gain: Component Sensitivity

The unity-gain Sallen and Key filter is noted for having low component sensitivities, as shown in Table 8.2. The sensitivities with respect to cutoff frequency are all −0.5, which makes excellent sense as if we wanted to deliberately reduce the cutoff frequency by 1%, we would, from the design equation given above, increase either both resistors or both capacitors by 1%.

Note that the resistor values have sensitivities of zero with respect to Q. As the design equations show, Q depends only on the capacitor ratio.

Table 8.2: Second-Order Sallen & Key Unity-Gain Lowpass Component Sensitivities

Component	Cutoff Frequency	Q
R1	−0.5	0
R2	−0.5	0
C1	−0.5	0.5
C2	−0.5	−0.5

8.9 Sallen & Key Second-Order Lowpass: Equal-Capacitor

In this version of the Sallen & Key configuration, the two capacitors are made equal and the Q is controlled by increasing the gain of the amplifier above unity; this avoids difficulties with awkward capacitor ratios. The gain must not exceed three or the filter becomes unstable

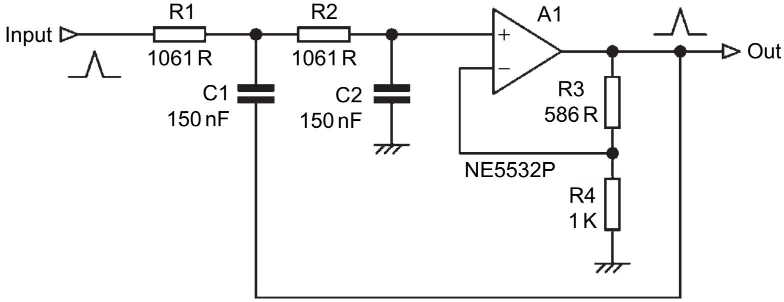


Figure 8.3: Equal-C second-order low-pass Butterworth filter with a cutoff frequency of 1 KHz. Gain must be 1.586 times for Butterworth maximally flat response. (Q = 0.7071).

and oscillates; this is shown in the analysis equation for Q, (Equation 8.11) where the bottom of the fraction becomes zero when A = 3, implying infinite Q.

The circuit is shown in Figure 8.3, with Q set to 0.7071 to give a second-order Butterworth response. The unity-gain buffer is now replaced with a voltage gain stage, and if the gain set by R3 and R4 is 1.586 times (+4.00 dB) for a Q of 0.707, then this allows C1 and C2 to be the same value. The gain introduced here is often an embarrassment in crossover designs, but there is another and more subtle snag to this neat-looking circuit; the component sensitivity is worse; this issue is dealt with in the next section. An equal-resistor-value highpass filter can be made in exactly the same way; this is also described later in this chapter.

An advantage of this circuit is that its filter characteristic can be smoothly altered from Linkwitz–Riley through to 3 dB-Chebyshev, passing through Bessel, Linear-phase, Butterworth, and all the lower-ripple Chebyshevs on the way, by making R3, R4 a variable potentiometer. Unfortunately the passband gain varies at the same time, though this could be solved by taking the output from the junction of R3 and R4; there will however still be a headroom issue due to the higher signal level at the opamp output.

Here are the design and analysis equations for the equal-C Sallen & Key filter. The design equations give the component values required for given cutoff frequency and Q; the analysis equations give the cutoff frequency and Q when the existing component values entered. The Qs and gains for various filter types are summarised in Table 8.3.

Design equations: Choose a value for C (= C1 = C2)

$$R_1 = R_2 = \frac{1}{(2\pi f_0)C}, \tag{8.7}$$

$$\text{Passband gain } A = 3 - \frac{1}{Q} \tag{8.8}$$

Then choose R3, R4 for required passband gain for the chosen Q

Table 8.3: Second-Order Sallen and Key Equal-C Lowpass: Qs and Gains for Various Filter Types

Type	FSF	Q	Gain A
Linkwitz–Riley	1.578	0.500	1.000
Bessel	1.274	0.578	1.270
Linear-Phase 0.05deg	1.210	0.600	1.333
Linear-Phase 0.5deg	1.107	0.640	1.437
Butterworth	1.000	0.707	1.586
0.5 dB-Chebyshev	1.231	0.864	1.842
1.0 dB-Chebyshev	1.050	0.956	1.954
2.0 dB-Chebyshev	0.907	1.129	2.114
3.0 dB-Chebyshev	0.841	1.305	2.234

Analysis equations:

$$f_0 = \frac{1}{2\pi RC} \text{ (R's and C's are both equal)} \tag{8.9}$$

$$\text{Passband gain } A = \frac{R3 + R4}{R4}, \tag{8.10}$$

$$Q = \frac{1}{3 - A} \tag{8.11}$$

Note that for the 3.0 dB Chebyshev case the high Q required means we are getting uncomfortably close to the gain limit of three times.

The Chebyshev filters have a peak in their response which it may be possible to utilise for equalisation purposes. The height and frequency of the peak is, however, fixed for a given Q, making it a very inflexible method. This sort of approach can make modifications to deal with driver changes, etc. problematic, compared with using a dedicated equalisation stage; it is very helpful if you can say "*this* modifies *this*." It is, however, economical and since less stages are required, noise and distortion may be somewhat improved.

8.10 Sallen & Key Second-Order Lowpass Equal-C: Component Sensitivity

The equal-C Sallen and Key configuration has sensitivities for cutoff frequency that are the same as for the unity-gain version, but it has a more complicated set of sensitivities for Q, with their actual value depending on the Q value chosen, as shown in Table 8.4.

Table 8.5 shows how the sensitivity for Q varies with the Q value chosen. For $Q = 0.7071$ $(1/\sqrt{2})$, which is the most popular value in crossover design, the capacitor sensitivities for Q are almost twice the 0.5 of the unity-gain version (compare Table 8.2 above). It may be a tricky decision as to whether the convenience of having equal capacitors makes up for the fact that they need to be twice as accurate for the same precision of response.

Table 8.4: Second-Order Sallen & Key Equal-C Lowpass Component Sensitivities (after Van Valkenburg)

Component	Cutoff Frequency Sensitivity	Q Sensitivity
R1	−0.5	−0.5 + Q
R2	−0.5	0.5 − Q
C1	−0.5	−0.5 + Q
C2	−0.5	0.5 − 2Q
R3	0	2Q − 1
R4	0	−(2Q − 1)

Table 8.5: Second-Order Sallen & Key Equal-C Lowpass Component Sensitivities for 3 Specific Q's (after Van Valkenburg)

Component	Q Sensitivity (Q = 0.7071)	Q Sensitivity (Q = 2)	Q Sensitivity (Q = 8)
R1	0.207	1.5	7.5
R2	−0.207	−1.5	−7.5
C1	0.914	2.5	15.5
C2	−0.914	−2.5	−15.5
R3	−1.00	−5.00	−15.00
R4	1.00	5.00	15.00

For Q = 2, the sensitivities become significantly greater. If you could get away with 5% capacitors for Q = 0.7071, now you will need 1% to get the same precision, and they may be significantly more expensive. The Linkwitz–Riley filter has the lowest Q and so the lowest sensitivities.

If you tried to generate a Q of 8, which is the sort of value that can turn up in high-order elliptical filters, things are dire indeed, with sensitivities of 15 or more. If you had 5% capacitors for Q = 0.7071, you now will need 0.3% to get the same precision, and they will be very expensive, if indeed they can be obtained at all.

If you do need this sort of Q, there are multiple opamp configurations which work much more dependably.

8.11 Sallen & Key Second-Order Butterworth Lowpass: Defined Gains

We saw in the previous section that we could select convenient capacitor values for the lowpass filter by using a particular gain in the filter. Looking at it another way, we can have whatever gain we want, and whatever Q we want, by altering the capacitor values. As I have said several times, gain in a filter is often unwanted, but there are occasions when a defined amount of gain *is* wanted, as in the HF path of the crossover design example in Chapter 19, and building it into a filter will save an amplifier stage and reduce cost, while possibly also reducing noise and distortion as the signal has gone through one less stage. On the other hand, distortion may increase somewhat because the filter stage giving the gain has less negative feedback; the outcome depends on the opamp types used and the exact circuit conditions.

It would take up an enormous amount of space to tabulate all the combinations of Q and gain, so I will concentrate on the ever-useful Butterworth filter. The exact resistor values required for gains from 0 to +8 dB are shown in Figure 8.4, and summarised in Table 8.6, which also includes the standard cases of unity gain and of equal capacitors. In the latter case the gain required is extremely close to +4 dB. All the filters have a 1 kHz cutoff frequency, and resistors set to 1 kΩ; other frequencies can be obtained by scaling the component values. It is assumed that R4 in the gain network is fixed at 1 kΩ, though in the higher-gain cases it would be advisable to reduce this to lower the impedance of the gain-determining network and so minimise noise.

There is no obvious reason why higher gains than this could not be obtained. The gain limit of 3 times given in the section just above on equal-capacitor lowpass filters does not apply here because we are adjusting the capacitor values to keep the Q constant at 0.707 instead of letting it shoot off to infinity. There is however the consideration that reducing the negative feedback factor of the opamp will worsen the distortion performance.

Similar design data for second-order Butterworth Sallen & Key highpass filters is given in the section below on highpass filters. Not surprisingly, the ratios between the capacitors are exactly the same as the ratios of the resistors in the highpass case.

8.12 Sallen & Key Second-Order Lowpass: Non-Equal Resistors

In lowpass Sallen and Key filters the resistors are almost always made equal. There is no absolute requirement that this be so—perfectly respectable filters can be made with non-equal resistors, but as you might imagine this somewhat complicates the calculations, and it seems as if there is little to be gained by doing it.

I did think at one point it might be possible to make useful second-order filters with equal capacitor values combined with unity-gain by choosing non-equal resistors, but this is regrettably not true. With equal capacitors and unity gain the maximum Q obtainable is 0.5 when the resistors are equal; this is useful for second-order Linkwitz–Riley filters but for no other type. As soon as the resistors are made non-equal in either direction the Q falls below 0.5.

Non-equal resistors are, however, useful for third- and higher-order filters constructed as a single stage because they allow all the capacitors to be the same value. There is more on this below.

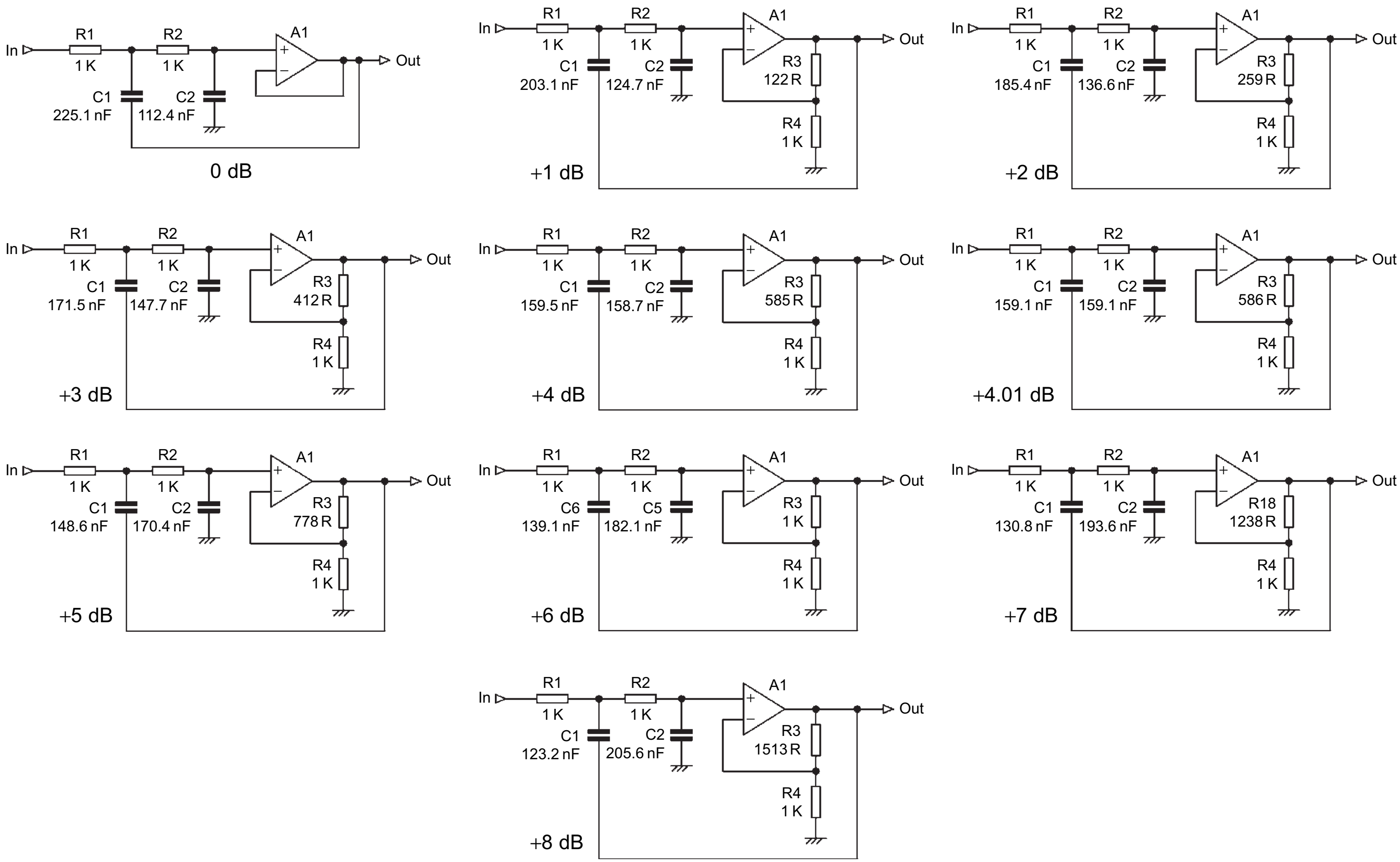


Figure 8.4: Low-pass Butterworth 1kHz S & K filters with defined passband gains from 0 to +8 dB.

Table 8.6: Component Values for Defined Gains in a Butterworth Second-Order S&K Lowpass Filter

Gain dB	Gain Times	C1 nF	C2 nF	C1/C2 Ratio Times	R3 Ohms	R4 Ohms
0.00	1.000	225.1	112.4	2.000	n/a	n/a
+1.00	1.122	203.1	124.7	1.628	122	1000
+2.00	1.259	185.4	136.6	1.357	259	1000
+3.00	1.412	171.5	147.7	1.161	412	1000
+4.00	1.585	159.5	158.7	1.004	585	1000
+4.01	1.586	159.1	159.1	1.000	586	1000
+5.00	1.778	148.6	170.4	0.871	778	1000
+6.00	2.000	139.1	182.1	0.764	1000	1000
+7.00	2.238	130.8	193.6	0.675	1238	1000
+8.00	2.512	123.2	205.6	0.599	1513	1000

8.13 Sallen & Key Third-Order Lowpass in a Single Stage

So far we have created first- and second-order responses by using a single stage containing a single op-amp. It is also possible to create third- and fourth-order responses in a single stage, though I should warn at once that the component sensitivities are worse.

Figure 8.5 shows third-order Butterworth lowpass Sallen & Key filters implemented as single stages. As for the second-order stages, this can be done in two different ways; with a unity-gain buffer or by using voltage amplification. By altering the capacitor ratios one can obtain any third-order characteristic with any gain. Figure 8.5a shows a unity-gain lowpass filter; starting with R = 1 kΩ, we find that there are no simple capacitor ratios, and C2 is inconveniently big at 562 nF, more than twice the size of C1. Figure 8.5b shows an alternative circuit with a gain of 1.5 times (there is nothing magical about this figure), which gives the same Butterworth response as Figure 8.5a. As usual, this gain may be an embarrassment rather than a help in a crossover filter.

There are no simple design equations for this sort of single-stage filter; and the values given are the result of some fairly serious mathematics which I am not going to inflict on you. The practical way to use these filters, and also the more complex ones that follow, is to take one of the examples given in this section and scale the resistor and capacitor values to get the cutoff frequency you want, bearing in mind that the resistor values must not be too high (or there will be excess noise) nor too low (because there will be excess distortion due to unduly heavy opamp loading). There is more guidance on this issue in Chapter 13 on the use of opamps.

The interesting thing about Figure 8.5b is that by using a gain of 1.5 times, we find that C2 is now smaller than C1, and a much more manageable 188.4 nF. This led me to ponder that there must be an intermediate value of gain that would give equal values for C1 and C2; this would

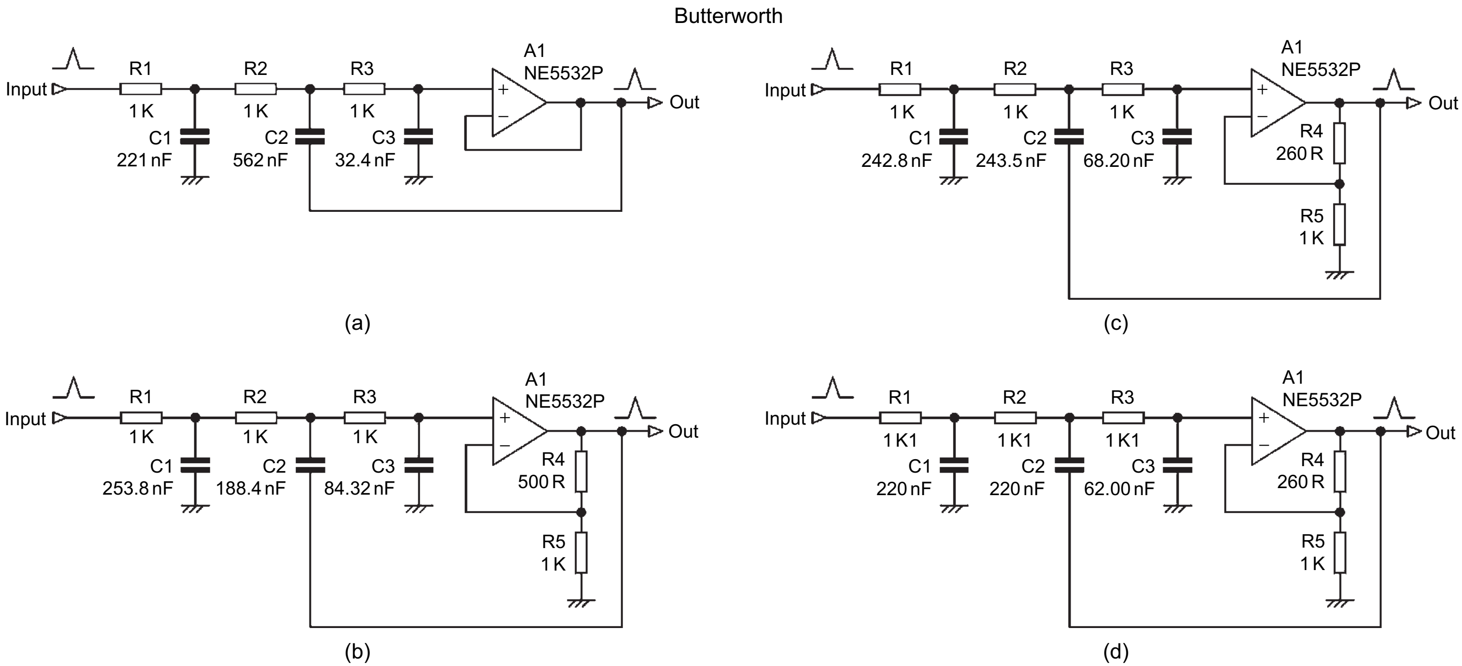


Figure 8.5: Third-order Sallen & Key lowpass Butterworth filter implemented as a single stage; (a) unity-gain; (b) with gain = 1.5; (c) equal-C with gain = 1.26; (d) equal-C with preferred values, gain = 1.26 cutoff frequency 1 kHz for all.

be very convenient. A good deal more mathematics leads to Figure 8.4c, where a gain of 1.26 times yields values for C1 and C2 that are very close to equality. Figure 8.5d shows the circuit with R and C values scaled to give C1 = C2 = 220 nF. By happy coincidence R1 = R2 comes out at the E24 value of 1.1k, and gives a very accurate Butterworth response. To the best of my knowledge this way of using identical capacitors in a third-order single-stage filter is a new idea.

Since there are no more degrees of freedom to explore, we are in general stuck with an awkward value for C3; here, by happy coincidence again, it comes out exactly as an E24 value. Capacitor series with E24 values are however not common, so we would use two E6 capacitors in parallel thus: 47 nF + 15 nF = 62 nF; yet another happy coincidence. I'm starting to think there is something spooky about this circuit.

Note that it is the two largest capacitors we have made equal and preferred values, leaving the much smaller C3 to be made up of a parallel combination; the smaller capacitors required to do this will be significantly cheaper.

I warned you that the component sensitivities would be worse, and the living proof is in Table 8.7 and Table 8.8.

The same process can be used to generate any third-order filter characteristic with equal-C values by choosing the correct gain. Figure 8.6 shows two examples. Figure 8.6a shows a unity-gain

Table 8.7: Third-Order Butterworth Sallen & Key Single-Stage (Unity-Gain) Lowpass Component Sensitivities

Component	Cutoff Frequency	Q
R1	0.9974	
R2	0.9994	
R3	1.2180	
C1	0.9943	
C2	1.1098	
C3	1.3445	

Table 8.8: Third-Order Butterworth Sallen & Key Equal-C Single-Stage (Gain = 1.26) Lowpass Component Sensitivities

Component	Cutoff Frequency	Q
R1	0.9979	
R2	0.9984	
R3	1.3357	
C1	0.9946	
C2	1.0050	
C3	1.7646	
R4	0.9200	
R5	0.9186	

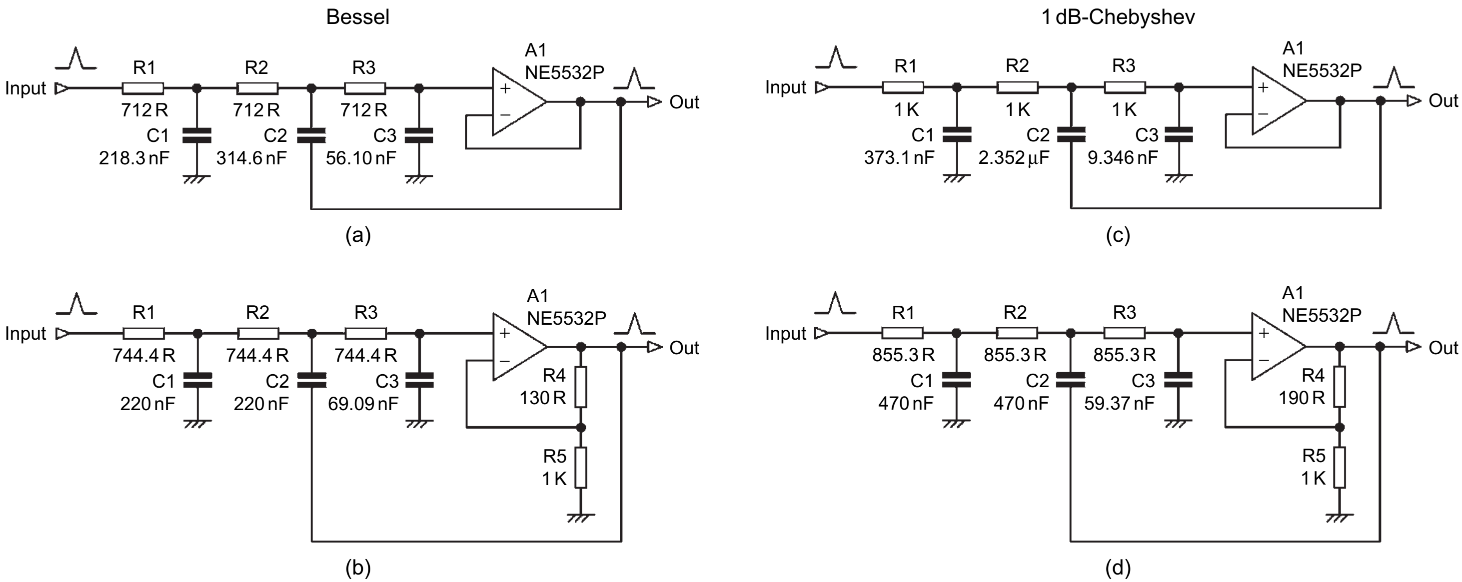


Figure 8.6: Third-order Sallen & Key lowpass filters implemented as a single stage; (a) unity-gain Bessel; (b) equal-C Bessel with gain = 1.13 and preferred C values; (c) unity-gain 1 dB-Chebyshev; (d) equal-C 1 dB-Chebyshev with gain = 1.19 and preferred C values. Cutoff frequency 1 kHz for all.

Table 8.9: Capacitor Ratios and Gains Required for Equal-C Third-Order Single-Stage Filters

Type	C1/C3 Ratio	Gain A	C3 Sensitivity for Freq
Linkwitz–Riley	0.326	1.16	1.377
Bessel	0.314	1.13	1.339
Linear-Phase 5% delay ripple	0.210	1.14	1.544
Butterworth	0.282	126	1.765
0.5 dB-Chebyshev	0.157	1.22	2.453
1.0 dB-Chebyshev	0.126	1.19	2.682
2.0 dB-Chebyshev	0.0949	1.15	3.072
3.0 dB-Chebyshev	0.0633	1.123	3.331

third-order Bessel filter, and Figure 8.6b a third-order Bessel filter with two equal capacitors. Figures 8.6c and Figure 8.6d show the same two variations for a 1 dB-Chebyshev filter.

Table 8.9 shows the ratio between the two equal capacitors C1, C2 and the remaining capacitor C3, and the gain A required for a wider range of filter types.

In every case in Table 8.9 the worst sensitivity for frequency is shown by C3. The values are tabulated in the last column, where it can be seen that as expected, the response demanding the highest Q has the worst sensitivity, and that this is noticeably worse than for the normal method of making a third-order filter by cascading suitable second-order and first-order filters, as indicated earlier. For the Butterworth response, the worst capacitor sensitivity for frequency is 1.765 rather than the 0.914 we saw for an equal-C second-order filter; almost twice as bad. This is the sort of thing you must expect when higher-order filters are implemented in a single stage.

This means that this approach is not very suitable for precise crossover filters, but it is not without value. There are some applications, such as subsonic or ultrasonic filtering, where a very accurate amplitude response is not essential. A third-order filter ensures there is very little roll-off in the passband, so minor variations in it are of negligible importance. A good example of the method is its use in the combined subsonic-ultrasonic filter described later in this chapter.

8.14 Sallen & Key Third-Order Lowpass in a Single Stage: Non-Equal Resistors

Earlier in this chapter we noted that there was little to be gained by using non-equal resistor values in a second-order Sallen & Key stage. However, it can be useful in a third-order stage. As we just saw, two of the capacitors in such a stage can be made equal by suitable choice of the amplifier gain. If we allow non-equal resistors, the extra degrees of freedom allow us to design third-order stages with *three* equal capacitors. Figure 8.7 shows

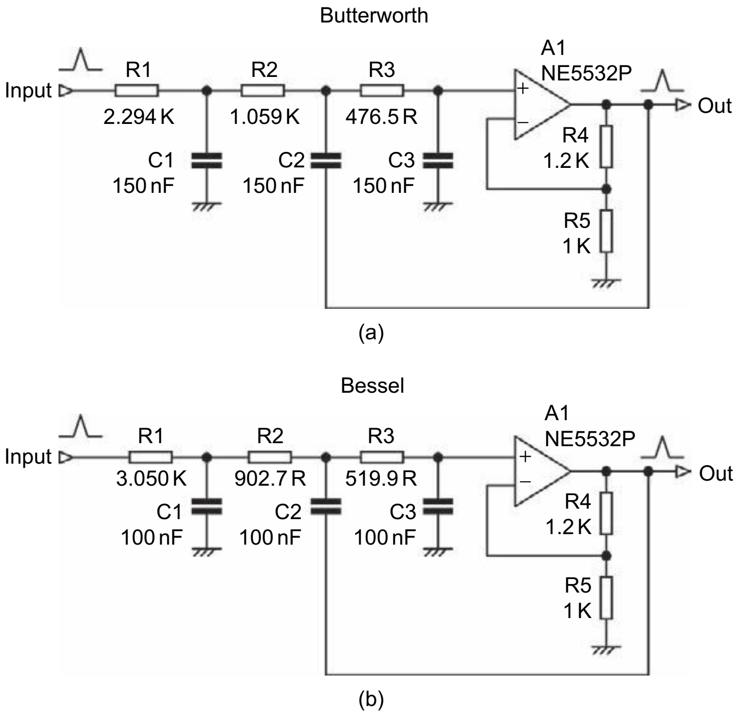


Figure 8.7: Third-order Sallen & Key 1 kHz lowpass filters implemented as a single stage, with three equal capacitors: (a) Butterworth with gain = 2.2 and exact values; (b) Bessel with gain = 2.2 and exact values.

Butterworth and Bessel versions; note that the required amplifier gain required is relatively high at 2.2 times (+6.8 dB), which may be inconvenient in crossover design. Chebyshev filters of this type can also be designed.

Once again there are no manageable design equations, and you should scale the component values given, which are as always for a 1 kHz cutoff, to get the filter you want.

Figure 8.7 also shows that the capacitor values are in general lower than for the previous filters—150 and 100 nF instead of 220 nF. This could give a significant saving when buying close-tolerance capacitors; another cost benefit is that three, and three only, capacitors are required, as it is never necessary to parallel components to achieve awkward capacitance values. Note that the resistor values range over a ratio of 4.81.

8.15 Sallen & Key Fourth-Order Lowpass in a Single Stage

Having seen how it is possible—if not always advisable—to make a third-order filter block using a single opamp, one's mind naturally turns to pondering if a fourth-order filter block can be constructed in the same kind of way. The answer is yes. As before, you can make any filter characteristic with any gain, and it is also still possible to choose a filter characteristic and then choose a gain which makes two of the capacitors equal. But there are now four capacitors, and the other two will in general be non-convenient values.

Figure 8.8a shows the exact values for R = 1 KΩ. The obvious step is to make C1 and C3 the nearest value E6 of 220 nF, and scale the other components accordingly to keep the cutoff frequency at 1 kHz. This gives us Figure 8.8b, where the capacitors are kept in the same ratio and the resistor values adjusted.

The component sensitivities for the filter in Figure 8.8b are shown in Table 8.10, and they are not as bad as might be feared, because of the relatively low Q's involved in a fourth-order Butterworth characteristic. Even so, the worst is 3.05 for C4, which is six times worse than a second-order Butterworth stage.

The most popular application of fourth-order filters is in Linkwitz–Riley crossovers, so let us have a quick look at the possibility of implementing a fourth-order Linkwitz–Riley filter

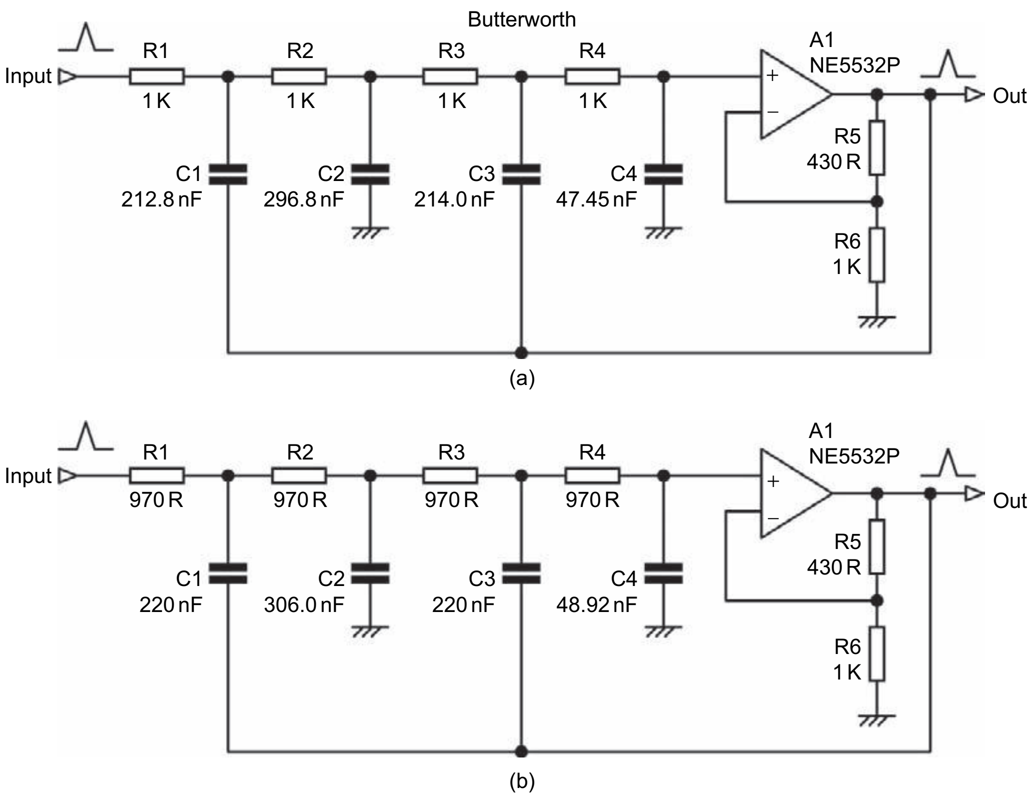


Figure 8.8: Fourth-order Sallen & Key Butterworth 1 kHz lowpass filters implemented as a single stage; (a) equal-C with gain = 1.43 and exact values; (b) equal-C with gain = 1.43 and preferred values for C1 and C3.

Table 8.10: Fourth-Order Sallen & Key Equal-C Butterworth Lowpass Component Sensitivities

Component	Cutoff Frequency
R1	1.14
R2	1.25
R3	1.54
R4	2.13
C1	1.04
C2	2.29
C3	2.83
C4	3.05
R5	2.99
R6	2.98

in a single stage. Figure 8.9a shows the precise values that emerge when, as before, the gain is adjusted to give the desired characteristic while making C1 and C3 very nearly equal. Figure 8.9b shows the result of scaling the circuit values to make C1 = C3 = 220 nF, while keeping the same 1 kHz cutoff frequency.

The overall action of the circuit is exactly the same as two cascaded second-order Butterworth filters with the same cutoff frequency, which is the usual way of making a fourth-order Linkwitz–Riley filter. We have saved the cost of an opamp, and probably reduced noise and distortion somewhat due to its absence, but the extra series resistances (four instead of two) may cause trouble with increased common-mode–distortion.

The component sensitivities for the filter in Figure 8.9b are shown in Table 8.11; they are slightly better than for the fourth-order Butterworth because of the lower Q's involved in a Linkwitz–Riley characteristic. The worst, for C4 again, is 2.15 for C4, which is now only about four times worse than a second-order Butterworth stage, and with appropriate capacitor sourcing may make this approach a viable alternative. However, overall it may not be the optimal strategy; we have saved an opamp but need more precise capacitors to achieve the same accuracy, and the net cost may well be greater.

It was earlier mentioned in passing that a disadvantage of the Sallen & Key lowpass configuration was that the frequency response at very high frequencies would start to come back up, because one of the shunt capacitors is terminated at the opamp output, which has a non-zero output impedance that rises with frequency. This effect is more of a problem with the fourth-order single-stage filter because there are now two capacitors terminated at the opamp output, and it is possible that significant inaccuracies might arise in the audio band.

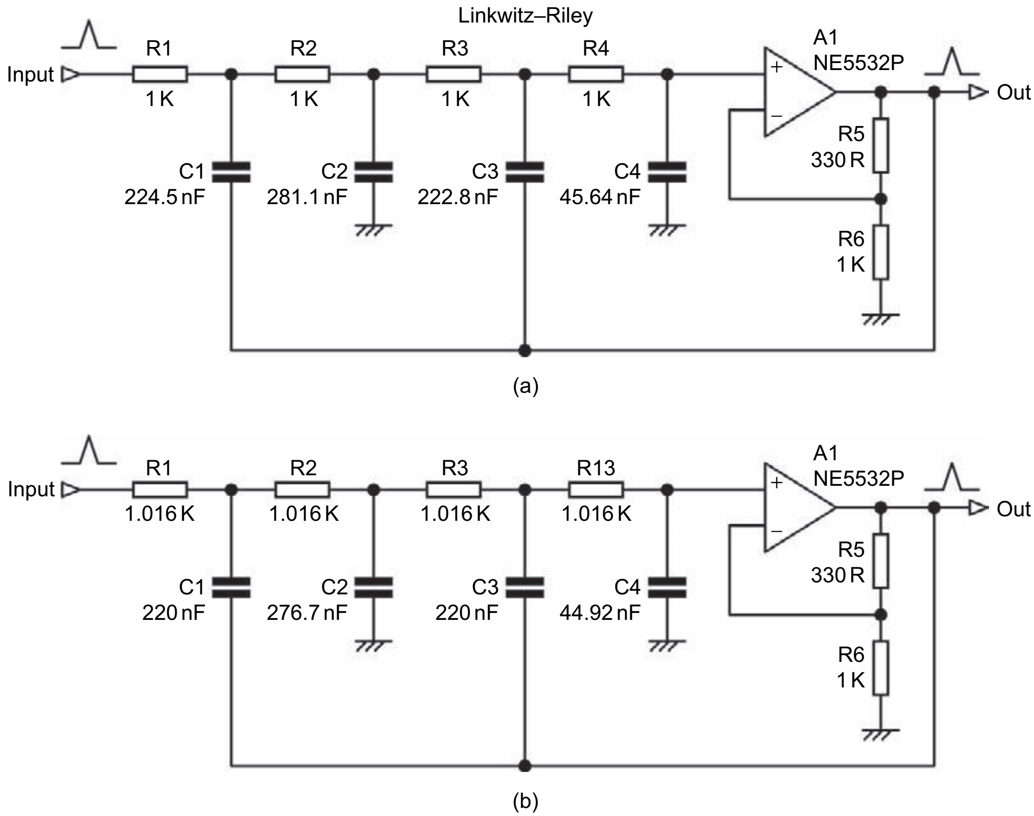


Figure 8.9: Fourth-order Sallen & Key Linkwitz–Riley lowpass filters implemented as a single stage; (a) equal-C with gain = 1.33 and exact values; (b) equal-C with gain = 1.33 and preferred values for C1 and C3.

Table 8.11: Fourth-Order Sallen & Key Equal-C Linkwitz–Riley Lowpass Component Sensitivities

Component	Cutoff Frequency	Q
R1	1.02	
R2	1.02	
R3	1.02	
R4	1.60	
C1	0.88	
C2	1.79	
C3	1.64	
C4	2.15	
R5	1.77	
R6	1.77	

Take the Butterworth filter in Figure 8.8b; if implemented with an elderly TL072 it will be found that the response has risen above the theoretical characteristic by 1 dB as low as 4.6 kHz. Admittedly this is at a theoretical attenuation of 52 dB, so will probably be of no consequence, but it is something to keep an eye on. This difficulty is much reduced by using opamps like the 5532 that can maintain a low output impedance up to considerably higher frequencies.

8.16 Sallen & Key Fourth-Order Lowpass in a Single Stage: Non-Equal Resistors

Earlier in this chapter we saw that by using non-equal resistors, it was possible to make a filter with three equal capacitors. This principle can be extended to fourth-order filters, as shown in Figure 8.10. The required gain is still 2.2 times, but note that the resistors now range over a ratio of 13.5 instead of 4.81.

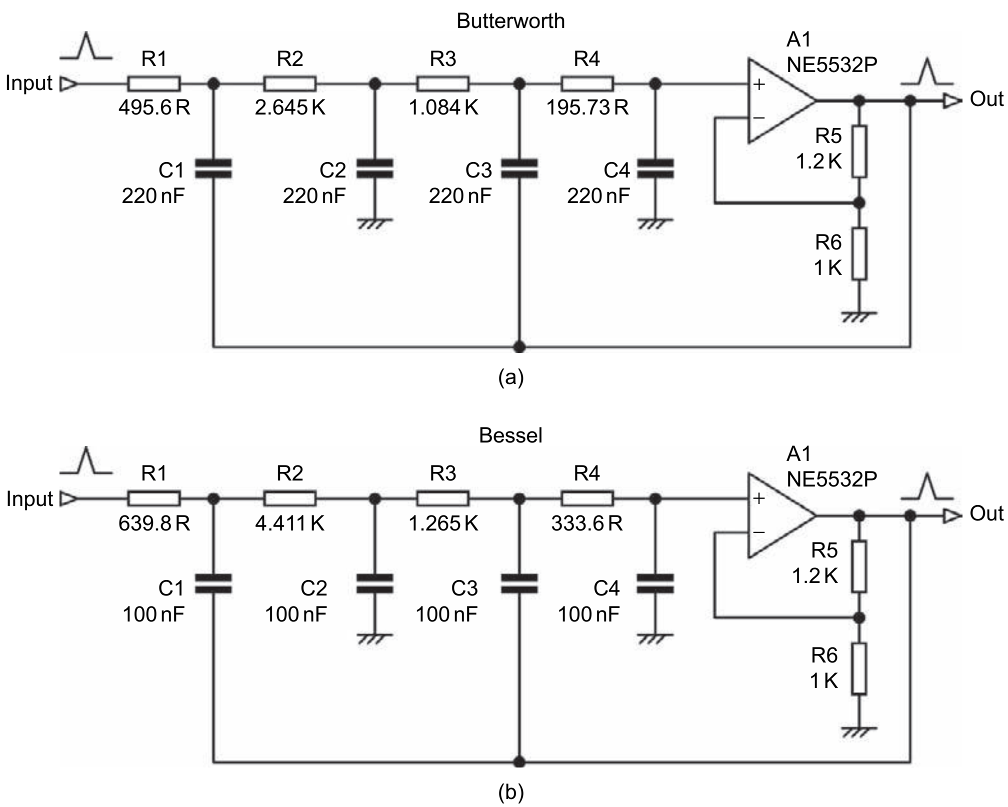


Figure 8.10: Fourth-order Sallen & Key lowpass filters implemented as a single stage, with four equal capacitors, gain = 2.2 and exact values; (a) Butterworth; (b) Bessel.

These circuits also have potentially worse problems with the response coming back up again in the audio band.

8.17 Sallen & Key Fifth- and Sixth-Order Lowpass in a Single Stage

Until very recently I had never seen an attempt to make a fifth-order active filter in one circuit block, and I had a dark suspicion that it might be mathematically impossible. However, in the stacks of a proper book-type library I discovered a 1972 paper by Aitken & Kerwin [3] that purports to show how to do it. Unfortunately some of the design parameters they give appear to be wrong, and it took me a good deal of hard work to come up with the circuit in Figure 8.11a, which definitely gives a 30 dB/octave roll-off slope, though I do not vouch for its exact Butterworthiness around the cutoff frequency. The sixth-order filter in Figure 8.11b was much easier, as the given design parameters were correct in this case, and it does give an accurate Butterworth response.

Calculating the sensitivities of these circuits is unlikely to be a rewarding experience, and I have not undertaken it. What you can always do, of course, is to simulate the filter and try

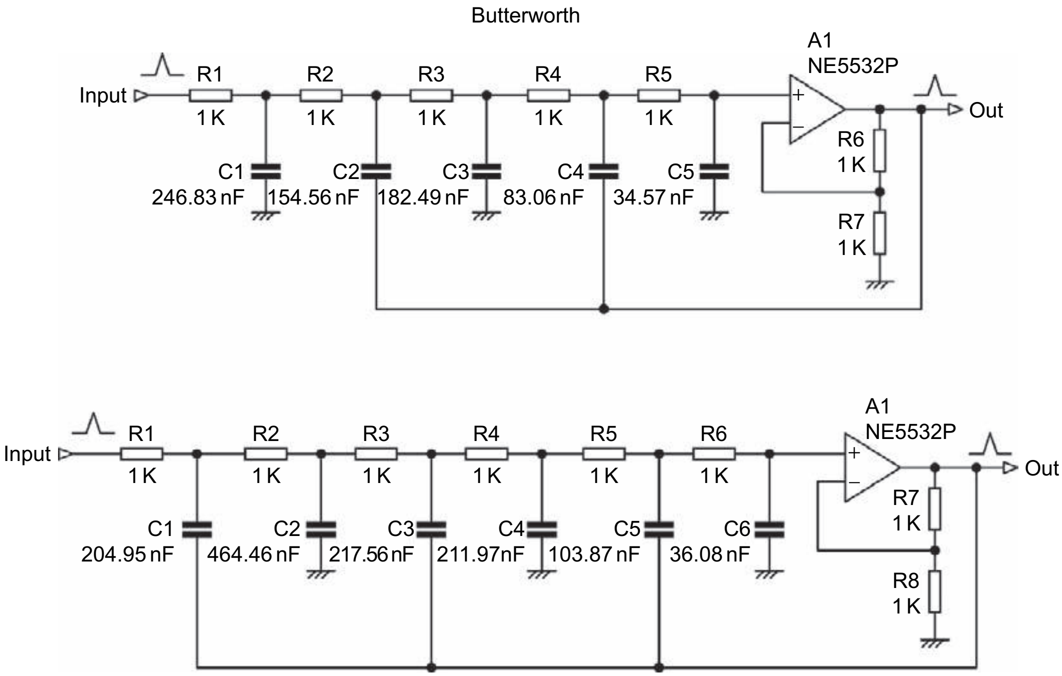


Figure 8.11: (a) Fifth-order and (b) Sixth-order Sallen & Key Butterworth 1-kHz lowpass filters, implemented as single stages, with gain = 2 and exact values; these are probably more curiosities than useful circuits as sensitivities will be poor, especially for the sixth-order version.

Table 8.12: Sixth-Order Sallen Key Butterworth Lowpass Component Sensitivities: See Figure 8.10b

Component	Frequency Sensitivity	Component	Frequency Sensitivity
R1	−0.8	C1	−0.8
R2	−0.2	C2	+0.4
R3	−0.6	C3	−1.2
R4	−2.7	C4	−4.1
R5	+2.7	C5	+5.2
R6	−2.2	C6	−4.5

out the effects of altering one component at a time by a small amount, such as 1%. The effect on the cutoff frequency is then easy to assess, though the effect on other filter parameters is rather harder to assess; you will have to examine the frequency response and see which of its features are of most interest to you. To give an example, I applied this method to the sixth-order circuit in Figure 8.11b, and the results are shown in Table 8.12. Please bear in mind that these figures are read from a cursor on a graph, and their accuracy is thereby limited.

Before I began I thought that the sensitivities would be very bad, and I imagined the value of C6 would be especially critical. In fact, C5 has the worst figure, and it is not as awful as I expected. Compare the fourth-order Sallen & Key equal-C Butterworth with a worst-case sensitivity of 3.05.

These high-order single-stage filters are in some ways clever designs, as no less than two opamp sections have been saved, but in actual fact they can, potentially at least, cause trouble. Attempts to make high-Q filters with them would be difficult simply because it would not be feasible to obtain components of the requisite accuracy for positions like C5. These filters should perhaps be regarded as a curiosities rather than practical designs.

8.18 Sallen & Key Highpass Filters

Highpass Sallen & Key filters are basically the same as their lowpass brothers, with the R's and C's swapped in their circuit positions. All the considerations described for the lowpass filters, such as the component sensitivities of particular configurations, are applicable so long as this is kept in mind. Rather than repeat all this material, the section below on highpass filters has been kept to a manageable length by providing only the design examples, with a minimum of commentary.

As mentioned earlier, highpass filters have the advantage that the capacitors are usually all the same value, and it is the resistors that come in awkward values, and paralleling them is cheap and takes up little PCB space. Nonetheless, in an ideal world we would have as many identical resistors as possible, and some of the highpass filters in this section are configured to achieve this.

8.19 Sallen & Key Second-Order Highpass: Unity-Gain

Sallen & Key highpass filters are very much the same as the lowpass filters with the Rs and the Cs swapped over.

Figure 8.12 shows a second-order highpass Sallen & Key unity-gain filter with a cutoff frequency of 1 kHz and a Q of 0.707 to obtain a Butterworth characteristic. Now the capacitors are equal, while the resistors have a ratio of two to define the required Q. The measured noise output of this filter is −115.2 dBu (corrected), with its design equations.

Design equations: Choose R2:

$$C = \frac{2Q}{(2\pi f_0)R_2} \tag{8.12}$$

$$R_1 = \frac{R_2}{4Q^2} \tag{8.13}$$

Analysis equations:

$$f_0 = \frac{1}{2\pi C\sqrt{R_1 R_2}} \tag{8.14}$$

$$Q = \frac{1}{2}\sqrt{\frac{R_2}{R_1}} \tag{8.15}$$

The input impedance of this circuit is relatively high at low frequencies, where it is dominated by the series impedance of the capacitors. It then falls with increasing frequency,

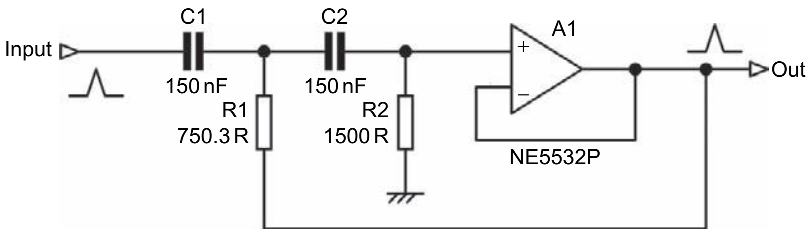


Figure 8.12: The classic second-order highpass Sallen and Key filter. Cutoff frequency is 1 kHz. Q = 0.7071 for a Butterworth response.

reaching a minimum just above the −3 dB cutoff frequency. Above this it levels out to the value of R2, as the capacitor impedance is now negligible. R1 does not add to the loading because it is bootstrapped by the voltage-follower.

The TL072 is a very poor choice for S&K highpass filters. Its common-mode problems are sharply revealed if it is used in highpass configurations; the output hits the top rail and then shoots down to hit the bottom rail as the opamp internals go into phase reversal, which is about the worst sort of clipping you could imagine. In the Bad Old Days when TL072s had to be used in audio paths for cost reasons, if a highpass filter was required the signal would be attenuated by 6 dB so the common-mode limits could never be reached, and the lost level recovered later. The TL072 also produces a lot of CM distortion, and generates considerable extra distortion when its output is loaded even lightly, and it must be regarded as obsolete for almost all audio applications.

8.20 Sallen & Key Second-Order Highpass: Equal-Resistors

You can make a lowpass Sallen and Key filter with conveniently equal capacitor values if you configure the amplifier to give voltage gain instead of acting as a unity-gain buffer. A highpass filter can be made in exactly the same way with equal-resistor values, though there is less of an advantage because awkward resistor values are much less of a problem than awkward capacitor values. Figure 8.13 shows such a second-order equal-R Butterworth highpass S&K filter; you will note that the component values are exactly the same as in the lowpass equal-C filter in Figure 8.3, and so is the gain required for a Q of 0.7071.

Due to the equal-value R and C components, the design and analysis equations are just the same as for the lowpass equal-C second-order Butterworth described earlier.

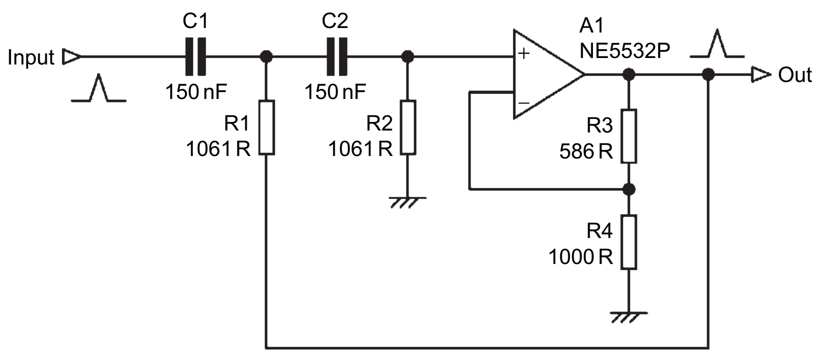


Figure 8.13: Equal-R second-order highpass Butterworth filter with a cutoff frequency of 1 kHz. Gain must be 1.586 times for maximally flat Butterworth response (Q = 0.7071).

Design equations: Choose a value for C (= C1 = C2)

$$R_1 = R_2 = \frac{1}{(2\pi f_0)C}, \tag{8.16}$$

$$\text{Passband gain A} = 3 - \frac{1}{Q} \tag{8.17}$$

Then choose R3, R4 to get the required passband gain A for the chosen Q

Analysis equations:

$$f_0 = \frac{1}{2\pi RC} \text{ (R's and C's are both equal)} \tag{8.18}$$

$$\text{Passband gain A} = \frac{R3 + R4}{R4} \tag{8.19}$$

$$Q = \frac{1}{3 - A} \tag{8.20}$$

8.21 Sallen & Key Second-Order Butterworth Highpass: Defined Gains

We saw in the previous section that we could select convenient resistor values for a second-order Sallen & Key highpass filter by selecting the gain in the filter. We can choose both the gain and the Q by altering the resistor values. Filter gain is often unwelcome, but sometimes required, as in the HF path of the crossover design example in Chapter 19, and building it into a filter will save an amplifier stage and reduce cost, while possibly also reducing noise and distortion as the signal has gone through one less stage. On the other hand, distortion may increase somewhat because the filter stage has less negative feedback; this depends on the opamp types used and the exact circuit conditions.

As for the lowpass case I will focus on the Butterworth filter characteristic. The exact resistor values required for gains from 0 to +8 dB are shown in Figure 8.14, where the cutoff frequency in each case is 1 kHz, and the capacitors have been set to 100 nF; other frequencies can be obtained by scaling the component values. The resistor values summarised in Table 8.13 also include the standard cases of unity gain and of equal resistors. In the latter case the gain required is extremely close to +4 dB. It is assumed that R4 in the gain network is fixed at 1 kΩ; in the higher-gain cases it would be advisable to reduce its value to lower the impedance of the gain network and thus minimise noise.

Higher gains than this can be obtained. The gain limit of 3 times given in the section just above on equal-resistor highpass filters does not apply here because we are adjusting the

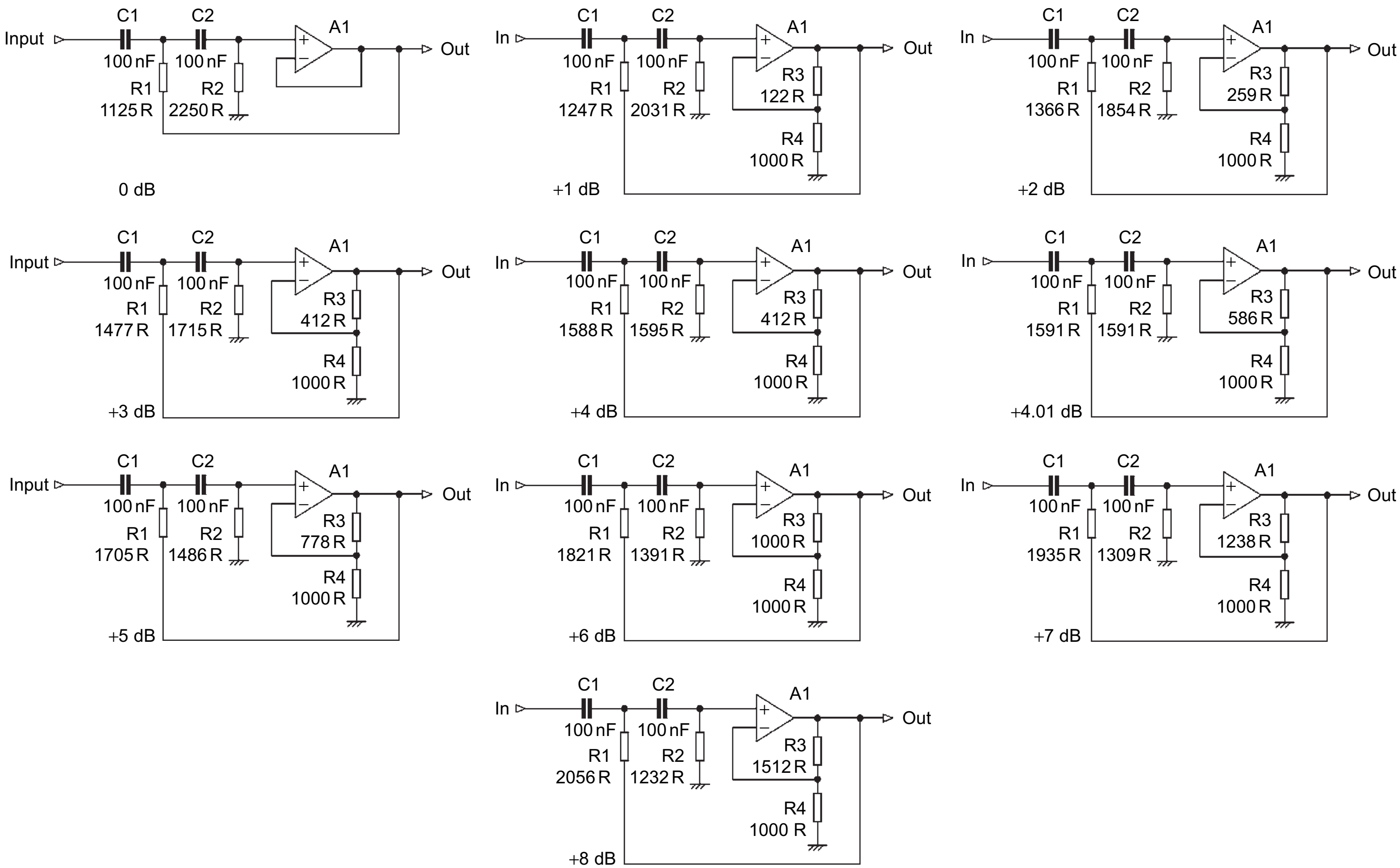


Figure 8.14: Highpass 1 kHz Butterworth S & K filters with defined passband gains from 0 to +8 dB.

Table 8.13: Component Values for Defined Gains in a Butterworth Second-Order S&K Highpass Filter

Gain (dB)	Gain (Times)	R1 (Ohms)	R2 (Ohms)	R2/R1 Ratio (Times)	R3 (Ohms)	R4 (Ohms)
0.00	1.000	1125	2250	2.000	n/a	n/a
+1.00	1.122	1247	2031	1.628	122	1000
+2.00	1.259	1366	1854	1.357	259	1000
+3.00	1.412	1477	1715	1.161	412	1000
+4.00	1.585	1588	1595	1.004	585	1000
+4.01	1.586	1591	1591	1.000	586	1000
+5.00	1.778	1705	1486	0.871	778	1000
+6.00	2.000	1821	1391	0.764	1000	1000
+7.00	2.238	1935	1309	0.637	1238	1000
+8.00	2.512	2056	1232	0.599	1512	1000

resistor values to keep the Q constant at 0.707 instead of letting it go off to infinity, or indeed beyond. There is however the consideration that further reducing the negative feedback factor of the opamp will worsen the distortion performance.

Similar design data for second-order Butterworth Sallen & Key lowpass filters is given in the section above on lowpass filters. Not surprisingly, the ratios between the resistors here are exactly the same as the ratios of the capacitors in the lowpass case.

8.22 Sallen & Key Second-Order Highpass: Non-Equal Capacitors

We noted that for lowpass Sallen & Key filters, the resistors are almost always made equal. In the same way, the capacitors in highpass filters are almost always made equal; this is not an essential property of Sallen & Key filters but does make the calculations simpler. There seems to be little point in using non-equal capacitors in any second-order Sallen & Key stage, and even less point in using them in third- and higher-order highpass filters constructed as a single stage; by analogy with the lowpass equivalent, it would allow all the resistors to be made the same value, but because of the sparse nature of the available capacitor values, having equal capacitors is much more useful than having equal resistors; both the total capacitor cost and the PCB area occupied can be minimised.

8.23 Sallen & Key Third-Order Highpass in a Single Stage

Third-order highpass filters can be made in a single stage, just as for the lowpass equivalent in Figure 8.5. In Figure 8.15a is shown a third-order Butterworth highpass using a unity-gain amplifier. As expected, this gives us three different resistor values for R1, R2, and R3.

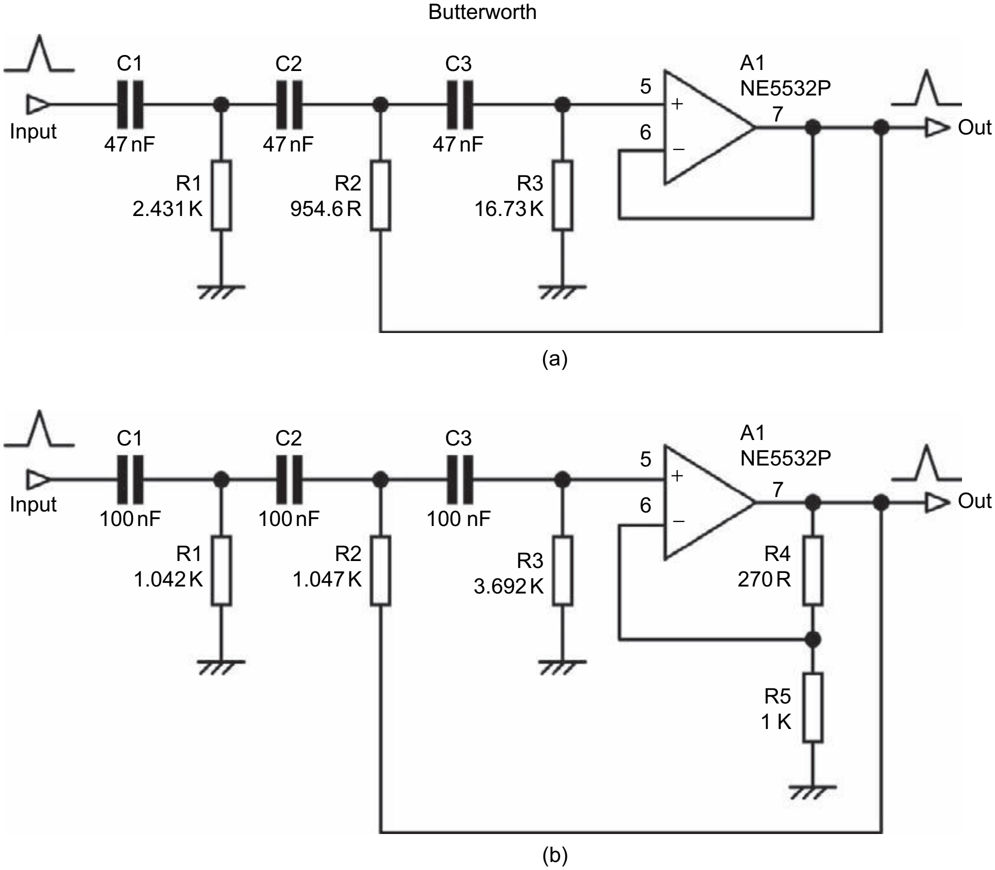


Figure 8.15: Third-order Sallen & Key highpass Butterworth filters implemented as a single stage; (a) unity-gain; (b) equal-R with exact values, gain = 1.27 times; cutoff frequency 1 kHz.

Figure 8.15b shows the alternative version where the amplifier gain is chosen to be 1.27 times (+2.08 dB) to make R1 and R2 almost identical in value. Note that this method requires larger capacitors if the lowest resistor value is to be kept in the range 700–1000 Ω. R4 very conveniently comes out as the E12 value of 270 Ω.

The alert reader will have spotted that the gain required for the most accurate resistor equality is 1.27 times, as opposed to the 1.26 times in the equivalent lowpass filter in Figure 8.5. This is because the calculation method used only allows gain to be specified to two decimal places, and has no other significance. Either value may be used with very small response errors.

8.24 Sallen & Key Fourth-Order Highpass in a Single Stage

As described for lowpass fourth-order filters, you can make any filter characteristic with any gain, and it is also still possible to choose a filter characteristic and then choose a gain that makes two of the resistors equal. There are four resistors, and the other two will in general be non-convenient values, though as explained before this is much less of a problem than awkward capacitor values. A fourth-order Butterworth highpass filter is shown in Figure 8.16a, with the exact component values that emerge from the design process.

You will have spotted that R1 and R3 in Figure 8.16a are both extremely close to the E12 value of 1.2 kΩ. Making them so introduces response errors of less than 0.05 dB from 10 Hz to 2 kHz, and above that the errors are negligible; the resulting circuit is shown in Figure 8.16b. R2 and R4 are much further from preferred values and would be best

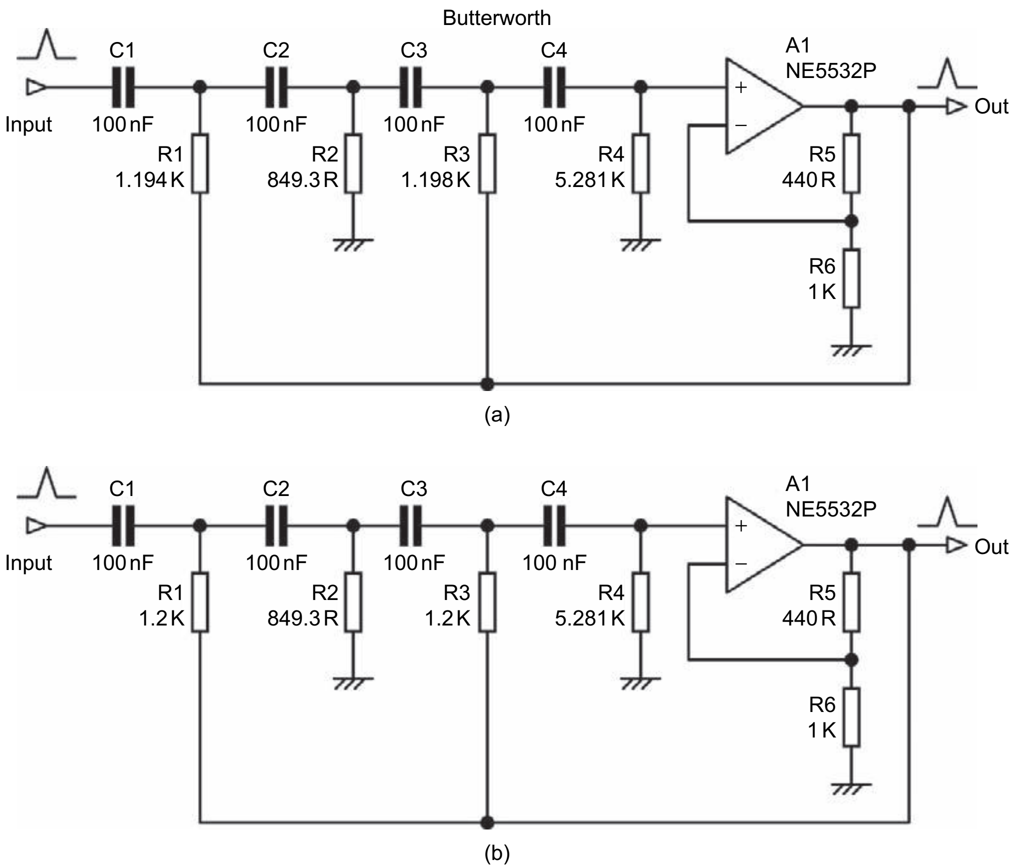


Figure 8.16: Fourth-order Sallen & Key Butterworth highpass filter in a single stage; (a) equal-C with gain = 1.43 and exact values; (b) equal-C with gain = 1.43 and preferred values for R1 and R3.

implemented as parallel combinations. Note that R5 could be very handily made up of two 220 Ω resistors in series.

The gain required for the most accurate resistor equality is 1.44 times, as opposed to the 1.43 times in the equivalent lowpass filter in Figure 8.8 above. Once again, this is because the gain is only specified to two decimal places. Either value may be used with very small response errors.

A fourth-order Linkwitz–Riley filter implemented in a single stage is shown in Figure 8.17, to complement the lowpass version in Figure 8.9 above. The gain for best equality of R1 and R3 is now 1.32 times (+2.48 dB) rather than 1.33, for the same reason as before. This time we have been very lucky with the values, as R1 and R3 in Figure 8.17a are both very close to the E12 value of 1.1 kΩ, R2 is close to 910R, and R4 is very close to 5.6 kΩ; R5

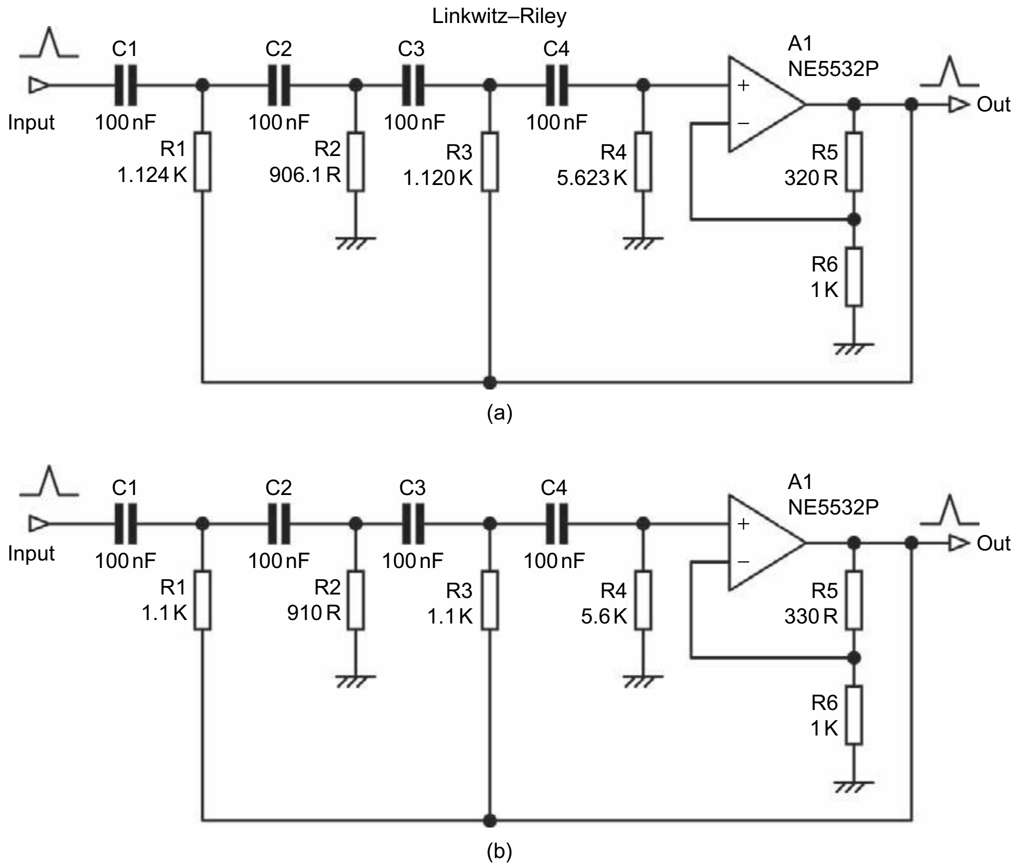


Figure 8.17: Fourth-order Sallen & Key Linkwitz–Riley highpass filters implemented as a single stages. (a) Equal-R with gain = 1.32 and exact values; (b) Equal-R with gain = 1.33 and E24 preferred values for R1, R3 and R5, R6.

has been set to 330 Ω, giving a gain of 1.33 times. A version using these values is shown in Figure 8.17b; the errors are +0.3 dB from 10 Hz to 900 Hz, but there is an error peak of −0.7 dB around 1.4 kHz, falling off to −0.1 dB at 10 kHz.

This example of the use of E24 preferred values shows that it is tricky with fourth-order single-stage filters because of the increased component sensitivity over low-order filters. The greatest error here is introduced by taking R5 to be 330 Ω rather than 320 Ω, so if you do decide to use this kind of filter (and I honestly don't think it will very often be a good option) you might consider making R5 and R6 as parallel combinations.

8.25 Implementing Linkwitz–Riley with Sallen & Key Filters: Loading Effects

The most common implementation of the popular fourth-order Linkwitz–Riley filter is made by cascading two second-order Butterworth Sallen & Key filters. This is why the Linkwitz–Riley is sometimes called a "Butterworth-squared" filter. A typical implementation of a 3 kHz Linkwitz–Riley lowpass filter is shown in Figure 8.18, where the resistor values have been reduced as much as appears to be prudent in the quest for low noise. (The resistor values just happen to come out as 797 Ω; this is not some sort of subtle post-modernist reference to the AD797 opamp.) It is therefore necessary to consider carefully the way that the loading from the second filter will affect the first filter that drives it. As we saw earlier in the chapter, the input impedance of a second-order Sallen & Key lowpass filter is high at low frequencies, where the capacitor impedance is high, but with increasing frequency, falls to the value of R1, ie 797 Ω. A previous 5532 stage should be able to drive this to full level without significant deterioration in the distortion performance but this depends on how much it is loaded in performing its own function.

The second Sallen & Key lowpass filter is identical so naturally acts in the same way. The current it draws from opamp A1 is, however, less because the amplitude of the signal from A1 is falling as frequency increases. If the input to R1 is 10 Vrms, the maximum current drawn from the previous stage at high frequencies is 12.5 mA rms. The current drawn by the second stage from A1 only reaches a maximum of 7.6 mA, peaking gently around 2.5 kHz. This is illustrated in Figure 8.19.

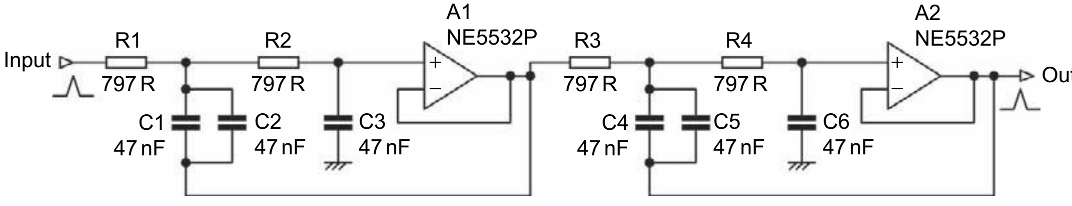


Figure 8.18: Fourth-order Linkwitz–Riley filter made by cascading two second-order Butterworth Sallen and Key filters. Cutoff is −6 dB at 3 kHz.

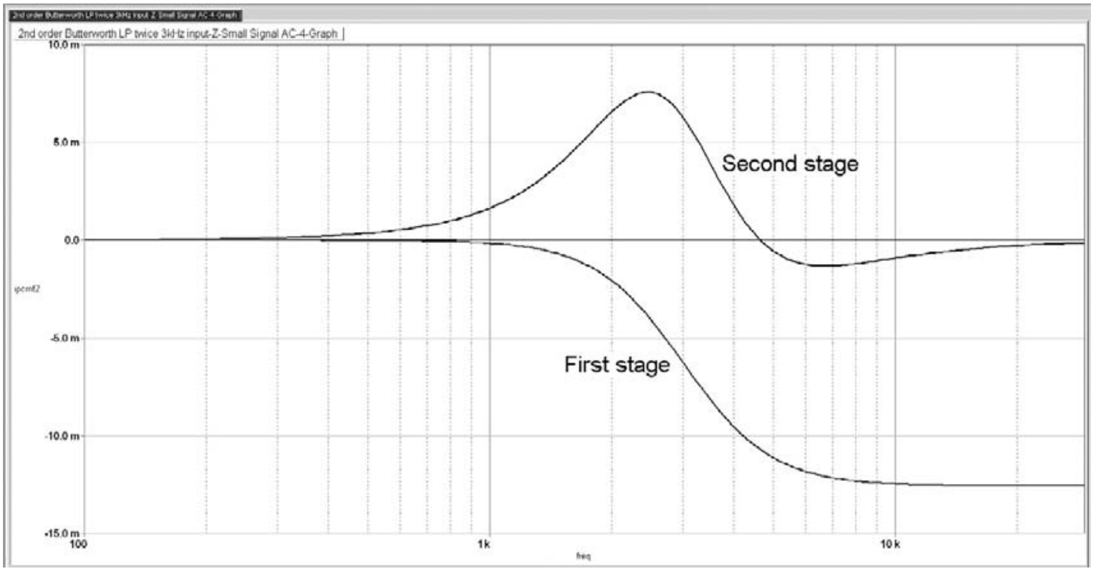


Figure 8.19: The current drawn at the inputs of the first and second stages of the Linkwitz–Riley filter in Figure 8.18. Input voltage was 10 Vrms.

The overall distortion of the two stages is therefore somewhat less than might be expected from a casual glance. Similar considerations apply to the equivalent Linkwitz–Riley highpass filter.

8.26 Lowpass Filters with Attenuation

It may occur that you need to attenuate a signal before applying it to a lowpass Sallen & Key filter. Since all the filter designs in this chapter require to be driven from a very low source impedance to get the desired response, you might think that the only solution is to use a potential divider with reasonably high values, so as not unduly load upstream circuitry, and then put a unity-gain buffer between that and the filter. In fact, with the Sallen & Key configuration it is usually unnecessary to indulge in an extra opamp.

If R1 and R3 in Figure 8.20b are chosen so that their ratio gives the required attenuation, but the output impedance of the divider (their value as a parallel combination) is equal to the R1 in Figure 8.20a, no buffer is required, with a saving in power consumption and parts cost, and hopefully some reduction in noise and distortion.

The equivalent circuit for the Sallen & Key highpass filter, which has a capacitor at its input, would be a capacitor potential divider with an output reactance equivalent to the original capacitor. This however is likely to lead to high-frequency overload

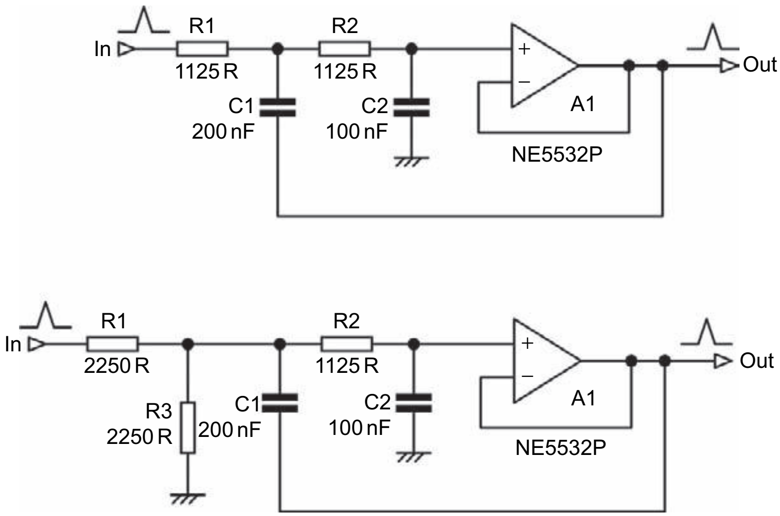


Figure 8.20: Building attenuation into a filter: (a) the original lowpass filter; (b) with a 6 dB input divider R1, R3 scaled to give the same driving impedance as R1 in the first filter. Cutoff frequency is 1 kHz.

problems in the previous stage as the impedance to ground of the divider will fall with frequency.

8.27 Bandwidth Definition Filters

This seems a good point to look at some real filter applications that show how our requirements drive the choice of filter type and its detailed design. An important feature of crossovers for sound reinforcement is that they implement bandwidth definition (or you might wish to call it bandwidth limitation, but that somehow sounds a bit less appealing) to keep subsonic and ultrasonic signals out of the amplifier/speaker system. Subsonic signals of sufficient level will cause mechanical and possibly thermal damage to LF drivers, but even smaller amplitudes will erode precious headroom. Ultrasonic signals can burn out tweeters and damage power amplifiers.

The design of lowpass filters to stop subsonic signals and highpass filters to stop ultrasonic signals is always something of a compromise because you must steer a course between the most effective filtering and intrusion on the audio bandwidth that you want to keep. A complicating factor is the variation of opinion on the importance of phase-shifts introduced near the edges of the audio band; a hi-fi application would be more likely to set the filter cutoff frequencies further away from the audio band edges, whereas a sound reinforcement crossover would be more likely to have the filters working closer in to maximise the protection and minimise the chance of damage to expensive banks of loudspeakers.

8.27.1 Bandwidth Definition: Butterworth versus Bessel Ultrasonic Filters

The lowpass filters used to define the upper limit of the audio bandwidth are usually second-order with roll-off rates of 12 dB/octave; third-order 18 dB/octave filters are rather rarer, probably because there seems to be a general feeling that phase changes might be more audible at the top end of the audio spectrum than the bottom. Either the Butterworth (maximally flat frequency response) or Bessel type (maximally flat group delay) can be used, and this gives us an excellent opportunity to make a real-life comparison of the two types. It is unlikely that there is any real audible difference between the two types of filter in this application, as just about everything happens above 20 kHz, but using the Bessel alignment does enforce a compromise in filtering effectiveness because of its slow roll-off. I will demonstrate.

The second-order Butterworth lowpass filter in Figure 8.21a has its −3 dB point set to 50 kHz, and this gives a loss of only 0.08 dB at 20 kHz, so there is minimal intrusion into the audio band; see Figure 8.22. The response is usefully down by −11.6 dB at 100 kHz and by an authoritative −24.9 dB at 200 kHz. C1 is composed of two 2n2 capacitors in parallel.

But let us suppose we are deeply concerned about linear phase at high frequencies and we decide to use a Bessel filter with the same cutoff frequency instead. The only circuit change is that C1 is now 1.335 times as big as C2 instead of 2 times, but the amplitude response is very different. If we design for −3 dB at 50 kHz once more, we find that the response is –0.47 dB at 20 kHz; a good deal worse than 0.08 dB, and not exactly a stunning figure for your spec sheet. If we decide we can live with −0.2 dB at 20 kHz from the Bessel filter then it has to be designed for −3 dB at 72 kHz. Due to its inherently slower roll-off the response is now only down −5.6 dB at 100 Hz, and −14.9 dB at 200 kHz, as seen in Figure 8.22; the latter figure is 10 dB worse than for the Butterworth. The measured noise output for both versions is −114.7 dBu (corrected).

If we want to keep the 20 kHz loss to 0.1 dB, the Bessel filter has to be designed for −3 dB at 100 kHz, and the response is now only −10.4 dB down at 200 kHz, more than 14 dB less

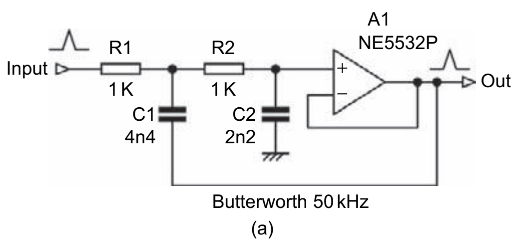


Figure 8.21: Second-order Sallen and Key lowpass circuits for ultrasonic filtering; (a) Butterworth; (b) Bessel. Both have a loss of less than 0.2 dB at 20 kHz.

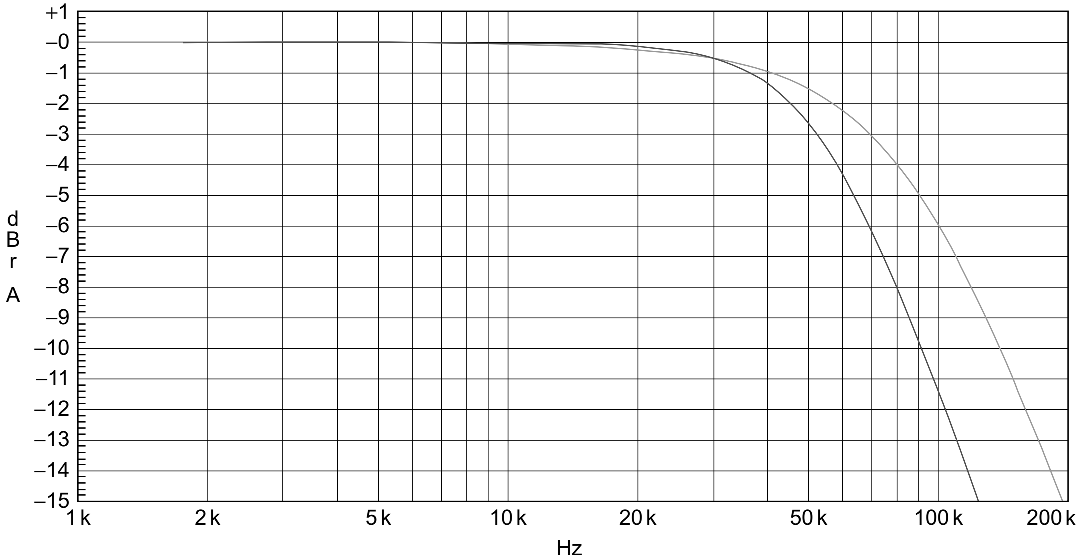


Figure 8.22: Frequency response of second-order 50 kHz Butterworth and 72 kHz Bessel filters.

Table 8.14: The Frequency Response of Various Ultrasonic Filter Options

Frequency	Butterworth 50 kHz	Bessel 50 kHz	Bessel 72 kHz	Bessel 100 kHz
20 kHz	−0.08 dB	−0.47 dB	−0.2 dB	−0.1 dB
100 kHz	−11.6 dB	−10.0 dB	−5.6 dB	−3.0 dB
200 kHz	−24.9 dB	−20.9 dB	−14.9 dB	−10.4 dB

effective than the Butterworth; this is the design shown in Figure 8.21b. These results are summarised in Table 8.14.

Discussions on filters always remark that the Bessel alignment has a slower roll-off, but often fail to emphasise that it is a *much* slower roll-off. You should think hard before you decide to go for the Bessel option in this sort of application.

It is always worth checking how the input impedance of a filter loads the previous stage, to make sure it is not loaded to the point where its distortion is significantly increased. The input impedance will vary with frequency. In this case, the input impedance is high in the passband, but above the roll-off point it falls until it reaches the value of R1, which here is 1 kΩ. This is because at high frequencies C1 is not bootstrapped, and the input goes through R1 and C1 to the low-impedance opamp output, which is effectively at ground. Fortunately, this low impedance only occurs at high frequencies, where one hopes the level of the signals to be filtered out will be low.

Another important consideration with low-pass filters is the balance between the R and C values in terms of noise performance. R1 and R2 are in series with the input and their Johnson noise will be added directly to the signal. Here the two 1 kΩ resistors generate −119.2 dBu of noise (22 kHz bandwidth, 25°C). The obvious conclusion is that R1 and R2 should be made as low in value as possible without causing excess loading (1 kΩ is a good compromise), with C1, C2 scaled to maintain the desired roll-off frequency. It is then the opamp voltage noise that dominates.

8.27.2 Bandwidth Definition: Subsonic Filters

Highpass filters used for subsonic protection are usually second-order or third-order Butterworth types with roll-offs at 12 dB/octave and 18 dB/octave. Fourth-order filters with 24 dB/octave slopes are less used, no doubt because of fears about the possible audibility of the faster phase changes generated by the steeper filter. If fourth-order filters are used the cutoff frequency is made lower to space any possible effects further away from the bottom of the audio band. I am not aware of any use of fifth- or sixth-order subsonic filters.

There is something of a consensus that a third-order Butterworth filter with a cutoff frequency of 25 Hz or a fourth-order filter with a cutoff at 15 Hz gives adequate protection. The usual Sallen & Key filters are normally used; one problem is that the low cutoff frequencies mean that large capacitor values are required if the circuit impedances are kept low to minimise noise. A third-order filter will require three, and a fourth-order filter four, of these components. While it is desirable to use polypropylene types to prevent capacitor distortion, this gets expensive, and polyester types are often used in subsonic filters, in the not unreasonable hope that the levels of subsonic disturbances will normally be low, and that very-low frequency distortion is not very audible.

Figure 8.23 shows a third-order Butterworth subsonic filter designed as a single stage. This must be fed from the usual low-impedance source to give an accurate response, but has the

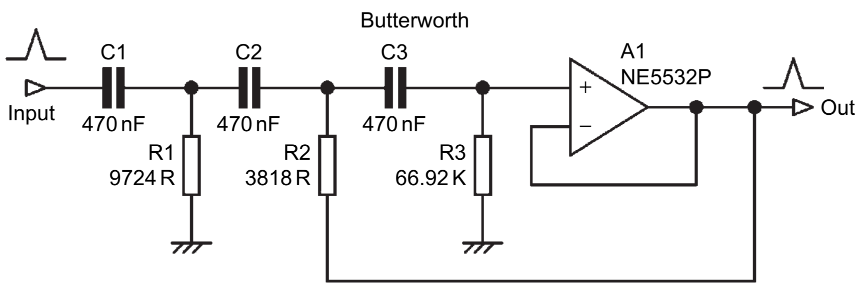


Figure 8.23: A third-order Butterworth single-stage subsonic filter with a cutoff frequency of 25 Hz.

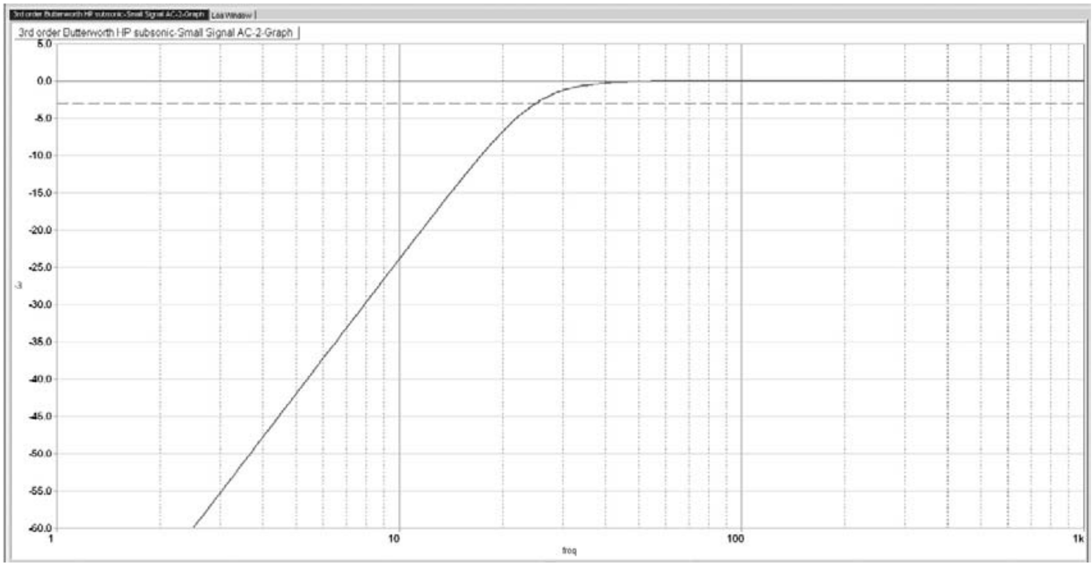


Figure 8.24: Frequency response of the third-order Butterworth single-stage subsonic filter, –3 dB at 25 Hz.

advantage over the second-order plus first-order configuration that the output impedance is low without the need for buffering by an extra opamp stage. The exact resistor values required are given.

Figure 8.24 gives the frequency response of the filter; it is 24 dB down at 10 Hz, and gives good protection against subsonic disturbances.

8.27.3 Bandwidth Definition: Combined Ultrasonic and Subsonic Filters

In some cases it is possible to economically combine highpass and lowpass filters into one Sallen & Key stage using only one opamp. I must say at once that this cunning plan is only workable when the highpass and lowpass turnover frequencies are widely different, and its usefulness for crossover filters as such is limited. However, it can be very handy when you wish to explicitly define the bandwidth of an audio signal path by using both subsonic and ultrasonic filters; a sophisticated active crossover will include such bandwidth definition. Combined filters have the advantage that the signal now passes through one opamp rather than two, which may reduce noise and distortion, and it also saves money. This approach can be extremely useful if you only have half of a dual opamp left.

Figure 8.25 shows a 20 Hz third-order Butterworth subsonic filter, much the same as that just described apart from a slightly lower cutoff frequency; note that this time E12 preferred values have deliberately been used for the resistors, to demonstrate that even so the cutoff

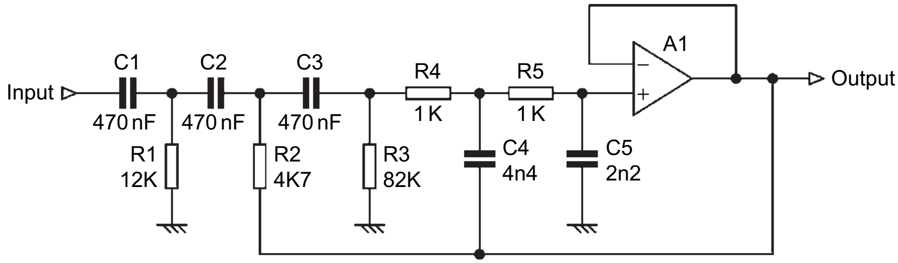


Figure 8.25: A third-order Butterworth subsonic filter combined with a second-order ultrasonic filter.

frequency comes out as 21 Hz, very close to the desired value, and such an error is not likely to be of much significance in a subsonic filter. This filter is combined with a second-order 50 kHz Butterworth lowpass ultrasonic filter, and the response of the combination is exactly the same as expected for each separately. The lowpass filter is cautiously designed to prevent significant loss in the audio band, and has a −3 dB point at 50 kHz, giving very close to 0.0 dB at 20 kHz; the response is −11.6 dB down at 100 kHz and −24.9 dB at 200 kHz. C1 is made up of two 2n2 capacitors in parallel.

The only compromise is that the midband gain of the combined filter is −0.15 dB rather than exactly unity; this tiny loss is not exactly a cause for alarm. It occurs because the series combination of C1, C2, and C3, forms a capacitive potential divider with C5, attenuating by 0.15 dB, and this is why the turnover frequencies need to be widely separated for filter combining to work well. If they were closer together then C1, C2, C3 would be smaller, C5 would be bigger, and the capacitive divider loss would be greater.

8.28 Distortion in Sallen & Key Filters: Highpass

When they have a signal voltage across them, many capacitor types generate distortion. This unwelcome phenomenon is described in Chapter 12. It afflicts not only all electrolytic capacitors, but also some types of non-electrolytic. If the electrolytics are being used as coupling capacitors, then the cure is simply to make them so large that they have a negligible signal voltage across them at the lowest frequency of interest; less than 80 mV rms is a reasonable criterion. This means they may have to be ten times the value required for a satisfactory frequency response.

However, when non-electrolytics are used to set time constants in filters they obviously must have substantial signal voltages across them and this simple fix is not usable. The problem is not a marginal one—the amounts of distortion produced can be surprisingly high. Figure 8.26 shows the frequency response of a conventional second-order

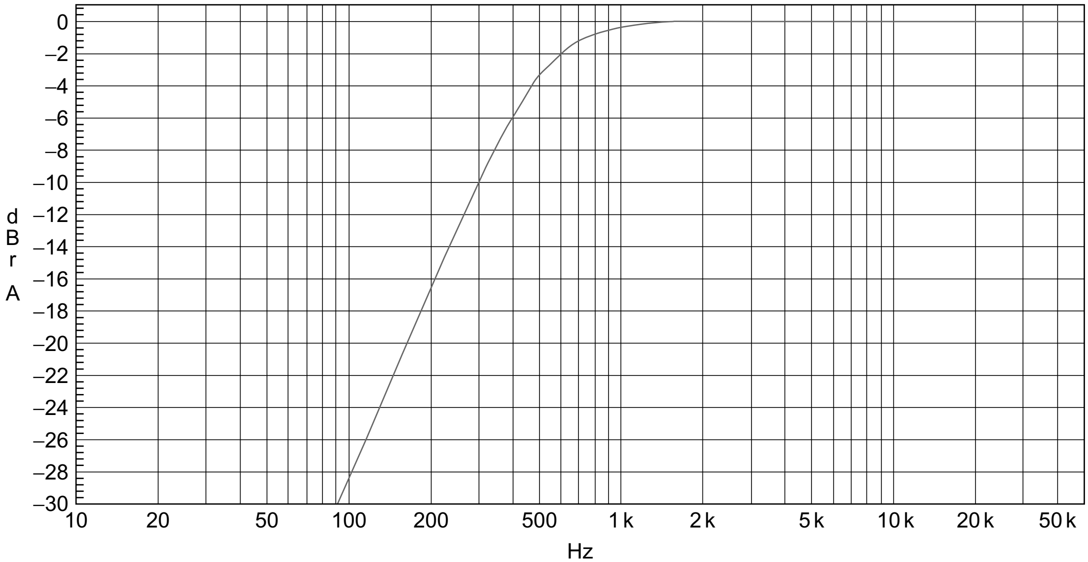


Figure 8.26: The frequency response of the second-order 520 Hz highpass S&K filter in Figure 8.26a.

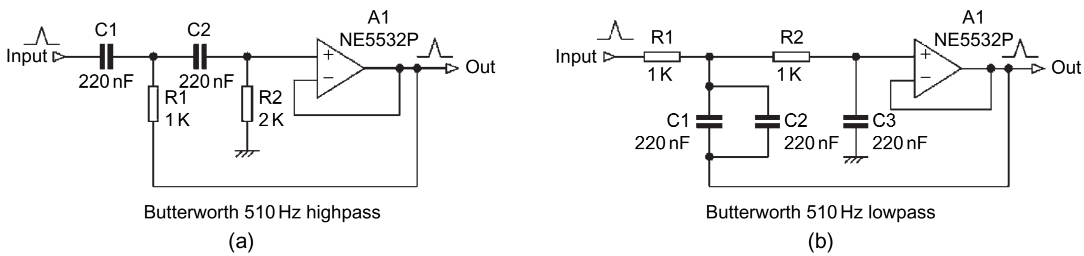


Figure 8.27: (a, b) Second-order highpass and lowpass Butterworth Sallen and Key filters for distortion and noise tests. Cutoff frequency is 510 Hz for both.

Butterworth Sallen & Key highpass filter with a –3 dB frequency of 510 Hz, as shown in Figure 8.27a. C1, C2 were 220 nF 100 V polyester capacitors, with R1 = 1 kΩ and R2 = 2 kΩ. The opamp was a Texas 5532. The distortion performance is shown by the upper trace in Figure 8.28; above 1 kHz the distortion comes from the opamp alone and is very low. However, you can see it rising rapidly below 1 kHz as the filter begins to act, and it has reached 0.015% by 100 Hz, completely overshadowing the opamp distortion; it is basically third order. The output from the filter has dropped to –28 dB by 100 Hz, and so the amplitude of the harmonics generated is correspondingly lower, but it still not a very happy outcome.

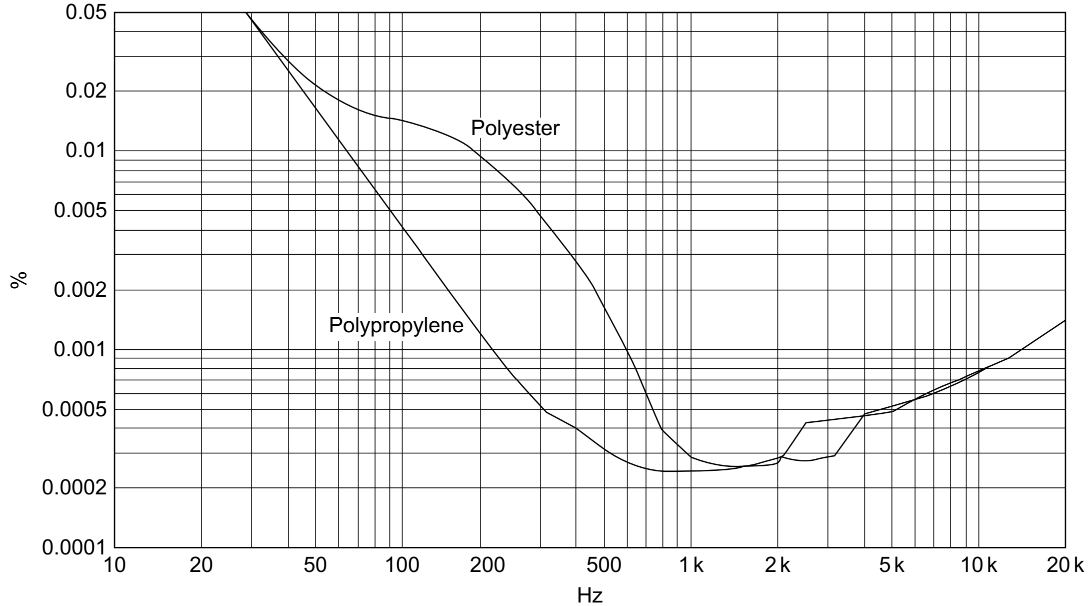


Figure 8.28: THD plot from the second-order 520 Hz high-pass S&K filter; input level 10 Vrms. The upper trace shows distortion from polyester capacitors; the lower trace, with polypropylene capacitors, shows noise only.

The test level used is high at 10 Vrms. It is chosen to be about as much voltage swing as you are likely to encounter in an opamp system, so it will give worst-case distortion results, but will be clear of any clipping effects. Most of the distortion tests in this book are run at 9 or 10 Vrms for this reason. The actual operating levels will naturally be significantly lower, to give some headroom, and the distortion levels produced will therefore be much lower. This is especially true in the case of capacitor distortion, which is essentially all third harmonic, and so the THD increases as the square of the signal voltage. (Second-harmonic THD increases more slowly, proportionally to the signal voltage.)

Thus, if the test circuit here was run at a nominal voltage of 3 Vrms, which is probably as high as is advisable (see the discussion in Chapter 14 on elevated operating levels), then the distortion levels would be lower by a factor of $(10/3)^2 = 11.1$ times. This is a big reduction, and explains why polyester capacitors are in practice acceptable in many applications where the highest possible quality is not being sought.

As explained in Chapter 12, polypropylene capacitors exhibit negligible distortion compared with polyester, and the lower trace in Figure 8.28 shows the improvement on substituting 220 nF 250 V polypropylene capacitors. The THD residual below 500 Hz is now pure noise, and the trace is only rising at 12 dB/octave because circuit noise is constant but the filter output is falling. The important factor is the dielectric, not the voltage rating; 63 V polypropylene capacitors are also free from distortion. The only downside is that polypropylene capacitors are larger for a given CV product and more expensive.

The non-linearity of polyester capacitors, even those of the same type and voltage rating, appears to be rather variable, and if you repeat this experiment the results for the upper trace may be different, but will always be much inferior to the polypropylene case. A further complication is that the non-linearity is time-dependent; if you set up a polyester capacitor in a simple RC circuit, with the frequency arranged so that there is a substantial voltage across the capacitor, the THD reading will slowly drift down over time, continuing to drop over 24 hours and longer. It will, however, never become as distortion-free as a polypropylene capacitor.

8.29 Distortion in Sallen & Key Filters: Lowpass

The capacitor-induced distortion behaviour of the lowpass Sallen & Key filter is rather different. The frequency response of the Butterworth 510 Hz lowpass filter in Figure 8.27b is shown in Figure 8.29, while the distortion performance is shown in Figure 8.30. With 220 nF 100 V polyester capacitors there is now a pronounced peak in distortion around

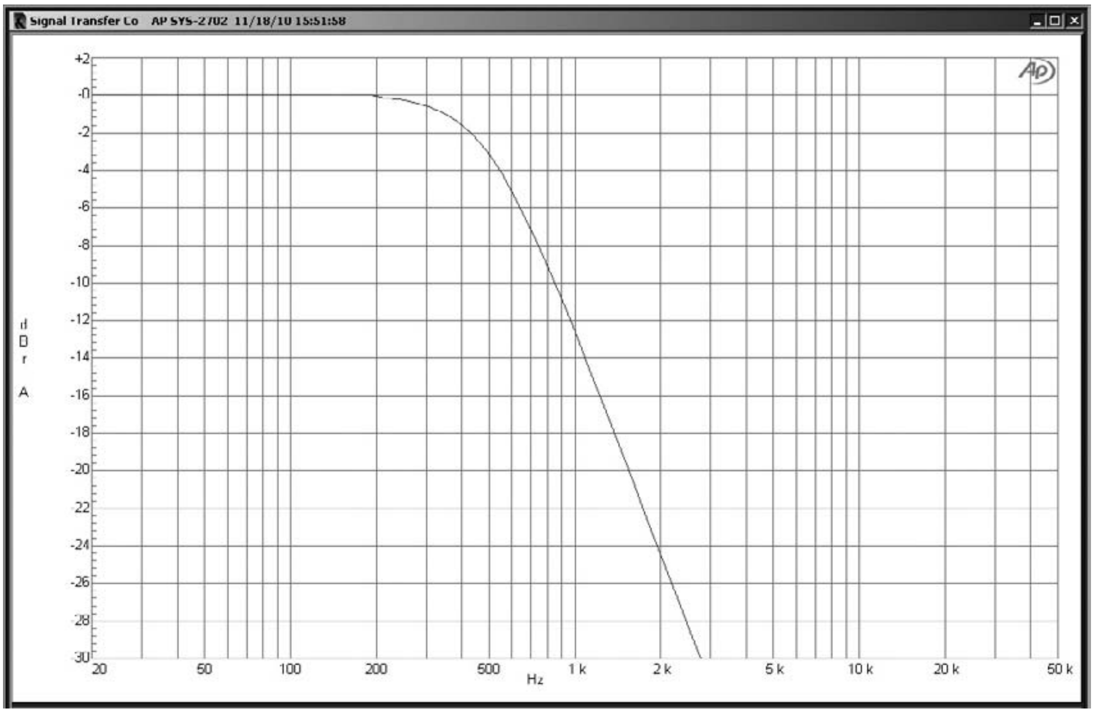


Figure 8.29: The frequency response of the second order 510 Hz lowpass S&K filter in Figure 8.27b.

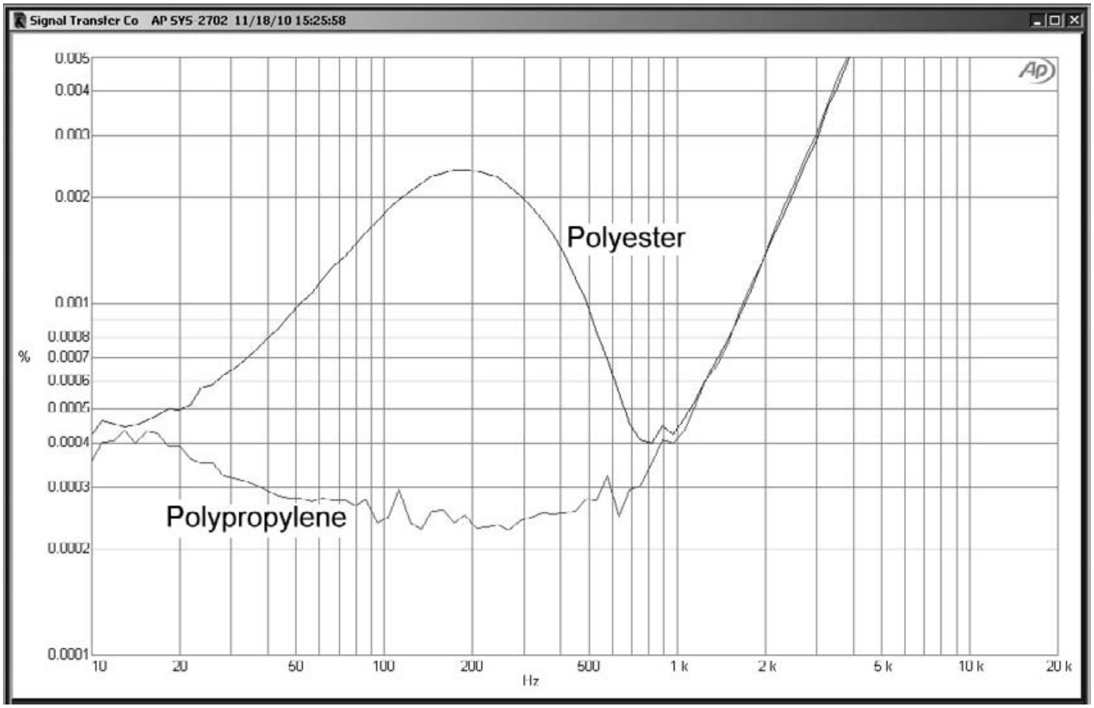


Figure 8.30: THD plot from the second-order 510 Hz S&K lowpass filter; input level 10 Vrms. The upper trace shows distortion from polyester capacitors; the lower trace, with polypropylene capacitors, shows noise and opamp distortion above 1 kHz.

200 Hz, more than an octave below the cutoff frequency. The distortion in this region is fairly pure third harmonic, and clearly comes from the capacitors and not the opamp. Replacing them with 220 nF 160 V polypropylene capacitors completely eliminates this distortion, as shown in Figure 8.30. Note that as before the test level is almost as high as possible at 10 Vrms, and practical internal levels such as 3 Vrms will give much lower levels of distortion.

As expected, the Noise + THD reading climbs rapidly as the cutoff frequency is exceeded and the output amplitude starts to fall at 12 dB/octave; in fact the Noise + THD reading climbs more rapidly than 12 dB/octave, which is a clear warning sign that the residual is not entirely composed of noise. At 1 kHz the reading is only 0.00037% but crossover-ish distortion from the opamp is clearly visible on the THD residual, well above the noise level. This condition persists as frequency rises, with opamp distortion still clear above the noise even at 10 kHz when the output has fallen below 24 mVrms. (This is quite different behaviour from that of the highpass filter where below the cutoff frequency the Noise + THD reading goes up at the 12 dB/octave rate, which would be expected if it was composed only of noise.) Removing the capacitors to get a flat frequency response greatly reduces the distortion seen at the higher

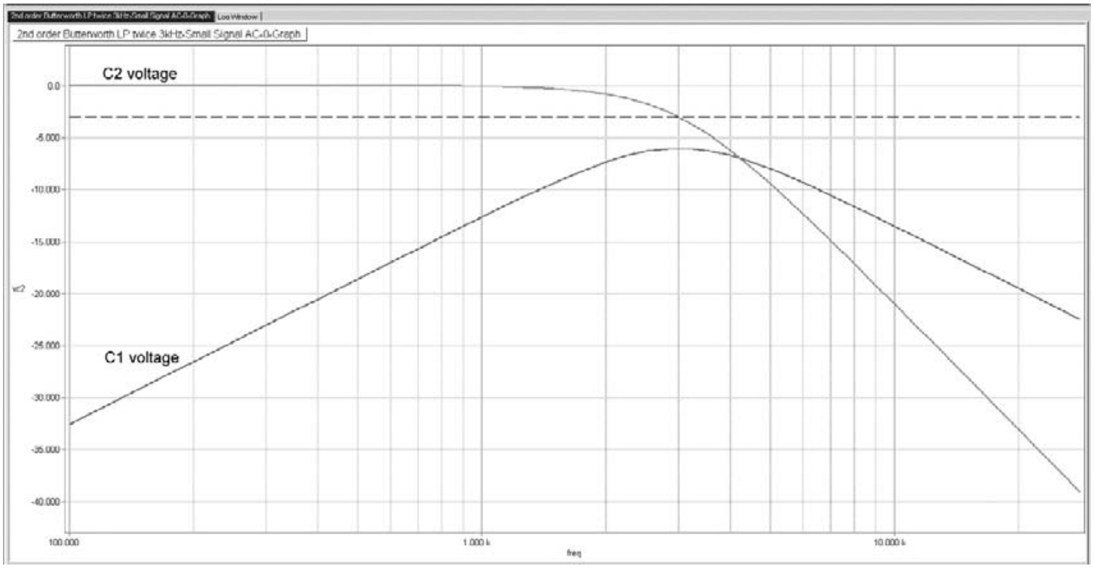


Figure 8.31: The voltages across the two capacitors in a Sallen and Key Butterworth lowpass filter with a 3 kHz turnover frequency. The voltage across C1 never reaches more than half the filter input voltage.

frequencies, even though the output level is much higher, so the effect appears to be something to do with the filtering action. These are deep waters, Watson.

The signal voltage across the capacitors in these filters varies strongly with frequency, and it is instructive to see how this relates to their capacitor-distortion behaviour. In a lowpass Sallen & Key filter the voltage across C1 peaks at one half of the output voltage at the turnover frequency, rolling off on either side at 6 dB/octave. The voltage across C2 is the input voltage up to the turnover frequency; it then rolls off at 12 dB/octave, as illustrated in Figure 8.31. Perhaps surprisingly, the capacitor voltages for the equivalent highpass filter are exactly the same, but with C1 and C2 swapped over.

What Figure 8.31 does not tell us is why the polyester capacitor distortion peaks more than an octave below the cutoff frequency; that is *not* where the signal voltage across the capacitors is a maximum. It is both one of the joys and one of the anguishes of electronics that a circuit made up of only five components can be so enigmatic in its behaviour.

8.30 Mixed Capacitors in Low-Distortion Sallen & Key Filters

But you have seen nothing yet. If you take the lowpass test filter, and change the capacitor dielectric type one at a time, (see Figure 8.27) you find that *only C2 has to be polypropylene for low distortion*. If the filter has polyester in both the C1 and C2 positions, then you get the

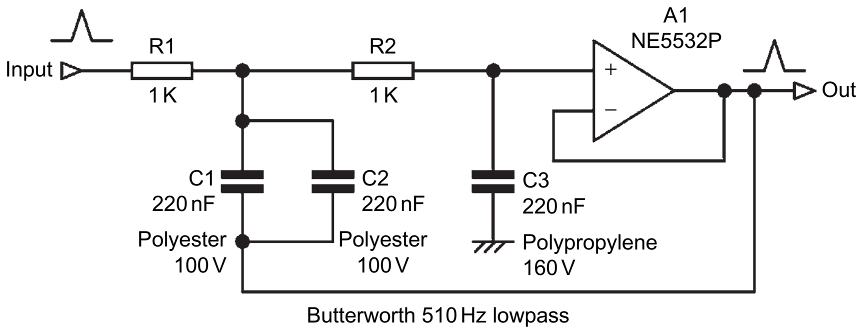


Figure 8.32: Mixed-capacitor Sallen and Key lowpass filter has same distortion performance as if all capacitors were polypropylene.

same peaked distortion trace seen in Figure 8.30; but if you replace just C2 with polypropylene, then the distortion performance is identical to the bottom trace obtained by making both capacitors polypropylene. This is, to the best of my knowledge, a novel observation, and it has the potential to much reduce capacitor costs in an active crossover; you will note it is the smaller of the two capacitors (C3 in Figure 8.32) which has to be the more expensive polypropylene type, which is just the way we would like it for economy.

This is clearly a most interesting effect, and I attempted to find out why the lowpass Sallen & Key circuit is so much more sensitive to the non-linearity of C2 than that of C1. Firing up the SPICE simulator, and putting an AC voltage source in series with C1 shows a maximum gain of unity from the inserted source to the output at its peak (at the turnover frequency), falling off at 6 dB/octave on either side. The same test on C2 shows there is actually a gain of 2 dB to the output just above the turnover frequency. This appears to give some good indication as to why C2 is the more critical component. It has a greater voltage across it, and the resulting distortion components appear amplified at the output. This does not, however, really answer the question, because repeated practical experiments show that using polyester capacitors for C1 does not produce a reduced amount of distortion—it produces none at all. More convincing results might be obtained by altering the capacitor models so their value was voltage-dependent, but there is regrettably no space to go any further into this issue here.

But what, you may ask, about the highpass filter? The components in the equivalent circuit positions to the capacitors are now resistors, which almost always have perfect linearity for our purposes, so there is little point in playing around with them. However, if you start swapping out polyester capacitors for polypropylene ones, it soon emerges that in this case C1 is the critical component for linearity. C2 has a very minor influence, but unless it is really bad, polypropylene for C1 only gives results very close to those obtained when both capacitors are polypropylene; the configuration is shown in Figure 8.33. Just think of the money it saves!

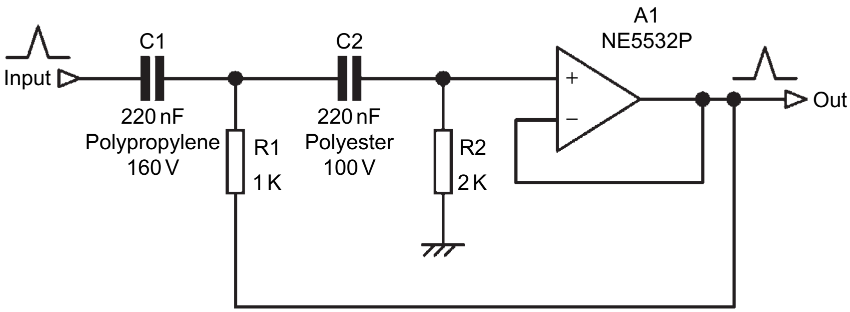


Figure 8.33: Mixed-capacitor Sallen and Key highpass filter has very nearly the same distortion performance as if both capacitors were polypropylene.

8.31 Noise in Sallen & Key Filters: Lowpass

It is difficult to give a comprehensive summary of the noise performance of even one type of filter in a reasonable space. There are a large number of variables, even if we restrict ourselves to the Sallen & Key configuration. There is the cutoff frequency, the Q, the impedance level at which it operates, whether it is highpass or lowpass, and whether it is second-order or higher.

If we start with the 510 Hz lowpass filter in Figure 8.27b above, it can be built as shown with 1 kΩ resistors and 220 nF capacitors (using two of them for C1), in which case the noise output (corrected by subtracting the internal noise of the measuring system, as are all noise data in this section) is −117.8 dBu, which is pretty low. The measurement bandwidth was 22 – 22 kHz as usual. If we rebuild our lowpass filter with 10 kΩ resistors and 22 nF capacitors, so the cutoff frequency remains 500 Hz but the impedances are all ten times as great, we rather disconcertingly get the same figure, to within the limits of measurement, which are here about ±0.1 dB. This is because the cutoff frequency is towards the lower end of the audio bandwidth, and so the Johnson noise from the resistors never makes it to the filter output. What we are seeing is the basic voltage noise of the opamp, which is not altered by changing circuit impedances. The only way to reduce it is to choose a quieter opamp or to use multiple opamps for noise reduction, as explained in Chapter 16. This at first seems to make nonsense of the recommendations throughout this book that resistance levels should be kept as low as practicable to minimise noise.

If, however, we are dealing with the same lowpass filter but with a ten times higher cutoff frequency of 5.1 kHz, things are very different. Our low-impedance version has 1 kΩ resistors and 22 nF capacitors; it produces −115.8 dBu of noise at its output; noticeably higher because of the higher cutoff frequency. If we once more increase the impedance level by ten times, by using 10 kΩ resistors and 2n2 capacitors, then the output noise increases to

−110.9 dBu, a sizable increase of 4.9 dBu. Clearly the higher the cutoff frequency of a lowpass filter, the greater the noise advantage to be gained by using low impedances. Raising the impedance ten times also makes the circuit noticeably more susceptible to electrostatically induced hum.

8.32 Noise in Sallen & Key Filters: Highpass

Turning our attention back to the highpass Butterworth filter, the noise output for a low-impedance version with R1 = 1 kΩ, R2 = 2 kΩ, and 220 nF capacitors is a very low −119.3 dBu. Raising the impedance level by ten times, using R1 = 10 kΩ, R2 = 20 kΩ, and 22 nF capacitors raises the noise output to −116.6 dBu, an increase of 2.7 dB. The low impedance approach is therefore well worthwhile, though the capacitors will cost a bit more.

8.33 Multiple-Feedback Filters

These are most familiar as bandpass filters working at modest Q's, but the basic configuration can also be used to make lowpass and highpass filters; the variations are shown in Figure 8.34.

Multiple-FeedBack (MFB) filters have the advantage that since they use shunt-feedback, with a virtual-earth at the inverting input, there is no common-mode voltage to cause distortion in the opamp. This is in contrast with the Sallen & Key filter where the opamp is working as a voltage follower, and so the full signal voltage appears at the opamp inputs; it is in fact the worst case for common-mode distortion. This potential problem is dealt with at length in Chapter 13, but it is worth pointing out here that if you use bipolar opamps such as the 5532 and relatively low source impedances, common mode distortion is not a serious difficulty.

While it is possible to make third-order highpass and lowpass filters in a single MFB stage, component sensitivity is much increased and it is probably not a good idea. I have never seen such filters used in a practical design.

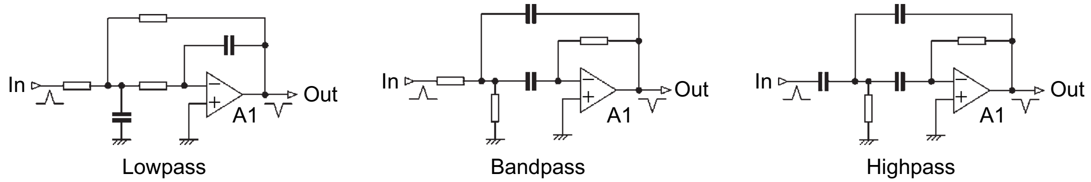
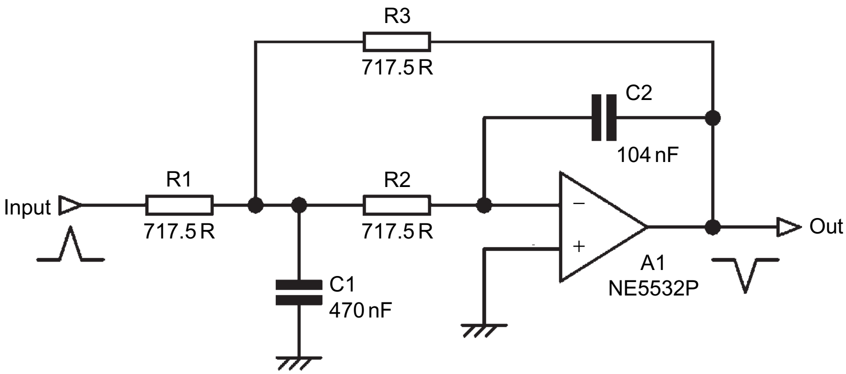


Figure 8.34: Multiple-feedback (MFB) filters: lowpass, bandpass, and highpass.

8.34 Multiple-Feedback Lowpass Filters

A practical version of an MFB lowpass filter is shown in Figure 8.35, designed for a Butterworth characteristic, (Q = 0.7071) and a cutoff frequency of 1 kHz; the passband gain is unity. Note that if C1 is a preferred value, C2 will in general not be; the value of 104 nF shown here would in practice be approximated by 100 nF in parallel with 4n7. As usual the resistor values do not work out conveniently, but very close approximations to these values can be cheaply made up using pairs of preferred values. MFB bandpass, highpass, and lowpass filters all inherently give a phase-inversion. This is no problem if they are used in pairs in a fourth-order Linkwitz–Riley crossover, but otherwise could be inconvenient, in the worst case requiring another stage that does nothing but invert the signal to get it in-phase again.

It is worth pointing out that the MFB lowpass filter does not depend on a low opamp output impedance to maintain stop-band attenuation at high frequencies, and so avoids the oh-no-it's-coming-back-up-again behaviour of Sallen & Key lowpass filters. It is doubtful, however, if this is of much relevance to crossover applications.



$$f_{-3\,dB} = \frac{1}{2\pi \sqrt{\frac{C2}{C1}}\, R\, C1}$$

$$Q = \frac{\sqrt{\frac{C1}{C2}}}{3}$$

Figure 8.35: A second-order lowpass Butterworth multiple-feedback (MFB) filter with the analysis equations. Cutoff frequency is 1 kHz, Q = 0.707, and passband gain is unity.

8.35 Multiple-Feedback Highpass Filters

The multiple-feedback highpass filter is the multiple-feedback lowpass filter with the resistors an capacitors interchanged. A practical version of an MFB highpass filter is shown in Figure 8.36, once more designed for a Butterworth characteristic (Q = 0.7071), a cutoff frequency of 1 kHz, and unity passband gain.

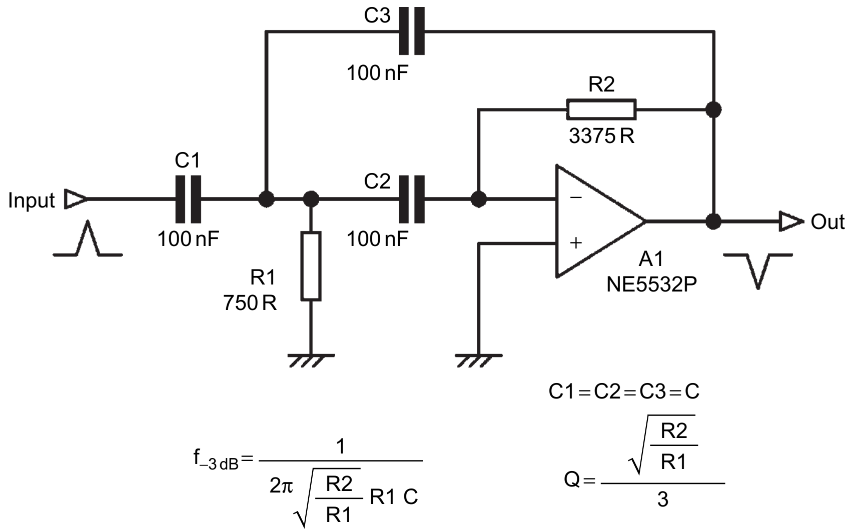


Figure 8.36: A second-order highpass Butterworth multiple-feedback (MFB) filter with the analysis equations. Cutoff frequency = 1 kHz, Q = 0.707, and passband gain is unity.

This time we have three identical capacitors which can be conveniently chosen from the E6 series, dealing with the awkward resistor values in the usual way; as it happens, in this case R1 comes out as the E24 value of 750 Ω.

8.36 Distortion in Multiple-Feedback Filters: Highpass

Figure 8.37 shows lowpass and highpass Butterworth MFB filters, each with a cutoff frequency of 510 Hz, and designed to work at a low impedance level. These filters were designed to allow direct comparison between them and the Sallen and Key 510 Hz filters of Figure 8.27 that were examined for distortion and noise earlier in this chapter.

Figure 8.38 demonstrates that replacing polyester capacitors with polypropylene once again gives a dramatic reduction in distortion. However, there were some unexpected issues with the distortion performance above 3 kHz which were not resolved at the time of going to press, and at the moment the verdict has to be that the highpass MFB filter is not to be designed into a system without careful scrutiny. In contrast, the lowpass MFB filter works very well and gives less opamp distortion than the Sallen and Key equivalent, as revealed in the next section.

You will recall that we discovered that it was in fact only necessary to make one capacitor polypropylene in the Sallen and Key lowpass and highpass filters, and capacitor distortion

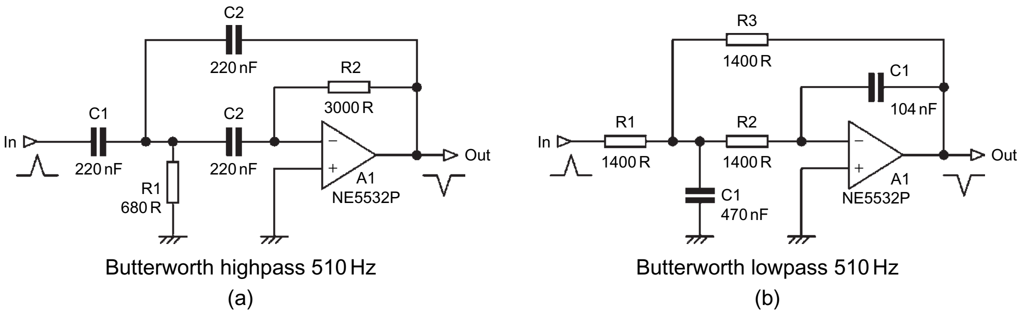


Figure 8.37: Second-order Butterworth highpass and lowpass MFB filters for distortion and noise tests. Cutoff frequency is 510 Hz for both.

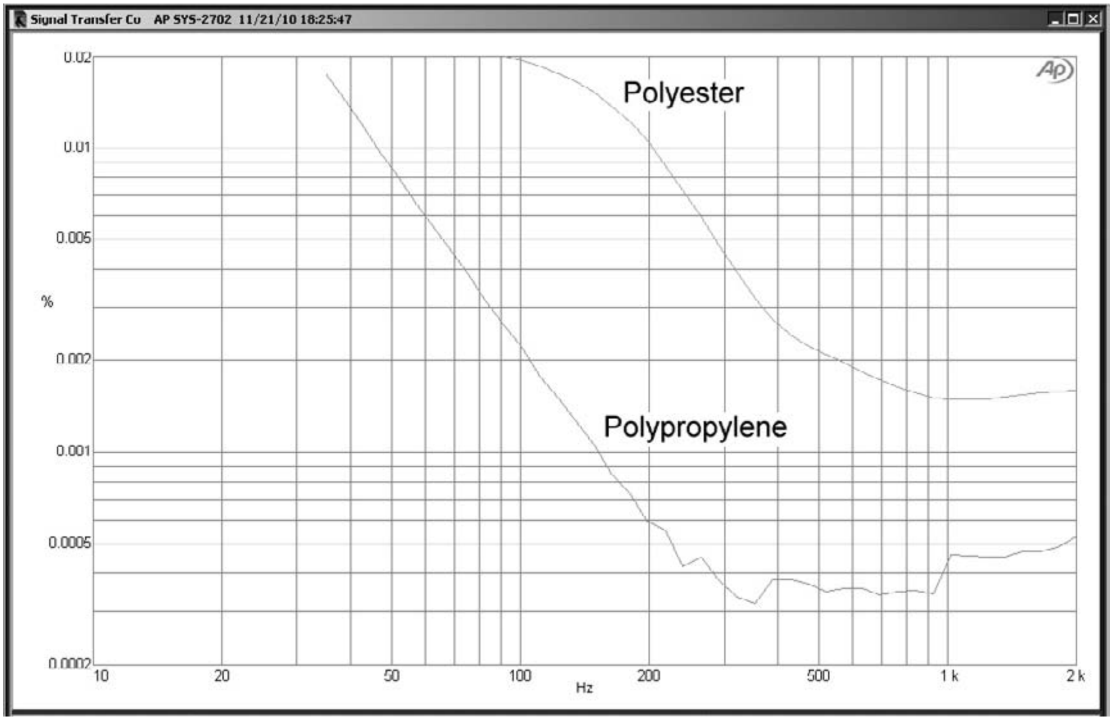


Figure 8.38: THD plot from the second order 510 Hz MFB highpass filter; input level 10 Vrms. The upper trace shows distortion from polyester capacitors; the lower trace, with polypropylene capacitors shows only noise.

was still eliminated. Obviously, we need to see if the same applies to the MFB filter. I was not optimistic, because of the different ways in which the two kinds of filters work, but it does more or less work. In the highpass case only C1 and C3 need to be polypropylene, both capacitors contributing distortion. C2, however, can be a cheaper and smaller polyester type.

8.37 Distortion in Multiple-Feedback Filters: Lowpass

Figure 8.39 shows the distortion performance of the second-order 510 Hz MFB lowpass filter with an input level of 10 Vrms. As for the Sallen & Key lowpass filter, the use of polyester capacitors causes a broad peak of extra distortion, though here it is centred on 240 Hz rather than 200 Hz. As before, the use of polypropylene capacitors eliminates the extra distortion.

Above 1 kHz, the Noise + THD reading climbs rapidly as the cutoff frequency is exceeded and the output amplitude starts to fall at 12 dB/octave; but this time it rises at 12 dB/octave because it is composed of noise only; some opamp distortion is visible around 100–500 Hz when polypropylene capacitors are used. Comparing Figure 8.39 with the graph for the Sallen & Key version (Figure 8.29), here we have a Noise + THD reading of only 0.0008% compared with 0.0015% for S&K. The MFB lowpass filter has the better linearity.

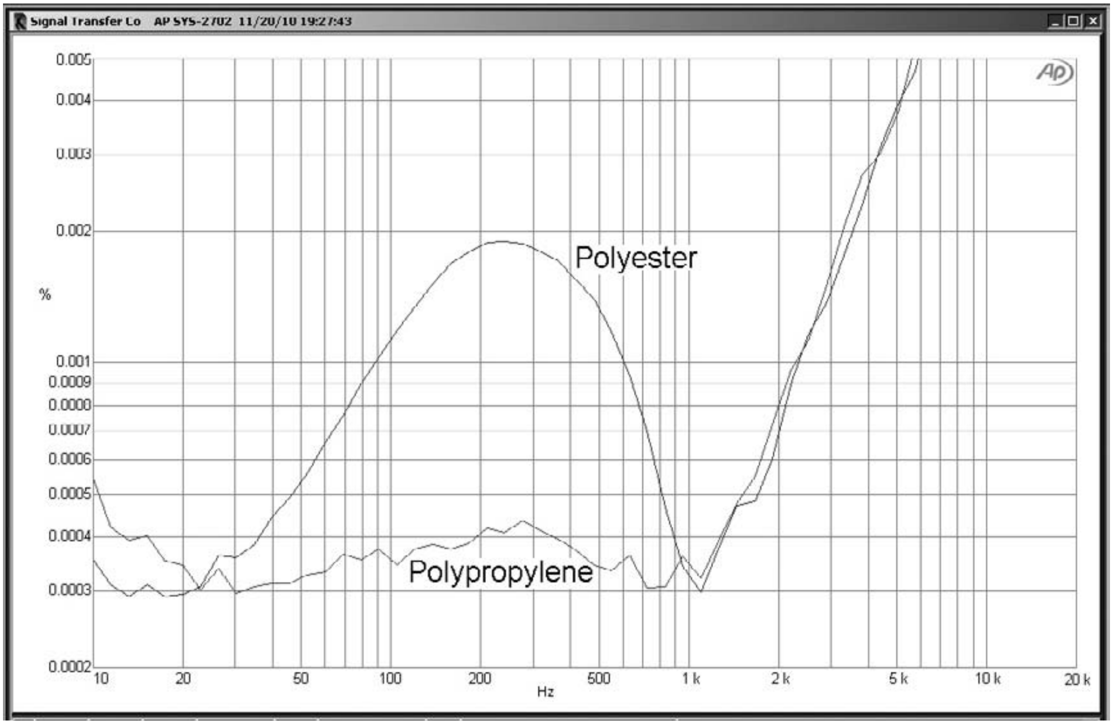


Figure 8.39: THD plot from the second-order 510 Hz MFB lowpass filter; input level 10 Vrms. The upper trace shows distortion from polyester capacitors; the lower trace, with polypropylene capacitors, shows only noise above 1 kHz. Unlike the S&K filter, there is no visible opamp distortion.

You will recall that we discovered that the full benefits of capacitor distortion elimination could be obtained in the Sallen & Key second-order lowpass filter by replacing only the second of the two capacitors with a polypropylene type. In the MFB lowpass filter only C2 needs to be polypropylene to get the full elimination of capacitor distortion. Extraordinary!

8.38 Noise in Multiple-Feedback Filters: Highpass

With the values shown in Figure 8.37a, the noise output was −112.8 dBu, which is substantially worse than that of the Sallen and Key highpass filter, which was −119.3 dBu. This is not promising, and reinforces the judgement that the highpass MFB filter is not to be recommended.

8.39 Noise in Multiple-Feedback Filters: Lowpass

The low-impedance MFB lowpass filter in Figure 8.37b has a noise output of −117.8 dBu, which is quite satisfyingly low. As we did for the Sallen and Key filters, we will raise the circuit impedance level by ten times and see what effect that has on the noise performance. Plugging in R1 = R2 = R3 = 14 kΩ, and C1 = 47 nF, C2 = 10.4 nF, the noise output rises to −114.0 dBu, an increase of 3.8 dBu. (As before, the measurement bandwidth was 22 – 22 kHz and the readings were corrected by subtracting the internal noise of the measuring system.) This is quite different behaviour from the Sallen and Key lowpass filter, which showed no measurable change in noise output with a 510 Hz cutoff frequency. Once again, raising the impedance ten times makes the circuit noticeably more susceptible to electrostatically induced hum.

The low-impedance version of the MFB lowpass filter has a noise output of −117.8 dBu, which, interestingly, is exactly the same noise output as the Sallen and Key lowpass filter examined earlier. From the noise point of view, there is nothing to choose between the two filter configurations.

8.40 State-Variable Filters

State-variable filters (SVFs) give highpass, bandpass, and lowpass outputs simultaneously. In crossover applications the bandpass output is not normally used, but might be employed in a filler-driver scheme. These filters show low component value sensitivity and are simple to design. They are called "state-variable" filters because in the most common second-order version they are made up of two integrators and an amplifier; signals from all three stages are used for feedback (the output of each integrator and the output of the amplifier) and completely define the state of the circuit. There are several variations on the basic configuration, but the version shown in Figure 8.40 is probably the most straightforward. As

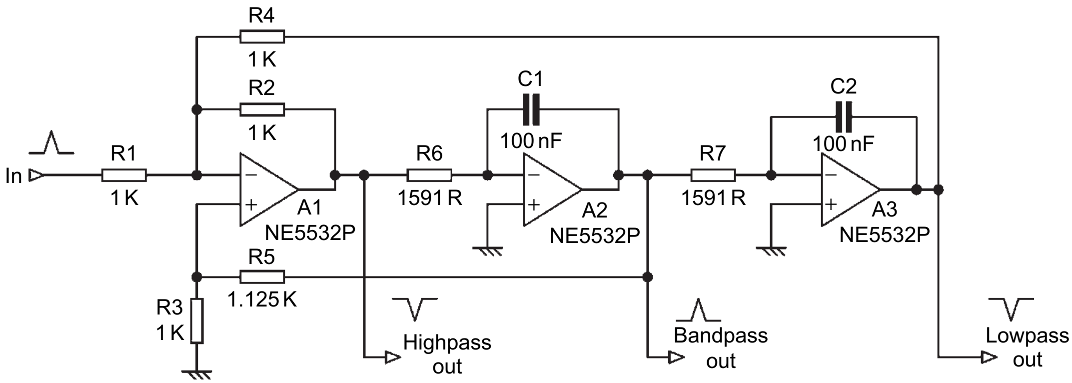


Figure 8.40: Second-order state-variable Butterworth filter. Set R1 = R2 = R3 = R4, R6 = R7, and C1 = C2. Centre frequency is 1 kHz.

shown, it has unity gain in the passbands of the highpass and lowpass outputs, and a gain of −3 dB at the passband centre of the bandpass output. The highpass output is phase-inverted in the passband, the bandpass output is in-phase in the centre of the passband, and the lowpass output is phase-inverted in the passband. The presence of two integrators shows that it is a second-order filter. Higher-order state variable filters are perfectly possible; for example, a fourth-order state variable filter has four integrators. Higher order SVFs are dealt with later in this chapter.

The amplitude responses of the three outputs can be seen in Figure 8.41, where the –3 dB points of the highpass and lowpass filters coincide with the band centre of the bandpass output at 1 kHz. The Q is set at 0.7071 to give a Butterworth filter characteristic at the lowpass and highpass outputs, such as might be used in a second-order crossover application. This low Q gives a very broad bandpass characteristic with no peaking at the centre. The lowpass and highpass outputs both have an ultimate slope of 12 dB/octave, as we would expect from a second-order filter, while the bandpass output has slopes of 6 dB/octave on each side of its peak. Since the cutoff frequencies of the highpass and lowpass filters are inherently the same, this is obviously a very convenient way to make a second-order crossover in one handy stage. This does, however, mean that it is not possible to employ frequency offset between the highpass and lowpass filters.

The state-variable filter may look more complex than the Sallen & Key or the MFB filter, but it is in fact delightfully simple to design. Choose a resistor value R = R1 = R2 = R3 = R4, and choose a capacitor value C = C1 = C2.

The required centre frequency *f* is then used to calculate R6 and R7:

$$R6 = R7 = \frac{1}{2\pi fC} \tag{8.21}$$

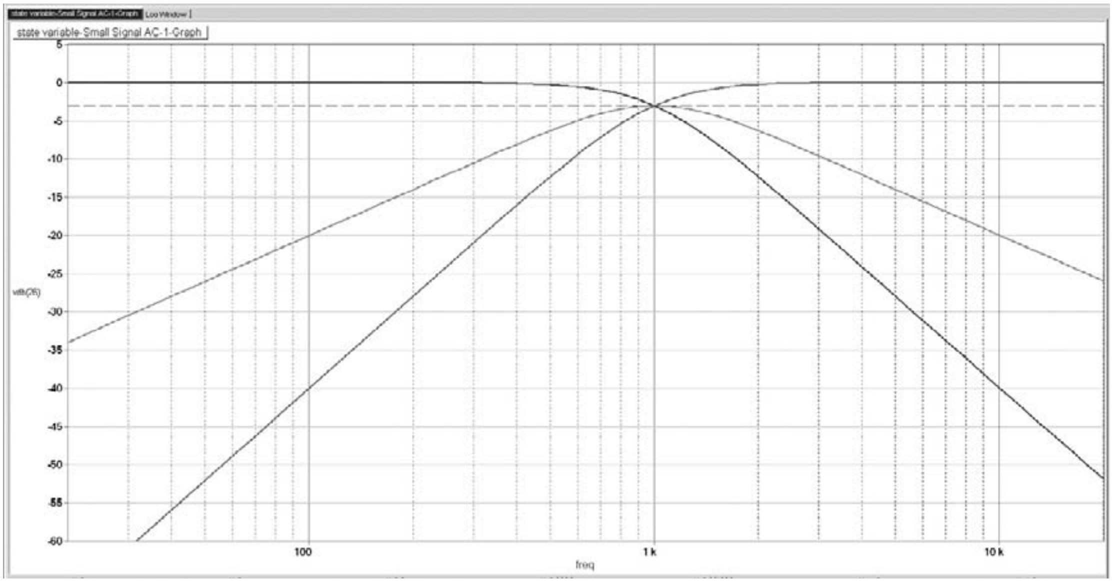


Figure 8.41: Amplitude response of the three outputs from the state-variable Butterworth filter, centre frequency 1 kHz.

Table 8.15: R5 Scaling Factor and Frequency Scaling Factor for Filter Characteristics

Filter Type	R5 Scaling Factor	Frequency Scaling Factor
Bessel	0.575	1.273
Butterworth	1.125	1.000
3 dB-Chebyshev	3.47	0.841

and Q is set by selecting an appropriate value for R5. For Butterworth it is 1.125 times the value of R.

To obtain other filter characteristics, it is necessary to set a different value of Q and use a frequency scaling factor to get the desired cutoff frequency, as we did with other forms of filter. So, to get a Bessel characteristic with a cutoff frequency R5 is set to 0.575 times R, and Equation 8.21 uses 1.273 times f as its input. The necessary values for the Q and scaling factor are given in Table 8.15.

You will note that the lower the value of R5, the lower the Q; this is because the feedback path from the first integrator controls the damping of the circuit; the more feedback, the more damping.

A notch output from a state-variable filter, with its centre at the same frequency as the bandpass output, can be obtained by the "1-bandpass" principle. The bandpass output is subtracted from the input signal by a fourth opamp. This is a relatively complicated way of making a notch filter, and is not usually justified unless the ease of changing the centre frequency which comes with a state-variable filter is important.

8.41 Variable-Frequency Filters: Sallen and Key

Active crossovers for sound-reinforcement applications commonly have variable crossover frequencies so they can be used with a wide range of loudspeaker systems. This presents problems when the usual Sallen & Key filters are used. A typical variable Sallen & Key lowpass filter is shown in Figure 8.42, where ganged variable resistors alter the cutoff frequency over a 10:1 range from 100 Hz to 1 kHz, such as might be used for the LF-MID crossover point. R1 and R2 are end-stop resistors to limit the frequency range at the upper end.

The variable resistors are normally of the reverse-log law so that the frequency calibration is approximately linear in octaves. One problem with this circuit is that it relies on the matching of the resistance values of the variable resistor tracks, and also matching of the value obtained at a given degree of rotation, which is worsened by the need for a log law. This gives rise to errors in the cutoff frequency; the effect on the filter Q is much less. Be aware that in Figure 8.42 no precautions have been taken to deal with the opamp input bias current flowing through the variable resistors. This could lead to noises as the controls are altered unless suitable DC-blocking capacitors are added.

A more serious difficulty is the very high degree of control-ganging that is required in a practical crossover. If we assume the filter in Figure 8.42 is being used in a fourth-order

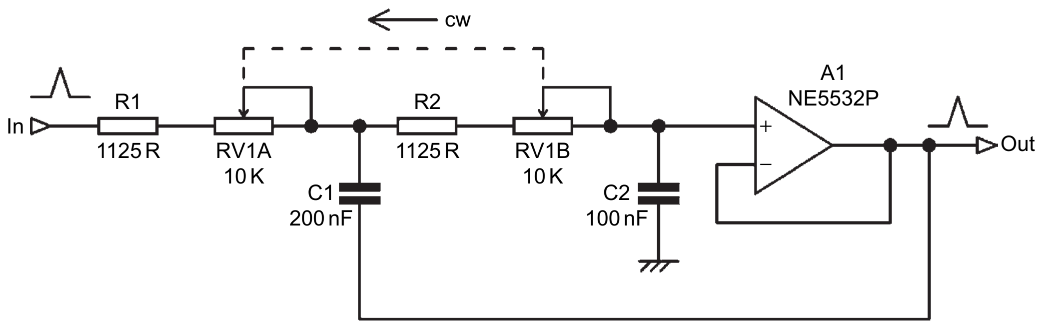


Figure 8.42: Variable-frequency second-order Sallen and Key filter, cutoff variable from 100 Hz to 1 kHz.

Linkwitz–Riley crossover, which will very often be the case, then to vary the frequency of the two cascaded Butterworth second-order filters in one crossover path we need a four-gang part, which is obtainable but costly compared with a two-gang. Since we need to vary the frequency of the another path simultaneously (to change the LF-MID crossover point we must alter the frequencies in both the LF and MID paths), then we must procure an eight-gang control, which is not only going to be relatively hard to get and expensive, but will also probably have a poor control feel due to the increased friction. Fortunately, this is the sort of control that is rarely altered so that is not the issue that it might be on a hi-fi preamplifier volume control.

We now pause and think nervously of the need to control a stereo crossover. That implies a sixteen-gang control, which is not really a practical component to specify in a design. Matters are even worse if we try to use Multiple FeedBack (MFB) filters. The highpass version still requires two resistors to be altered, while the lowpass demands three.

Lowpass filters similar to that in Figure 8.42 are most commonly used in mono subwoofer crossovers where only a single second-order filter is required, and so an easily-obtained dual control is all that is necessary.

8.42 Variable-Frequency Filters: State-Variable Second Order

The number of gangs in a crossover frequency control can be halved by abandoning the Sallen & Key configuration, and instead using state-variable filters that produce outputs for two paths simultaneously; this technology is widely if not universally used in variable-frequency crossovers. We saw earlier that two variable resistors control the cutoff frequencies of both the lowpass and highpass outputs simultaneously.

Figure 8.43 shows a variable-frequency, second-order, state-variable filter based on the design of Figure 8.40 above. The frequency-control method is slightly more sophisticated than that of the Sallen & Key filter; it aims to reduce the effect of control mismatches. The variable resistors are now being use as potentiometers, so to a first approximation the effect of differing total track resistances will cancel out. To make this effective the loading on the potentiometer needs to be light relative to the track resistance, and you will observe that R6, R7 have been increased in value by ten times, while C1 and C2 have been reduced in value by ten times, to raise the impedances without altering the centre frequency. This will of course have some consequences in terms of increased noise. Log controls can be used to get a suitable frequency calibration law.

Altering the potentiometer setting alters the effective value of R6 and R7, and so the filter centre frequency. R8 and R9 are end-stop resistors to limit the frequency range at the lower end. Note that the bandpass output of A2 is not used.

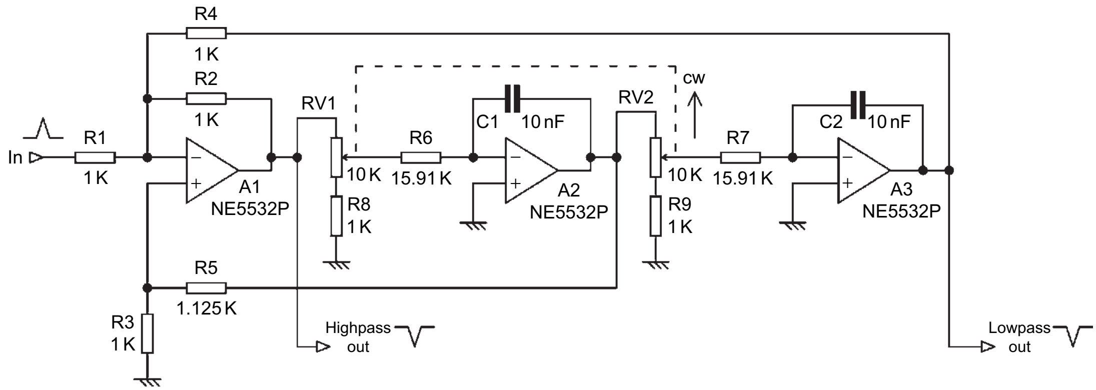


Figure 8.43: Variable-frequency, second-order, state-variable filter, with crossover frequency variable from 85 Hz to 1 kHz.

8.43 Variable-Frequency Filters: State-Variable Fourth Order

While the second-order variable-frequency filter works very nicely, second-order crossovers are not exactly the most popular choice, because of their relatively shallow slopes and considerable band overlap. What is much more desirable is a variable-frequency fourth-order Linkwitz–Riley state-variable crossover; precisely this was provided by Dennis Bohn in 1983 [4], and this very clever implementation has seen extensive use in the crossover business. Figure 8.44 shows a fourth-order variable filter with a crossover frequency range from 210 Hz to 2.10 kHz. Note that there are now four integrators plus the summing/differencing stage, to which feedback from all four integrators is returned. As usual, the capacitors have been made preferred values, letting the resistor values fall where they may.

Figure 8.45 shows the lowpass and highpass outputs from the fourth-order filter. The two outputs are both 6 dB down at the crossover frequency, as expected for a Linkwitz–Riley crossover, and the ultimate slopes are at 24 dB/octave. The design of the filter for various frequencies is not hard. It follows the same process as the second-order SVF described above. Ignoring the variable resistors, and treating R8, R9, R10, and R11 as the total resistance feeding each integrator from the previous stage, the centre frequency *f* is then set by choosing R6 (= R7):

$$R8 = R9 = R10 = R11 = \frac{1}{2\pi fC} \tag{8.22}$$

Note that the resistors around A1 have to be in the ratio shown for correct operation and to get the correct Q for a Linkwitz–Riley alignment. The required ratios are shown in brackets in Figure 8.44.

Figure 8.46 attempts to give some insight into how the filter works. The output from A1, the summing/differencing stage, is the highpass output, with an ultimate roll-off slope of

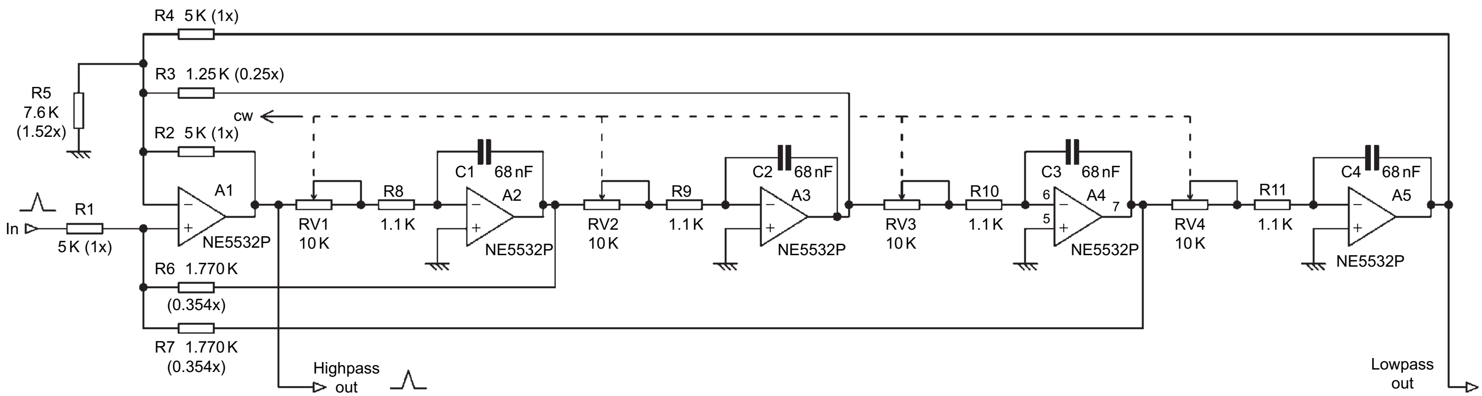


Figure 8.44: Fourth-order variable-frequency state-variable filter, with crossover frequency variable from 210 Hz to 2.10 kHz.

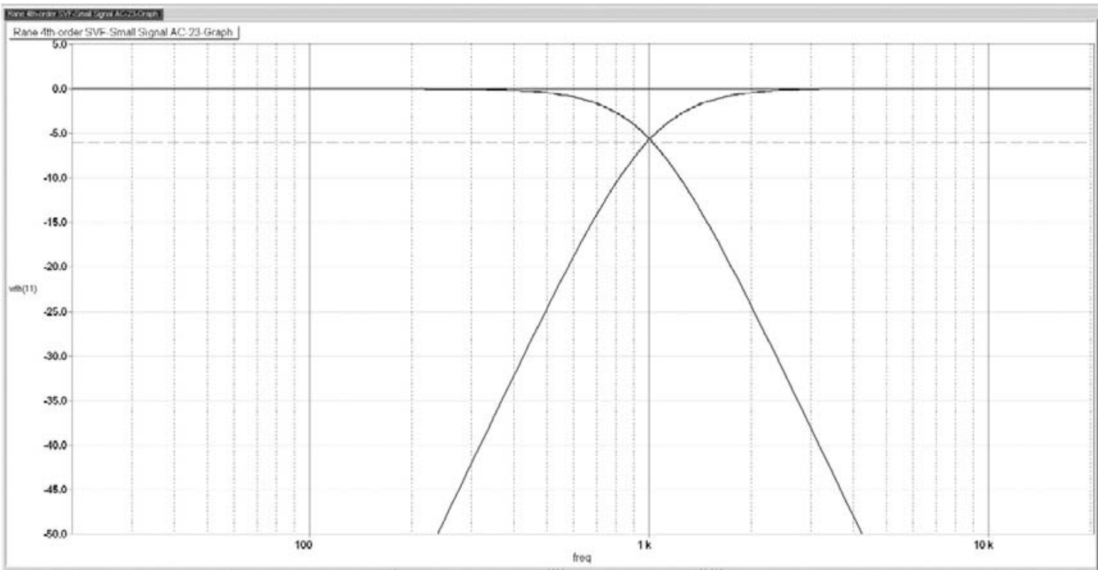


Figure 8.45: Lowpass and highpass outputs from the fourth-order variable-frequency state-variable filter, crossing over at −6 dB for a Linkwitz–Riley alignment. Crossover frequency is set to 1 kHz.

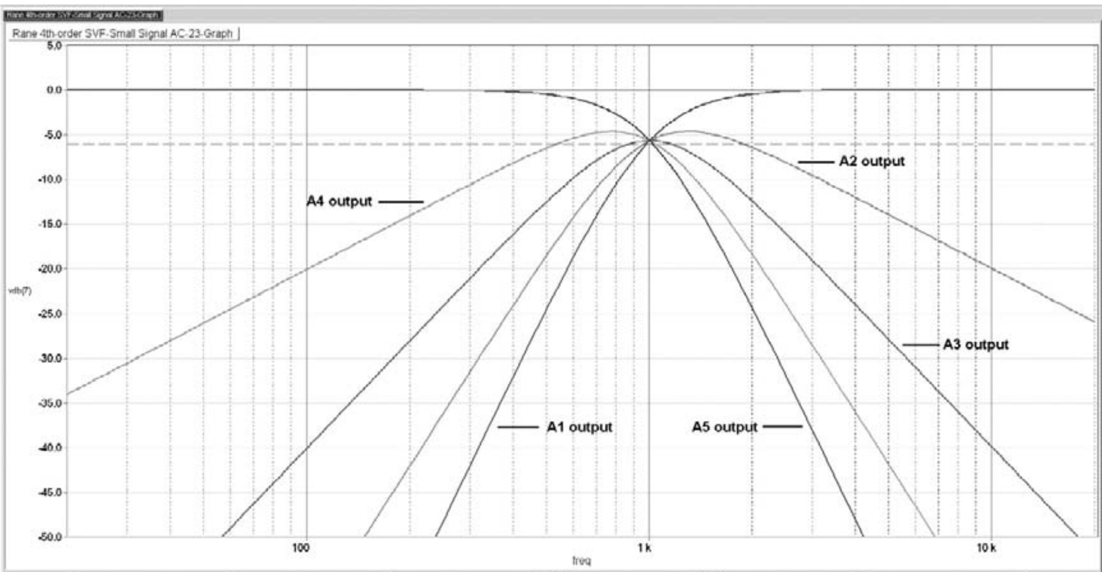


Figure 8.46: The outputs from all five stages of the fourth-order variable-frequency state-variable filter, with crossover frequency set to 1 kHz.

24 dB/octave below crossover as frequency decreases. The output of the first integrator A2 is the same but with a 6 dB/octave slope decreasing with frequency applied across the whole range, so the part of the response that was flat now slopes downward at 6 dB/octave slope with frequency, while the 24 dB/octave section has its slope reduced by 6 dB/octave to give 18 dB/octave. The second integrator A3 performs the same process again, so its output is a fourth-order bandpass response with skirt slopes of 12 dB/octave. The third integrator A4 does the same thing, its output having a 6 dB/octave slope in its low-frequency section, and a 18 dB/octave slope in its high-frequency section. The fourth integrator completes the process, and its output is the lowpass signal, with a flat low-frequency section and a 24 dB/octave roll-off above the crossover frequency. Note that all the plots in Figure 8.46 pass through the −6 dB point at the crossover frequency.

This configuration could of course be adapted for fixed-frequency operation by removing the variable elements.

8.44 Variable-Frequency Filters: Other Orders

You may at this point be wondering what happened to third-order state variable filters; there appears to be no reason why these should not be constructed in exactly the same way, using three integrators plus a summing/differencing stage. In view of the lesser desirability of third-order crossovers compared with fourth-order, I have chosen not to investigate these at this time. It is certainly possible to make eighth-order variable-frequency state-variable filter crossovers by using the same approach, as demonstrated by Dennis Bohn in 1988 [5], and so there is every reason to assume that fifth-, sixth-, and seventh-order versions could be implemented without great difficulty. Eighth-order crossovers are sometimes used in sound reinforcement applications, but seem unlikely to become popular in domestic hi-fi.

References

[1] Sallen, Key, A practical method of designing RC active filters, IRE Trans. Circuit Theory 2 (1) (1955) 74–85.

[2] S. Linkwitz, Active crossover networks for non-coincident drivers, J. Audio Eng. Soc 24(1) (1976) 2–8.

[3] Aitken, Kerwin, Single amplifier, minimal RC, Butterworth Thomson and Chebyshev filters to sixth order, in: Proceedings of International Filter Symposium, Santa Monica press, California, 1972.

[4] D.A. Bohn, A Fourth-Order State Variable Filter for Linkwitz-Riley Active Filters, 74th AES Convention, Audio Engineering Society, Newyork, 1983, Preprint 2011 (B-2).

[5] D.A. Bohn, An 8th-Order State Variable Filter for Linkwitz-Riley Active Crossover Design, 85th AES Convention, Audio Engineering Society, Newyork, 1988, Preprint 2697 (B-11).

CHAPTER 9

Bandpass & Notch Filters

This chapter deals with bandpass and notch filters. Bandpass filters as such are rarely used in performing the basic band-splitting functions of a crossover (the filler-driver concept being the notable exception), but they can be useful for equalisation purposes, and are essential for putting together high-order allpass filters for time correction. Notch filters can also be useful for equalisation, but their most important use is in the construction of notch crossovers, whether based on elliptical filters or other sorts of filtering.

9.1 Multiple-Feedback Bandpass Filters

When a bandpass filter of modest Q is required, the Multiple FeedBack (MFB) or Rauch type shown in Figure 9.1 has many advantages. The capacitors are equal and so can be made any preferred value. The opamp is working with shunt feedback and so has no common-mode voltage on the inputs, which avoids one source of distortion. It does, however, phase-invert, which can be inconvenient. Phase inversions are no problem in passive crossovers—you simply swap over the wires to the speaker unit—but in an active crossover an extra inverting stage may be needed to undo the first inversion.

Figure 9.1 shows the normal configuration of an MFB or Rauch filter. The minimum Q that can be normally achieved is 0.707 ($1/\sqrt{2}$), which requires R2 to be set to infinity, that is, not fitted at all. If you want a lower Q than that, you still omit R2, but the vital point is that a different set of design equations are used, which give different values for the other components and allow lower Q's to be realised. These lower Q's are required for time-delay compensation allpass filters, and for filler-driver crossovers. The design process for allpass filters is fully described in Chapter 10 on time-domain filtering, where low Q's are especially necessary. This variation is on the standard MFB filter is sometimes called a Deliyannis filter.

Note that the Deliyannis filter is not suitable for high Q's, because as the Q is increased, the passband gain increases with it; it cannot be controlled independently as it can in the standard MFB filter. This will lead to either headroom problems, or alternatively noise problems if the signal is attenuated before it reaches the filter.

The filter response is defined by three parameters—the centre frequency f_0, the Q, and the passband gain (i.e., the gain at the response peak) A. The filter in Figure 9.1 was designed

The Design of Active Crossovers

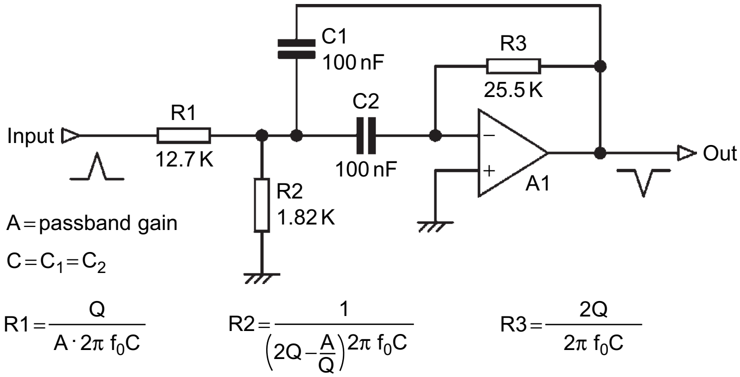


Figure 9.1: An MFB or Rauch bandpass multiple-feedback filter with $f_0 = 250$ Hz, Q = 2, and a gain of 1.

for $f_0 = 250$ Hz, Q = 2, and A = 1 using the equations given, and the usual awkward resistor values emerged. The resistors shown are the nearest E96 values, and the simulated results come out as $f_0 = 251$ Hz, Q = 1.99, and A = 1.0024, which, as they say, is good enough for rock'n'roll.

The Q of the filter can be quickly checked from the response curve as Q is equal to the centre frequency divided by the −3 dB bandwidth, that is, the frequency difference between the two −3 dB points on either side of the peak. This configuration in either Rauch or Deliyannis form, is not suitable for Q's greater than about 10, as the filter characteristics become unduly sensitive to component tolerances. If independent control of f_0 and Q are required the state-variable filter should be used instead.

Similar configurations can be used for lowpass and highpass filters; see Chapter 8. The lowpass version does not depend on a low opamp output impedance to maintain stop-band attenuation at high frequencies, and so avoids the oh-no-it's-coming-back-up-again behaviour of Sallen & Key lowpass filters.

9.2 High-Q Bandpass Filters

As we have just seen, the simple MFB/Rauch filter is not suitable for high Q's. When these are required (which is not likely to be very often in crossover design) there are many, many kinds of active filter that can be used. We will just take a quick look at one of the most useful, the Double-Amplifier BandPass or DABP filter; this is a good example of the way that active filter performance can sometimes be transformed by adding one more inexpensive opamp section. Figure 9.2 shows the circuit with values for a centre frequency of 1 kHz and an impressively high Q of 70, with A2 is being used to provide one of the feedback paths. Note the high value of R1 with respect to the rest of the circuit; this derives from the high Q.

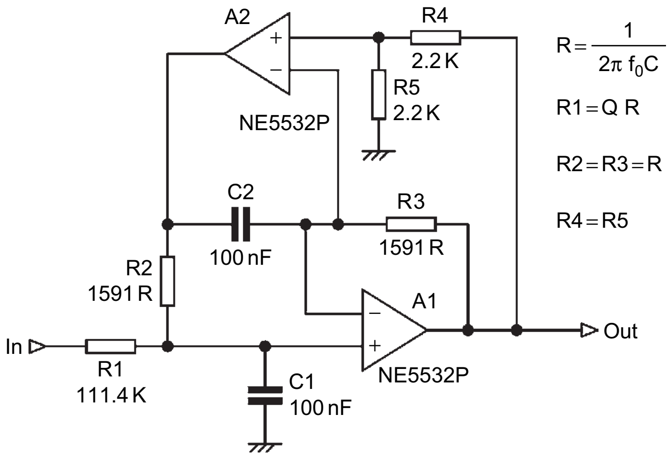


Figure 9.2: Dual-Amplifier BandPass (DABP) filter with a centre frequency of 1 kHz and a Q of 70. The gain at resonance is always +6 dB.

The design equations are straightforward; just calculate the value of R and then derive the actual resistor values from it. R4 and R5 are non-critical so long as they are of equal value; as usual they should be made as low as possible without overloading A1 output, in order to keep down both their Johnson noise and the effect of the current noise flowing from the A2 non-inverting input.

The gain at resonance is always +6 dB, which is very often not wanted. The most convenient way to introduce a compensating 6 dB of attenuation is to split R1 into two resistors which have in parallel the same value as R1, and connect one of them to ground. This technique is described in Chapter 8 on lowpass and highpass filters.

Figure 9.3 shows the satisfyingly sharp resonance obtained with the circuit of Figure 9.2. Away from the peak, the filter slopes slowly merge into straight lines at 6 dB/octave, as is normal for second-order bandpass responses. Fourth-order bandpass filters with 12 dB/octave skirt slopes can be made by cascading two second-order bandpass filters.

9.3 Notch Filters

There are a very large number of notch filters, each with their own advantages and disadvantages. The width of a notch is described by its Q. As for a resonance peak, the Q is equal to the centre frequency divided by the −3 dB bandwidth, that is, the frequency difference between the −3 dB points on either side of the notch. Figure 9.4 demonstrates this procedure for a notch with a centre frequency of 1.00 kHz and a Q of 1.5; the bandwidth between the two −3 dB points is 0.667 kHz, so Q = 1.00/0.667 = 1.5. Be aware that the Q of a notch has

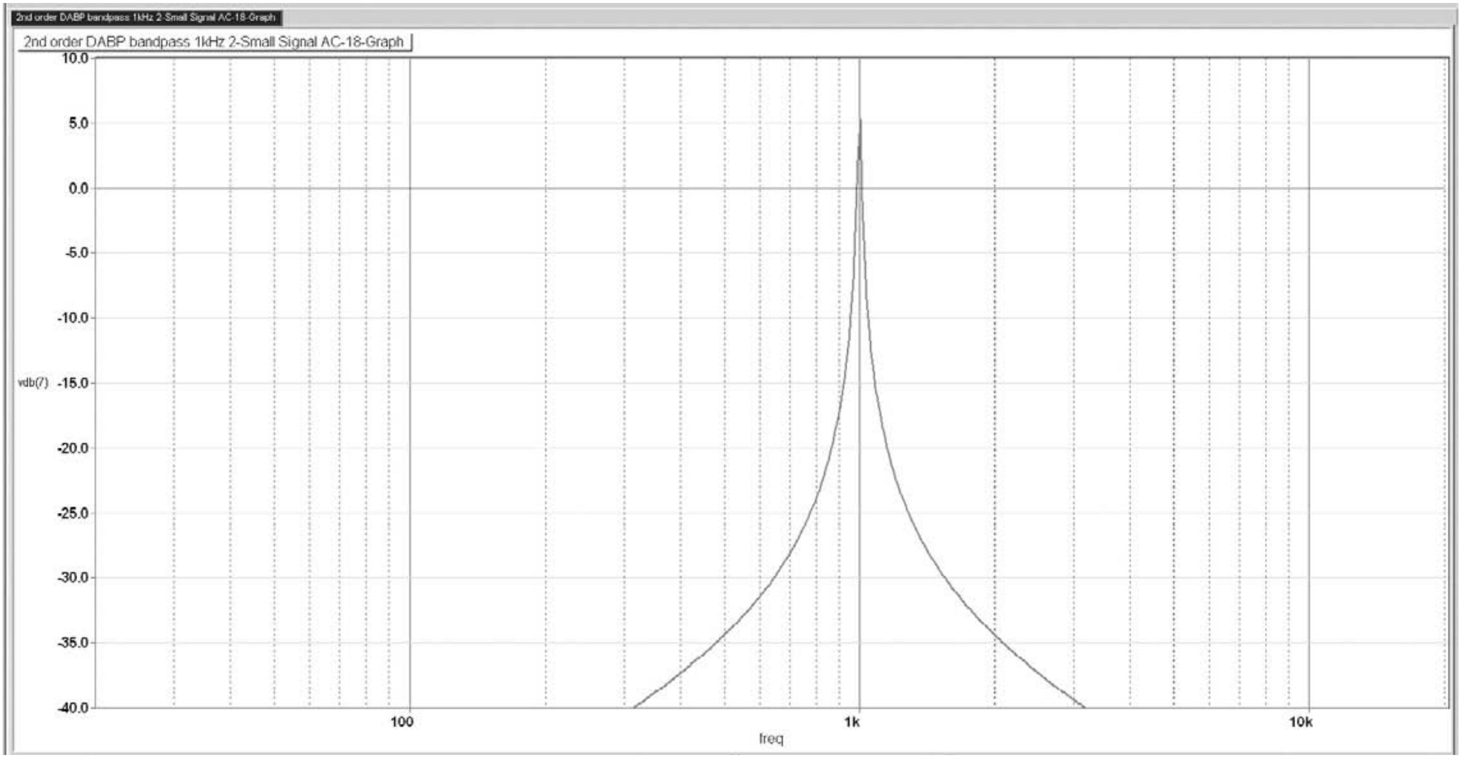


Figure 9.3: Dual-Amplifier BandPass (DABP) filter with a centre frequency of kHz and with a centre frequency of 1.00 kHz and a Q of 70. The gain at resonance is +6 dB.

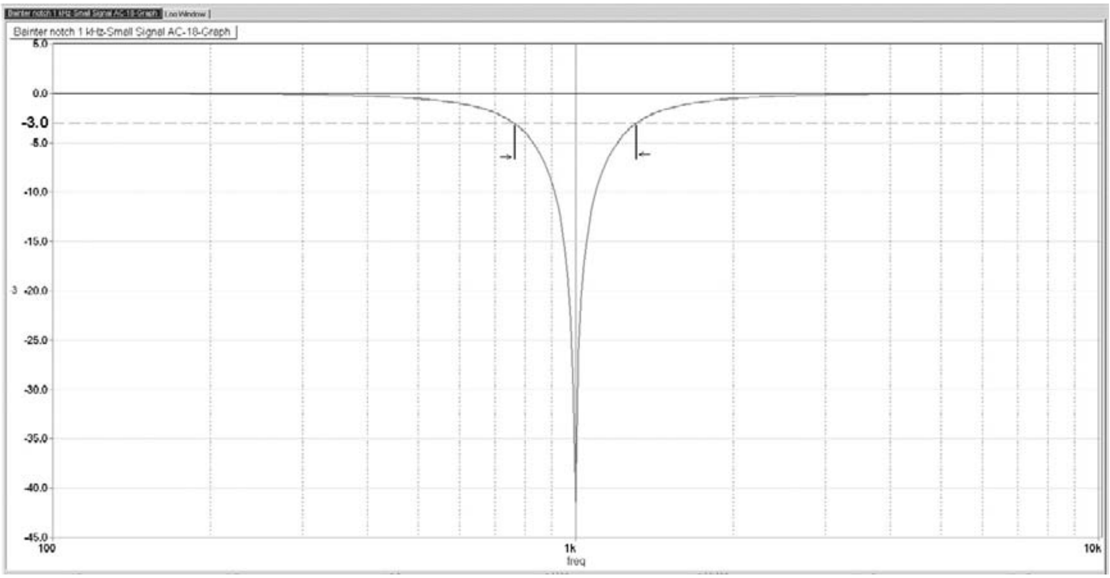


Figure 9.4: Determining the Q of a notch filter, which is equal to the centre frequency divided by the bandwidth between the two –3 dB points.

no special relation to the depth, but if the notch is really shallow—more of a dip than a notch—then the response at the –3 dB points may be affected.

It is important to understand that while the symmetrical notch shown in Figure 9.4 is the best-known sort, there are also highpass notches and lowpass notches. There is more information on the use of lowpass and highpass notches to create elliptical filters in Chapter 5 on notch crossovers, and in Chapter 8 on highpass and lowpass filters.

9.4 The Twin-T Notch Filter

The best-known notch filter is the Twin-T notch network shown in Figure 9.5a, invented in 1934 by Herbert Augustadt [1]. The notch depth is infinite with exactly matched components, but with ordinary ones it is unlikely to be deeper than 40 dB. It requires ratios of two in component values, such as 100 Ω – 200 Ω, and there are only six such pairs in the E24 resistor series; see Chapter 12. Capacitors have sparser value series and two will need to be paralleled to get the 1:2 ratio required.

The Twin-T notch when used alone has a Q of only ¼, which is too wide to be useful. It is therefore normally used with positive feedback via an opamp buffer A2, as shown. The proportion of feedback K and hence the Q-enhancement is set by R4 and R5, which here give a Q of 1. A great drawback is that the notch frequency can only be altered by changing three components simultaneously.

9.4.1 The 1-Bandpass Notch Filter

Another way of making notch filters is the "one-minus-bandpass" principle, usually called '1-BP'. The input goes through a bandpass filter, typically the Multiple FeedBack type described earlier, and is then subtracted from the original signal. The accuracy of the cancellation and hence the notch depth is critically dependent on the mid-band gain of the bandpass filter. Figure 9.5b shows an example that gives a notch at 50 Hz with a Q of 2.85. The subtraction is performed by A2, as the output of the MFB filter is phase-inverted. The MFB filter is designed for unity passband gain, but the use of E24 values as shown means that the actual gain is 0.97, limiting the notch depth to −32 dB. The value of R6 can be tweaked to deepen the notch; the nearest E96 value is 10.2 kΩ which gives a depth of −45 dB. The final output is inconveniently phase-inverted in the passband.

9.4.2 The Bainter Notch Filter

A most useful notch filter is the Bainter configuration [2, 3] shown in Figure 9.5c, where the values shown give a notch at 700 Hz with a Q of 1.29. The design equations can be found in [3], but I suggest you scale the values given here. This filter is non-inverting in the passband, and has the advantage that two out of the three opamps are working at virtual earth and will give no trouble with common-mode distortion. It has the important feature that the notch depth does not depend on the matching of components, but only on the open-loop gain of the opamps, being roughly proportional to it. With TL072-type opamps the depth is from −40 to −50 dB. Having said that, deep notches are not normally required for crossover design; even notches that look worryingly shallow usually have a negligible effect on the summed response. The only real downside to the Bainter is that it is one of those enigmatic configurations that on inspection give very little clue as to how they work.

A property of the Bainter filter that does not seem to appear in the textbooks is that if R1 and R4 are altered together, that is, having the same values, then the notch frequency is tuneable with a good depth maintained, but the Q does change proportionally to frequency. To get a standard notch with equal gain either side of the crevasse R3 must equal R4. R4 greater than R3 gives a lowpass notch, while R3 greater than R4 gives a highpass notch; these responses are useful for making elliptical filter, and for other applications in crossover design; see Chapter 8 for more details.

The Bainter filter is usually shown with equal values for C1 and C2. This leads to values for R5 and R6 that are a good deal higher than other circuit resistances and this will impair the noise performance. I suggest that in Figure 9.5c, C2 is made ten times C1, that is, 100 nF, and R5 and R6 are reduced by ten times to 5.1 kΩ and 6.8 kΩ; the response is unaltered and the Johnson noise much reduced.

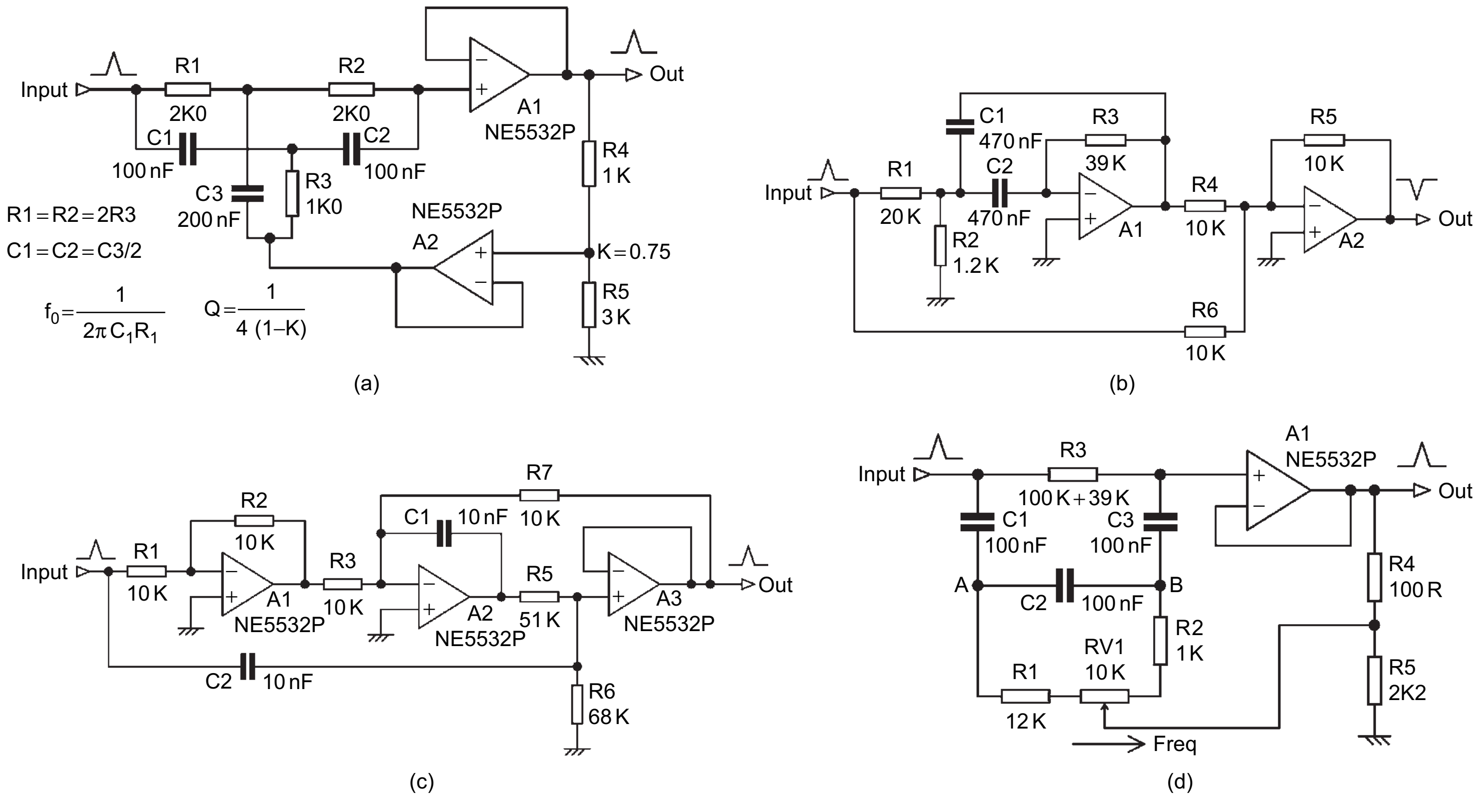


Figure 9.5: Notch filters: (a) Twin-T with positive feedback, notch at 795 Hz and a Q of 1; (b) "1-bandpass" filter with notch at 50 Hz, Q of 2.85; (c) Bainter filter with notch at 700 Hz, Q of 1.29; (d) Bridged-differentiator notch filter tuneable 80 to 180 Hz.

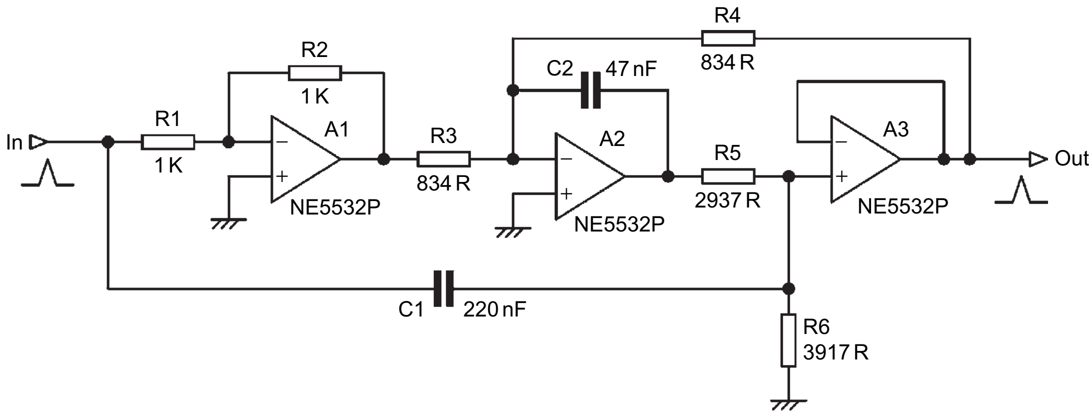


Figure 9.6: Bainter filter with notch at 1 kHz and a Q of 2.3.

As an example of this process, a low-noise Bainter filter with a symmetrical notch at 1.0 kHz and a Q of 2.3 is shown in Figure 9.6. Starting from the circuit of Figure 9.5c, both capacitors were scaled by the same ratio to change the notch centre frequency to 1 kHz. The value of R6 was then changed to give the desired Q of 2.3. The first stage of impedance reduction was to then alter R3, R4 and C2 so that the resistance values were decreased by the same ratio as C2 was increased; this leaves the frequency-dependent behaviour of this network unchanged. The second stage of impedance reduction reduces R6 while increasing C1 by the same ratio, once again keeping the frequency-dependent behaviour the same.

The limits of this procedure are set by the value of R3, which directly loads the output of A1, the value of R4 that loads A3, and the combined value of R5, R6 in series, which loads A2. The last condition is less critical; as you can see from Figure 9.6, the values of R5 and R6 are such as to present only a light load on A2 output, despite the fact that C1 is considerably bigger than C2. A further reduction in R5, R6 would be possible, but C1 then starts to get expensive, and the loading on whatever stage is driving this filter must also be considered.

9.4.3 The Bridged-Differentiator Notch Filter

A notch filter that can be tuned with one control can be useful in development work. Figure 9.5d shows a bridged-differentiator notch filter tuneable from 80 to 180 Hz by RV1. R3 must theoretically be six times the total resistance between A and B, which here is 138 kΩ, but 139 kΩ gives a deeper notch, about −27 dB across the tuning range. The downside is that Q varies with frequency from 3.9 at 80 Hz to 1.4 at 180 Hz.

9.4.4 Boctor Notch Filters

Another interesting notch filter is the Boctor circuit, which uses only one opamp [4, 5]. Versions exist that can give either a highpass or lowpass notch. The design equations are complicated but can be found in the two references given. Figure 9.7 shows the highpass and lowpass versions.

Figure 9.8 shows the highpass notch produced by the circuit of Figure 9.7b above; the passband gain is +12 dB, with a gain of +6 dB on the low-frequency side of the notch. The capacitor values may be scaled to change the notch frequency but they must be the same.

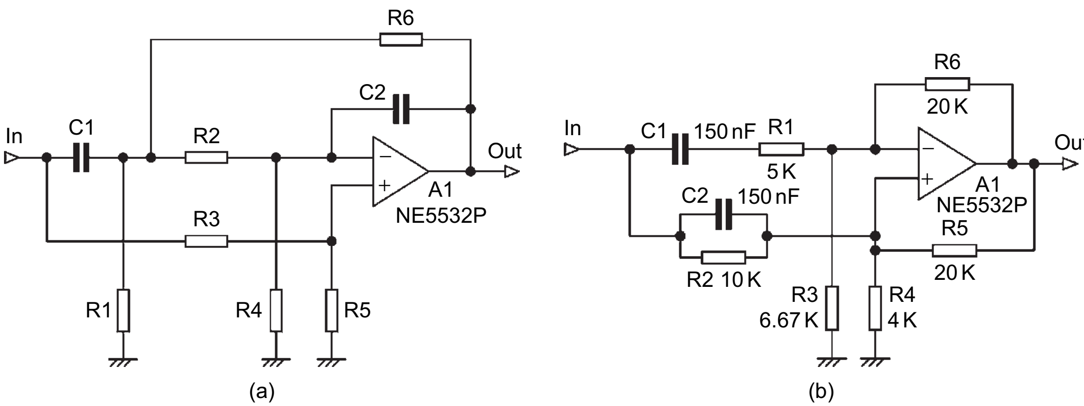


Figure 9.7: Lowpass-notch (a) and highpass notch (b) Boctor filters. The design at (b) has a notch at 150 Hz with a Q of 1.

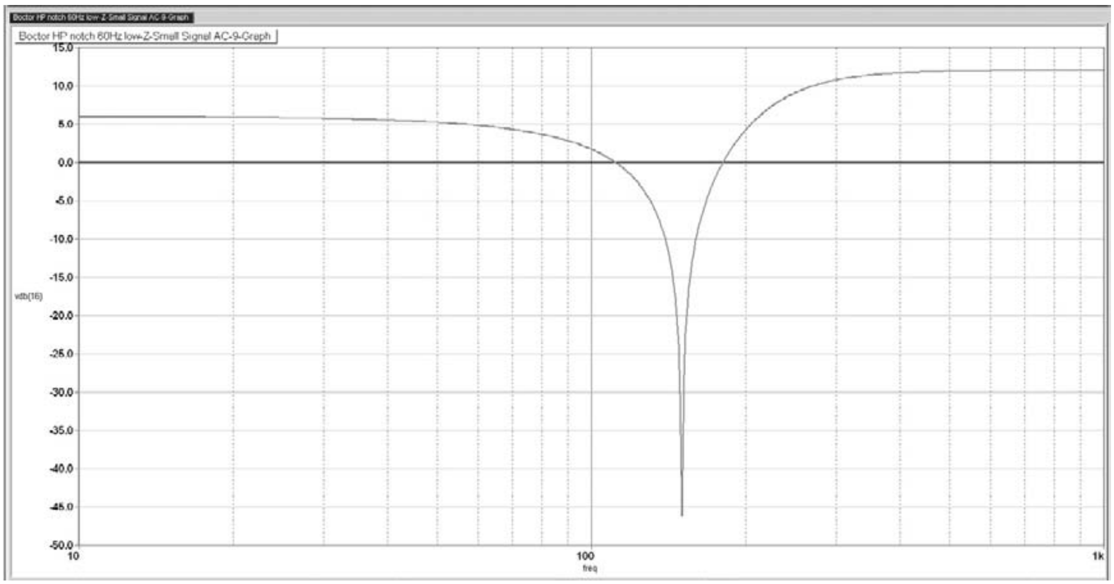


Figure 9.8: Response of Boctor filter in Figure 9.7b, giving a highpass notch at 150 Hz with a Q of 1. Boctor highpass or lowpass notch filters are frequently used as part of elliptical filters; see Chapter 7.

A lowpass notch is the mirror image of this sort of response, with the low-gain section on the high-frequency side.

9.4.5 Other Notch Filters

The field of active filters is rich in possibilities. Other notch filters that there is no space to examine here, but which can be found in the filter textbooks, are the Fliege filter [6], the Berka-Herpy filter, the Akerberg-Mossberg filter and the Natarajan filter. Filters that can generate highpass or lowpass notches are Friend's SAB circuit [7] and the Scultety filter. Both the biquad and state-variable filters can be configured to give notch outputs.

9.4.6 Simulating Notch Filters

When simulating notch filters, assessing the notch depth can be tricky. You need a lot of frequency steps to ensure you really have hit bottom with one of them. For example, in one run, 50 steps/decade showed a −20 dB notch, but upping it to 500 steps/decade revealed it was really −31 dB deep. In most cases having a stupendously deep notch is pointless. If you are trying to remove an unwanted signal then it only has to alter in frequency by a tiny amount and you are on the side of the notch rather than the bottom and the attenuation is much reduced. The exception to this is the THD analyser, where a very deep notch (120 dB or more) is needed to reject the fundamental so very low levels of harmonics can be measured. This is achieved by continuously servo-tuning the notch so it is kept exactly on the incoming frequency.

References

[1] H. Augustadt, Ron Mancini, M.E. van Valkenburg, Electric Filter, US Patent 2,106,785, Feb 1938 (Assigned to Bell Telephone Labs).
[2] J.R. Bainter, Active filter has stable notch, and response can be regulated, Electronics (1975) 115–117.
[3] M. E. van Valkenburg, Analog Filter Design, Holt-Saunders International, p. 346.
[4] S.A. Boctor, Single amplifier functionally tuneable low-pass notch filter, IEEE Trans. Circuits Syst. CAS-22 (1975) 875–881.
[5] M. E. van Valkenburg, Analog Filter Design, Holt-Saunders International, p. 349.
[6] B. Carter, R. Mancini (Eds.), Opamps for Everyone, third ed., Chap. 21, Elsevier Inc., Boston, MA, 2009, p. 447, ISBN 978-1-85617-505-0.
[7] M. E. van Valkenburg, Analog Filter Design, Holt-Saunders International, p. 356.

CHAPTER 10

Time Domain Filters

10.1 The Requirement for Delay Compensation

A lot of discussion on loudspeaker behaviour is based on mathematically simple point sources, but real loudspeaker drive units have a physical size and must be mounted with some space between them. If they are mounted on a flat baffle as in Figure 10.1, their acoustic centres (the position from which the sound effectively radiates) have different horizontal distances from the baffle position. If we take our listening position as in-line with the tweeter axis, as is usually done, then there is a greater distance from the ear to the MID unit and an even greater one to the bass unit. To minimise the differences, the drive units are normally mounted as close together as physically possible. We therefore have two effects giving rise to different distances from drive unit to ear and hence differing delays in the arrival of sound:

1. The vertical spacing of the drive units
2. The different distances between the drive-unit acoustic centres and the front baffle

I am assuming here that the drive units are mounted in a vertical line. This is almost always the case as it gives the best horizontal directivity.

These delays lead to frequency-dependent variations in the response, due to cancellation and reinforcement at a given point in space. The dimensions shown in Figure 10.1 are intended to be reasonably typical, but I will tell you now that they have also been carefully chosen to require a compensation delay of exactly 80 usec for the tweeter and 400 usec for the MID unit; these figures are used extensively as examples throughout this chapter. It must not be thought that a spacing of a few millimetres is insignificant. The 22 mm difference between the MID drive unit and the tweeter should be compared with the wavelength of sound at 4 kHz, which is 86 mm. At this frequency, 22 mm therefore gives a 90° phase-shift, turning a +6 dB fully in-phase reinforcement into a +3 dB partial reinforcement. This sort of thing is obviously going to cause frequency response irregularities.

This is why delay compensation is highly desirable for multi-way loudspeakers. The technique is sometimes called time equalisation or time alignment. Siegfried Linkwitz says that crossovers that are not time-compensated are only marginally usable [1], and, for what it's worth, I agree with him. Figure 10.1 shows a simplified diagram of the situation; please be aware that the drawing is very much not to scale. First, we have to decide how far away

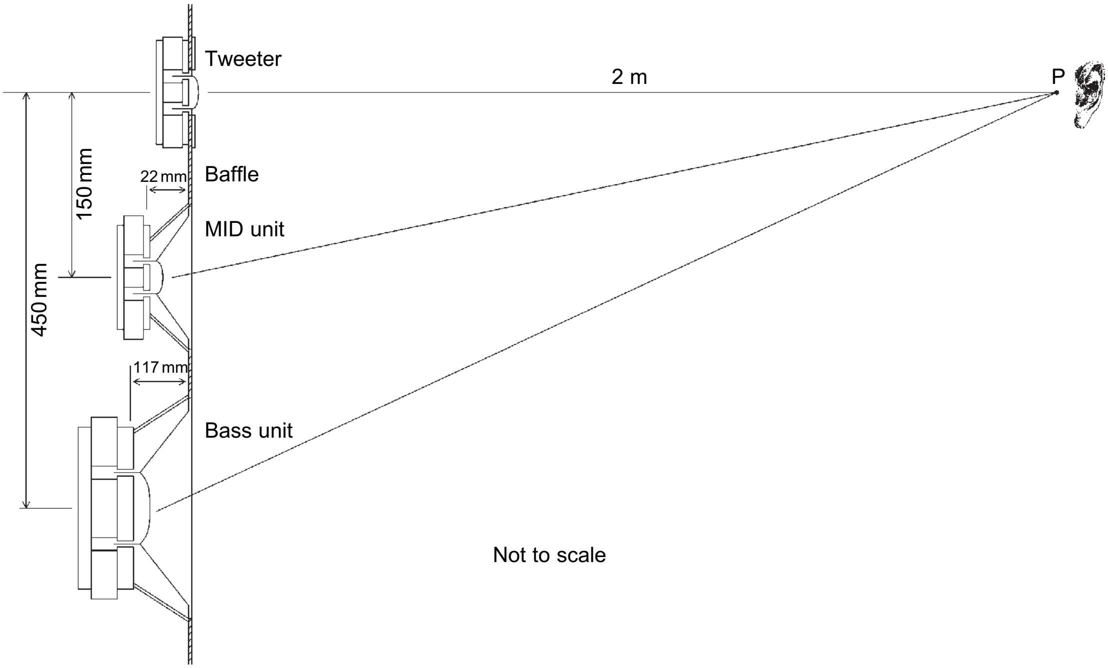


Figure 10.1: The need for delay compensation; the differences in path lengths to the ear when three drive units are mounted in a common vertical plane.

to put our reference listening point P, marked by that disembodied ear floating in space. This distance is usually set at 1 or 2 metres; I have here chosen 2 metres as the more realistic listening distance but of course this will vary in practice, upsetting our careful calculations somewhat. We can but press on and do our best.

The vertical distance between the drive-unit centre-lines is easy to determine, but the position of the acoustic centres is less easy, not least because it usually varies somewhat with frequency. Some people use the voice-coil cap, but others claim that the centre of the voice-coil itself should be regarded as the acoustic point of origin, as the speed of sound in the average coil assembly is roughly the same as in air; this seems more than a bit dubious as the speed of sound is in fact quite different in solids, and there are actually two different speeds of sound, transmitted in one case by volumetric deformations and in the other by shear deformations.

Add to this the composite construction of the voice-coil, with copper windings bonded to the coil former, and the whole situation gets very complicated. The position of the acoustic centre is not normally part of a drive-unit specification.

An important consideration is that we only need to consider the delay that must applied to the tweeter with respect to the MID unit, and the delay to be applied to the MID unit with respect to the LF unit. It is not necessary to delay the tweeter to match the LF unit, because the LF unit output should be negligible when the tweeter is operating, and vice versa. This greatly reduces the amount of delay that needs to be used on the HF unit.

10.2 Calculating the Required Delays

The initial step is to calculate what the path-length differences are. We need to find the path lengths MP and BP in Figure 10.2. MP is the hypotenuse of the right-angled triangle MYP, and the two shorter sides are YM and (YT + TP). Using Pythagoras:

$$\mathrm{MP} = \sqrt{\mathrm{YM}^2 + (\mathrm{YT} + \mathrm{TP})^2} \tag{10.1}$$

Likewise, BP is the hypotenuse of right-angled triangle BXP, and the two shorter sides are XB and (XT + TP). Flexing our Pythagoras again:

$$\mathrm{BP} = \sqrt{\mathrm{XB}^2 + (\mathrm{XT} + \mathrm{TP})^2} \tag{10.2}$$

For the dimensions in Figure 10.2 and a listening distance TP of 2 metres, we find the path lengths are MP = 2.027 metres and BP = 2.165 metres. It is the difference between these path lengths and the listening distance that causes the delay we need to compensate for; the extra MID path length is thus 2 − 2.027 = 0.027 metres = 27 mm, and the extra LF path length is the difference between the MID path and the LF path, in other words 2.027 − 2.165 = 0.137 metres = 137 mm.

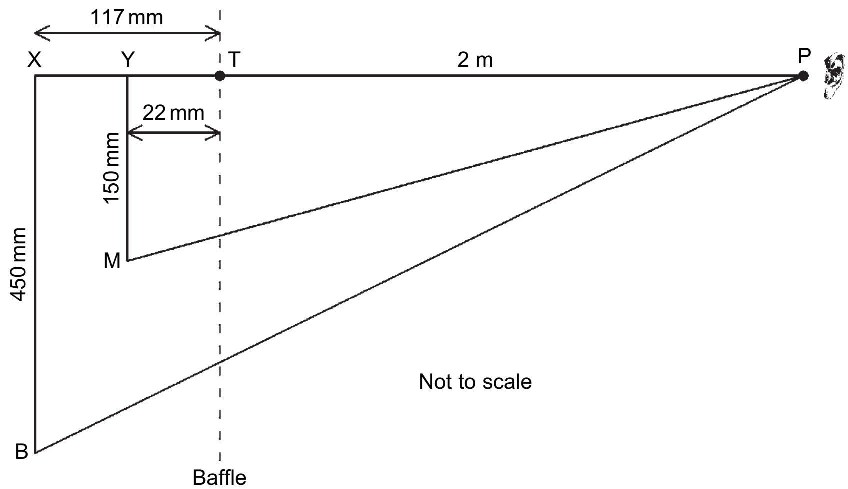


Figure 10.2: The delay situation of Figure 10.1 reduced to its elements. We need to calculate MP and BP.

Table 10.1: How the Speed of Sound Varies with Temperature

Temp deg C	Speed m/sec	% Ref 20°C
35	351.9	2.5
30	349.0	1.7
25	346.1	0.8
20	343.2	0.0
15	340.3	−0.9
10	337.3	−1.7
5	334.3	−2.6
0	331.3	−3.5

Having calculated the path length differences, the delays themselves are simply calculated by using the speed of sound. This is 343.2 metre/sec (equal to 767.7 mph or 1.13 feet per msec) at 20°C. It varies slightly with air temperature, being proportional to the square root of absolute temperature; this is measured from absolute zero at −273°C, which is why the speed variation is relatively small at the kind of temperatures we are used to. Table 10.1 shows this, together with the percentage variation. The speed of sound varies by ±2 % between 10 and 30°C, so for this reason alone there really is no point in being overly precise with delay calculations. Sound velocity varies very slightly with barometric pressure, and humidity also has a small but measurable effect (it can cause an increase of 0.1% – 0.6%), because some of the nitrogen and oxygen molecules in the air are replaced by lighter water molecules.

Dividing the path length differences by the speed of sound, we find that the tweeter must be delayed by exactly 80 usec with respect to the MID unit, and the MID unit must be delayed by exactly 400 usec with respect to the LF unit. As I mentioned earlier, these nice round figures are no accident.

At the start of these calculations we had to choose a listening distance, and we picked 2 metres. In reality, the listening distance will vary, and we'd better find out how much difference this makes to the delays required for exact delay compensation. A bit of automated Pythagoras on a spreadsheet gives us the rather worrying Table 10.2, which gives the delays required for listening distances between 1 and 10 metres, and for infinity (but not beyond).

Table 10.2 also gives the percentage variation of each delay with respect to the 2-metre listening distance. You can see at once that picking 2 metres as the reference rather than 1 metre has made big differences to the delays needed; 19% less for the tweeter delay and a thumping great 25% less for the MID delay. One metre is not a very practical distance for listening to a big hi-fi system of the sort that is likely to use active crossovers, and I would have thought that 1.5 to 5 metres is a likely range for normal use. This assumption gives us a tweeter delay that varies from +6.6% to −12.1%, while the MID delay varies from +9.0% to −17.7%. This makes it clear that if fixed amounts of delay are used there is, once more, no call for extreme precision. I am not aware that any manufacturer has ever produced an active crossover with a "Listening distance" dial that would alter the two delays appropriately to

Table 10.2: The Variation of the Delays Required with Listening-Point Distance

Listening Distance Metres	Tweeter Delay usec	Tweeter % Ref 2 Metres	MID Delay usec	MID % Ref 2 Metres
1.0	95.7	19.6%	500.6	25.2%
1.1	92.9	16.1%	483.8	21.0%
1.2	90.5	13.2%	469.2	17.3%
1.5	85.3	6.6%	435.9	9.0%
2.0	80.0	0.0%	400.0	0.0%
2.5	76.8	−4.0%	377.3	−5.7%
3.0	74.6	−6.8%	361.7	−9.6%
3.5	73.1	−8.6%	350.3	−12.4%
4.0	72.0	−10.0%	341.7	−14.6%
4.5	71.1	−11.1%	334.9	−16.3%
5.0	70.3	−12.1%	329.4	−17.7%
6.0	69.2	−13.5%	321.1	−19.7%
7.0	68.5	−14.4%	315.1	−21.2%
8.0	67.9	−15.1%	310.6	−22.4%
9.0	67.4	−15.8%	307.1	−23.2%
10.0	67.1	−16.1%	304.3	−23.9%
Infinity	63.8	−20.3%	278.4	−30.4%

match the listening position; it's an interesting idea. It would only work with one model of loudspeaker as the variations required would depend on the horizontal and vertical offsets.

The required delays alter more slowly as the listening distance increases and the vertical offsets become relatively less significant. The need to compensate for the horizontal offsets is, however, unchanged, and the delays required asymptote to these values at an infinite listening distance. The situation is illustrated in Figure 10.3, which plots how tweeter and MID delays change with the listening distance.

To put this into perspective, let's look at the effect of uncertainty in the position of the acoustic centre of a drive unit. We currently have to apply a delay of 400 usec to the MID unit to match the LF unit output. But if we hold the LF unit in our hands and solemnly contemplate it, we might conclude that the true acoustic centre could be 30 mm either side of the position we have assumed. If it is 30 mm forward, then the MID delay required falls to 313 usec; if is 30 mm to the rear, we need 484 usec. This shows that errors in finding the true acoustic centre can have just as much effect as altering the listening distance, and it is a bit worrying.

We have so far glibly assumed that the position of the acoustic centre is constant with frequency. I currently have no information on this, but if it does prove to vary with frequency then it should be possible to design a delay filter system that compensates for this.

Sound reinforcement applications are rather different in the distances involved, but the delay calculations work the same way. The listening distances are much greater, but then the physical separation between banks of HF units and banks of MID units, for example, will also be greater.

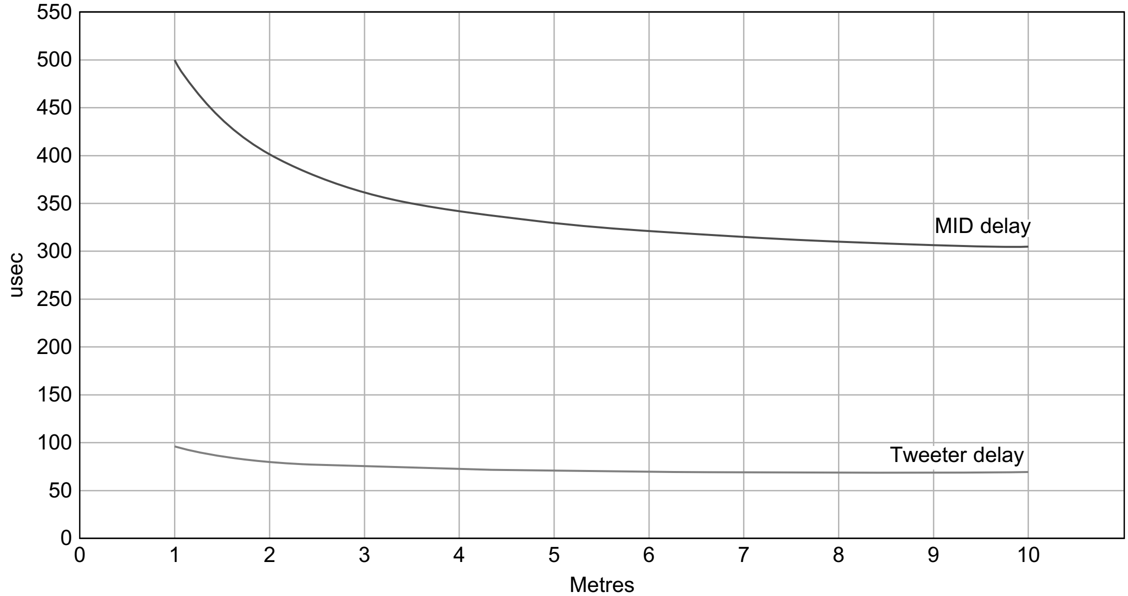


Figure 10.3: The variation of the required tweeter delay (bottom) and the required MID delay (top) with listening distance, for the dimensions given in Figure 10.1.

Delay compensation is only really practicable with active crossovers. The electrical losses involved in LC delay lines in passive crossovers, and the cost of the components to implement, makes it an unattractive proposition. So far we have only dealt with 3-way loudspeakers. The same principles are of course applicable to 4 ways or more.

10.3 Signal Summation

I shall shortly be showing you how the difficulty of designing a delay filter, and the cost of its components, depends very much on how much delay is required and over what range of frequency it must remain constant. At the crossover frequency, the delay must be constant at the required value, but at some frequency above this the amount of delay will begin to fall, and we need to know how much the acoustical summation of the two signals, which combine in the air in front of the speaker, is going to be affected by a fall in the time delay above the crossover frequency. Figure 10.3 gives the rules for adding sine waves of arbitrary magnitude and phase.

There is, as far as I am aware, no published work on the delay-flatness required for loudspeaker delay compensation. I have here assumed that a reduction of not more than 10% in the group delay is acceptable so long as it is well away from the crossover

With amplitudes a, b, and phase shift α

We get $a\sin x + b\sin(x+\alpha) = c\sin(x+\beta)$,

Where, magnitude $c = \sqrt{a^2+b^2+2ab\cos\alpha}$

and phase $\beta = \arctan\left(\dfrac{b\sin\alpha}{a+b\cos\alpha}\right) + \begin{cases} 0 & \text{if } a+b\cos\alpha \geq 0 \\ \pi & \text{if } a+b\cos\alpha < 0 \end{cases}$

Figure 10.4: The summation of two sinusoidal signals of arbitrary amplitude and phase to give a single sinusoidal output (Equation 10.3).

frequency, at which point the contribution of the drive unit being compensated is negligible.

Let us assume that the delay falls by −10% at a frequency two octaves above the relevant crossover frequency. If we are using a Linkwitz–Riley 4th order crossover then the ultimate filter slopes will be 24 dB/octave. The higher frequency unit of the drive-unit pair, the one with the compensation delay added, will therefore be something like 48 dB down, though this figure is likely to be affected by drive-unit response irregularities. From the equations in Figure 10.4, we can quickly determine that even using the worst-case phase-shift for the smaller signal (0°, 180°, 360°, etc.), the output will only be altered in amplitude by ±0.035 dB, which is completely negligible compared with other possible errors. I therefore suggest that the 10% fall-off in time delay two octaves away from the crossover frequency will have no audible consequences.

Table 10.3 show how the amplitude of the combined signal is increased for varying attenuations of the delayed signal, assuming a worst case of signals exactly in phase. It is clear that delayed signals below −40 dB will have an absolutely negligible effect, no matter what phase-shift is imposed on them by a fall-off of delay time.

Table 10.3: The Result of Combining Two In-Phase Sinewaves, with One at a Lower Amplitude

Amplitude dB	Combined dB	Amplitude dB	Combined dB
0	6.021	−25	0.475
−1	5.535	−30	0.270
−5	3.876	−35	0.153
−10	2.387	−40	0.086
−15	1.422	−45	0.049
−20	0.828	−50	0.027

10.4 Physical Methods of Delay Compensation

Before we leap into the fascinating world of time domain filters, we will take a quick look at some of the physical methods used for delay compensation. These are shown in two-way loudspeaker form in Figure 10.5. Clearly it is best to mount the drive units as close together vertically as physically possible to minimise the path difference, but loudspeaker manufacturers do not appear to always do this, possibly because of the aesthetic drawbacks of two drive units that look "crammed together." Possibilities are:

1. Tilting the front baffle. This compensates the delay for a listener on the horizontal axis. The problem is that the drive units will have been designed for optimal performance on the drive-unit axis, and now you are listening to off-axis output. This method is relatively cheap to manufacture, but people are used to having the front of a loudspeaker facing them directly, and as a result there is a general feeling that it looks wrong. It may compensate for the horizontal offsets (though for it to work with a three-way system you will have to have the three acoustic centres in a straight line, which may not be easy to arrange), but only reduces the effective vertical offsets slightly.

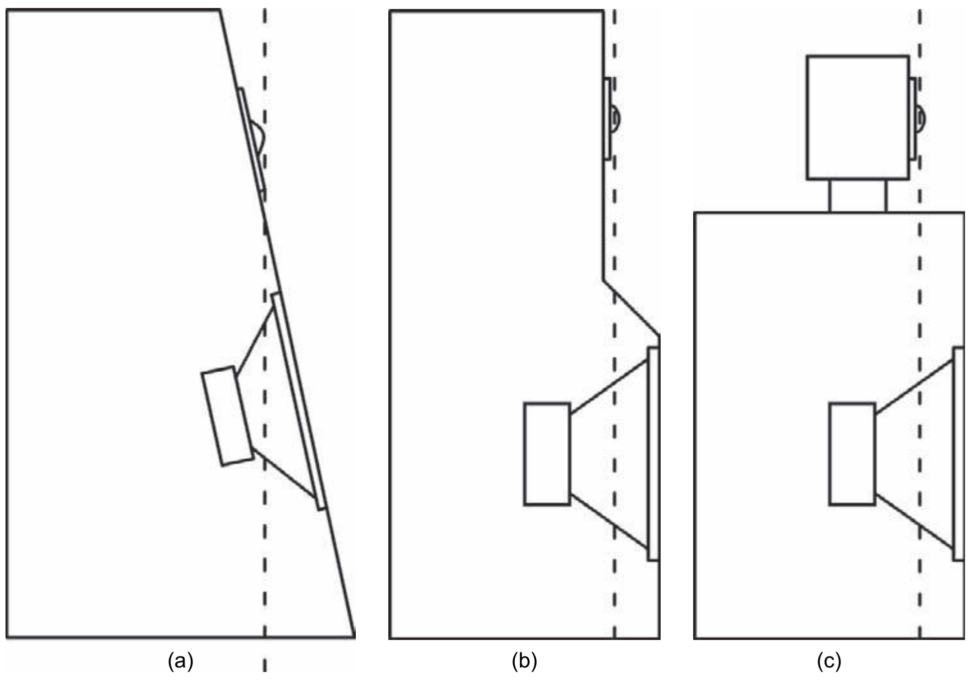


Figure 10.5: Physical methods of compensating for drive-unit delay: (a) Tilting the front baffle; (b) stepping the front baffle; (c) separate box for the tweeter.

2. Putting a step or steps in the front baffle. This enables you to listen on the intended axis, but costs more to implement. There may be diffraction problems with the step or steps, which will need to be carefully shaped, or possibly compensated for by equalisation in the crossover. It also makes the enclosure more complex and significantly more expensive to build. It can correct the horizontal offsets exactly but does nothing at all to reduce the vertical offsets—in fact they will have to be bigger to make room for the steps between the drive units.
3. A separate box for the tweeter. More expensive again, and there may be more diffraction problems with the second box. A three-way system will require three boxes, and once more the horizontal offsets may be corrected exactly, but the vertical offsets are likely to be increased.

While these methods can compensate for the differing distances of the drive-unit acoustic centres from the front baffle—the horizontal offsets, none of them can help much with the other half of the problem—the vertical offsets. This appears to be a serious inherent problem in the vast majority of loudspeaker designs. The only way to eliminate the vertical offsets is to mount two loudspeakers in effectively the same physical position by using a dual-concentric design, where typically a tweeter is mounted in front of an LF unit voice-coil or behind it. Current versions of this are the British KEF Uni-Q driver, and the French Cabasse co-axial drive unit, which has an annular diaphragm surrounding a small horn-loaded dome tweeter. The famous Tannoy Monitor Red Loudspeaker had a rear-mounted tweeter that used the LF cone itself as a horn.

The dual-concentric approach has the great benefit that there are no vertical offsets that require delay compensation that varies with listening distance. The horizontal offsets may or may not remain, depending on the details of construction, but can be easily compensated for by constant delays. It would at first appear that a major limitation of dual-concentric technology is that it is restricted to two-way loudspeakers, as it is hard to mechanically fit LF, MID, and tweeter units *all* on the same axis. However, Cabasse has recently announced not only the TC23 three-way coaxial driver, which combines a tweeter and separate annular high-MID and low-MID diaphragms, but also the QC55 four-way unit that combines the TC23 with a 22-inch LF unit. This is built into a spherical enclosure, which is the ideal shape for minimising diffraction effects. This technology clearly wants watching.

10.5 Delay Filter Technology

In the analogue domain, delay equalisation is performed by using allpass filters, which have a flat amplitude-frequency response but a phase-shift that varies with frequency. Allpass filters, like their more familiar lowpass and highpass relatives, come in first-order, second-order, third-order, and higher-order variants. The higher orders have increasing complexity but give a faster phase-change at the turnover frequency, allowing the group delay to remain flatter for further up the frequency range. This is analogous to the frequency response of a filter for amplitude; the higher order the filter the flatter the response will be until the turnover frequency is reached.

Allpass filters are sometimes used to flatten the delay curve of a high-order filter. This often means matching the combined delay of two or three allpass stages with a delay curve that looks like a dog's back leg, and typically this requires computer optimisation. Here, mercifully, the situation is much simpler. We just want a constant delay with frequency.

You can also obtain delays with lowpass filters such as the Bessel sort, which is optimised for flat group delay, but the bandwidth is much lower, as you might expect. There is more on the topic of obtaining delay with lowpass filters at the end of this chapter.

10.6 Sample Crossover and Delay Filter Specification

This is the basic specification for a loudspeaker crossover that we will use as an example for this chapter.

No of bands	Three
Type	Linkwitz–Riley 4th order
MID/HF crossover frequency	2.5 kHz
LF/MID crossover frequency	400 Hz
HF delay	80 usec, tolerance +/−5%
MID delay	400 usec, tolerance +/−5%

Most of the delay filter examples in this chapter will be designed for the 80 usec delay.

10.7 Allpass Filters in General

When I first encountered the phrase "allpass filter," my immediate reaction—and here I suspect I am not alone—was "What's the point of that, then?" The point is, of course, that while the frequency response is flat, the phase-shift varies with frequency. A less enigmatic name would be "variable-phase-shift-only filter" but I'm quite sure that won't catch on.

The general form of an allpass filter is shown in Figure 10.6. There are two signal paths, one of which goes through the box labeled T(s), which just means that it has a frequency response that varies with frequency. It will be either a simple RC first-order filter or a second-order bandpass filter; its output is amplified by two and subtracted from the original signal.

It is important to be aware that allpass filters do not inherently have an absolutely flat frequency response. Due to the finite gain-bandwidth-product of the opamps, and

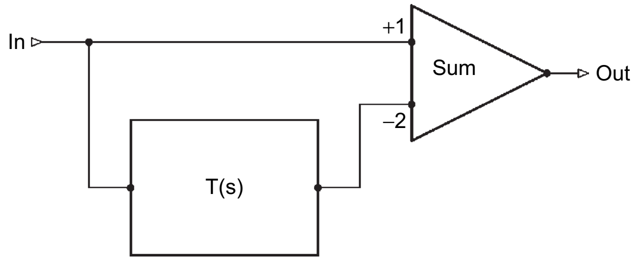


Figure 10.6: The general form of an allpass filter. The path bypassing T(s) must be inverted and have half the gain.

components that do not have mathematically exact values, the magnitude response is likely to show small deviations from perfect flatness. These deviations are, however, usually very small and of no consequence compared with speaker unit tolerances.

10.7.1 First-Order Allpass Filters

Figure 10.7 shows the two versions of the basic first-order allpass filter. This deceptively simple circuit is analogous to a first-order RC filter and so gives a slow phase change as frequency changes.

The first version in Figure 10.7a is non-inverting at low frequencies, in other words the phase shift is 0°. As the frequency rises the phase shift increases until it approaches 180° (inverting) at high frequencies, as shown in Figure 10.10 below. This changing phase-shift is equivalent to a delay. The sharp-eyed reader will note that in Figure 10.10 the phase has actually reached 180°, and is clearly headed for more. This plot was produced by simulating the circuit with a model of a real opamp, rather than just evaluating a mathematical equation, and the extra phase-shift has accumulated because of the finite open-loop bandwidth of that opamp. This is what will happen when you build real crossovers, but with any modern opamp the effects are negligible. Looking at Figure 10.10 and Figure 10.11 we can see that the phase is still changing quite quickly in the 10 kHz – 20 kHz range (from −135° to −158°, i.e., 21°) when the delay has fallen to a quarter of its maximum value. This is not going to give us a minimum-phase crossover.

When the positions of the resistor and capacitor are exchanged, the same phase change is obtained but phase inverted, so the second version in Figure 10.7b inverts at low frequencies but has an in-phase output at high frequencies. An allpass filter gives twice the maximum phase-shift of an ordinary filter of the same order. The non-inverting version of Figure 10.7a could also be called the RC version, as the resistor comes before the capacitor, forming a lowpass filter. This means the common-mode voltage will fall with frequency. The inverting version of Figure 10.7b is correspondingly the CR version, and the capacitor-resistor arm now acts as a highpass filter.

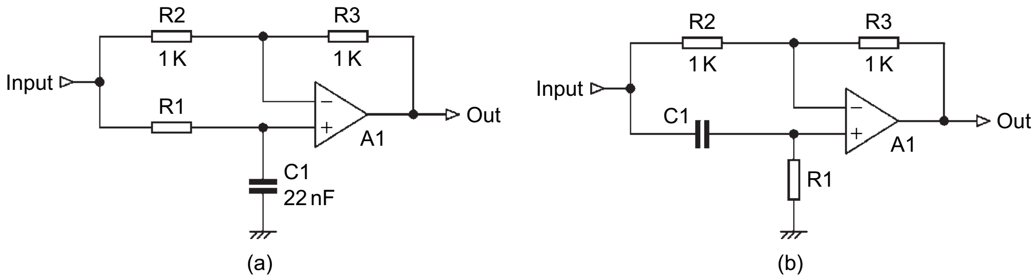


Figure 10.7: Two versions of the first-order allpass filter: (a) Non-inverting or RC type (b) Inverting or CR type.

A first-order allpass filter has only one parameter to set; the RC time-constant. The delightfully simple calculation for the low-frequency delay is shown in Equation 10.3. There is no such thing as the Q of a first-order allpass. The output of an allpass filter does not have a "turnover" as such. According to some authors its operation is defined by the frequency at which the group delay has fallen to 1/√2 of its low frequency value; while this corresponds in a way to "−3 dB" for a amplitude-frequency filter, it has no real significance in itself. A more useful description is the frequency at which the phase-shift reaches 90°, half-way between the extremes of 0° and 180°, because it is derived very easily from the circuit values, as in Equation 10.4. Note that these are not the same frequencies.

$$\text{Delay} = 2\text{RC} \tag{10.4}$$

$$f_{90} = \frac{1}{2\pi \text{RC}} \tag{10.5}$$

You will recall that in our example loudspeaker, the tweeter signal had to be delayed by 80 usec. The resulting RC allpass circuit is shown in Figure 10.8.

If you trustingly feed a step waveform into Figure 10.8, what emerges is an immediate negative spike followed by a long slow approach to the steady input voltage as shown in Figure 10.9. "Doesn't look much like a delay to me!" I hear you cry, which was pretty much what I cried when I first tried this experiment.

The reason for this unhelpful-looking output is that allpass filters do not give a constant time delay with frequency, and this one is no exception. Figure 10.11 shows how as the frequency goes up the group delay is indeed 80 usec initially, but begins to decrease early as the frequency increases. Figure 10.10 gives the phase response for comparison. The delay is down by 10% at 2.5 kHz, and down to 50% at 9.3 kHz, slowly approaching zero above 100 kHz. This is clearly not much use for equalising the delay in the HF path. Since there is only one variable—the time-constant R3, C1—the only way to keep the delay constant to

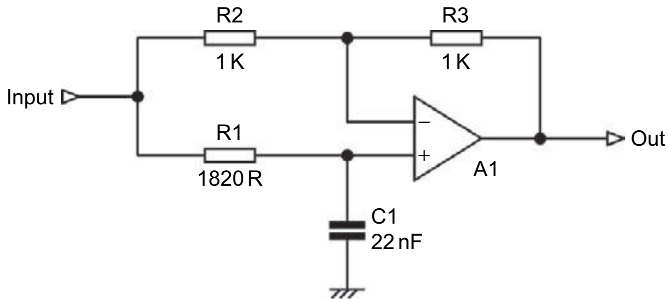


Figure 10.8: A non-inverting RC first-order allpass filter designed for a group delay of 80 usec.

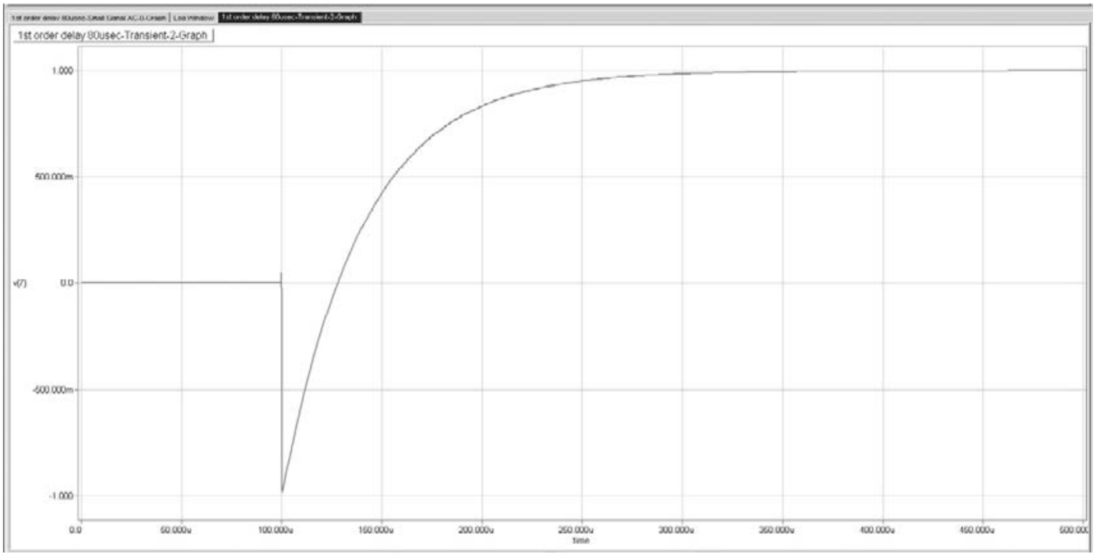


Figure 10.9: The disconcerting response of the 80 usec first-order allpass filter to a 1 V step input.

higher frequencies would be to reduce its value, which would make the filter useless for implementing the required delay compensation.

The fall-off in delay with frequency explains Figure 10.9; the high frequencies that make up the edge of the input step-function are hardly delayed at all, and since they are subject to a 180 degree phase-shift, give the immediate inverted spike. The lower frequencies are delayed but get through eventually, causing the slow rise. The mistake we have made is applying a stimulus waveform with a full frequency range.

Figure 10.12 shows the rather more convincing result obtained if the 1 V step input is band-limited before it is applied to our allpass filter. The step input (Trace 1) has been put

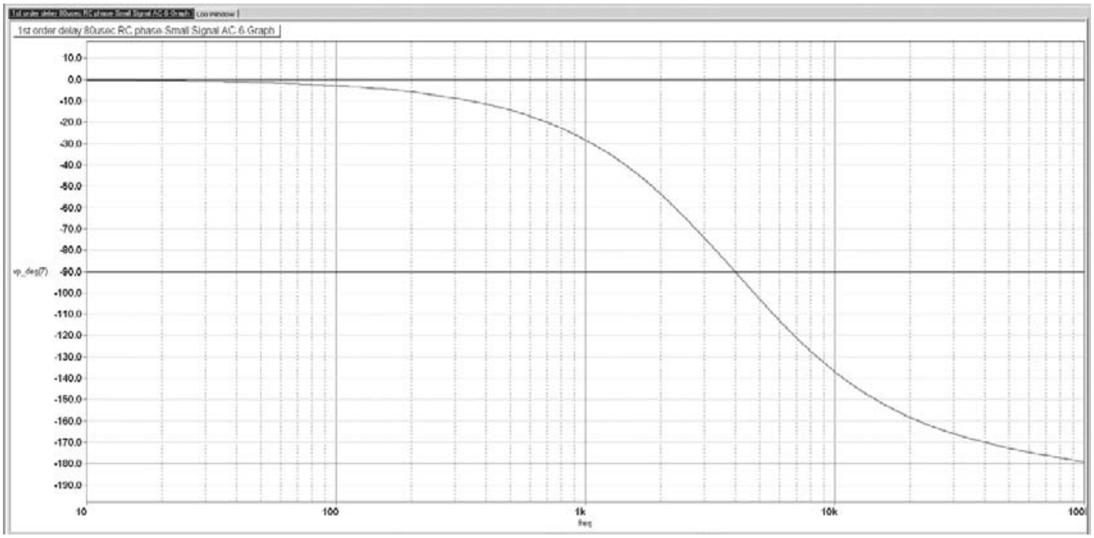


Figure 10.10: The phase response of the 80 usec first-order RC-type allpass filter. The phase shift is still changing at the top of the audio spectrum.

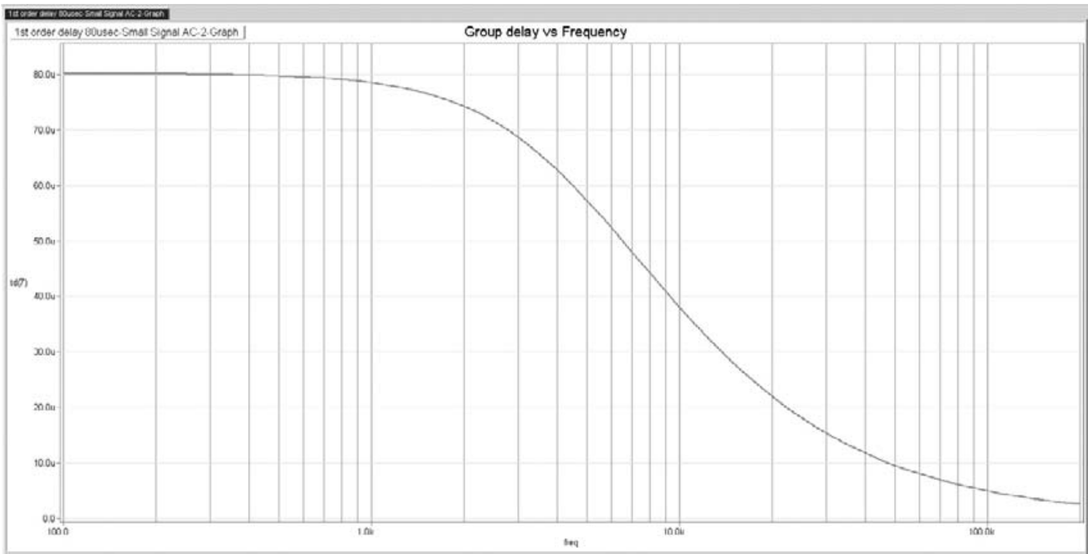


Figure 10.11: The group delay response of the first-order 80 usec allpass filter. The delay is down by 10% at 2.5 kHz, and 50% at 9.3 kHz.

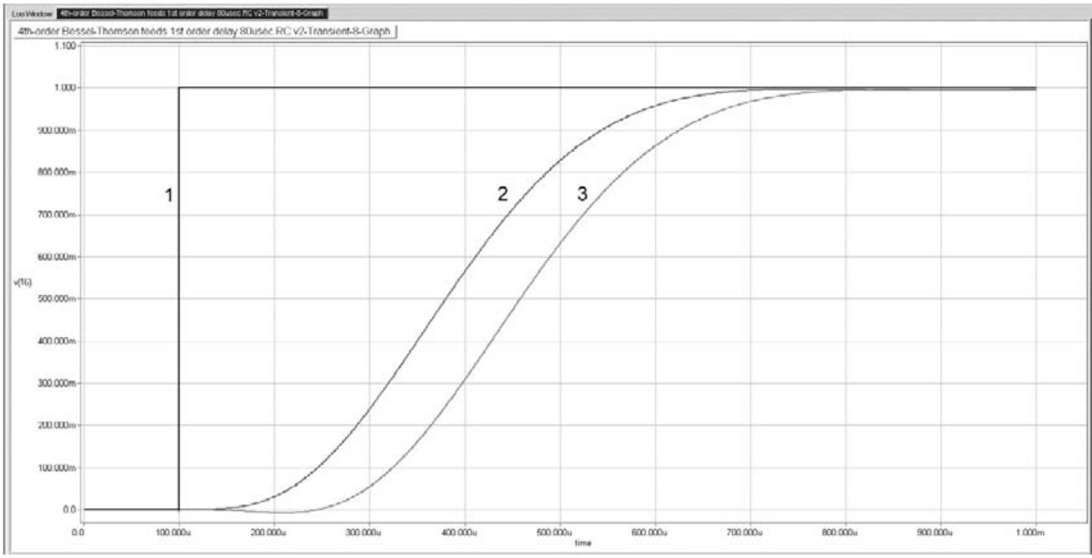


Figure 10.12: The 1 V step input is Trace 1, putting it through a fourth-order Bessel-Thomson filter gives Trace 2, and the 80 usec first-order allpass filter delays it to give Trace 3.

through a fourth-order Bessel lowpass filter with a −3 dB frequency of 1.2 kHz, the Bessel characteristic being chosen to prevent overshoot in the filtered waveform; a Butterworth filter would have given a bit of overshoot and confused matters. The Bessel filter output is the leisurely rising waveform of Trace 2. When this goes through the allpass filter it emerges as Trace 3 with almost exactly the same shape but delayed by 80 usec. That looks a bit more like a delay, eh?

If you look closely at Figure 10.12 you will see that the output of the allpass filter does show a very small amount of undershoot before it rises. This could have been further suppressed by reducing the cut-off frequency of the lowpass filter. You will also note that the output of the fourth-order Bessel lowpass filter, Trace 2, has already been delayed a good deal compared with the step input Trace 1; in fact far more than by the simple first-order allpass filter. The use of Bessel lowpass filters to delay signals is examined at the end of this chapter.

The circuit in Figure 10.7a must be non-inverting at low frequencies, because when C1 is effectively an open-circuit, we get a non-inverting stage because of the direct connection to the non-inverting input. Likewise, in Figure 10.7b, when C1 is effectively open-circuit the configuration is clearly a unity-gain inverter.

The input impedance of the RC version in Figure 10.8 is 2.7 kΩ at 10 Hz, falling to 646 Ω at 1 kHz, whereafter it remains flat (it happens to be 666 Ω at 100 Hz but I don't think you should try to read too much into that). The equivalent CR version (with all component values the same) has an input impedance that is flat at 1 kΩ from 10 Hz to 1 kHz. Above that the impedance rises slowly until it levels off at 1.8 kΩ around 30 kHz. It is pretty clear that the CR version will be an easier load for the preceding stage, especially at high audio frequencies, and this will have its effect on the distortion performance of that stage. There is more on that in the section below on the performance of third-order allpass filters.

The group delay plot in Figure 10.11 shows an elegantly sinuous curve, quite unlike the tidy straight-line approximation roll-offs we are used to when we look at filter amplitude responses. These latter of course are plotted logarithmically with dB on the vertical axis, whereas here we have group delay time as a linear vertical axis. This is how it is normally done; dB are useful for measuring amplitude, not least because they follow the logarithmic nature of how we perceive loudness, but there is no perceptual analogue for how we experience small time delays. Still, I thought it might be interesting to plot group delay with log-of-delay as the vertical axis, and the result is seen in Figure 10.13. I don't recall seeing this done before.

You can see that it looks very much like the amplitude plot of a 6 dB per octave roll-off, with a linear roll-off to the right. Is this a more useful way of plotting group delay? Probably not, because it tends to de-emphasise variations in the region of maximum group delay, which is where we are most concerned. Interesting picture, though.

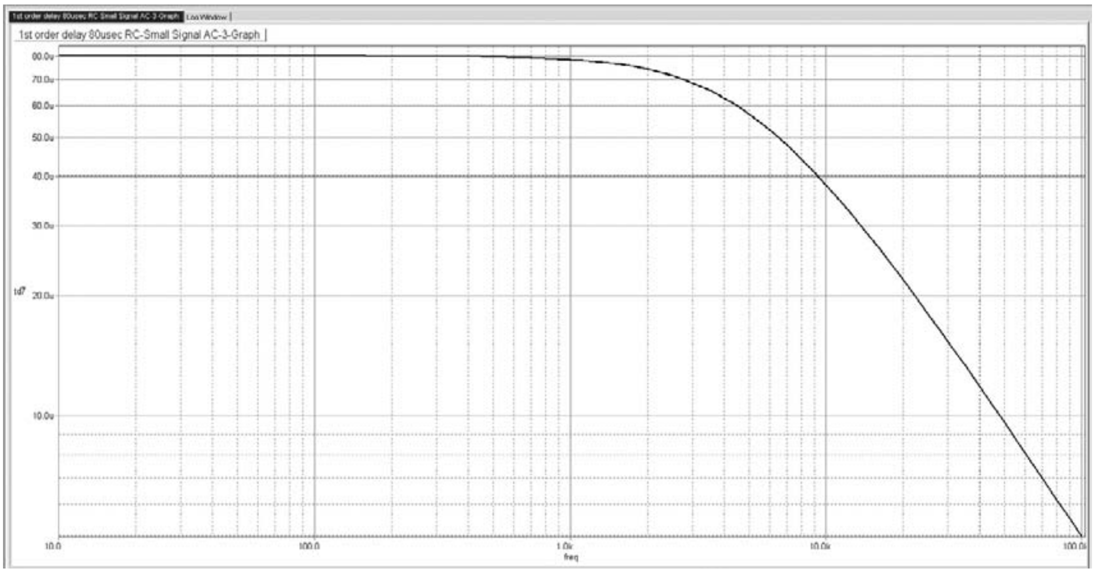


Figure 10.13: The group delay response of the first-order 80 usec allpass filter plotted with a log-delay vertical axis. The delay is down by 10% at 2.5 kHz.

Distortion and Noise in First-Order Allpass Filters

While the two versions of the first-order allpass may appear to be functionally identical, apart from the phase inversion, they do in fact have different distortion behaviour at high frequencies. The non-inverting version or RC version (Figure 10.7a), has the resistor before the capacitor, forming a lowpass filter, so the common-mode (CM) voltage will fall with the frequency. The inverting or CR version (Figure 10.7b) has a capacitor-resistor arm that acts as a highpass filter, so the CM voltage will rise with the frequency. This makes a very significant difference to the HF distortion, as shown in Figure 10.14; at 20 kHz the CR version gives 0.0026% as opposed to 0.00070% for the RC version. The allpass filter measured was the 80 usec design, the non-inverting version of which is shown in Figure 10.8. Polypropylene capacitors were used to eliminate any capacitor distortion. Note that the 9 Vrms test level is close to the maximum possible, and practical internal levels such as 3 Vrms will give significantly lower levels of distortion.

When the input level is 9 Vrms, the RC version has a CM voltage of 1.67 Vrms at 20 kHz. The CR version has a CM voltage of 8.84 Vrms at 20 kHz, more than five times as much,

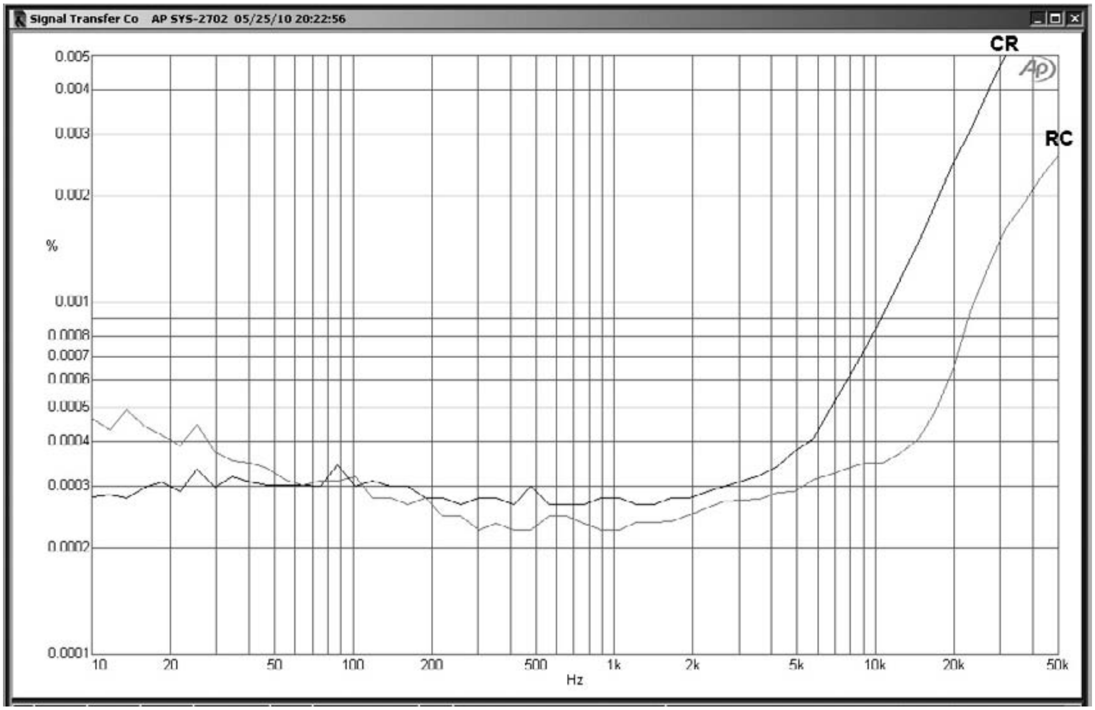


Figure 10.14: Distortion plots for RC and CR first-order 80 usec allpass filters using 5532s. The CR version has more HF distortion because the common-mode voltage is high at high frequencies, and vice versa for the RC version. Polypropylene capacitors, input 9 Vrms.

leading to much-increased HF distortion. First-order allpass filters with longer delays have larger capacitors and so the difference extends over a wider frequency range. The measurements in Figure 10.14 were taken using 63 V polypropylene capacitors; however, replacing these with 100 V polyester microbox capacitors in both RC and CR versions made no measurable difference to the THD plots. This was somewhat unexpected.

It is therefore well worthwhile using the non-inverting (RC) version wherever possible. If there is a need for a phase-inversion, then it may be worth looking for some other part of the signal path in which to place it, where it will not cause degradation of the distortion performance.

The noise output for the non-inverting 80 usec version is −110.3 dBu (22–22 kHz). The noise output for the inverting 80 usec version is −110.7 dBu (22–22 kHz). The difference is small but real.

Cascaded First-Order Allpass Filters

The delay required in a typical crossover is too long to be effectively provided by a single first-order allpass filter, because the longer the delay used, the lower the frequency at which that delay starts to roll off. See Chapter 19 for a real example. The Siegfried–Linkwitz crossover design in [2] addresses this problem by using three first-order filters in series, as shown in Figure 10.15. This spreads the delay over three sections, allowing each one to be set to a lower delay which can be sustained up to higher frequencies. The RC version of the filter is used here because it is non-inverting and has a markedly superior distortion performance.

As Figure 10.16 shows, this approach greatly improves the delay flatness, which is now 10% down at 7.2 kHz, and down 50% at 28 kHz, outside the audio band. However, 7.2 kHz is only 1.5 octaves clear of the 2.5 kHz MID-HF crossover frequency. The signal to the MID drive unit will, allowing for the initially shallower filter slope, be 35 dB down at that point. Its acoustic contribution will probably be less than that as it is hardly likely to have a flat response 1.5 octaves above the crossover frequency, and may be significantly down; on the other hand, it might have an ugly peak. In view of this uncertainty, it is hard to judge if we have sufficient delay flatness to avoid audible effects, but we can see from Table 10.3 that a −35 dB signal is

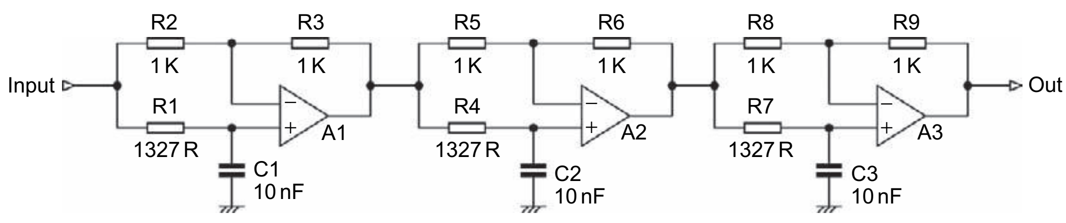


Figure 10.15: Three first-order allpass filters in series, designed for an overall group delay of 80 usec.

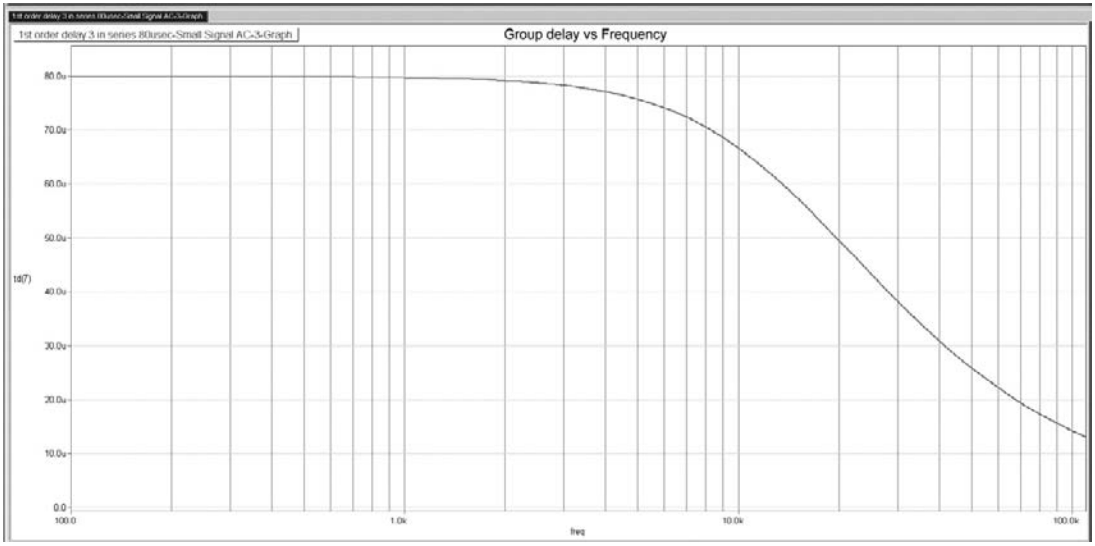


Figure 10.16: The group delay response of the three first-order allpass filters in series. The 80 usec delay is now down 10% at 7.2 kHz, a higher frequency by a factor of three.

only going to increase the total amplitude by 0.15 dB; I am inclined to think we are fairly safe. It would of course be easy to divide the required group delay over more stages, to further extend the flat delay response, but since every stage adds a certain amount of noise, distortion, and cost, and requires power, this is not an attractive route.

While the amplitude perturbations resulting from this solution may be tiny, a further potential worry is the varying phase-shift of the filter above the frequencies at which we are trying to keep the group delay constant.

Figures 10.10 and 10.11 above, for a single first-order section, showed that the phase is still changing quite quickly in the 10 kHz – 20 kHz range. Our triple-allpass scheme has flattened the delay/frequency curve, but has it helped with the phase? No. This now changes considerably more from 10 kHz to 20 kHz, going from −240° to −357°, a change of 157°. This very large increase is not just because we now have three filters all phase-shifting, but also because the part of the phase curve with the most rapid changes is now moved up over the 10 kHz – 20 kHz range. The audibility of this sort of thing is still a matter of debate, but anyway it's not clear there's very much we can do about it. The only solution would be make the group delay effectively constant up to a very high frequency, well out of the audio band. Let's assume we grit our teeth and accept a phase-shift that goes up to 10° at 20 kHz. That implies raising the group delay −10% frequency by no less than 57 times, to 410 kHz. This is hardly practical—here we need three stages just to achieve 7.2 kHz—and so it looks like we are going to have to live with frequency dependent phase-shifts if we use this sort of delay technology.

The iterated nature of the multiple-first-order circuit means we are using three times as many components. A more efficient approach is to use a higher-order allpass filter to create the desired delay.

10.7.2 Second-Order Allpass Filters

Second-order allpass filters have the advantage that you can get a flatter delay response using less components than the cascaded-first-order approach. The disadvantage is that they are conceptually more complex.

The usual method of making a second-order allpass filter is the 1–2 BP configuration, where the signal is fed to a conventional second-order bandpass filter with a passband gain (i.e, gain at the response peak) of two times, and then subtracted from the original signal, as in Figure 10.17. It is not exactly intuitively obvious, but this process gives a flat amplitude response and a second-order allpass phase response. This method should not be confused with the 1-BP configuration, where the bandpass filter has unity gain; this gives a notch amplitude response. To obtain a maximally flat delay response the Q of the bandpass filter must be 0.5. (Note that this differs from the amplitude response of highpass and lowpass filters, where a Q of 1/√2, i.e, 0.707, is required to get a maximally flat Butterworth response.)

A second-order allpass filter has two parameters to set instead of one—we now have both frequency and Q to set for the bandpass filter. A convenient configuration is the Multiple-FeedBack (MFB) type as shown in two forms in Figure 10.18, which has the useful property of inverting the signal, so the required subtraction from the original signal can be implemented simply by summing original and filter output together. A vital point to be aware of is that if you use the more common form of MFB filter in Figure 10.18a, with its

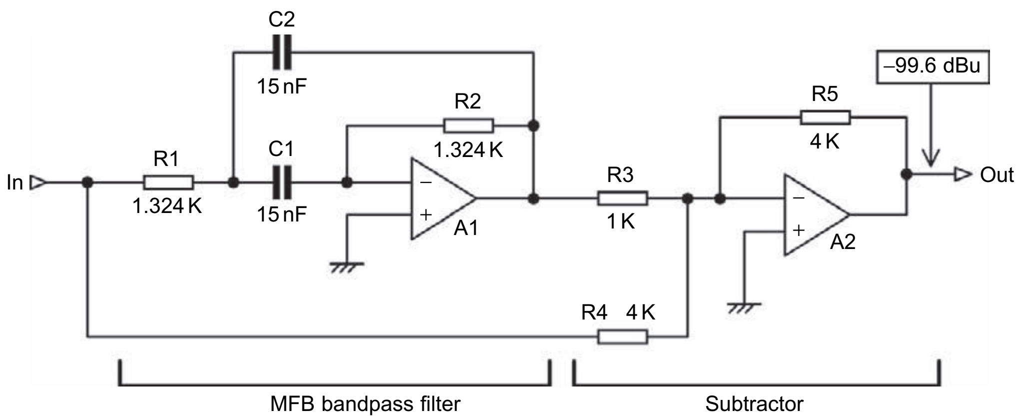


Figure 10.17: A second-order allpass filter using the 1–2 BP principle, designed for a group delay of 80 usec. The MFB filter has a gain of –6 dB so R3 = R4/4.

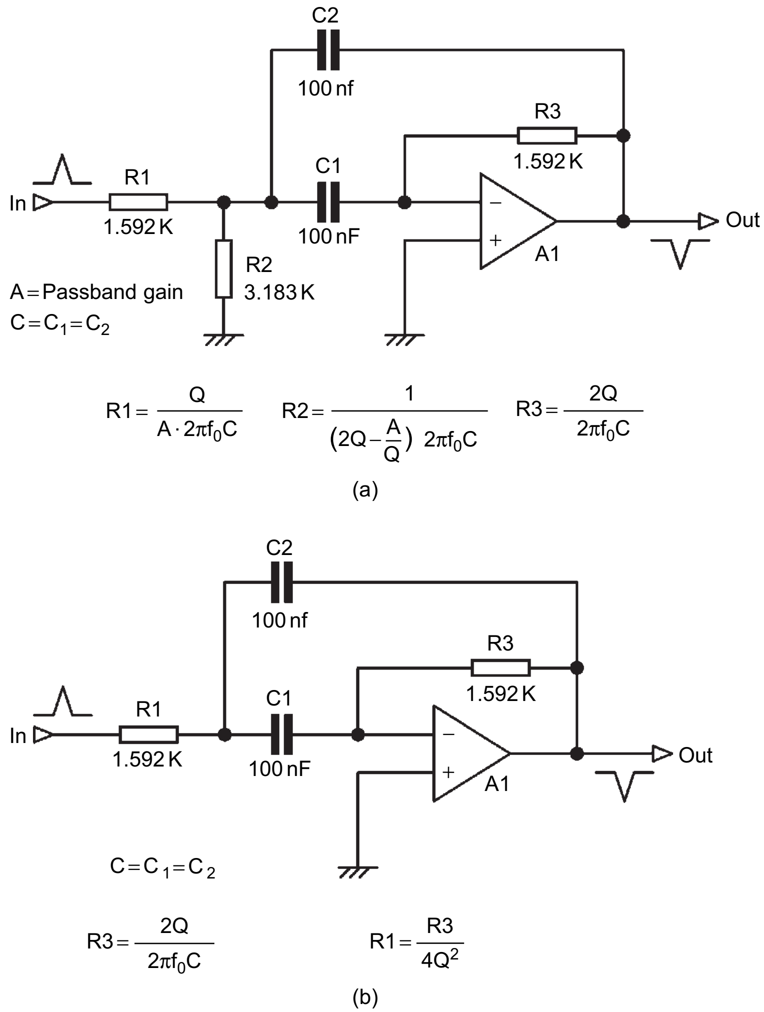


Figure 10.18: Two Multiple-FeedBack (MFB) bandpass filters with centre frequencies of 1 kHz; (a) The more usual version, with design equations. Passband gain = 1 and Q = 1; (b) The low-Q Deliyannis version with design equations. Passband gain = 0.5 and Q = 0.5. Note the similarity in component values.

associated design equations, it is not possible to get a Q of less than 0.707, and we need here a Q of 0.5. If you put a Q value of 0.707 into these equations, then R2 comes out as infinite. This is not a problem—you just leave it out altogether—but it also means that putting in values of Q less than 0.707 yield negative values for R2, which is a little harder to implement (though I hasten to add I have tried it, and it does work).

A simpler approach is to use the Deliyannis version of the MFB filter shown in Figure 10.18b; the circuit is the same as Figure 10.18a with infinite R2, but the vital point is that different design equations are used. It is now not possible to set passband gain independently of Q. The values shown give a Q of 0.5, which also sets the passband gain at 0.5 (−6 dB). This is only true in this case and in general the passband gain is not equal to Q.

To design the filter, first choose a value for C (C1 and C2 are equal) and put that and the required centre frequency into Equation 10.6 to get R3, then use Equation 10.7 to calculate R1.

$$R2 = \frac{2Q}{2\pi fC} \tag{10.6}$$

$$R1 = \frac{R2}{4Q^2} \tag{10.7}$$

A practical implementation of this method is shown in Figure 10.17, which gives an 80 usec delay at low frequencies. To achieve this the MFB filter, consisting of R1, R2, C1, C2 and A1, is designed for a centre frequency of 7.91 kHz and a Q of 0.5. Its inverted output allows the required subtraction from the original signal to be done simply by summing original and filter output together by the shunt-feedback stage A2. Note that R3 is one quarter the value of R4; this implements the 2 in the 1–2 BP. Because of the low Q required in the MFB for a maximally flat delay, its passband gain is −6 dB. This means that to achieve 1–2 BP, the resistor R3 must be a quarter of the value of R4 and R5. In this case R4, R5 can conveniently be made up of two 2.0 kΩ resistors in series. The group delay is shown in Figure 10.19; it is down 10% at 4.79 kHz, which is predictably better than the first-order filter (10% down at 2.5 kHz), but worse than the triple first-order filter (10% down at 7.2 kHz).

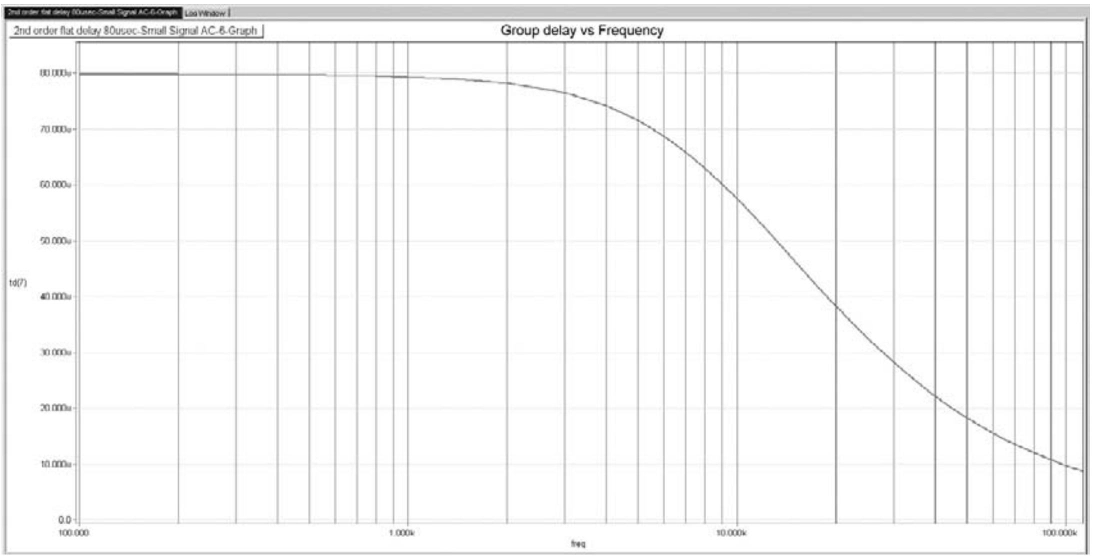


Figure 10.19: The group delay response of the second-order 80 usec allpass filter. The delay is now down 10% at 4.79 kHz.

Second-order allpass filters of this type give a phase inversion at low frequencies, and also at high frequencies; as mentioned before, an allpass filter gives twice the phase-shift of a conventional filter of the same order. The only time when the phase is non-inverting (0° phase-shift) is as the phase plot goes through 0° as it moves smoothly from 180° to −180°.

It is possible to make a second-order allpass filter that only requires one opamp, but there are compromises on noise, gain, and component sensitivity that make this unattractive for our purposes. For reference, two kinds of one-opamp second-order allpass filter are shown in Figure 10.20.

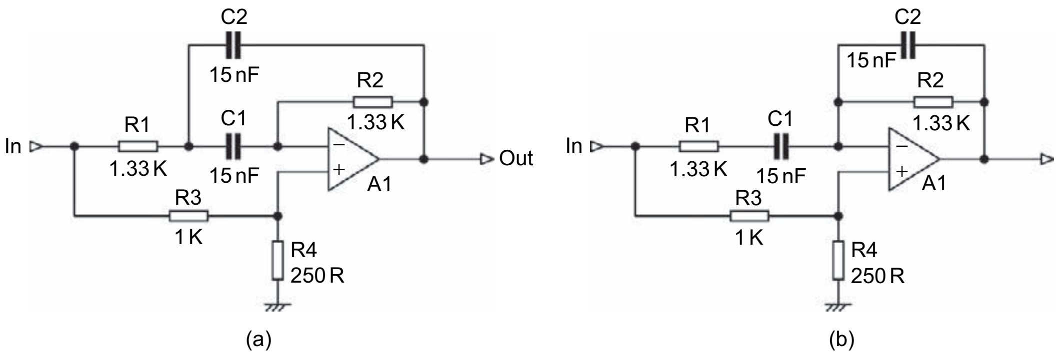


Figure 10.20: Two second-order allpass filters using only one opamp: (a) The Delyiannis circuit; (b) The Lloyd circuit.

The first version in Figure 10.20a, is also generally known as the Delyiannis circuit [3], and is essentially an MFB bandpass filter (R1, R2, C1, C2) with an additional subtraction path via the attenuator R3, R4; the basic aim is to implement the circuit of Figure 10.17 in one stage. The values shown give an 80 usec delay that is 10% down at 4.8 kHz with a maximally flat delay response. While this circuit is undoubtedly minimalist, it has the great drawback that the lower the Q, the lower the gain. Setting the Q to the required value of 0.5 results in a stage gain of 1/5, or −14 dB, and the considerable amount of amplification required to get the output back to the input level will introduce a lot of additional noise; the extra amplifier stage required to do it also completely undoes the cost advantage of a one-amplifier allpass filter.

Figure 10.20b shows another second-order allpass filter, known as the Lloyd circuit [4]. R1, C1, R2, C2 form a second-order bandpass response; this configuration is only capable of a low Q but that is all we need. Once more there is a subtraction path via R3, R4. The values shown give an 80 usec delay that is 10% down at 4.8 kHz. It is notable that the circuit values are the same as for the Delyiannis circuit, despite the different configuration. This circuit has a delay response equivalent to two cascaded first-order stages, but uses only opamp. However, once again the stage shows a considerable loss of −14 dB which must be made up somewhere, and my conclusion is that pursuing economy in this way is not a very hopeful quest.

Two other second-order allpass configuration were thoroughly investigated by Robert Orban in 1991 [5]; these were the Bruton circuit [6] using two opamps, and the Steffen circuit, using one [7]. Neither seem to have any great advantages over the 1–2 BP method, but the Bruton beats the Steffen if a very flat amplitude response is required.

Distortion and Noise in Second-Order Allpass Filters

The second-order filter in Figure 10.17 was built using 5532 opamps and polypropylene capacitors to eliminate any capacitor distortion; the measured THD results are shown in Figure 10.21. The THD falls rapidly from 0.0022% at 20 kHz as frequency decreases, but then peaks in the range 2 to 10 kHz, as this is the region where the MFB filter output, with its own distortion content, makes a significant contribution. Distortion from the MFB filter is low because of its shunt configuration, which eliminates CM voltages, and because the output voltage is low due to its −6 dB passband gain; the MFB THD does not exceed 0.0004%.

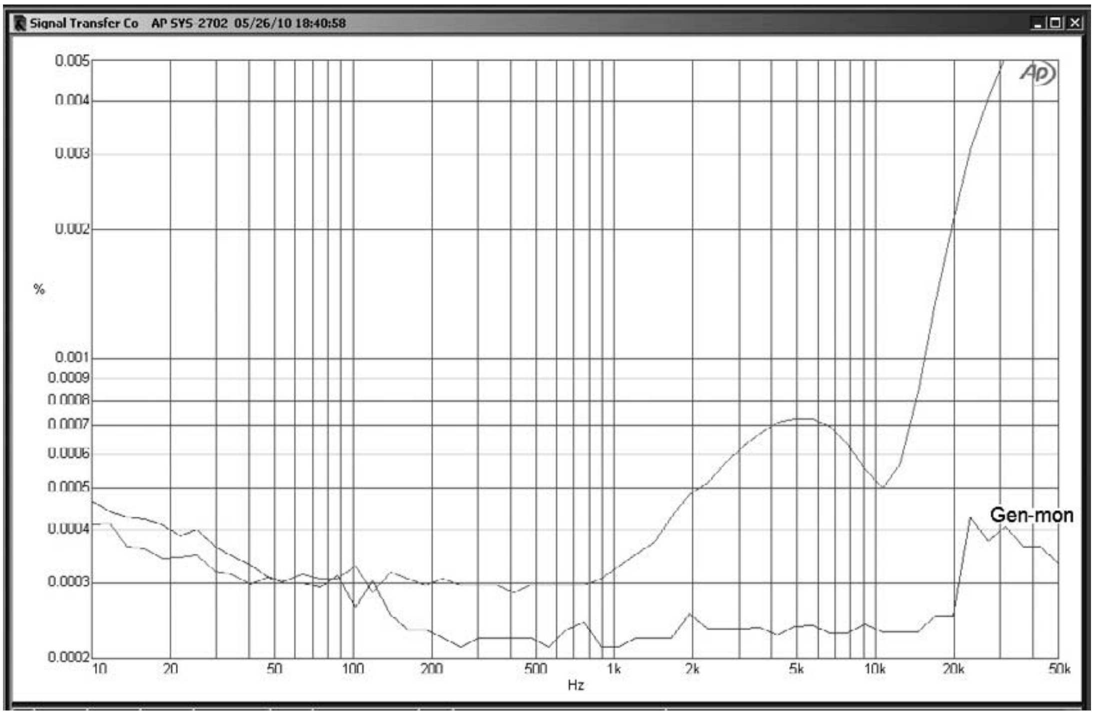


Figure 10.21: Distortion plot for the second-order 80 usec allpass filter in Figure 10.16, using 5532s. The THD drops rapidly from 0.0022% at 20 kHz as frequency falls, but peaks around 2 to 10 kHz where the MFB makes a significant contribution. High-frequency distortion above 10 kHz is solely from the second opamp. Polypropylene capacitors, input 9 Vrms.

The high-frequency distortion above 10 kHz comes solely from the second opamp, as the MFB filter output is 10 dB down at 20 kHz.

The circuit was then remeasured using 63 V polyester microbox capacitors, to find out how much extra distortion they introduced. The answer is, not much. The THD in Figure 10.20 now peaked at 0.0009% at 5 kHz instead of 0.0007% for the polypropylene capacitors. The modest increase is due to the low signal level in the MFB filter.

The noise output is −99.6 dBu (22–22 kHz), which is more than 10 dB noisier than the first-order stage measured above. The greatest proportion of this extra noise comes from the second stage A2, which because of the low value of R3, is working at a noise gain of 15.6 dB. When R3 is disconnected, the noise output drops precipitately to −109.5 dBu; when, however, R3 is connected to ground instead of the MFB filter output, noise only increases to −102.1 dBu, showing that about half of the extra noise is produced by the filter.

The high noise gain in the second stage is inherent in the 1–2 BP mode of operation, because of the subtraction involved, and it is not easy to see a way around that.

10.7.3 Third-Order Allpass Filters

A third-order allpass filter is made up of a second-order allpass cascaded with a first-order allpass, in the same way that third-order frequency/amplitude domain filters are constructed. The second-order allpass is arranged to peak in its delay, but the first-order allpass is then designed to cancel out the peaking, giving a maximally flat delay overall. The arrangement is shown in Figure 10.22; note that the first-order stage is now of the inverting or CR type, which undoes the inversion performed by the second-order stage. (As we noted earlier in the chapter, this version has an inferior distortion performance to its non-inverting (RC) equivalent; however the overall outcome for distortion is not obvious, as related in the next section on the performance of this filter.)

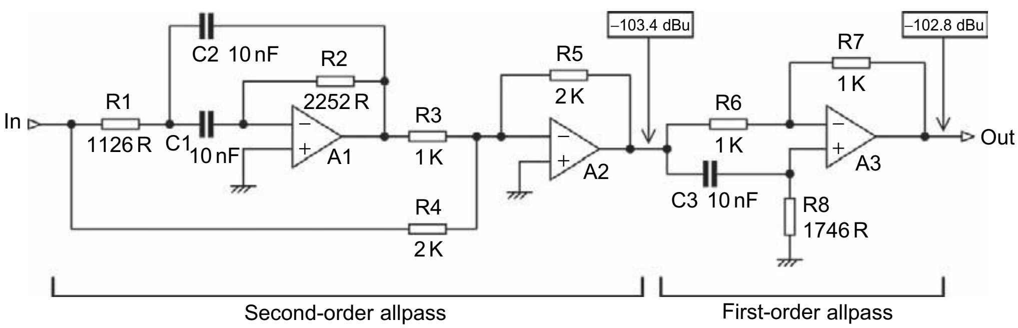


Figure 10.22: A third-order 80 usec allpass filter, made up of a second-order allpass followed by a first-order allpass. Noise at each stage output is shown for 5532s.

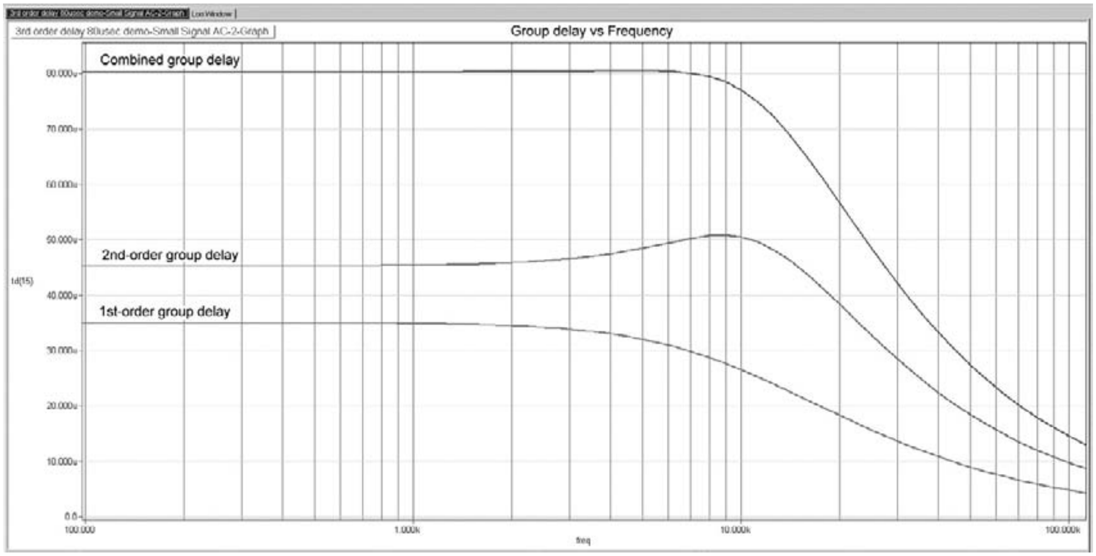


Figure 10.23: The group delay response of the third-order 80 usec allpass filter. The delay is now down 10% at 12.7 kHz.

The output of the second-order allpass is the middle trace in Figure 10.23, peaking just above 10 kHz. When this is combined with the slow roll-off of the following first-order stage, (the lower trace) the final result is a flat response that then rolls-off quickly. This is the top trace in Figure 10.23, where the group delay remains flat and then rolls off rapidly, being 10% down at 12.7 kHz, and 50% down at 31.9 kHz. This gives almost twice the frequency range of the three cascaded first-order networks in Figure 10.15, but actually uses one less resistor. This third-order filter is clearly the better solution.

The delay −10% point is now 2.3 octaves above the 2.5 kHz crossover frequency, so assuming a 24 dB/octave slope, the MID speaker output will be about 55 dB down and so I suggest that the fall-off in time delay will have no audible consequences.

Looking at the low-frequency area to the left, you will see that the second order stage gives a delay of 45 usec; to this is added the 35 usec of the first-order stage, giving a total delay of 80 usec.

The third-order allpass filter has input and output in phase at the low-frequency end. At the high-frequency end the phase shift reaches a total of 540° lagging (180° × 3).

Distortion and Noise in Third-Order Allpass Filters

The third-order filter in Figure 10.22, with its CR final stage, was built using 5532 opamps and polypropylene capacitors. Figure 10.24 shows the distortion at the second-order stage output and the final output. An interesting point is that between 200 Hz and 6 kHz, the final

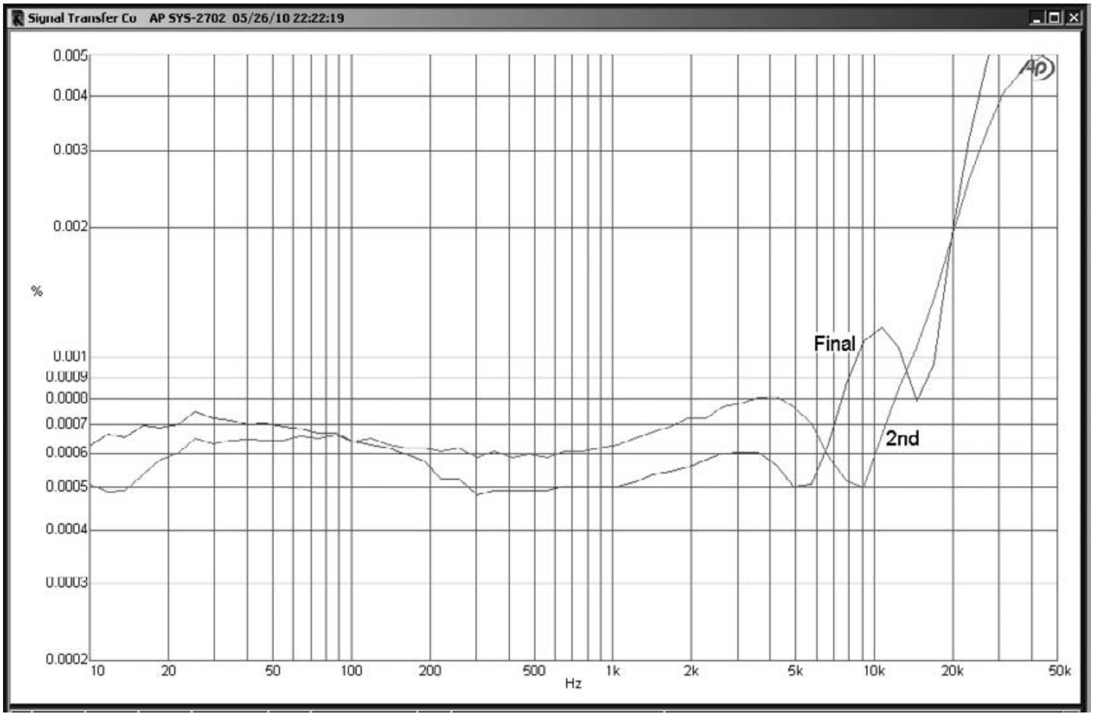


Figure 10.24: Distortion plots for third-order 80 usec allpass filter in Figure 10.22, showing the second-order stage output and the final output. Polypropylene capacitors, 5532s, input 9 Vrms.

output shows a lower THD reading than the intermediate point in the circuit. This is because there is some cancellation of the second-harmonic from the second-order filter in the final stage. Both points show 0.002% at 20 KHz.

Figure 10.25 shows the result of swapping the CR first-order stage for its RC equivalent, which as we saw earlier has in itself lower distortion. We might hope that that would give us a better overall distortion performance, but actually it is markedly worse; we now suffer 0.003% THD at 20 kHz, and perhaps more worryingly, we have THD around 0.0014% in the audible region from 6 kHz to 10 kHz, whereas before we were well below 0.001% in this region.

The alert reader (and I trust you all are) will have noticed that the THD plots for the intermediate second-order output in Figures 10.24 and 10.25 are not the same, and you may very well wonder why this is the case when the second-order stage has not been altered. The answer is that the CR and RC first-order stages differ considerably in how much loading they are putting on the previous stage. The RC version has a three times lower input impedance at high frequencies, and in this circuit that outweighs the fact that it has

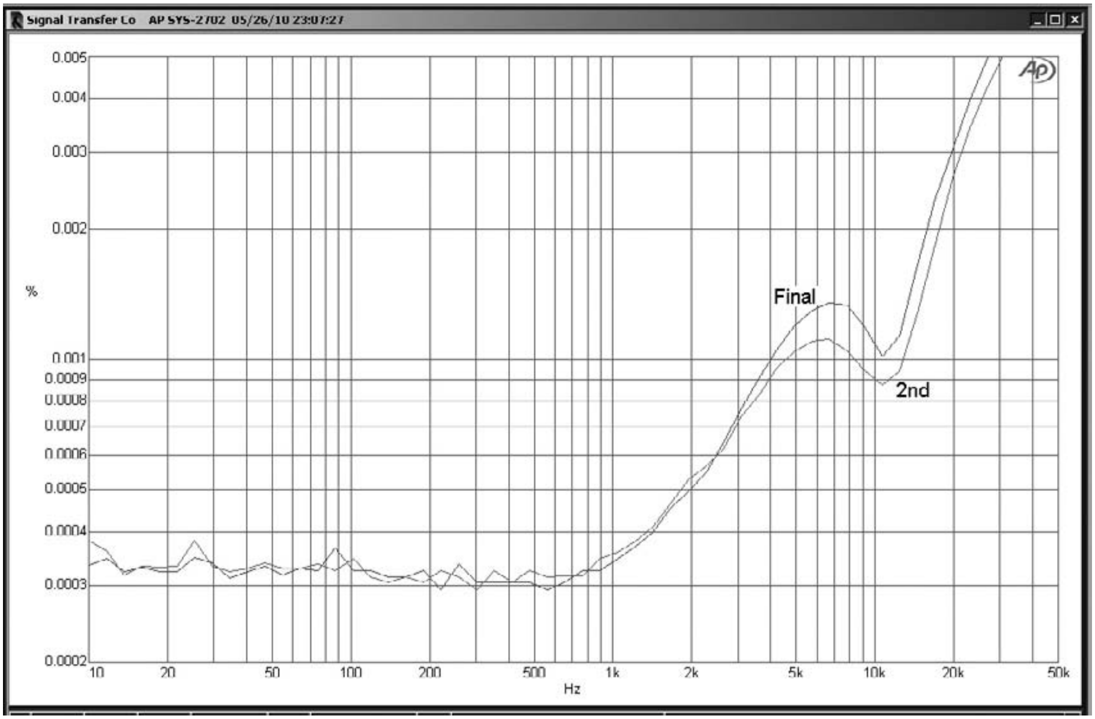


Figure 10.25: Distortion plot for the third-order 80 usec allpass filter in Figure 10.22, but with an RC type first-order filter as the final stage. The final distortion output is unexpectedly worse. Polypropylene capacitors, 5532s, input 9 Vrms.

lower distortion of its own. The input impedances of first-order allpass filters were described earlier in this chapter.

I hope you don't think that I am belabouring a trivial point here. There is in fact a fundamental message—that you must never neglect the loading that one stage imposes on its predecessor if you are aiming for the lowest practicable distortion. To ram the point home I will show one more THD plot. Figure 10.26 shows the THD at the second-order stage output for the RC and CR cases, and also for the unloaded condition (NL) with the first-order stage disconnected to remove its loading altogether.

You can see that the traces for the RC and CR cases are the same as those in Figures 10.24 and 10.25, but the unloaded (NL for No Load) trace is a good deal lower than either of them, except around 5 to 10 kHz where the CR plot gains from some cancellation effect. Only the CR case shows elevated LF distortion from 10 Hz to 5 kHz; this is characteristic of loading on a 5532 output, and it is caused by the relatively low input impedance of the CR version at low frequencies. At higher frequencies the input impedance rises, which is why the CR trace takes a nose-dive at 5 kHz, before coming back up again.

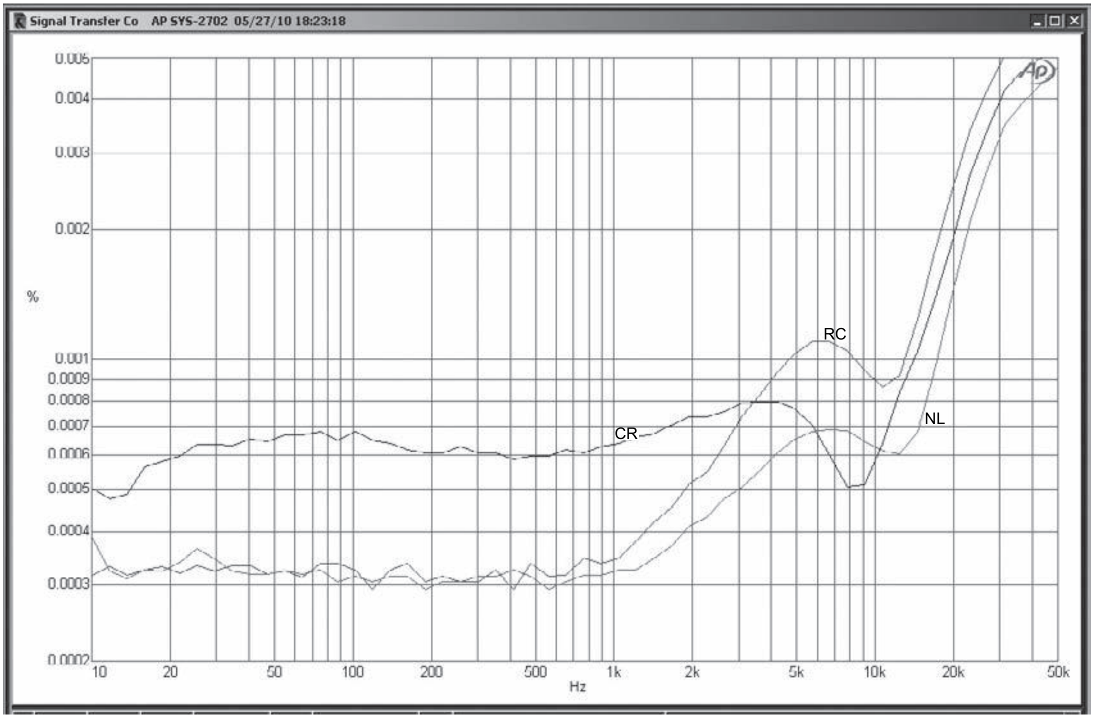


Figure 10.26: Distortion plot for the second-order stage output (A2) of the third-order 80 usec allpass filter in Figure 10.22, when it is loaded by the RC or CR final stage, and also with No Load (NL). Polypropylene capacitors, 5532s, input 9 Vrms.

The 20 kHz THD at the second-order stage output (A2) in Figure 10.26 is really not very good in either the CR or RC case. There are two reasons for this:

1. Opamp A2 in Figure 10.22 is heavily loaded by the input impedance of the first-order stage, as described above.
2. A2 is working at a relatively high noise gain and so has less negative feedback than usual available for linearisation.

Chapter 13 describes how the distortion performance of a heavily loaded 5532 can be improved by biasing the output. This looks like a good place to apply that technique. Adding 3K3 to V+ on the second-order output reduces the LF distortion from around 0.0006% to around 0.0005%; not exactly a giant improvement, but then it is almost free. See Figure 10.27, where there is a general reduction in LF distortion, but we have lost the dip around 8 kHz.

In Figure 10.22 the noise at the output of the second-order stage is −103.4 dBu (22–22 kHz), and the noise at the final output is −102.8 dBu.

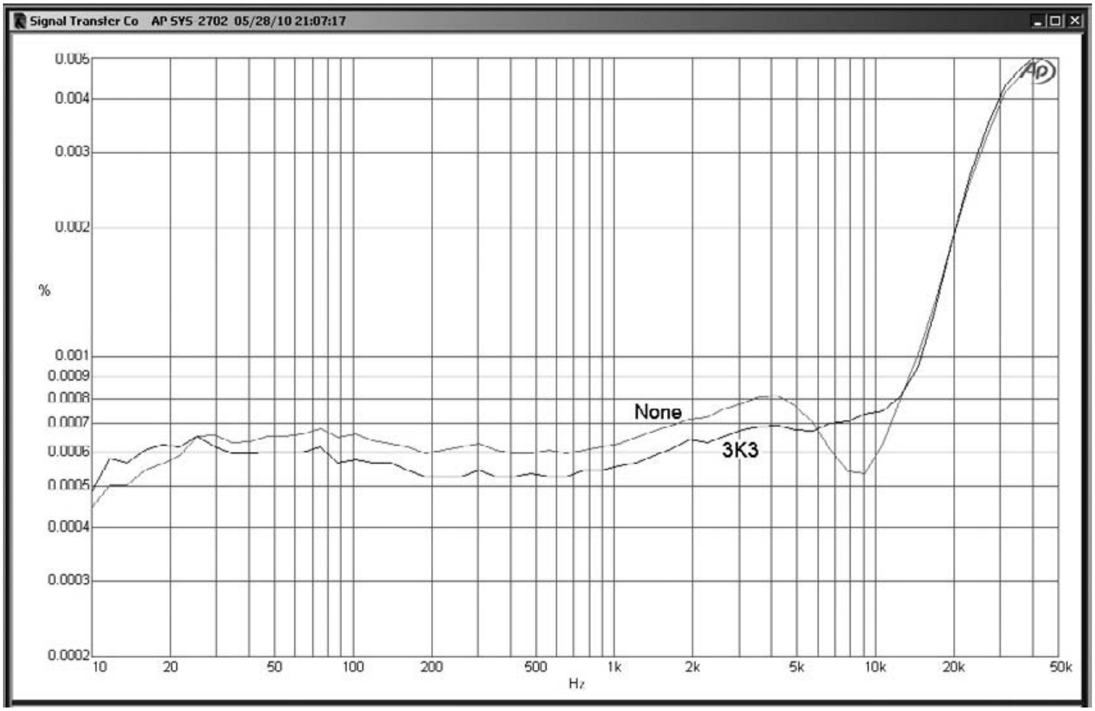


Figure 10.27: Distortion plot for the second-order-stage output of the third-order 80 usec allpass filter in Figure 10.22 is loaded by the CR final stage, only. Adding the 3K3 output biasing resistor reduces distortion below 6 kHz. Polypropylene capacitors, 5532s, input 9 Vrms.

Why is this circuit 3.2 dB quieter than the −99.6 dBu we get from the second-order allpass filter, when we have more opamps in the signal path? It's a subtle chain of reasoning. Firstly, since the second-order stage is now peaking in delay, the MFB filter has a higher Q, and therefore a higher passband gain, here 0 dB. Thus R3 is half the value of R4, R5 instead of a quarter. Therefore the noise gain in the subtractor stage A2 is 3.6 dB lower at 12.0 dB.

This is another reason why a third-order allpass is better than a second-order allpass delay filter; it is not only flatter, it is quieter. Is this an original observation? I suspect so.

10.7.4 Higher-Order Filters

If third-order filters do not provide a sufficiently extended flat-delay characteristic, then higher order filters might be considered. We have just seen that a third-order allpass filter provides the most extended flat-delay response for a given component count, and also has usefully lower noise. If greater delays, or flatter delay characteristics than this configuration can provide are required, higher-order filters may be the answer.

The basic principle, as we saw for the third-order filter, is to take one or more second-order stages with a high Q, giving a peak in the delay response, and then cancel out the peak by cascading it with one or more stages, either first- or second-order, with a low Q. Figure 10.28 shows how higher-order filters are built up. For each stage the Q is that of the overall allpass function, and *not* the Q of the MFB filter, which is a different quantity. For example, in a second-order allpass filter the Q for maximal flatness is 0.58, but the Q of the MFB filter is 0.50. Here the first-order sections have been put at the end of the chain; this makes no difference as all the stages have unity gain.

Designing high-order filters is not something to try by hand, and the usual design procedure is to consult a table of allpass filter coefficients, as in Table 10.4. This assumes you are using the 1–2 BP approach to making allpass filters.

To design your filter, pick an order, and the cut-off frequency *f* required. For each stage of the complete filter first select the value of C = C1 = C2 in the MFB filter here I am using

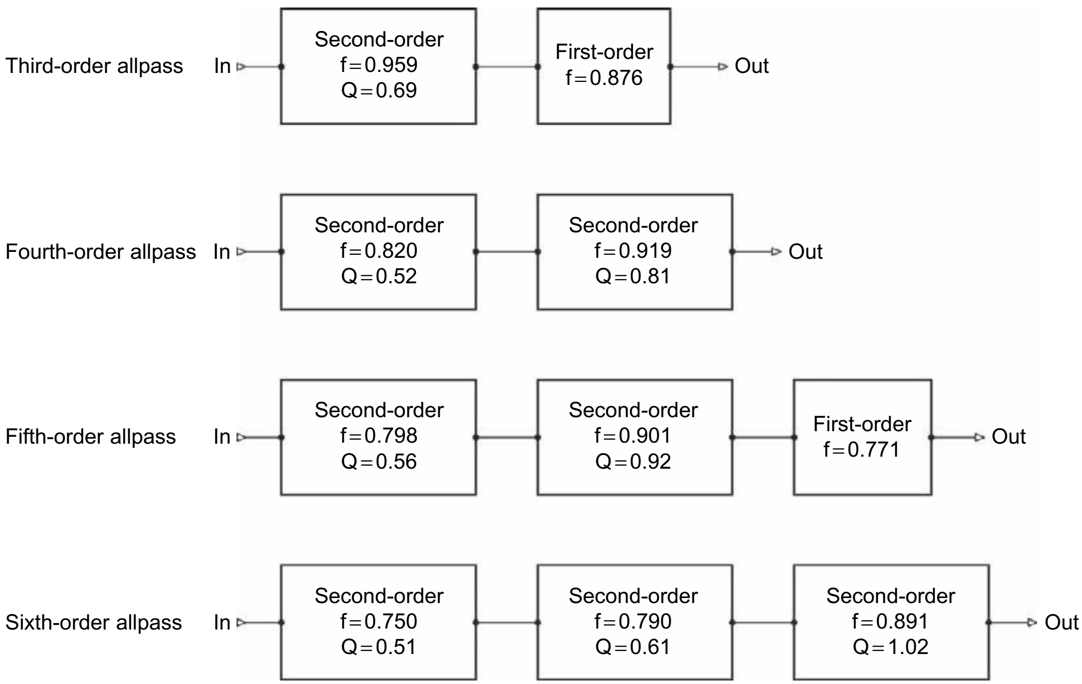


Figure 10.28: Constructing allpass filters from third to sixth order by cascading first- and second-order stages. The relative cut-off frequency and the Q is shown for each stage.

Table 10.4: Filter Coefficients for Allpass Filters (After Kugelstadt)

Filter Order	Filter Section	Coefficient		Section Q	R3 Factor α	Stage Gain
		a_i	b_i			
1st	1	0.6436	n/a			1.000
2nd	1	1.6278	0.8832	0.58	3.000	0.667
3rd	1	1.1415	n/a			1.000
	2	1.5092	1.0877	0.69	2.094	0.955
4th	1	2.3370	1.4878	0.52	3.671	0.545
	2	1.3506	1.1837	0.81	1.541	1.298
5th	1	1.2974	n/a			1.000
	2	2.2224	1.5685	0.56	3.149	0.635
	3	1.2116	1.2330	0.92	1.191	1.680
6th	1	2.6117	1.7763	0.51	3.840	0.521
	2	2.0706	1.6015	0.61	2.677	0.747
	3	1.0967	1.2596	1.02	0.955	2.095

the component designations in the first stage of Figure 10.20; if this results in resistor values that are undesirably low or high then you will have to go back and rethink the value of C. R1 and R2 are then calculated using the coefficients a_i and b_i thus:

$$R1 = \frac{a_i}{4\pi fC} \tag{10.8}$$

$$R2 = \frac{b_i}{a_i \pi fC} \tag{10.9}$$

This sets the frequency and Q of the stage. We now need to work out the parameter α:

$$\alpha = \frac{(a_i)^2}{b_i} \tag{10.10}$$

which lets us set R3 at R/α, where R is the value of R4 and R5, to get the correct subtraction to create an allpass filter; parameter α includes the 2 in the 1–2 BP. Job done, but your very next move should be to run the complete filter on a simulator to make sure you get what you want—it's all too easy to make an arithmetical error.

Figure 10.29 shows the delay response of first- to sixth-order allpass filters all set to the same cutoff frequency. This is the sort of diagram you often see in filter textbooks; but here it is simulated rather than derived mathematically.

More relevant to our needs is to set the group delay to the same value (80 usec again) for each order of filter and you can see how much more extended the flat portion of the delay curve is with the higher-order filters, as in Figure 10.30. It is also painfully obvious that the extra flat-delay bandwidth you gain with each increase of filter order decreases; this is true on a log-frequency plot because with each higher order you actually gain a constant extra bandwidth of

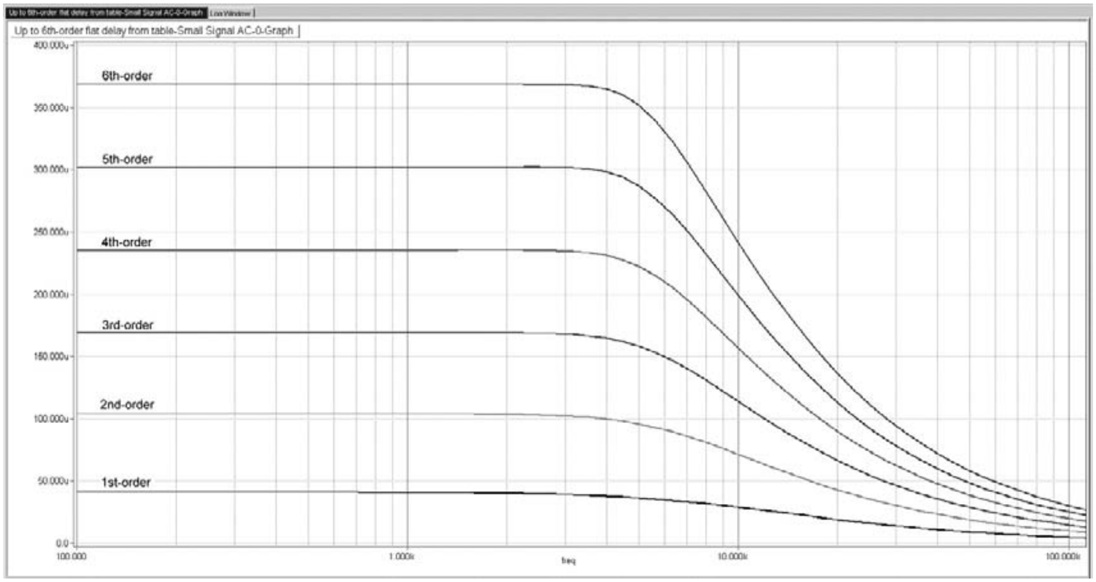


Figure 10.29: Delay response of first- to sixth-order allpass filters, all set to same cut-off frequency. The delay increases with filter order.

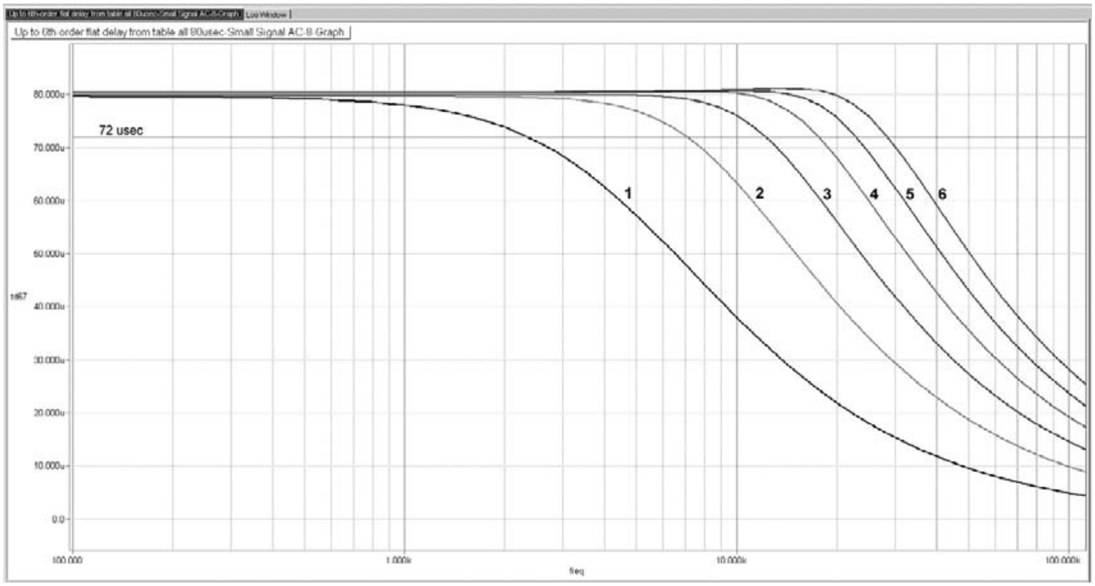


Figure 10.30: Delay response of first- to sixth-order allpass filters, all set to 80 usec delay. The 72 usec line shows where each response is 10% down.

about 5 kHz. There is clearly a limit to how high an order it is sensible to use. A powerful consideration here is that component sensitivity (i.e., how much the filter behaviour changes with component value tolerances) gets worse for higher-order filters. There are also issues with dynamic range, as higher-order allpass filters use at least one section with high Q—this means its MFB filter has gain well above unity and this is a potential clipping point. The sixth-order allpass filter described below illustrates this problem.

I will now give some practical designs for fourth-, fifth-, and sixth-order allpass filters, all set for a 80 usec LF delay; these were the circuits used to generate the curves labeled 4, 5, and 6 in Figure 10.30. The capacitor values have been chosen to keep the resistor values in the MFB filters above 1 kΩ to minimise loading problems. The resistor values given are exact, and not part of any series of preferred values.

Figure 10.31 shows a fourth-order allpass filter made by cascading two second-order allpass stages. The delay is now 10% down at 17.2 kHz, rather than the 12.7 kHz we got with the third-order filter examined earlier in this chapter, a useful increase in flat-delay bandwidth.

The lower the Q of the stage, the lower the gain of the MFB filter, and so the lower the value of the resistor feeding into the summing stage must be to get the correct gain. It must be remembered that this resistor constitutes a load to ground on the output of the MFB opamp, and its value must not be allowed to fall below 500 Ω if 5532 or similar opamp types are being used. In the first stage here this means that R4 and R5 have to be 2 kΩ to get R4 up to 544.8 Ω. In the second stage the Q is higher and the MFB has more gain so R9, R10 can be reduced to 1 kΩ, for better noise performance, while R8 is a respectably high 648.9 Ω.

The values of C1, C2 and C3, C4 can be selected independently for each second-order stage, and C3, C4 have therefore been set lower in value at 4n7 to keep R6, R7 above 1 kΩ in value.

Figure 10.32 shows a fifth-order allpass filter made by cascading two second-order allpass stages and one first-order stage. Because of the higher cut-off frequency, both C1, C2 and

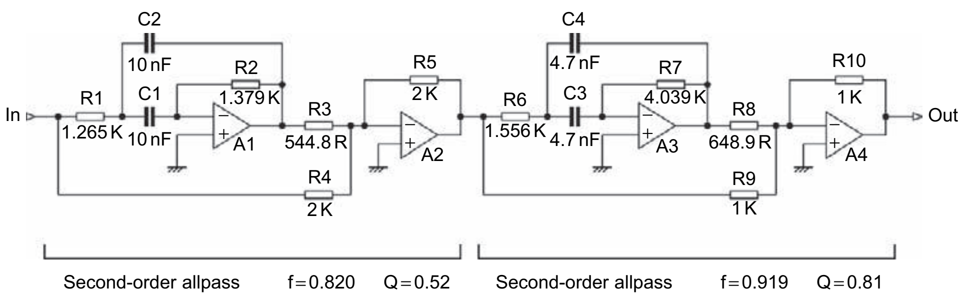


Figure 10.31: Fourth-order allpass filter built by cascading two second-order stages. Delay = 80 usec.

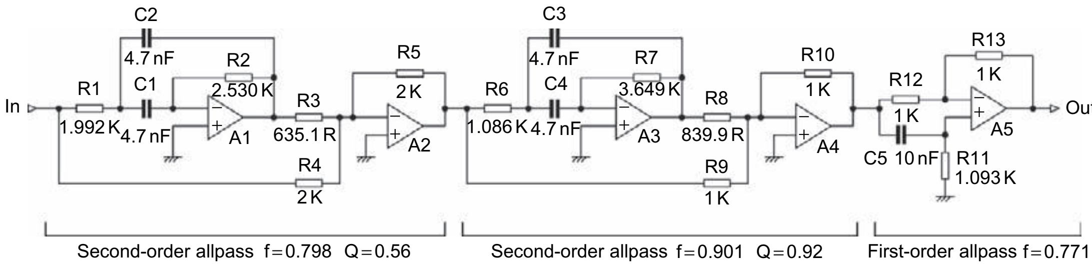


Figure 10.32: Fifth-order allpass filter built by cascading two second-order stages and one first-order stage. Delay = 80 usec.

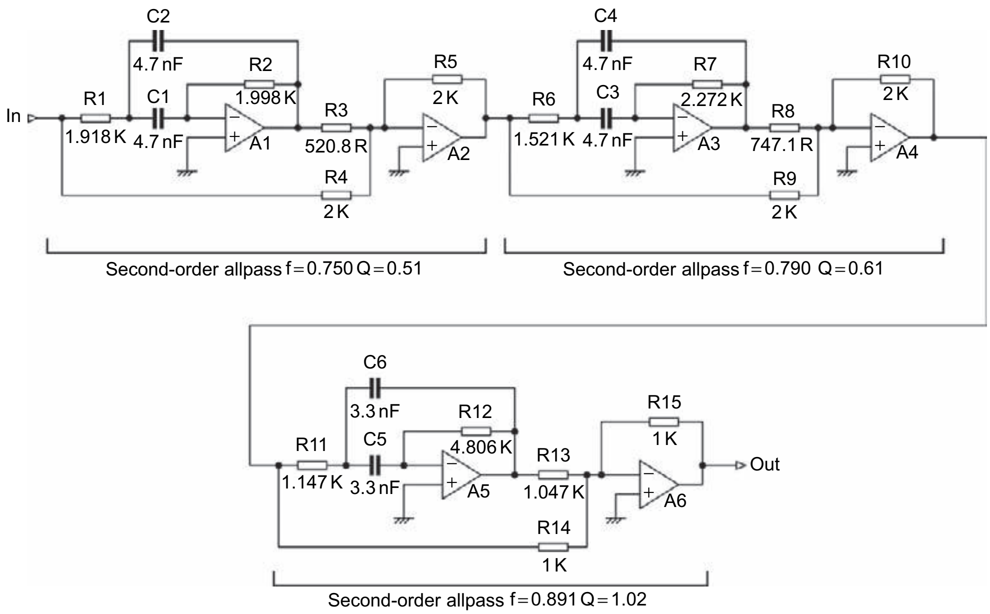


Figure 10.33: Sixth-order allpass filter built by cascading three second-order stages. Delay = 80 usec.

C3, C4 have been set to 4n7 to keep the associated resistors above 1 kΩ. C5 can also be set independently and has been made 10 nF to give a suitable value for R11; as low as possible for noise purposes, but not so low that it puts an excessive load on A4.

Figure 10.33 shows a sixth-order allpass filter made by cascading three second-order allpass stages. In the third high-Q stage it has been necessary to reduce C5, C6 to 3n3 to keep the associated resistors above 1 kΩ.

The final second-order section has a Q of 1.02, which means a passband gain of 2.095 times in the MFB filter. This could cause headroom problems. It is another reason not to embrace higher-order allpass filters without careful consideration.

Since higher-order allpass filters are made up of combinations of first- and second-order stages, the distortion and noise characteristics that apply have been dealt with earlier in this chapter. Noise will clearly increase as more stages are cascaded, and life being what it is, it is highly likely that distortion will also increase additively, though you may benefit from some degree of cancellation.

10.8 Delay Lines for Subtractive Crossovers

There is another use for delay filtering in crossover design; it can be a fundamental part of the crossover operation rather just compensating for mechanical displacements. Lipshitz and Vanderkooy [8] published in 1983 a crossover topology in which a time delay in one signal path allowed the construction of linear-phase crossovers with high slopes; a constructional project based on this was put forward by Harry Baggen in Elektor in 1987 [9]. Both are described in Chapter 6 on subtractive crossovers. Lipshitz and Vanderkooy put forward a specimen crossover design, in which a delay of 289.2 usec was required, and they suggested the following possibilities:

1. Three cascaded 9th-order Bessel allpass filters, each giving a delay of 96.4 usec, accurate to 1% up to 20 kHz. Total delay 289.2 usec.
2. Four cascaded 6th-order Bessel allpass filters, each giving a delay of 72.3 usec, accurate to 1% up to 15 kHz, and to 10% up to 20 kHz. This will have less component sensitivity because of the lower-order filters. Total delay 289.2 usec.
3. Three cascaded 6th-order Bessel allpass filters, each giving a delay of 96.4 usec, accurate to 1% up to 10 kHz, and to 20% up to 20 kHz. Total delay 289.2 usec.
4. Four cascaded 4th-order Bessel allpass filters, each giving a delay of 72.3 usec, accurate to 1% up to 8 kHz, and to 50% up to 20 kHz. Total delay 289.2 usec.

The trade-off here is between how far up the audio band you want linear-phase behaviour to extend, and how complex and costly a delay-line you are prepared to pay for. (When you have many cascaded stages it is more usual to talk of a "delay-line" rather than a "delay filter.") It is clearly impractical to implement any of these options in a passive crossover—the cost, the resistive losses, and the non-linearity of the many inductors required would be crippling. Such a delay-based crossover can only be realised by active circuitry, and even then it presents some serious challenges in the analogue domain. In a digital crossover, of course, manipulating time delays is very simple indeed.

I decided to have a go at constructing such a delay line, adopting the simplest option, that numbered 4 above. I used maximally flat allpass filters rather than Bessel types, in order to explore how much more of a cost-effective solution it was in terms of greater bandwidth for the same number and order of filters. The resulting schematic is shown in Figure 10.34; it consists of four cascaded identical fourth-order filters similar to that in Figure 10.31, but with the resistor and capacitor values in the MFBs altered to theoretically give a delay of 72.3 usec each.

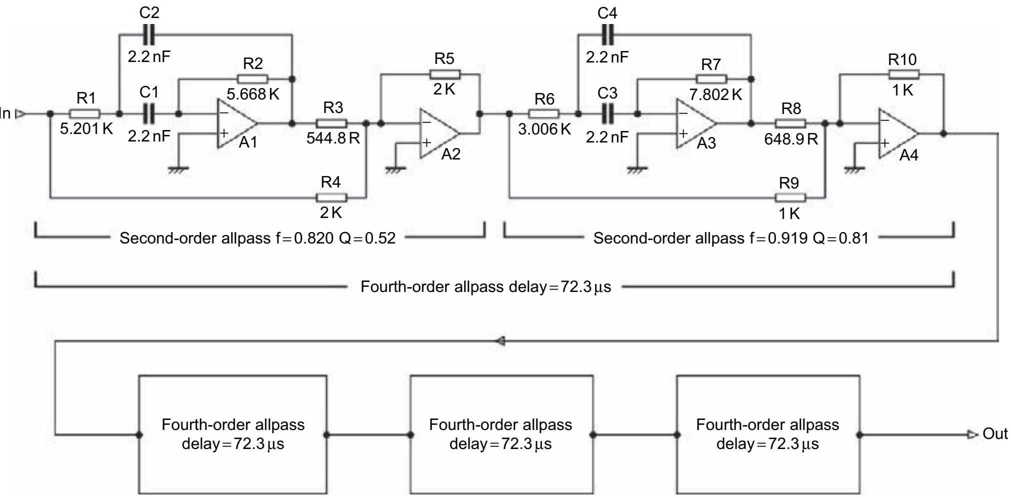


Figure 10.34: Four cascaded maximally flat fourth-order allpass filters for Lipshitz and Vanderkooy time-delay crossover.

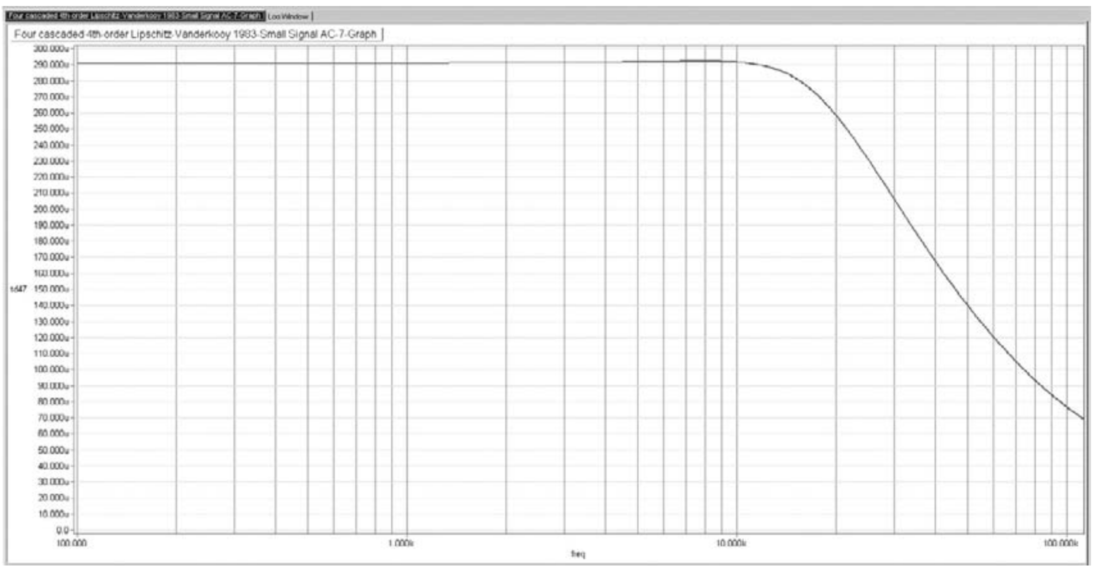


Figure 10.35: Four cascaded fourth-order allpass filters for Lipshitz and Vanderkooy time-delay crossover. Delay = 291 usec, 1% down at 12.5 kHz, 11% down at 20 kHz.

The actual delay came out at 291 usec rather than 289.2, showing that I should have used a bit more precision in my calculations; there is also a very small amount of delay peaking around 10 kHz for the same reason. The delay response is shown in Figure 10.35, and it is down by 1% at 12.5 kHz rather than 8 kHz, and down 11% at 20 kHz rather than 50%. The

50% point is not reached until 47.8 kHz. These results would appear to be notably superior to the figures put forward by Lipshitz and Vanderkooy for the Bessel approach.

This delay-line design may hopefully be of use to those wishing to explore delay-based crossovers. The component cost is not excessive; you need 16 capacitors (all of the same size in this case), 40 resistors, and 8 opamps. If a greater delay is needed then more stages could be added on.

It is a good question as to whether the delay-line could be implemented more efficiently by using third-order allpass filters. As we saw earlier in the chapter, a third-order allpass is some 3 dB quieter if everything else is equal. A possible design solution would be to use five cascaded third-order filters (using 15 amplifiers) instead of four cascaded fourth-order filters (using 16 amplifiers). We should therefore get a noise improvement as well as a slight saving on parts.

Please note that there are serious questions about the practicality of subtractive crossovers, because they do seem to require very precise delays to maintain the desired slope steepness. This issue is looked into in Chapter 6.

10.9 Variable Allpass Time Delays

Variable time delays are often required for subwoofer applications (see Chapter 15) because the distance between the main loudspeakers and the subwoofer is not fixed, and so the delay needs to be variable to get the best integration of responses.

It can also be useful to have a time delay configuration which allows tweaking during the design stage—it would be particularly useful in cases where the effective acoustic centre of a drive unit is not easy to determine.

Unfortunately, making variable the most efficient delay filters, such as the third-order 1–2 BP configuration, requires multiple components to be changed in various ways and is not easy to implement. A more promising approach is to use cascaded first-order allpass filters, as described earlier, in which a two- or four-gang pot with equal-value sections (anything else will be hard to come by), can be used to alter the frequency-setting resistor in each stage. This method has been used in commercial crossovers such as those by Rane.

Figure 10.36 shows two cascaded first-order allpass filters with a two-gang control pot. The delay is variable from 40 usec (down 10% at 10 kHz) to 440 usec (down 10% at 900 Hz).

10.10 Lowpass Filters for Time Delays

Lowpass Bessel filters are often used to preserve the waveform of a signal, typically being used to remove out-of-band noise before A-to-D conversion. Because in the analogue world any real-life filter must be causal (in other words, you can get an output

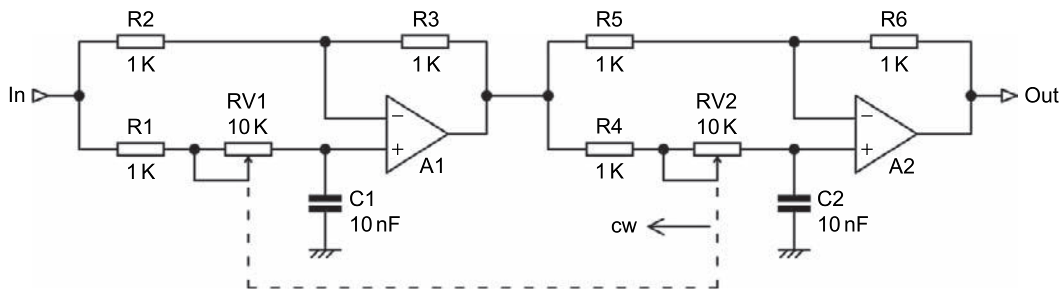


Figure 10.36: Variable time-delay using cascaded first-order allpass filters. Moving the control clockwise increases the delay.

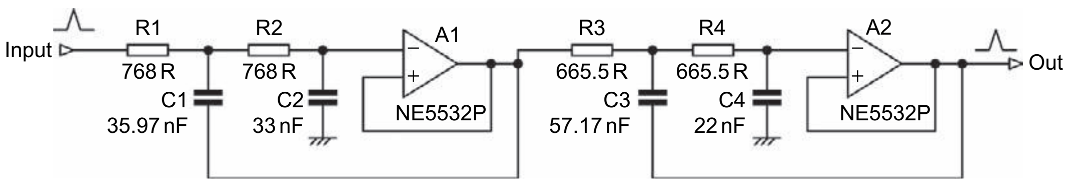


Figure 10.37: Fourth-order Bessel filter giving a delay of 80 usec and an amplitude roll-off −3 dB at 4.2 kHz.

after the input, but never *before* the input); the output is inevitably a delayed version of the input. It seems reasonable to take a look at how effective Bessel lowpass filters are at implementing delays. There is at least the possibility that this sort of filter could implement delay compensation at the same time as giving a crossover roll-off for a MID drive unit.

Figure 10.37 shows a typical fourth-order Bessel lowpass filter with a −3 dB point at 4.2 kHz, which corresponds with an 80 usec delay, as used before. A similar lowpass filter (with the capacitors scaled up to give –3 dB at 1.2 kHz) was use to demonstrate first-order allpass delay action earlier in this chapter. Like almost all fourth-order filters, it is composed of two second-order stages, with the later stage having the higher Q. The design of Bessel filters was described in Chapters 7 and 8.

Figure 10.38 shows that the delay is 10% down at 8.5 kHz, compared with 10% down at 17.2 kHz for a fourth-order allpass filter. This is less than half the bandwidth, and indicates that Bessel filters are not an efficient way of obtaining a flat delay.

The frequency response is shown in Figure 10.39. As with all Bessel filters, the roll-off is relatively slow. The amplitude response is perceptibly drooping at 2 kHz, but the −3 dB point does not occur until 4.2 kHz, more than an octave away.

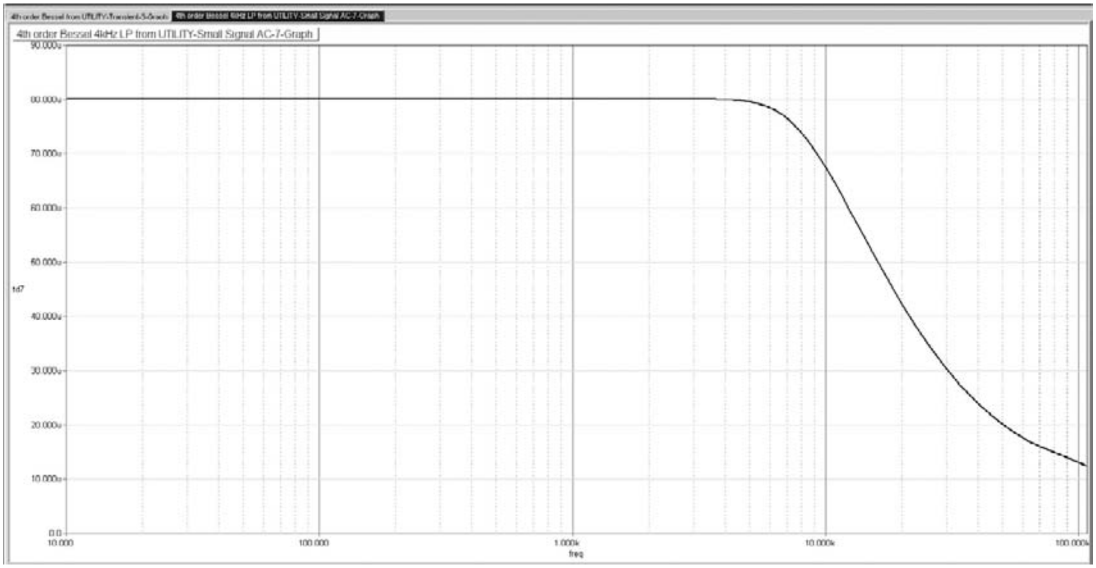


Figure 10.38: Group-delay of fourth-order Bessel filter of Figure 10.28.
The −10% point is at 8.5 kHz.

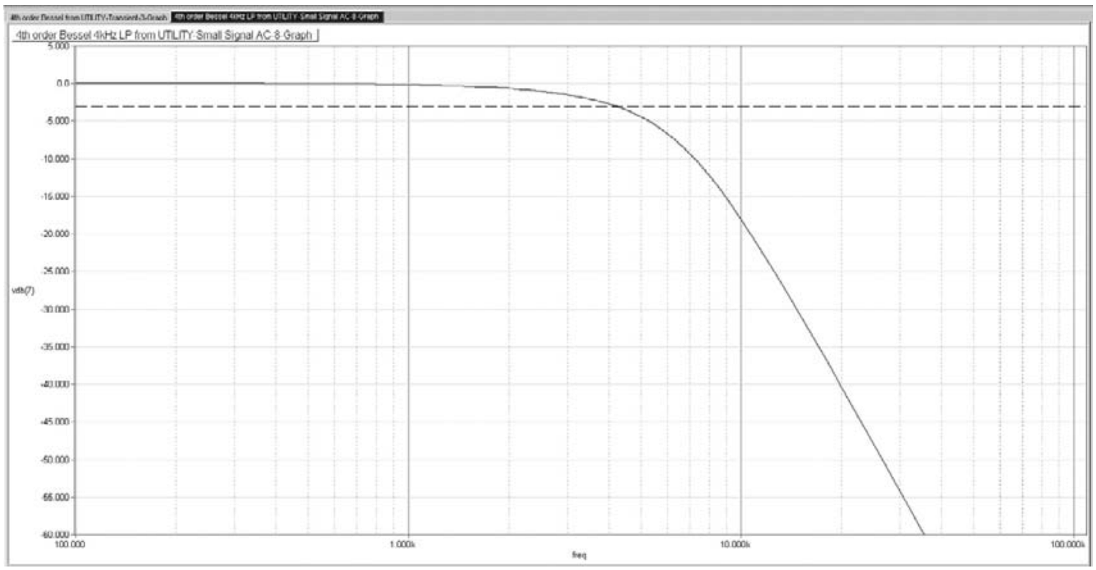


Figure 10.39: Frequency response of fourth-order Bessel filter of Figure 10.28.

References

[1] S. Linkwitz, http://www.linkwitzlab.com/crossovers.htm, 2010 (accessed 12.10).
[2] S. Linkwitz, Loudspeaker system design, Electronics World, London, 1978.
[3] T. Delyiannis, Y. Sun, J.K. Fidler, Continuous-Time Active Filter Design, Chap. 4, CRC Press, 1999, p. 134.
[4] A.G. Lloyd, Here's a better way to design a 90° phase-differencing network, Electron. Des. (15) (1971) 78.
[5] R. Orban, Active all-pass filters for audio, JAES 39 (4) (1991) 250–260.
[6] L. Bruton, RC-Active Circuits: Theory & Design, Prentice-Hall, Englewood Cliffs, NJ, 1980, p. 351.
[7] S.S. Haykin, Synthesis of Active RC Circuits, Chap. 3, McGraw-Hill, London, 1969.
[8] S.P. Lipshitz, J. Vanderkooy, A family of linear-phase crossover networks of high slope derived by time-delay, JAES 31 (1/2) (1984) 2–20.
[9] H. Baggen, Active phase-linear crossover network, Elektor (9) (1987) 61.

This page intentionally left blank

CHAPTER 11

Equalisation

11.1 The Need for Equalisation

Equalisation to attain a desired flat frequency response may be applied to correct problems in the loudspeaker itself, or, moving along the audio chain, to modify the interaction of the loudspeaker with the room it is operating in. Moving along the audio chain still further, equalisation can also be used to modify the response of the room itself, by cancelling resonances with dips or notches in the overall amplitude response. However, it is not normally considered a good idea to try to combine an active crossover with a room equaliser, not least because they are doing quite different jobs. Moving the loudspeakers from one listening space to another will not require adjustment of the crossover, except insofar as the loudspeaker placement with regard to walls and corners has changed, but would almost certainly require a room equaliser to be re-adjusted unless the room dimensions, which determine its resonances, happen to be the same.

At low audio frequencies, normal rooms (i.e., not anechoic chambers with enormous sound absorption) have resonances at a series of frequencies where one dimension of the space corresponds to a multiple number of half wavelengths of the sound being radiated. The half-wavelength is the basic unit because there must be a node, that is a point of zero amplitude, at each end. Sound travels at about 345 metres/second, so a room with a maximum dimension of 5 metres will have resonances from 34.5 Hz upwards. This is simply calculated from velocity/frequency = wavelength, bearing in mind that it is the half-wavelength that we are interested in. We might therefore expect a resonance at 34.5 Hz, and another at about 69 Hz; this is twice the frequency because we now need to fit in two half-wavelengths between the two reflecting surfaces. This continues for three and four times the lowest frequency, and so on. These "resonant modes" cause large peaks and dips in response, the height of which depends on the amount of absorbing material. A room with big soft sofas, thick carpeting, and heavy curtains will be acoustically fairly "dead," and the peaks and dips of the frequency response will typically vary by something like 5–10 dB. A bare room with hard walls and an uncarpeted floor will be much more acoustically "live," and the peaks and dips are more likely to be in the range of 10 to 20 dB, though larger excursions are possible. Resonant modes at low frequencies cause the greatest problems, because they cannot be effectively damped by convenient absorption material such as curtains or wall hangings. Room equalisation that attempts to deal with this situation is a very different subject from active crossover design and is not dealt with further here.

This chapter deals only with the equalisation of the frequency response, but there is another very important form of equalisation; this is commonly called time equalisation, time-delay compensation, or phase equalisation, and while its results can be seen in the form of an improved frequency response, the circuitry involved essentially works in the time domain, and in itself has a flat frequency response. This branch of active crossover design is dealt with in detail in Chapter 10, and is not referred to further here.

11.2 What Equalisation Can and Can't Do

When the word "equalisation" is used without qualification, it almost always refers to correcting the amplitude/frequency response, without attempting to simultaneously correct the phase/frequency response. Correcting both is much more difficult, but it can be done. It is important to realise that if you use the right sort of equaliser, you can put a peak into the frequency response, and then cancel it out completely by using another equaliser with the reciprocal characteristic. The high-Q peak/dip equaliser described later in this chapter can perform this, as it is reciprocal—in other words, its boost and cut curves are exact mirror-images of each other about the 0 dB line. Figure 11.18 below shows the symmetrical 6 dB peak and dip at 1 kHz that this circuit creates. If one equaliser is set to maximum boost, and it is then followed by an otherwise identical equaliser set to maximum cut, the final frequency response is exactly flat, as one might hope. What is much less obvious, but equally true, is that the phase shifts introduced by the first equaliser are also cancelled out by the second equaliser, so a square-wave input will emerge as a square-wave output. This process is demonstrated in Figure 11.1, where the square-wave with added ringing is the output from the first (peaking) equaliser.

The visually perfect square-wave is not the input waveform; it is the output from the second (dip) equaliser. The rise and fall times of the input square-wave were deliberately slowed down to 10 usec to avoid distracting effects due to the finite bandwidth of the opamps used. You will note that the height of the ringing takes a little while to settle down after the start at 0 msec. When applying square-waves and the like to filters and other circuits with energy-storage elements, you need to allow enough cycles for the circuit to reach equilibrium, otherwise you may draw some wildly false conclusions.

Having reassured ourselves on this point, things are generally very much otherwise in crossover equalisation. If a drive unit has, say, a 2 dB peak or hump in its amplitude/frequency response, this will probably be due to some under-damped mechanical resonance or other electro-acoustic phenomenon, and while it is, in principle at least, straightforward to cancel this out, it is very unlikely that the associated phase-shifts will also cancel, because the physics behind the drive unit peak and the equaliser dip are completely different.

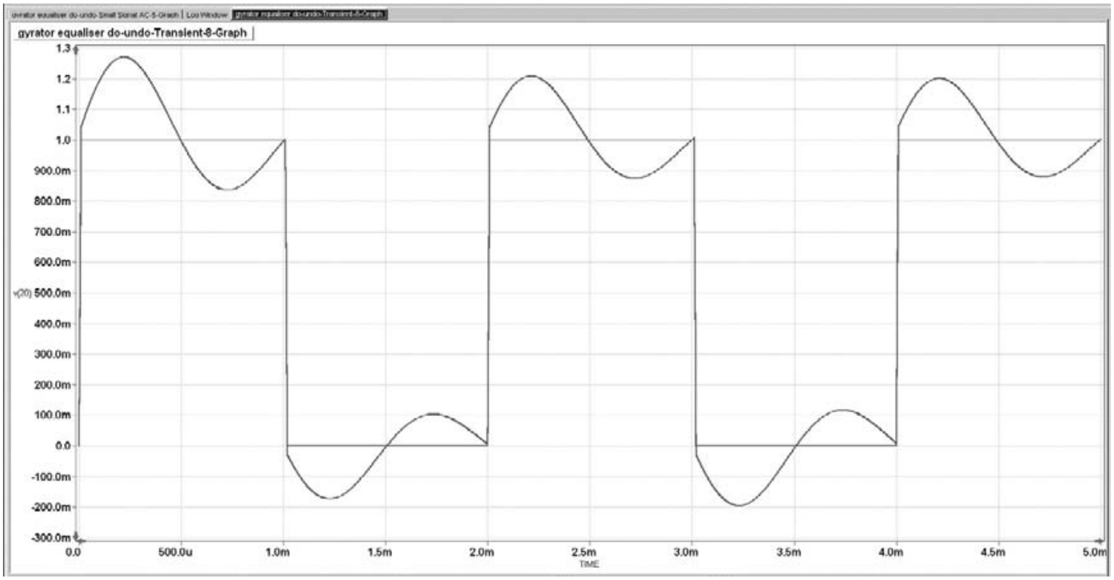


Figure 11.1: Demonstrating that two reciprocal equalisers cancel their phase-shifts as well as amplitude/frequency response changes.

11.3 Loudspeaker Equalisation

Active crossovers may consist simply of frequency-dividing filters, but very often also include equaliser circuits with a carefully tailored frequency response which are used to counteract irregularities in the overall response of the system. Since the active crossover and its power supply, etc. is already present, relatively little extra circuitry is needed and the cost is low. A large number of different equaliser circuits are described later in this chapter.

There are several reasons why equalisation might be necessary:

1. Correcting irregularities in the frequency response of the drive units themselves.
2. Correcting frequency response features inherent in the driver/enclosure system, for example 6 dB/octave equalisation of dipole LF loudspeaker units.
3. Using equalisation to evade the physical limits of driver/enclosure system; most commonly extending the bass response of any kind of loudspeaker.
4. Compensation for other unwanted interactions between the loudspeaker and its enclosure. The most common application is compensation for enclosure front panel diffraction effects.
5. Compensation to deal with interactions between the loudspeaker and the listening space, such as when a loudspeaker is mounted against a wall.

11.3.1 Drive Unit Equalisation

Drive unit equalisation usually involves the correction of minor humps and dips in the frequency response. For example, a consistent 2 dB dip in drive unit response might be cancelled out by a 2 dB peak introduced by an equalisation circuit. Exact cancellation will not be possible, either because the shape of the dip is too complex to be mimicked by practical amounts of circuitry, or because of variation in the shape of the dip due to driver tolerances or aging. This kind of correction is commonly performed by peak/dip equalisers.

A classic example of drive unit equalisation is Constant Directivity orn Equalisation. Following the work of Don Keele at ElectroVoice [1], constant directivity (CD) high-frequency horn loudspeakers appeared in the late 1970s. Basically, an initial exponential section was combined with a final conical flare, causing the shorter high-frequency wavelengths to be dispersed more effectively off axis; purely conical horns, as seen on old gramophones and phonographs, are not satisfactory because of their poor low-frequency response. Because CD horns direct more high-frequency energy off axis, the amount of high-frequency energy available directly on axis is reduced. Therefore, the CD horn no longer measures flat directly on axis unless it is given equalisation that is the inverse of the horn high-frequency roll off response. This is called Constant Directivity Horn Equalisation or CDEQ.

Constant Directivity horns roll off the higher frequencies at about –6 dB per octave from around 2 to 4 kHz, continuing to 20 kHz and beyond, so the equalisation therefore takes the form of a +6 dB per octave boost starting at the appropriate frequency. The equalisation curve is arranged to shelve to prevent excessive boost at frequencies above 20 kHz, which could lead to inappropriate amplification of ultrasonic signals and imperil the sound system stability. The frequency at which the equalisation begins may be derived from measurements or from the –3 dB point of the CD horn frequency response as provided by the manufacturer. The equalisation is typically performed by a HF-boost equaliser as described later in this chapter.

11.3.2 6 dB/oct Dipole Equalisation

The operation of dipole loudspeakers is described in Chapter 15 on sub-woofers. To summarise it quickly, a dipole loudspeaker has a drive unit mounted on a flat panel, usually called a baffle, which prevents sound from simply sliding straight round from the front to the back of the drive unit and cancelling out. The larger the baffle, the lower the frequency down to which this is effective. The panel may be folded in various ways to save space, so it remains large acoustically but is physically smaller. The name “dipole” is derived from the way that the polar response consists of two lobes, which have equal radiation forwards and backwards (in opposite polarities), and none at right angles to the front-back axis of the drive unit. Because the baffle is of finite size, there is a frequency at which the response begins to fall off as sound goes round the edge of the baffle. (Loudspeakers in sealed boxes are sometimes referred to as being mounted in an "infinite baffle," but that does not of course mean that their response goes down to infinitely low frequencies.)

Dipole speakers are therefore commonly regarded as needing equalisation in the form of a 6 dB/oct boost as frequency falls, starting from an appropriate frequency. This has to be done with great caution as even apparently modest amounts of equalisation greatly increase both cone excursion and the amplifier power required; a dipole loudspeaker is particularly vulnerable to this because unlike a drive unit with a sealed box, there is no loading on the cone at low frequencies. It is essential that the bass boost ceases at a safe frequency, and it is wise to arrange for an effective subsonic roll-off if drive unit damage is to be avoided. There is often also a need to equalise away a peak in the loudspeaker response where the LF roll-off begins. The biquad equaliser, described later, can do the task effectively and economically, but may need to be supplemented with further subsonic filtering to make sure that no accidents occur.

11.3.3 Bass Response Extension

As described in Chapter 2, the bass response of a loudspeaker is essentially that of a highpass filter whose characteristics are determined by the type of enclosure and the parameters of the LF drive unit. A sealed-box loudspeaker consists of a mass-compliance-damping system that gives a classical second-order highpass response and can for some purposes be simulated by a highpass active filter of the Sallen and Key or other configuration.

Much thought has been given to extending the low-frequency response of such systems by adding bass boost in a controlled fashion. The presence of an active crossover makes adding this facility straightforward, though its design is not so simple. The basic principle is to counteract the downward slope of the loudspeaker response with the upward slope of equaliser boost, but there are two major complicating factors. Firstly, when the LF response of a loudspeaker is rolling off, the cone excursion is at its greatest. Adding equalisation increases the excursion further, and it is all too easy to exceed the safe limits of the drive unit. Secondly, a much increased amplifier power capability is required.

It has been claimed that not only is the LF response extended, but the time response may also be improved, because the ringing and overshoot caused by an underdamped LF response can be cancelled by a matching dip in the response of the equalisation circuit, since we are dealing with a simple minimum-phase system. The final LF rolloff is determined by that of the equaliser. It is questionable if this can really be counted as "equalisation" as such, because the intent is not so much to correct an error as to extend the performance beyond what would otherwise be physically possible. The biquad equaliser is a good choice for this form of equalisation, giving great freedom of parameter variation.

Because of the low frequencies at which this kind of equaliser operates, large capacitor values are required if impedance levels are to be kept suitably low, and this puts up the cost.

11.3.4 Diffraction Compensation Equalisation

We saw back in Chapter 2 that the diffraction of sound from the corners of a loudspeaker enclosure can have profound effects on the frequency response. The somewhat impractical spherical loudspeaker is free from most response irregularities, but still has a gentle 6 dB rise with frequency, because at low frequencies the long sound wavelengths diffract around the sides and rear of the enclosure, so radiation occurs into "full-space," while at high frequencies the drive unit cone radiates mostly forwards, into what is called "half-space." This rise in response is sometimes called the "6 dB baffle step," though it is actually a very smooth transition between the two modes. Since the frequencies that define it are constant, being derived from the enclosure dimensions, it can easily be cancelled out by first-order shelving equalisers, such as those described later in this chapter.

Spherical loudspeakers are, however, very rare and for practical reasons the vast majority of loudspeaker enclosures are rectangular boxes, as seen in Figure 11.2, which is derived from the all-time classic paper by Olson in 1969 [2]. The lengths of the edges of the rectangular box were 2 ft and 3 ft, the drive unit being mounted midway between the two side edges and 1 ft from the top short edge. Olson comments that the dips at 1 kHz and 2 kHz are induced by diffraction from the top and side edges, the frequencies being relatively high because the distances from the drive unit to these edges are small; the broader response minimum just below 600 Hz is due to the longer distance to the bottom edge. It is obvious

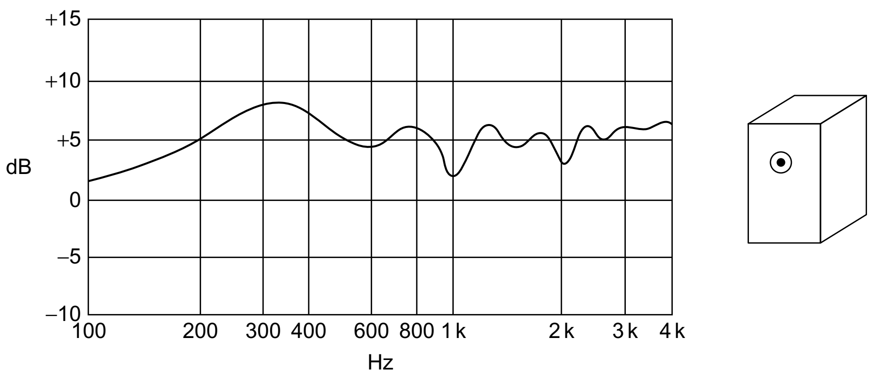


Figure 11.2: Response disturbances, due to the sharp corners of a rectangular box, are superimposed on the inherent 6 dB response rise (after Olson, 1969).

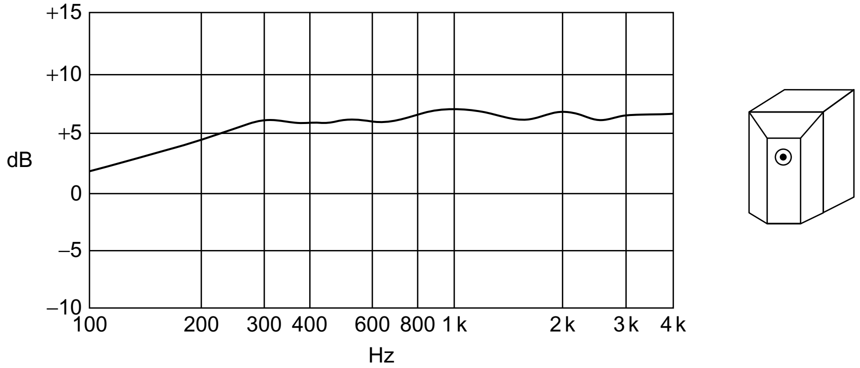


Figure 11.3: Much reduced response disturbances due to the blunter corners at the front of the box, added to the inherent 6 dB rise (after Olson, 1969).

that equalising away these three minima, as well as compensating for the inherent 6 dB rise is a fairly ambitious undertaking, requiring at least four equaliser stages. There are also minor dips of about 1 dB at 1.5 kHz and 2.5 kHz, which are at multiples of the frequency of the minimum due to the distance to the bottom edge, and are due to the path containing whole wavelengths.

Olson suggested a more sophisticated enclosure as in Figure 11.3, that would give blunter edges and much reduce the response irregularities, while still fitting into a living room better than a sphere would. The graph shows response deviations reduced to about 2 dB. This is a great improvement and this sort of box has had some popularity, though it is of course more difficult to make and not everybody likes the shape. The response could be made almost flat by correcting the inherent 6 dB rise and then applying a little cut at 1 kHz and 2 kHz.

It should be pointed out that these famous graphs show a rather smoothed version of the actual frequency response. There must have been many other minor irregularities that were not reproduced as they were irrelevant to the central argument about the importance of diffraction effects. You cannot assume that the equalisation options suggested above would result in a ruler-flat frequency response. Loudspeakers just don't work that way. In addition, diffraction effects vary according to the listening or measuring position, because the 'virtual' sources are displaced from the position of the actual driver. Correcting the on-axis response may make the off-axis coloration worse.

11.3.5 Room Interaction Correction

In Chapter 2, we saw how the placement of a loudspeaker with respect to walls and corners could have a significant effect on the frequency response. A loudspeaker in free air (perhaps on a high pole in the middle of a large field, which can be useful for measurements but is less so for actual listening) is said to be working into "whole-space." As we saw in the previous section, the low-frequency output will tend to diffract around the loudspeaker enclosure and travel backwards away from the listener, while the high-frequency output will mostly be radiated forwards.

If we now mount the loudspeaker in the middle of a large vertical wall, with the front baffle flush with the wall, it is working in "half-space" and the low-frequency output can no longer travel backwards—it has to go forwards, and so more low-frequency energy reaches the listener. The high-frequency output is unchanged because it was all going forwards anyway, so the relative rise in LF level is about 6 dB. If we put the loudspeaker at the junction between a vertical wall and the floor, the effect is enhanced and this is called "quarter-space" operation. Finally, we can put our loudspeaker in a corner, where the floor and two walls at right-angles meet, and this is known as "eighth-space" operation. If this progression was taken further by adding more enclosing surfaces, we would end up with something like a horn loudspeaker.

From the point of view of crossover design, the important thing is that the low-frequency acoustic output is boosted relative to the high-frequency output each time we move from whole-space to half-space to quarter-space and then to eighth-space. A loudspeaker/crossover system designed to give a flat "free-space" response, typically for an outdoor sound-reinforcement application, will sound very bass-heavy indoors. Some studio monitor speakers have "half-space," and "quarter-space" settings that switch in a low-frequency rolloff, the frequency at which it starts depending on the enclosure size. A suitable equaliser might be the LF-cut circuit described later in this chapter.

While studio monitor speakers are mounted very carefully, often flush with a surface to give true "half-space" operation, more compromise is usually required in the domestic environment, and it is quite possible to encounter a situation where one loudspeaker is against a wall (quarter-space) while the other has to be in a corner (eighth-space). This is obviously undesirable, but sometimes in life one must make the best of a non-optimal situation, and providing separate open/wall/corner equalisation switches for the left and right channels of a crossover might be worth considering.

Things get more complicated when we contemplate a loudspeaker standing on the floor and not flush (or nearly so) with a wall but some distance from it. If the loudspeaker is one quarter of a wavelength away from a reflective wall at a given frequency, the low-frequency energy that diffracts backwards is reflected, so that its total path length is one half-wavelength and it will reach the loudspeaker again in anti-phase so that cancellation occurs. If the front of a loudspeaker is 1 metre from a wall the first cancellation notch will be at about 86 Hz, as this frequency has a quarter-wavelength of 1 metre. How complete the cancellation is depends on how accurately the speaker-wall distance is one quarter-wavelength and on the reflection coefficient of the wall. A lower frequency means a greater distance for a quarter-wavelength, and the amplitude of the reflected signal reaching the speaker will be lower, because there is more opportunity for diffusion, and so the cancellation will be less effective.

If the loudspeaker is half a wavelength away from a the reflective wall, the total path length is a whole wavelength and it will in arrive in-phase and reinforce the direct sound, theoretically giving a +6 dB increase in level. The effect on the sound reaching the listener will vary from complete cancellation to +6 dB reinforcement depending on the relationship between the sound wavelength and the total path length via reflection, and on the wall reflection coefficient. This effect is not confined to one frequency, because there can be any number of whole wavelengths in the go-and-return path; Figure 11.4 shows how this gives a “comb-filter” response with the reinforcement peaks and the cancellation notches reducing in amplitude with increasing frequency because more sound energy is being radiated forward and less is diffracting to the rear. The peaks and notches get closer together because the graph is drawn in the usual way with a logarithmic frequency axis, but the reinforcement/cancellation process is dependent on a linear function of frequency.

Calculating the actual path length is complicated by the fact that the low-frequency radiation has to go round a 180-degree corner, so to speak, as it diffracts around the front of the enclosure, and it is a question as to how sharp a turn it makes in a given situation.

This loudspeaker-room interaction has been examined here because it is a very good example of a mechanism that it is *not* possible to correct completely by equalisation, for both theoretical and practical reasons. Practically, boosting the gain at all the notch frequencies would be very hard to do, because of the need to line multiple boost equalisers

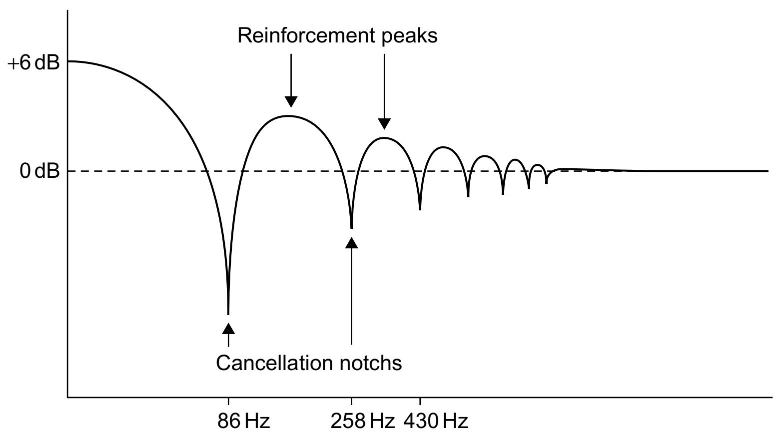


Figure 11.4: Comb-filter effect produced by reflection from wall behind loudspeaker. The amplitude of the peaks and the notches reduces with increasing frequency because more sound is radiating forward and less is diffracting to the rear.

up with multiple narrow notches; an alignment that would become grotesquely incorrect as soon as the loudspeaker was moved by a few inches. You will, however, note that the gain variations become less as the frequency increases, so what can be done is to make some compensation for the really big response variations at the LF end.

The other important point that stands out here is that the loudspeaker-spaced-from-a-wall situation is extremely common, giving rise to the sort of frequency response shown in Figure 11.4, and yet we still listen quite happily to the result. It is not necessary to have a ruler-flat frequency response to enjoy music.

Another important property of a listening space is the amount of high-frequency absorption it contains. A room with hard walls and floors will reflect high-frequency energy, and a proportion of this will reach the listener as reverberation. On the other hand, a room with wall hangings, thick carpets, and comfy sofas will absorb some of high-frequency energy and the effect will be less treble. The more absorbent room will give the more accurate sound as a greater proportion of the energy at the listener will be directly from the loudspeaker, with an accurate frequency response, and will be subjected to less room colouration. Correcting for high-frequency absorption is a job for the preamplifier rather than the active crossover, and this is just one reason why the concept of preamplifiers without tone-controls is a daft idea.

11.4 Equalisation Circuits

There are many ways of obtaining a desired equalisation response, and any attempt to examine them all would probably fill up the whole book, so I have had to be very selective in picking those looked at here. I have aimed to provide circuits that are easy to design, predictable in their response, easy to configure for good noise and distortion performance, and well-adapted for dealing with common loudspeaker problems. I have included some where the fixed resistors that set the response can be temporarily replaced with variable controls to speed the optimisation of a crossover design.

11.5 HF-Boost and LF-Cut Equaliser

This is also known as a shelving highpass equaliser. It gives a frequency response that at low frequencies is flat, but begins to rise when the frequency passes the boost frequency f_b. It continues to rise at a basic rate of 6 dB/octave until the shelf frequency f_s is reached, at which point it shelves or levels out to a fixed gain. Unless the boost frequency and the shelf frequency are spaced by several octaves the transition slope between the gain regimes will not have time to develop an actual 6 dB/octave slope. Figure 11.5 shows the inverting form of the circuit. A non-inverting version also exists but is less flexible because at no frequency can it have a gain of less than unity.

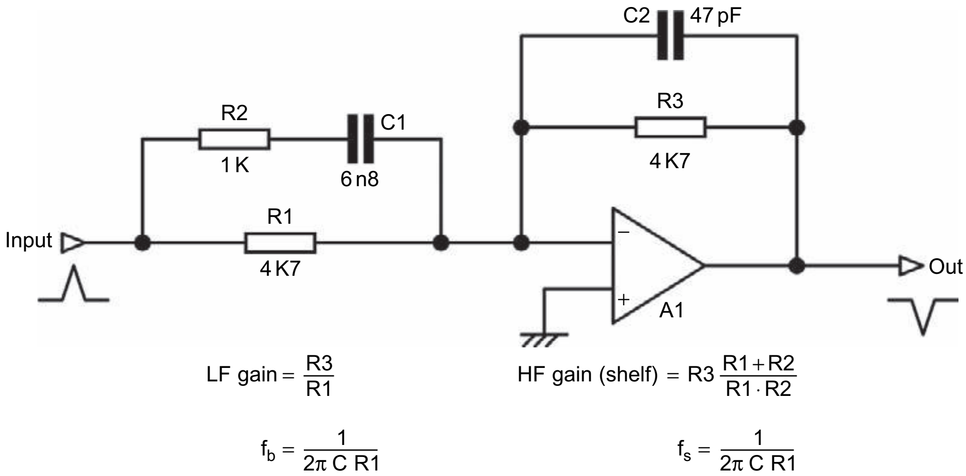


Figure 11.5: Typical application of HF-boost equaliser for Constant Directivity Horn Equalisation; with the values shown the boost starts around 2 kHz and begins to shelve to +15 dB above 20 kHz.

The circuit shown here is essentially a shunt-feedback amplifier with unity gain at low frequencies because R1 = R3. To make it a HF-boost equaliser the extra network R2, C1 is added. As the frequency rises the impedance of C1 falls and allows a greater input current to flow into the virtual-earth point at the inverting input of A1. As the frequency increases further, this current is limited by R2, causing the gain at high frequencies to reach a maximum of R3 divided by the value of R1 and R2 in parallel. The stabilisation capacitor C2 across R3 has no effect on the response at audio frequencies and is included only to emphasise that it is always good practice to include such a measure. The shunt-feedback configuration has no common-mode signal voltage; this is handy if you are using an opamp prone to common-mode distortion. The downside is that it introduces a phase-inversion which will need to be reversed somewhere else in the crossover system.

The design equations for the circuit are given in Figure 11.5, and the frequency response with the values given is shown in Figure 11.6. The equations for the LF gain and the HF or shelf gain are straightforward, but the expressions for the two frequencies require a little explanation. The boost frequency f_b is the frequency at which the gain would have increased by 3 dB, just as the cutoff frequency of a lowpass filter is usually specified as the frequency at which the amplitudes response has fallen by 3 dB. Likewise, the shelf frequency f_s is the frequency at which the gain is 3 d below its final shelving value. As with the response slope, in practice the interaction between the boost and shelving actions is such that the equaliser response will only show these 3 dB figures in its response if the boost frequency and the shelf frequency are a long way apart—much farther apart than is likely in any practical crossover design. This needs to be kept in mind when examining simulator

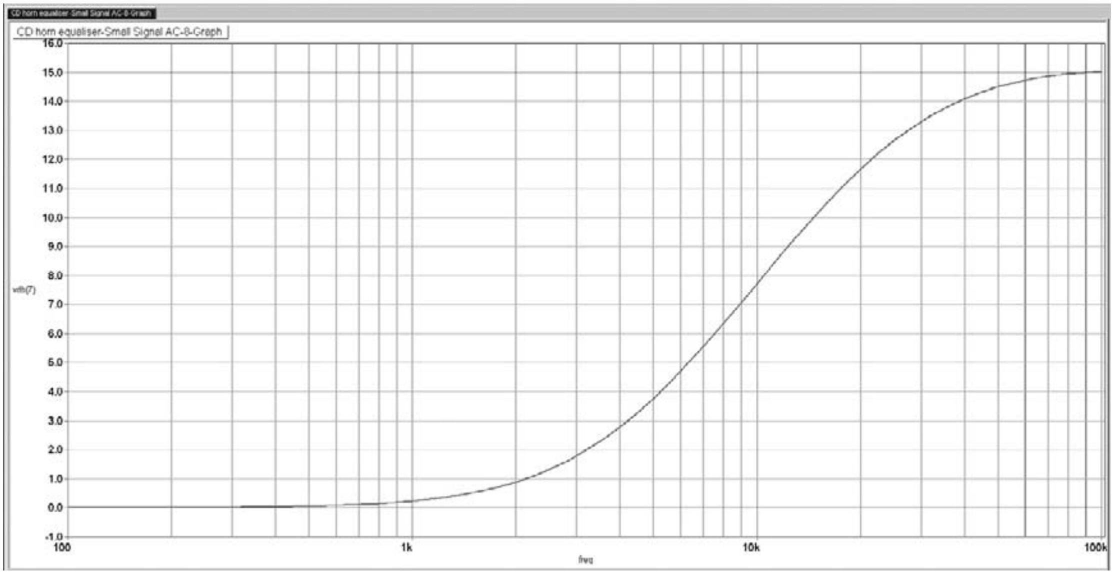


Figure 11.6: The frequency response of the HF-boost equaliser for constant directivity horn equalisation shown in Figure 11.5.

outputs and measured frequency responses. It is perfectly possible to use component values that give the same boost and shelving frequencies; this does not mean the equaliser is doing nothing, it means that the LF and shelf gains are 6 dB apart.

If a basic gain of unity is not what is required, it can be set to any value above or below by altering the value of R3. As with active filters and other frequency-dependent circuits, it is best to decide on a capacitor value first and derive the resistor values from that, given the much greater variety of resistor values available and the ease with which non-standard values can be obtained by combining two of them.

In choosing component values, the resistors should be kept as low as possible to minimise current noise and Johnson noise, but they must not be so low that opamp distortion is increased, either in A1 or in the preceding stage. Particular care is needed with the latter because the input impedance of the circuit falls to R1 in parallel with R2 at high frequencies, where opamp distortion is most troublesome. The circuit values shown give an HF input impedance of 824 Ω, and if 5532 opamps are being used you will not want to go much lower than this.

This type of HF-boost equaliser can be used to deal with response irregularities of many kinds. A common application in the sound-reinforcement field for is Constant Directivity Horn Equalisation; the requirement for this was explained earlier in this chapter. Siegfried Linkwitz also recommends this configuration to smooth the transition between a floor-mounted woofer and a free-standing midrange/tweeter assembly [3].

It is important to understand that while this circuit has so far been described as an HF-boost equaliser, it can also be regarded as an LF-cut equaliser; it is simply a matter of how you look at it. Figure 11.7 shows a circuit that has unity gain across most of the audio spectrum, but gives a gentle cut from about 200 Hz down, as in Figure 11.8. The important point is that to set the normal or unequalised gain, where the impedance of C1 is so low it has no effect, to unity you must choose R1 and R2 so their combined value in parallel is equal to that of R3. (Other unequalised gains greater or lesser than unity can be chosen.) As the frequency falls, the impedance of C1 increases until the current flow through R2 is negligible and the shelving gain of the circuit is set by R1 alone, to –2.5 dB. In this case, the boost frequency is *higher* than the shelving frequency, and not lower as it was with the CD horn example above.

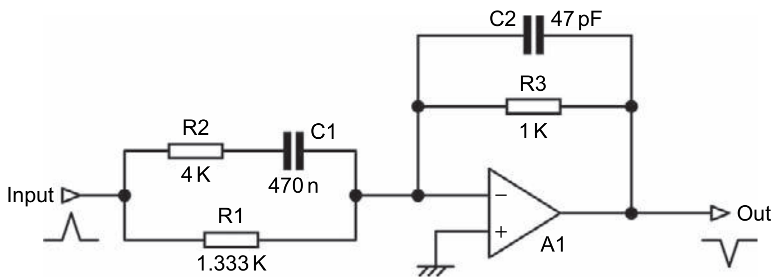


Figure 11.7: The same circuit treated as an LF-cut rather than an HF-boost equaliser.

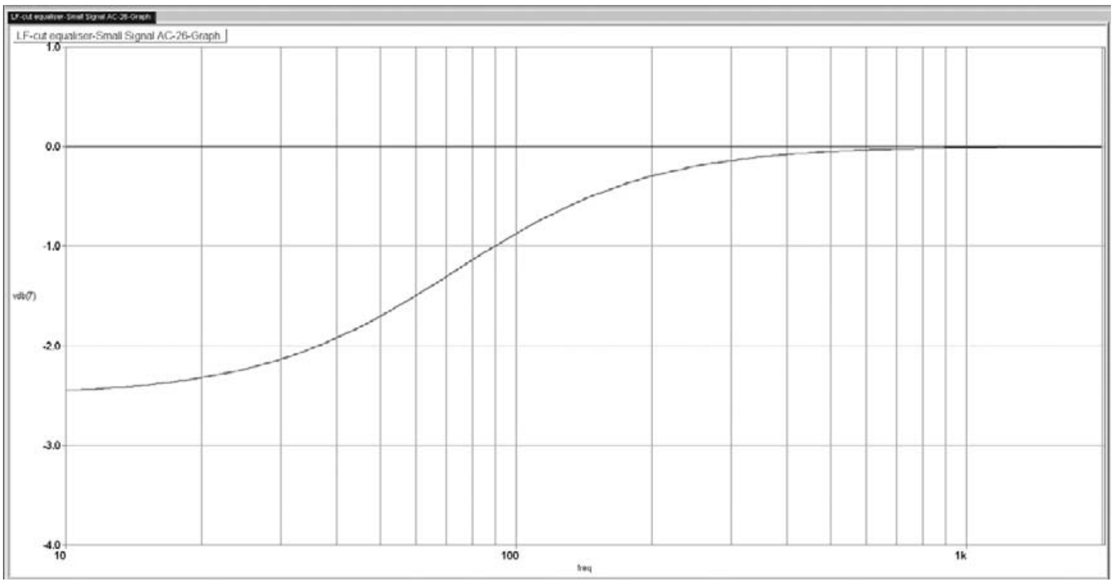


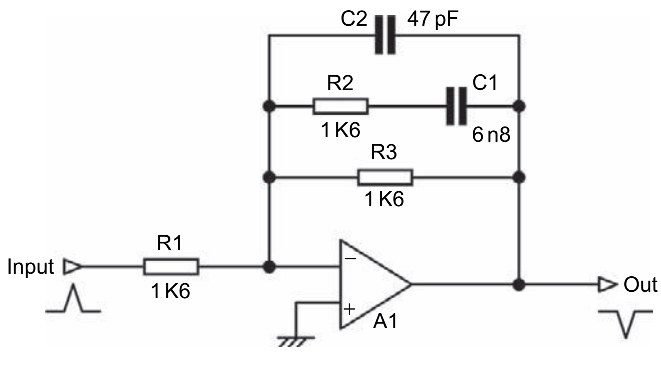
Figure 11.8: The frequency response of the LF-cut equaliser in Figure 11.7.

11.6 HF-Cut and LF-Boost Equaliser

This is also known as a shelving lowpass equaliser. It gives a frequency response that at low frequencies is flat, but begins to fall as the frequency rises and approaches the cut frequency f_c. It continues to fall at a basic rate of 6 dB/octave until the shelf frequency f_s is reached, at which point it shelves, leveling out to a fixed gain. As with the HF-boost equaliser, unless the boost and the shelf frequency are spaced by several octaves the slope between the gain regimes will be much less than 6 dB/octave. Figure 11.9 shows the inverting form of the circuit, with its design equations. A non-inverting version of this equaliser also exists but is less flexible because it cannot have a gain of less than one.

The circuit is a shunt-feedback amplifier with unity gain at low frequencies because R1 = R3. To make it a HF-cut equaliser the extra network R2, C1 is added. As the frequency rises the impedance of C1 falls and allows a greater feedback current to flow into the virtual-earth point at the inverting input of A1, reducing the gain. As the frequency increases further, this feedback current is limited by R2, causing the gain at high frequencies to reach a minimum of R2 and R3 in parallel, divided by the value of R1. A small stabilisation capacitor C2 is placed across R3 as before. Once again, the shunt-feedback configuration has the advantage of no common-mode signal voltage, but it introduces a phase-inversion which will need to be reversed elsewhere.

A typical use of this kind of equaliser is compensating for the high-frequency boost resulting from diffraction around the edges of the front panel of a loudspeaker. With the values shown in Figure 11.9 the basic gain at low frequencies is unity, and the fall in



$$\text{LF gain} = \frac{R3}{R1} \qquad \text{HF gain (shelf)} = \frac{1}{R1} \cdot \frac{R2 \cdot R3}{R2 + R3}$$

$$f_c = \frac{1}{2\pi\, C\, R3} \qquad f_s = \frac{1}{2\pi\, C\, R2}$$

Figure 11.9: Typical example of HF-cut equaliser set up for diffraction compensation; with the values shown the response falls from about 200 Hz and shelves to –6 dB around 5 kHz. The middle –3 dB point is at 1 kHz.

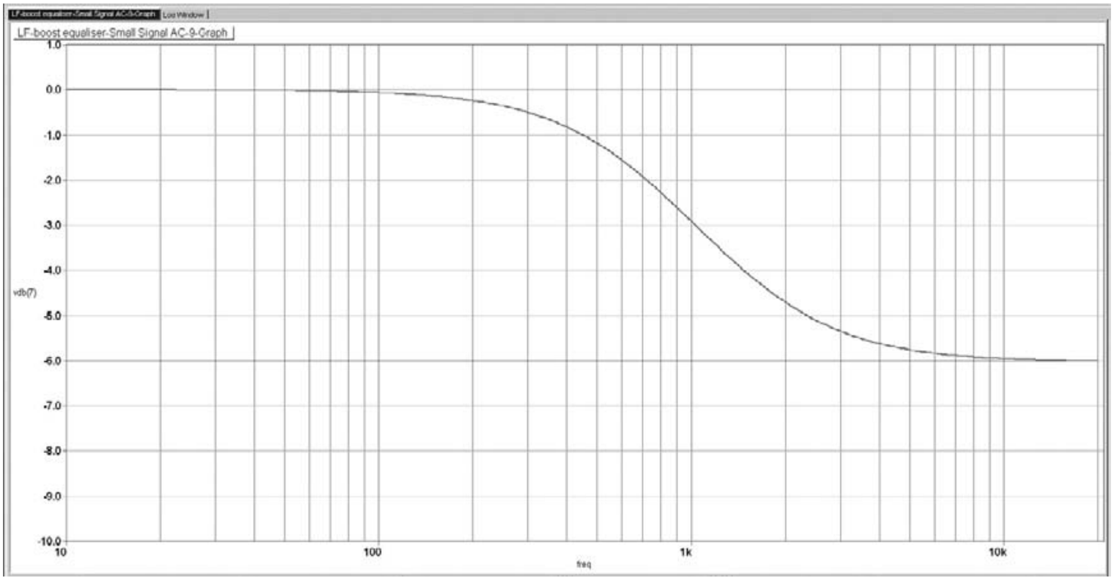


Figure 11.10: The frequency response of the HF-cut equaliser in Figure 11.9, showing a typical curve for diffraction correction.

response starts at around 200 Hz, with the gain shelving to –6 dB around 5 kHz, as in Figure 11.10. The middle –3 dB point is at 1 kHz.

As for the previous HF-boost/LF-cut equaliser, this HF-cut equaliser can also be regarded as a LF-boost circuit. To obtain a normal gain of unity at high frequencies, R1 is set equal to the value of the parallel combination of R2 and R3. Then, as frequency falls, the point is reached where the impedance of C1 becomes significant, and reduces the feedback current through R2; the gain therefore rises, and shelves when the impedance of C1 becomes large compared with R3.

This form of equaliser is useful for compensating for the low frequency roll-off from a loudspeaker fitted to an open baffle; the smaller the baffle, the greater is the loss of low-frequency response, and the greater the amount of boost required. Since the drive unit cone is not loaded at low frequencies, care must be taken to avoid excessive cone excursions. The biquad equaliser provides a more complex but much more versatile alternative.

11.7 Combined HF-Boost and HF-Cut Equaliser

The first of the equaliser types examined has its frequency-dependent network in the input arm of the shunt-feedback amplifier; the second has its frequency-dependent network in the feedback arm. It is possible to put a frequency-dependent network in both input and feedback arms, thus getting two equalisers for the cost of one opamp. Since the inverting opamp input is at virtual earth, there is no interaction between the two networks.

11.8 Adjustable Peak/Dip Equalisers: Fixed Frequency and Low Q

It is often desirable to include an equaliser that can put a peak (like that of a resonant circuit) or a dip (essentially a broad notch) in the frequency response. This is particularly useful for correcting amplitude response irregularities in drive units. The circuit shown in Figure 11.11 is based on the Baxandall tone control concept, and is commonly used in low-end mixing consoles. It can implement a low-Q peak or dip of variable height, giving a flat response if the control is set centrally. The centre frequency can only be altered by changing the capacitor values. This equaliser is described as "adjustable" because it can be set to either peak or dip by any desired amount within its limits. It is unlikely you would want to incorporate a peak/dip control potentiometer in a production crossover, though it could be extremely useful during the development phase. In manufacture RV1 would be replaced by a pair of fixed resistors that give the desired response.

The operation of this versatile circuit is very simple. As frequency increases from the low end of the audio band, the impedance of C2 falls and the position of the pot wiper begins to take effect. When RV1 is in the boost position more of the input signal is passed to the inverting input of A1 and adds to the output, giving a peaking response. When RV1 is in the cut position more of the output signal is passed to the inverting input, giving more negative feedback, and the gain is reduced, causing a dip in the response.

At a still higher frequency, the impedance of C1 becomes low enough to effectively tie the two ends of the pot together, and the position of the wiper no longer has any effect, the circuit reverting to a fixed gain of unity at high frequencies. Thus the circuit only acts over a limited band of frequencies, giving the pleasingly symmetrical response curves in Figure 11.12 for varying control settings. It must be said that the benefits of symmetry are

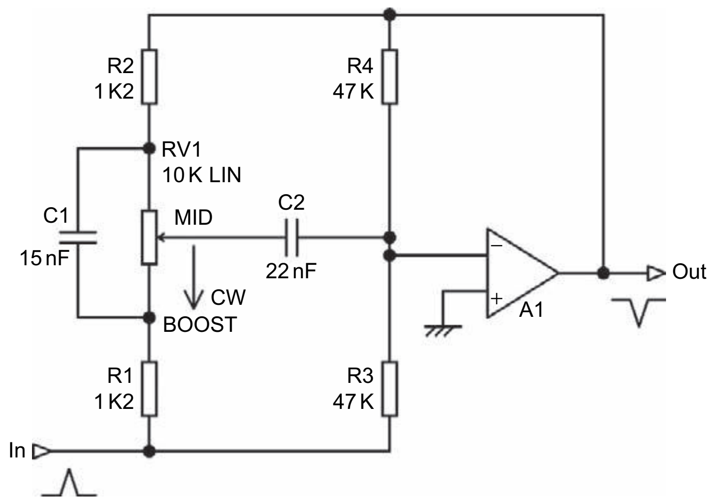


Figure 11.11: A adjustable peak/dip equaliser based on the Baxandall tone control concept.

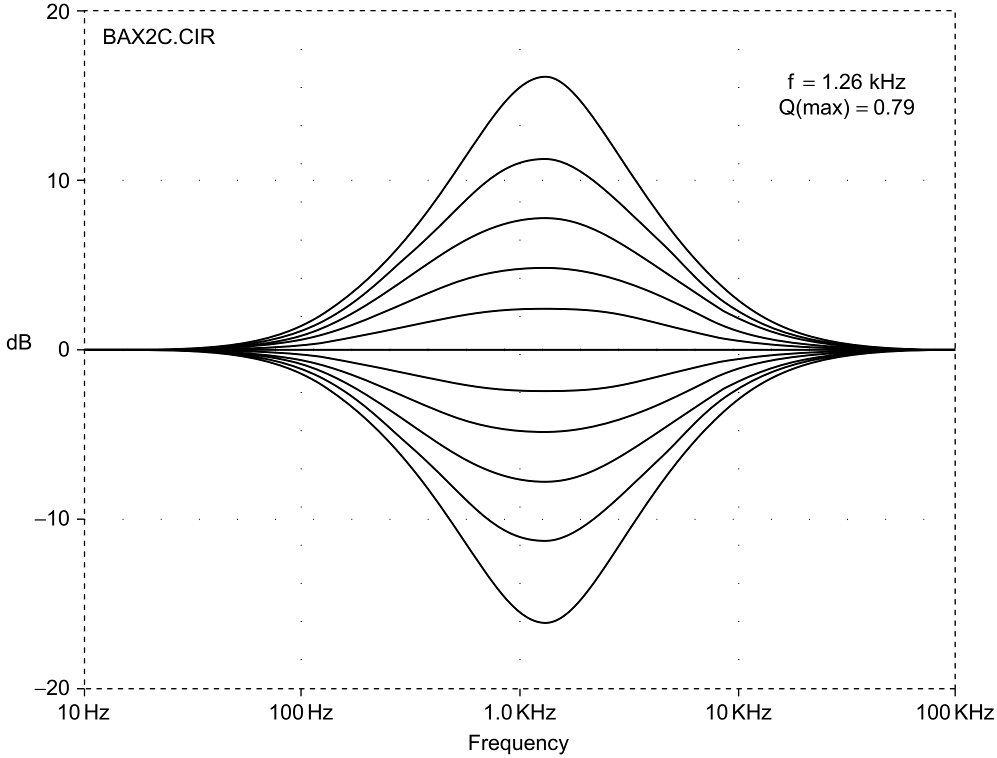


Figure 11.12: Frequency response of the adjustable peak/dip equaliser, for different boost/cut settings.

here visual rather than audible. Much more information on this type of equaliser can be found in my book *Small Signal Audio Design* [4].

The component values shown in Figure 11.11 give a centre frequency of 1.26 kHz, which can be simply altered by scaling the values of C1 and C2 while keeping them in the same ratio. This apparently random frequency is a consequence of the fact that both potentiometers and capacitors come in relatively few values. The maximum Q is 0.79, though this is only obtained at maximum boost or cut. At intermediate settings the curves are flatter and the Q considerably lower; with a boost of +8 dB the Q is only 0.29. The boost/cut limits are +/–15 dB, though this hopefully a much greater range than will be required for equalisation in practice; the range can be reduced by increasing R1, R2. Note that R4 and R5 are needed to maintain negative feedback for unity gain at DC, and to keep the stage biased properly. They must be high in value compared with the impedances in the rest of the circuit.

It is not possible to obtain high values of Q with this configuration, and that is probably its major drawback. The capacitor ratio in Figure 11.11 gives the maximum possible Q. This equaliser is therefore mainly useful for dealing with large-scale trends in the frequency response.

As with the previous equalisers, the shunt-feedback configuration used here means there is no common-mode voltage on the opamp inputs, but with that comes an inconvenient phase-inversion which must be taken into account in the system design.

11.9 Adjustable Peak/Dip Equalisers: Variable Centre Frequency and Low Q

The fixed-frequency peak/dip equaliser we have just looked at gives excellent control over the amount of boost or cut applied, but the centre frequency can only be altered by changing the capacitors. In the development phase of crossover design it is extremely useful to be able to temporarily include an equaliser that also has continuously variable control of its centre frequency as well as the amount of peak or dip. When optimisation of the crossover is complete, it is replaced in the final design by a fixed equaliser like that in the previous section.

The circuit shown in Figure 11.13 is also based on mixing console technology, where it is usually called a "sweep-middle EQ." It is based on a modified Wien-bridge network of the sort sometimes used in oscillators; this acts as a low-Q bandpass filter, so that only a selected band of frequencies reach the non-inverting input of A1. When RV1 is in the boost position the input signal passes through the Wien network and adds to the output, giving a peaking response. When RV1 is in the cut position the signal through the Wien network constitutes extra negative feedback, and the gain is reduced, causing a dip in the response.

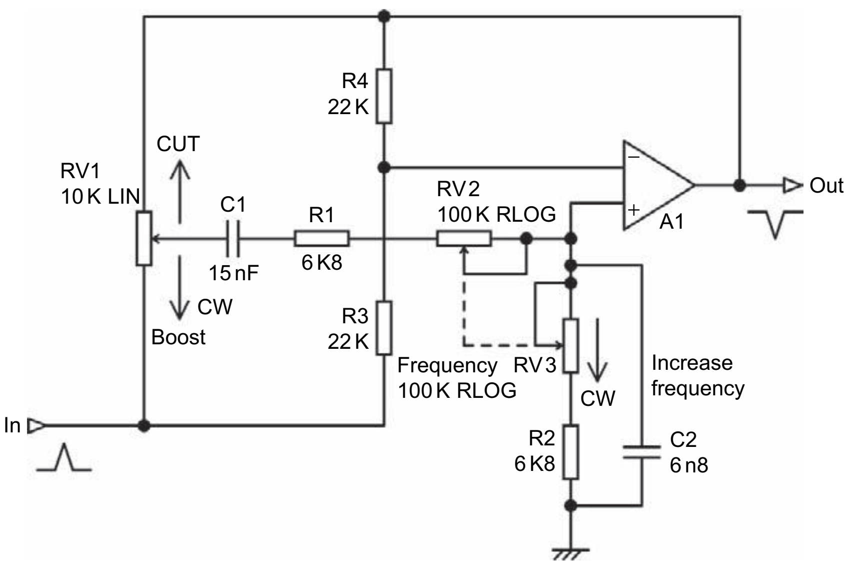


Figure 11.13: A peak/dip equaliser with variable centre frequency, intended for crossover development work.

The variable load that the Wien network puts on the cut/boost pot RV1, and the variable source impedance from its wiper cause a small amount of interaction between boost/cut and centre frequency settings. This is not likely to cause any significant problem, but if necessary it could be eliminated by putting a unity-gain buffer stage between RV1 wiper and the Wien bandpass network. There will also be minor inaccuracies due to imperfect matching of the two sections of the frequency control.

The combination of 100K pot sections and a 6K8 end-stop resistors gives a theoretical centre frequency range of 15.7 to 1, which is about as much as can be usefully employed when using reverse-log Law C pots. Greater ranges will give excessive cramping of the frequency calibrations at the high-frequency end of the scale. Such calibrations should only be used as a rough guide; it is much more accurate to measure the response of the circuit after you have completed the optimisation of the crossover system. The measured frequency responses at the control limits are shown in Figure 11.14. The frequency range is from 150 Hz to 2.3 kHz; the ratio is slightly adrift from theory due to component tolerances. To obtain different frequencies scale C1 and C2, keeping the ratio between them the same. More information on this kind of equaliser can be found in *Small Signal Audio Design* [5].

This variable-frequency circuit is relatively complex compared with fixed equalisers, and is noisier because of the extra Johnson noise from the high-value frequency-determining resistors.

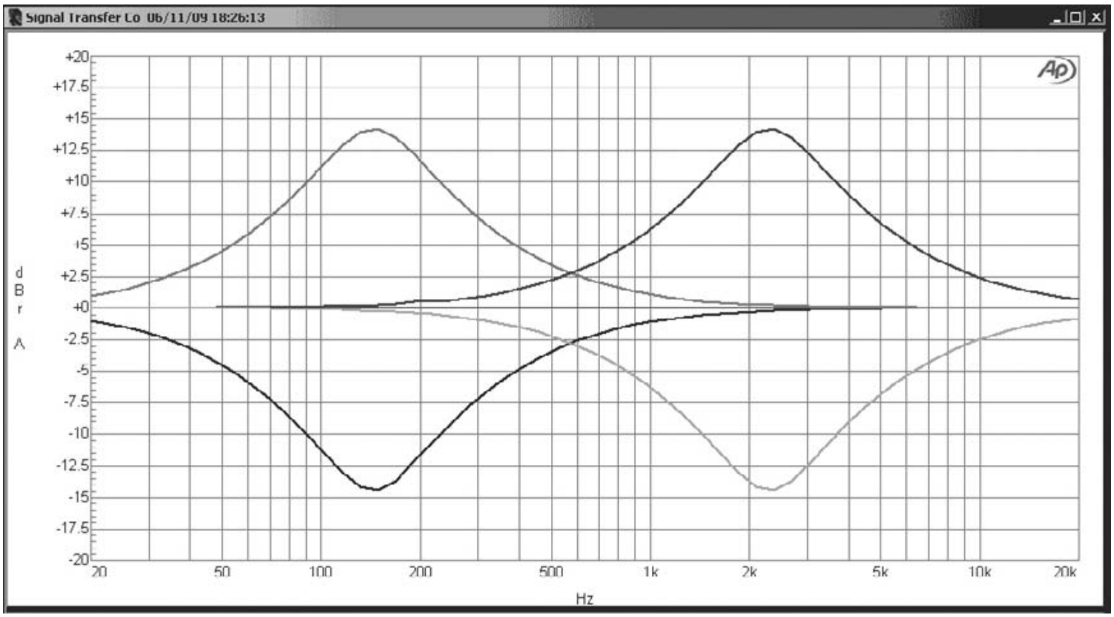


Figure 11.14: Frequency response of the variable-frequency peak/dip equaliser, showing extremes of frequency setting.

In production it would be replaced by a fixed equaliser like that in the previous section, with component values set to give the desired response.

11.10 Adjustable Peak/Dip Equalisers with High Q

The two peak/dip equalisers we have just examined have low Qs and are not suitable for dealing relatively narrow response irregularities; obtaining higher Qs requires more complex circuitry. There are several different approaches that might be taken; for example a state-variable filter would give the most flexibility, with control over centre frequency and Q as well as the amount of peak or dip. While it could be very useful for optimisation, it would, however, be excessively complex and costly for permanent inclusion in a crossover design. The approach I have chosen here is based on using a gyrator to simulate a series LC resonant circuit.

The essence of the scheme is shown in Figure 11.15, which the alert reader will spot as the basic concept behind graphic equalisers [6]. L1, C1 and R3 make up an LCR series resonant circuit; this has a high impedance except around its resonant frequency; at this frequency the reactances of L1 and C1 cancel each other out and the impedance to ground is that of R3 alone. (A parallel LC circuit works in the opposite way, having a low impedance at all frequencies except at resonance.) At the resonant frequency, when the wiper of RV1 is at the R1 end of its track, the LCR circuit forms the lower leg of an attenuator of which R1 is the upper arm; this attenuates the input signal and a dip in the frequency response is therefore produced. When the RV1 wiper is at the R2 end, an

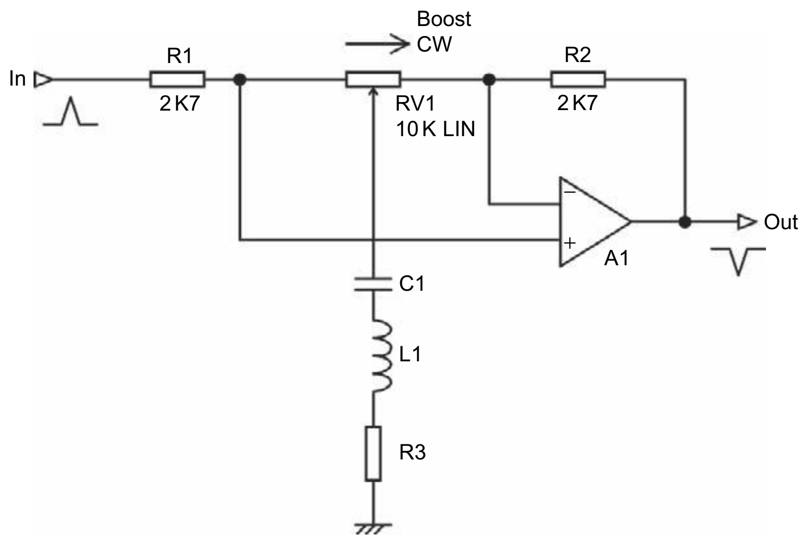


Figure 11.15: The basic idea behind the peak/dip equaliser; gain is unity with the wiper central.

attenuator is formed with R2 that reduces the amount of negative feedback at resonance and so creates a peak in the response. It is not exactly intuitively obvious, but this process does give absolutely symmetrical cut/boost curves. At frequencies away from resonance the impedance of the RLC circuit is high and the gain of the circuit is unity.

Inductors are always to be avoided if possible; they are relatively expensive, often because they need to be custom-made. Unless they are air-cored (which limits their inductance to low values) the ferromagnetic core material will cause non-linearity. They can crosstalk to each other if placed close together, and can be subject to the induction of interference from external magnetic fields. In general they deviate from being an ideal circuit element much more than resistors or capacitors do.

Gyrator circuits are therefore extremely useful, as they take a capacitance and "gyrate" it so it acts in some respects like an inductor. This is simple to do if one end of the wanted inductor is grounded—which fortunately is the case here. Gyrators that can emulate floating inductors do exist but are far more complex.

Figure 11.16 shows how it works; C1 is the normal capacitor as in the series LCR circuit, while C2 is made to act like the inductor L1. As the applied frequency rises, the attenuation of the highpass network C2, R1 falls, so that a greater signal is applied to unity-gain buffer A1 and it more effectively bootstraps the point X, making the impedance from X to ground increase. Therefore, we have a part of the circuit where the impedance rises proportionally to frequency—which is just how an inductor behaves. There are limits to the Q values that can be obtained with this circuit because of the inevitable presence of R1 and R2. The remarkably simple equation for the inductor value is shown; note that this includes R2 as well as R1.

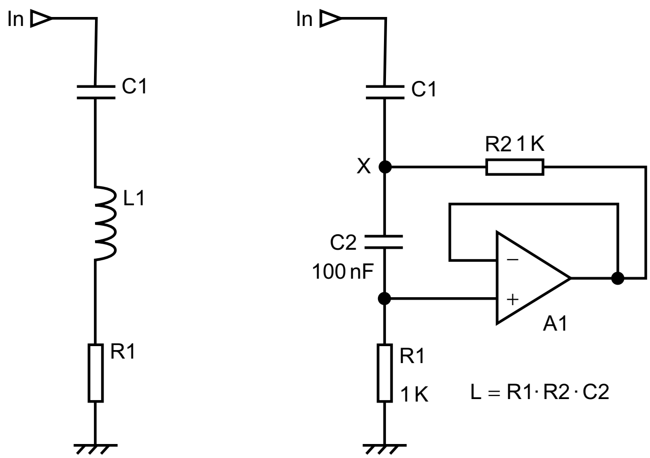


Figure 11.16: Synthesising a grounded inductor in series with a resistance using a gyrator.

The gyrator example in Figure 11.16 has values chosen to synthesise a grounded inductor of 100 mH in series with a resistance of 2 kΩ; that would be quite a hefty component if it was a real coil, but it would have a much lower series resistance than the synthesised version.

Figure 11.17 shows a gyrator-based high-Q peak/dip equaliser, with a centre frequency of 1 kHz. The Q at the maximum boost or cut of 6.3 dB is 2.2, considerably higher than that of the previous peak/dip equaliser we looked at, and much more suitable for correcting localised response errors. The maximal cut and boost curves and some intermediate boost values are seen in Figure 11.18. The +4 dB peak results from the values R2 = 9 KΩ and R3 = 1 KΩ. The +1.5 dB peak results from the values R2 = 7 KΩ and R3 = 3 KΩ. For development work R2 and R3 can be replaced by a 10 KΩ pot RV1. To obtain different centre frequencies scale C1 and C2, keeping the ratio between them the same.

The beauty of this arrangement is that two, three, or more LCR circuits, with associated cut/boost resistors or pots, can be connected between the two opamp inputs, giving us an equaliser with pretty much as many bands as we want. It is, after all, based on the classic graphic equaliser configuration.

This configuration can produce a response dip with well-controlled gain at the deepest point, but it is not capable of generating very deep and narrow notches of the sort required

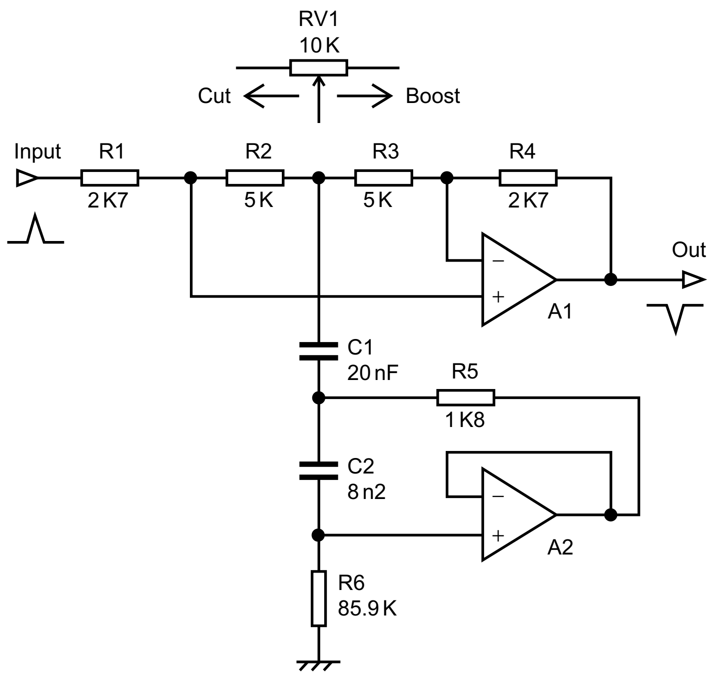


Figure 11.17: Gyrator-based high-Q peak/dip equaliser, with centre frequency fixed at 1 kHz. For development work R2 and R3 can be replaced by a 10 KΩ pot RV1.

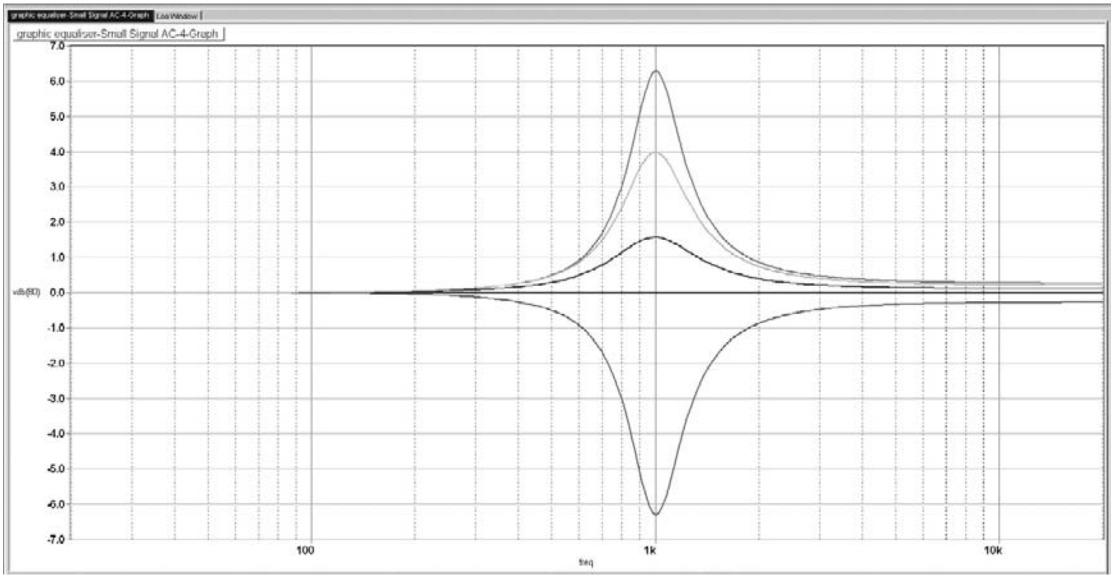


Figure 11.18: Frequency response of gyrator high-Q equaliser with various boost/cut settings.

for notch crossovers (see Chapter 5). However, notches for equalisation purposes are not normally required to be particularly deep or narrow. The implementation of filters that do have deep and narrow notches is thoroughly dealt with in Chapter 9.

11.11 The Bridged-T Equaliser

There are many types of bridged-T equalisers [7], but the configuration shown in Figure 11.19 is probably the best known; it is described by Linkwitz in [8]. This equaliser is potentially useful for modifying the LF roll-off of loudspeakers, but it has serious limitations compared with the biquad equaliser described in the next section.

One of these limitations is that no design equations are in common use. Figure 11.19 shows the values used for the attempt at LF equalisation in Figure 11.20, using the constraint R1a = R1b = R1, and I quite happily admit that they were obtained by twiddling values on a simulator and not by any more sophisticated process. The LF response of the loudspeaker shows a fairly high Q of 1.20 and a –3 dB cutoff point of 74 Hz. This apparently random value derives from the fact that if the Q was reduced to 0.7071, then the –3 dB point would be at the nice round number of 100 Hz. As you can see, it is possible to convert the peaking response to something approximating maximally flat, but in the process the –3 dB frequency has actually increased from 74 Hz to 85 Hz, and there is a rather unhappy-looking bend in the combined response around 60 Hz. Not until 50 Hz does the equalised response exceed the unmodified loudspeaker response. Despite this, the equaliser gain at 40 Hz is +4.0 dB,

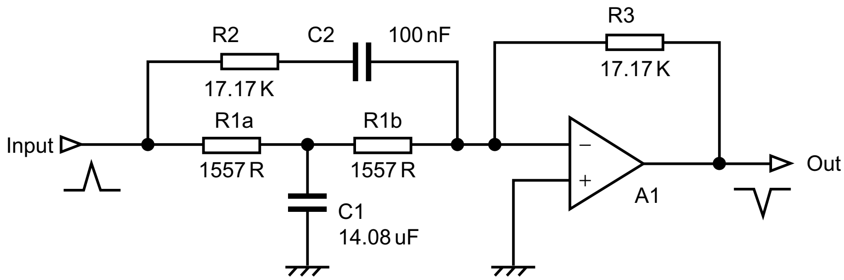


Figure 11.19: Bridged-T equaliser circuit designed for loudspeaker LF extension.

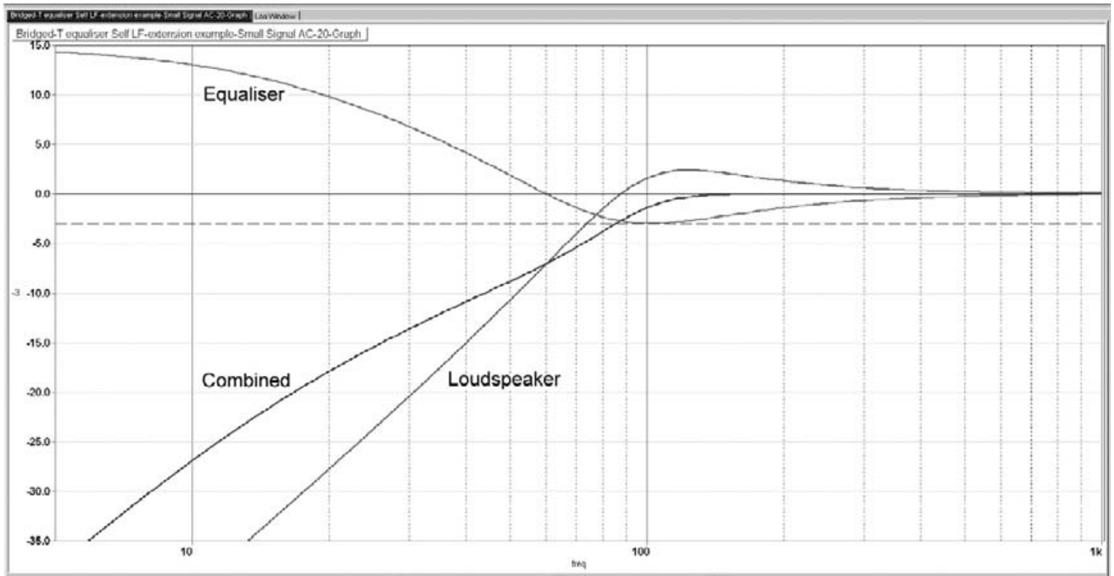


Figure 11.20: Low-frequency extension using a biquad equaliser for a loudspeaker with an unequalised Q of 1.20. The dotted line is at –3 dB.

and so the amplifier power will have to be more than doubled to maintain the maximum SPL down to this frequency. All in all, the result is not very satisfactory.

This bridged-T configuration has the disadvantage of high noise gain—in other words, the gain of the circuit for the opamp noise is greater than the gain for the signal. This is because at high frequencies C1 has negligible impedance so we have R1b connected from the virtual-earth point at the opamp inverting input to ground. R2 also affects the noise gain but by a small amount. In this case the noise gain is (R3 + R1b)/R1b, which works out at a rather horrifying twelve times or +21.8 dB. This stage will be much noisier than the active filters and other parts of the crossover system, and this is a really serious drawback.

Neville Thiele described several different bridged-T networks for loudspeaker LF extension in 2004 [7], but these work rather differently since they are placed in the feedback path of an opamp. He noted that he was surprised that he could find no earlier analysis of bridged-T networks. I can testify that there appears to be no useful material on these configurations on the Internet in December 2010.

11.12 The Biquad Equaliser

An especially effective equaliser is a combination of two bridged-T equalisers, as shown in Figure 11.21. It is much better adapted to equalising drive unit errors, being able to simultaneously generate a peak and an independently controlled dip in the response. It is referred to as a biquad equaliser because its mathematical description is a fraction with one quadratic equation divided by another.

While it is possible to alter all of the ten passive components independently, in practice it is found that the easiest way to get manageable design equations is to set:

$$R1a = R1b = R1 \tag{11.1}$$

$$R2a = R2b = R2 \tag{11.2}$$

$$R3a = R3b = R3 \tag{11.3}$$

$$C2a = C2b = C2 \tag{11.4}$$

This approach is illustrated in Figure 11.21; it reduces the number of degrees of freedom to four, and the amplitude response is conveniently defined by f_0, Q_0 (the frequency and Q of the dip in the response) and f_p, Q_p (the frequency and Q of the peak in the response).

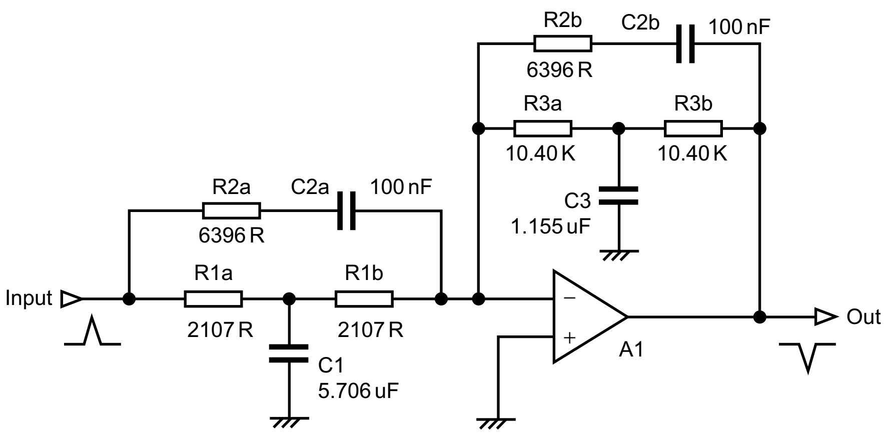


Figure 11.21: Biquad equaliser circuit designed for $f_0 = 100$ Hz, $Q_0 = 1.5$, and $f_p = 45$ Hz, $Q_p = 1.3$.

Note that Q values greater than 0.707 ($1/\sqrt{2}$) are required to give peaking or dipping. The circuit in Figure 11.21, derived from a crossover design I did for a client, has $f_0 = 100$ Hz, $Q_0 = 1.5$, and $f_p = 45$ Hz, $Q_p = 1.3$, and so helpfully shows both a peak and a dip with different Q's; its response is shown in Figure 11.22. You will see that the gain flattens out at +13.9 dB at the LF end, considerably greater than the HF gain of 0 dB. The further apart the f_0 and f_p frequencies are, the greater in difference the gain will be, and care is needed to make sure it does not become excessive. The actual peak frequency in Figure 11.22 is 36 Hz and the dip frequency 115 Hz, differing from the values put into the design equations because f_0 and f_p are sufficiently close together for their responses to interact significantly.

The value of R2a and R2b (i.e., R2) has been kept reasonably low to reduce noise but, as a consequence of the low frequencies at which the equaliser acts, the capacitors C1 and C3 are already sizable, and it is not very feasible to reduce R2 further. C1 in particular will be expensive and bulky if it is a non-electrolytic component.

The gain of this circuit always tends to unity at the HF end of the response, so long as R2a = R2b, because at high frequencies the impedance of C2a, C2b becomes negligible and R2a, R2b set the gain. At the LF end all capacitors may be regarded as open-circuit, so gain is set by R3/R1, and may be either much higher or much lower than unity, depending on the values set for f_0 and f_p. Note that the stage is phase-inverting, and this must be allowed for in the system design of a crossover.

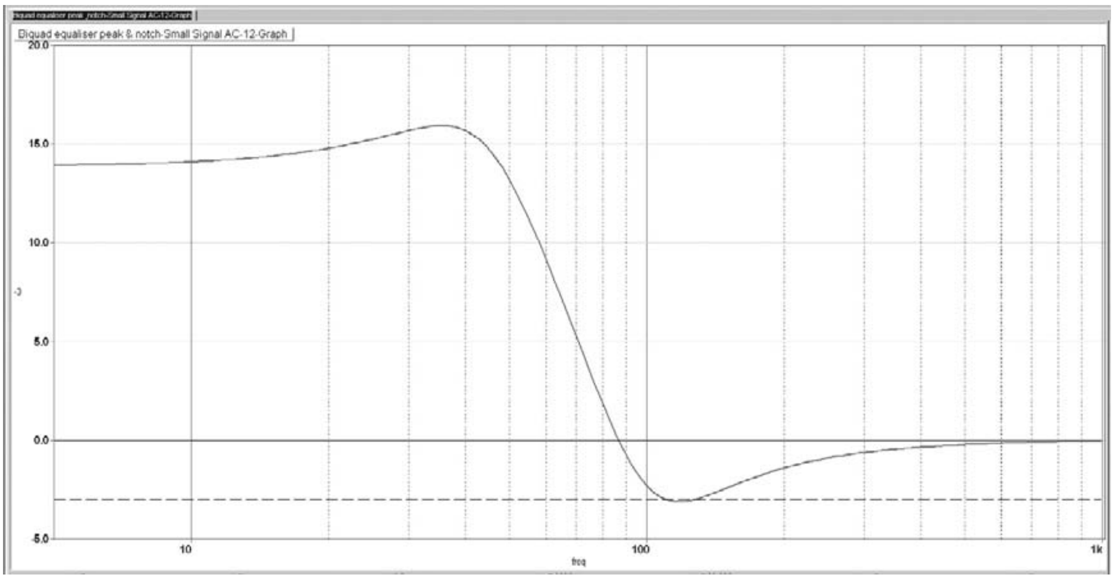


Figure 11.22: Frequency response of the biquad equaliser in Figure 11.21.

We saw in the previous section that the bridged-T configuration has the disadvantage of high noise gain because at high frequencies C1 has negligible impedance and R1b is effectively connected from the virtual-earth point to ground. In this case the value of R2a cannot be neglected as it is comparable with R1b, so the parallel combination of the two is used to give a more accurate equation for the noise gain:

$$\text{Noise gain} = \frac{(\mathrm{R2a} + \mathrm{R1b})\mathrm{R2b} + (\mathrm{R2a} \cdot \mathrm{R1b})}{\mathrm{R2a} \cdot \mathrm{R1b}} \tag{11.5}$$

Thus the noise gain for Figure 11.21 comes to 5.04 times or +14.0 dB, which is certainly uncomfortably high but nothing like as bad as the +21.8 dB given by the bridged-T equaliser in the previous section.

The design procedure given here is based on the equations introduced by Siegfried Linkwitz [8]. The procedure is thus:

1. First calculate the design parameter k

$$k = \frac{\dfrac{f_0}{f_p} - \dfrac{Q_0}{Q_p}}{\dfrac{Q_0}{Q_p} - \dfrac{f_p}{f_0}} \tag{11.6}$$

   The parameter k must come out as positive or the resistor values obtained will be negative.
2. Choose a value for C2 (a preferred value is wise, because it's probably the only one you're going to get, and it occurs twice in the circuit) then calculate R1:

$$\mathrm{R1} = \frac{1}{2\pi \cdot f_0 \cdot \mathrm{C2} \cdot \left(2Q_0(1+k)\right)} \tag{11.7}$$

3. Calculate R2, C1, C3 and R3

$$\mathrm{R2} = 2 \cdot k \cdot \mathrm{R1} \tag{11.8}$$

$$\mathrm{C1} = \mathrm{C2}\left(2Q_0(1+k)\right)^2 \tag{11.9}$$

$$\mathrm{C3} = \mathrm{C1}\left(\frac{f_p}{f_0}\right)^2 \tag{11.10}$$

$$\mathrm{R3} = \mathrm{R1}\left(\frac{f_0}{f_p}\right)^2 \tag{11.11}$$

The HF gain is always unity, because R2a = R2b, but the LF gain in dB is variable and is given by:

$$\mathrm{Gain_{LF}} = 40\log\left(\frac{f_0}{f_p}\right) \tag{11.12}$$

These equations can be very simply automated on a spreadsheet, which is just as well as several iterations may be required to get satisfactory answers.

The topology in Figure 11.21 appears to have been first put forward by Siegfried Linkwitz in 1978 [3]; it is not clear if he invented it, but it seems he was certainly the first to develop useful design equations. Its use for equalising the LF end of loudspeaker systems was studied by Greiner and Schoessow in 1983 [9]. Rather surprisingly, Greiner and Schoessow laid very little emphasis on the flexibility and convenience of this topology. The shunt-feedback configuration means that the two halves of the circuit are completely independent of each other and there is no interaction of the peak and dip parameters. However, it is important to understand that if f_0 and f_p are close together the summation of their responses at the output may make the centre frequencies appear to be wrong at a first glance.

We will now see how this equaliser can be used for LF response extension. Figure 11.23 shows the low-end response of the loudspeaker we used in the previous section, with the relatively high Q of 1.20 and a –3 dB cutoff point of 74 Hz, which corresponds to a –3 dB point would be at 100 Hz if the Q was reduced to 0.7071 for maximal flatness. The loudspeaker response peaks by 2.4 dB at 124 Hz. We now design a biquad equaliser with

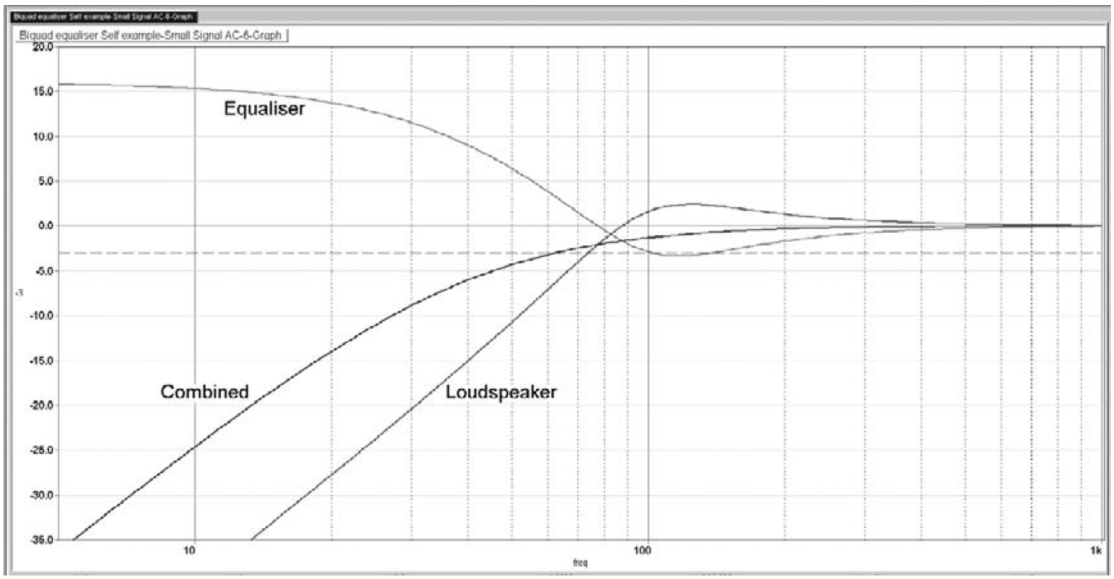


Figure 11.23: Low-frequency extension using a biquad equaliser for a loudspeaker with an unequalised Q of 1.20. The dotted line is at –3 dB.

$f_0 = 100$ Hz—note that is the frequency for the Q = 0.7071 version of the loudspeaker—and a Q of 1.20. This will cancel out the peaking in the loudspeaker response. The choice of f_p and Q_p depends on how ambitious we are in our plan to extend the LF response, but in this case f_p has been set to 40 Hz and Q_p to 0.5.

Figure 11.23 shows that the response peak from the original loudspeaker has been neatly cancelled, and the overall LF response extended. The –3 dB point has only moved from 74 Hz to 62 Hz, which does not sound like a stunning improvement, but the gentler roll-off of the new response has to be taken into account. The –6 dB point has moved from 63 Hz to 40 Hz, a rather more convincing change. The LF roll-off takes on the Q of Q_p, which at 0.5 would be considered over-damped by some critics, but I adopted it deliberately as it shows how equalisation can turn an under-damped response into an over-damped one.

It is worth noting that it is not possible to set f_p to 50 Hz instead of 40 Hz while leaving the other setting unchanged. This gives a negative value for the parameter *k* and a negative value of –3537 Ω for R2. While it is possible to construct circuitry that emulates floating negative resistors, it is not a simple business. If you run into trouble with this you should use a different kind of equaliser.

The improvements may not be earth-shattering, but bear in mind that the equaliser has a gain of 9.0 dB at 40 Hz, so if the maximum SPL is going to be maintained down to that frequency, the output of the associated power output amplifier will have to increase by almost eight times. Driver cone excursion increases rapidly—quadrupling with each octave of decreasing frequency for a constant sound pressure, so great care is required to prevent it becoming excessive, leading to much-increased non-linear distortion and possibly physical damage. Thermal damage to the voice-coil must also be considered; a sobering thought.

The equaliser circuit for Figure 11.23 is shown in Figure 11.24. The noise gain is now a much more acceptable 2.1 times or +6.4 dB, which emphasises that the biquad equaliser is much superior to the bridged-T in this respect. C3 could be a 100 nF capacitor in parallel with 1.5 nF without introducing significant error. C1 is a really awkward value, and if you are restricted to the E6 capacitor series, would have to be made up of 330 nF + 220 nF + 68 nF + 15 nF in parallel, which adds up to 633 nF, an error of +0.32%. This is a relatively expensive solution, especially if you are using polypropylene capacitors to avoid distortion, and will also use up a lot of PCB area. A solution to this problem is provided in the next section.

It is instructive to try some more options for LF extension. Once the design equations have been set up on a spreadsheet and the circuit built on a simulator (including a block to simulate the response of the loudspeaker alone, which can in many circumstances be a simple second-order filter of appropriate cutoff frequency and Q) variations on a theme can be tried out very quickly.

Figure 11.25 shows an optimistic attempt to further extend the LF response further by setting the equaliser f_p to 50 Hz and keeping the Q_p at 0.5. The –3 dB point is now moved from 74 Hz to 31 Hz, which would be a considerable improvement were it practical, but

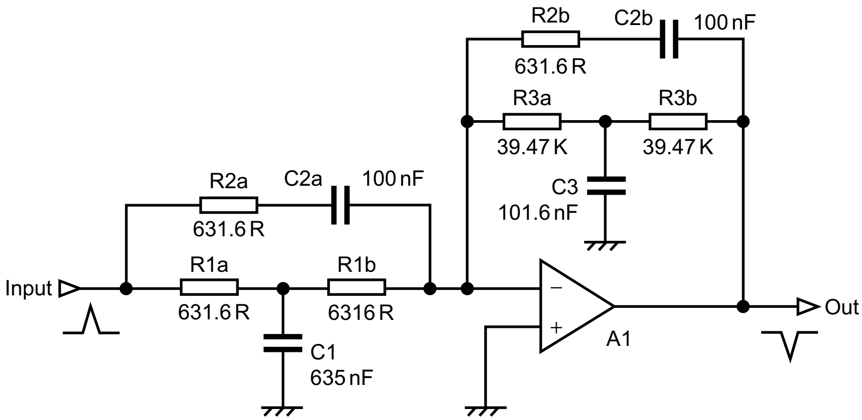


Figure 11.24: Schematic of a biquad equaliser designed for LF extension with exact component values.

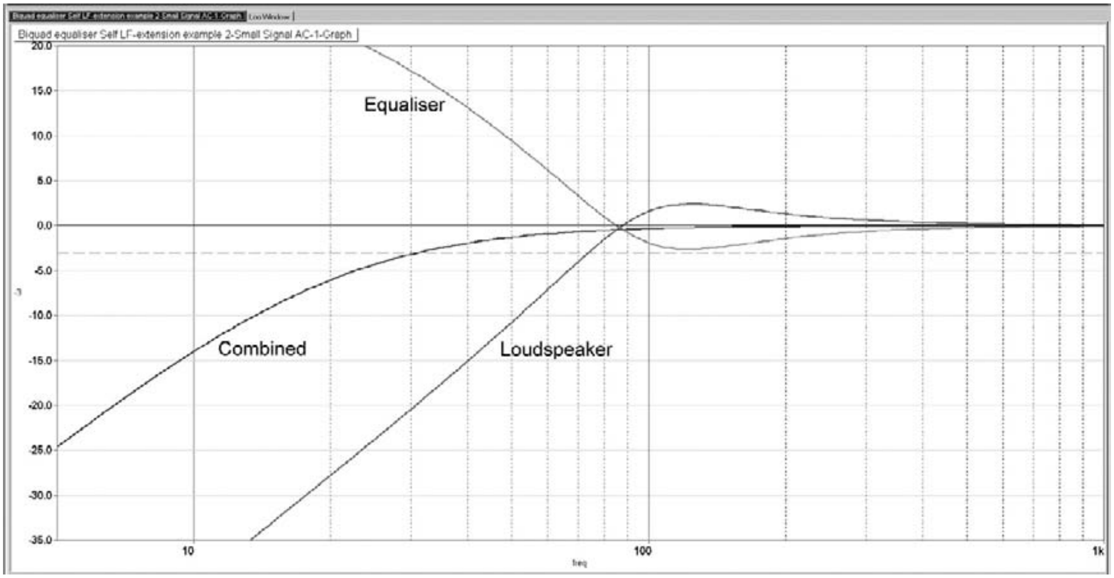


Figure 11.25: An ambitious attempt at low-frequency extension with the equaliser $f_p = 20$ Hz and $Q_p = 0.5$. The combined response is now –3 dB at 31 Hz but the equaliser gain to achieve this is excessive. The dotted line is at –3 dB.

note that the equaliser gain is now heading off the top of the graph, and does not begin to level out until around 5 Hz, at which point the gain has reached 27 dB. At the 31 Hz –3 dB point, equaliser gain is +17.4 dB, which would require no less than 55 times as much power from the amplifier and a truly extraordinary drive unit to handle it. It is important to realise that there is only so much you can do with LF extension by equalisation.

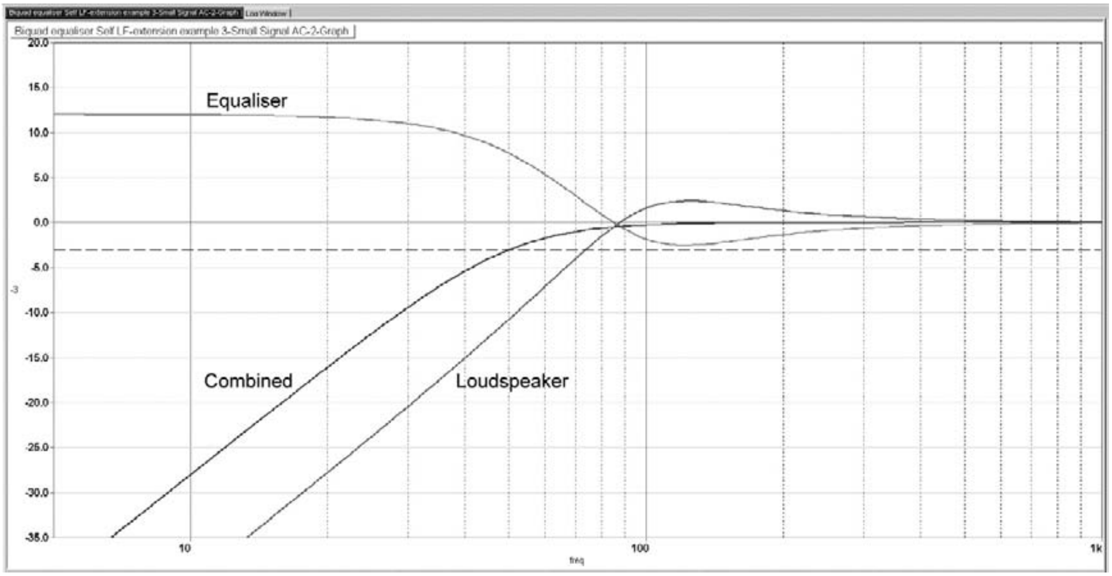


Figure 11.26: A more realistic plan for low-frequency extension, with the equaliser $f_p = 50$ Hz and $Q = 0.707$. The modified response is now –3 dB at 50 Hz and the equaliser gain is now acceptable. The dotted line is at –3 dB.

Figure 11.26 shows a more reasonable approach. The aim is to make the combined response maximally flat ($Q = 0.707$) rather than overdamped, and keep the demands for extra amplifier power and increased cone excursion within practicable limits. We therefore set the equaliser f_p to 50 Hz and the Q_p at 0.707. The –3 dB point is now moved from 74 Hz to 50 Hz, which is well worth having, but the equaliser gain never exceeds 12 dB and so the amplifier power increase is limited to a somewhat more feasible but still very substantial 16 times. If it is accepted that maximum SPL will only be maintained down to the new –3 dB frequency, only six times as much power is required. In practice the drive unit cone excursion constraints and the thermal performance of the voice-coil assembly will probably set the limits of what is achievable.

An example of the use of the biquad equaliser for LF response extension is shown in the Christhof Heinzerling subtractive crossover [10]. This design also includes a boost/cut equalisation control for the low LF based on the 2-C version of the Baxandall tone-control.

11.13 Capacitance Multiplication for the Biquad Equaliser

Since this equaliser is commonly used at the LF end of the spectrum and must work at low frequencies, if the resistors are to be kept low to minimise noise, then the capacitors can become inconveniently large in value, physical size, and cost. This is particularly true if polypropylene capacitors are used to prevent capacitor distortion. The use of a capacitance-multiplier architecture in an LF biquad equaliser means that the capacitors can be kept to a reasonable size, and any desired capacitance value can be obtained without paralleling components, simply by varying the multiplication factor.

The capacitance multiplication is very straightforward because the capacitors in question are grounded at one end in the basic circuit. Applying it to a floating capacitor would be much more difficult. The basic plan is to connect the normally grounded end of the capacitor to an opamp output, which is then driven so that as the voltage at the top of the capacitor (point A in Figure 11.27) rises, then the voltage at the bottom of the capacitor falls, increasing the current through the capacitor for a given voltage at point A. If the fall is equal in voltage to the rise, then as far as the rest of the circuit is concerned, the value of the capacitor has doubled. Varying the amount by which the bottom of the capacitor is driven allows the multiplication factor to be set to any desired value by varying a single resistor.

This technique is implemented by the circuitry in Figure 11.27, which implements the circuit of Figure 11.24, without the need to find and pay for a 635 nF capacitor. A3 is a shunt-feedback inverting stage that drives the bottom of capacitor C1 in anti-phase to the voltage at it top end. R4 and R5 are kept low in value to minimise current noise and Johnson noise, and the resulting low input impedance of the A3 stage is therefore buffered

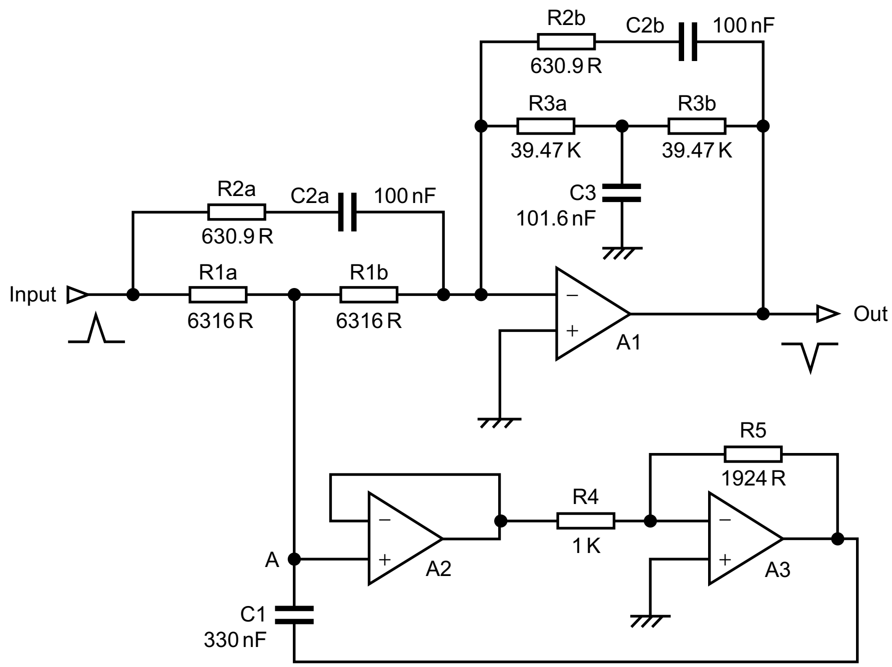


Figure 11.27: Capacitance multiplication applied to the biquad equaliser of Figure 11.24. Here a 330 nF capacitor acts like a 635 nF capacitor as its lower terminal is driven in anti-phase.

from the rest of the circuit by the voltage-follower A2. The multiplication factor, determined by R5/R4, is 1.924 times. so the 330 nF capacitor will appear to the rest of the equaliser circuit as 635 nF.

Any ingenious new circuitry that adds more active devices needs to be carefully scrutinised to make sure that neither the noise or distortion performance is unduly compromised. The capacitance multiplier method shown here does not significantly degrade either noise or distortion. The lack of extra noise is explained by the fact that if a signal is injected into the inverting input of A3 (which is a virtual-earth point) through a resistor of equal value to R5, then the gain to the output of the stage is –11 dB, and so the extra noise contribution from A2 and A3 is of little significance.

The technique does require careful consideration of signal levels. Given the condition R1a = R1b referred to above, the signal voltage at the top of C1 is half that at the input of the equaliser. Therefore if multiplication factors greater than two are used, clipping may occur in the capacitance multiplier at the output of A3 before it does in any other part of the circuit.

11.14 Equalisers with Non-6 dB Slopes

Since lowpass and highpass filter slopes come only in multiples of 6 dB/octave, creating any other slope requires some approach other than a straightforward filter. Lesser slopes can be made over small frequency ranges by putting single highpass and lowpass time-constants close together, but this will not work over larger ranges. There are more sophisticated ways of combining time-constants to get a non-standard response slope, which I shall illustrate with a classic non-loudspeaker example. Pink noise is much more useful than white noise for audio measurement because it gives a flat line on a spectrum analyser, but the various methods of noise generation all give white noise. White noise is therefore converted into pink noise by what is called a "pinkening filter," which has a –3 dB/octave slope over the whole audio band from 20 Hz to 20 kHz.

A 3 dB/octave slope is proverbially difficult, or at least non-obvious, to obtain as the ultimate slopes of filters come in multiples of 6 dB/octave only. The standard solution when you need a pinkening filter is a series of overlapping lowpass and highpass time-constants (i.e., alternate poles and zeros) that approximate to the required slope. This gives a response that wiggles up and down around a –3 dB/octave line, and the more pairs of poles and zeros used, the less the wiggle and the more accurately the response approximates to the line.

Two possible versions are shown in Figure 11.28; in both cases capacitor values have been restricted to the E6 series. The simpler version uses three RC networks to create pole-zero pairs, with a final unmatched pole introduced by C4. Its response is shown in Figure 11.29 with an exact –3 dB/octave line (dotted) added for comparison; the error trace at the top has

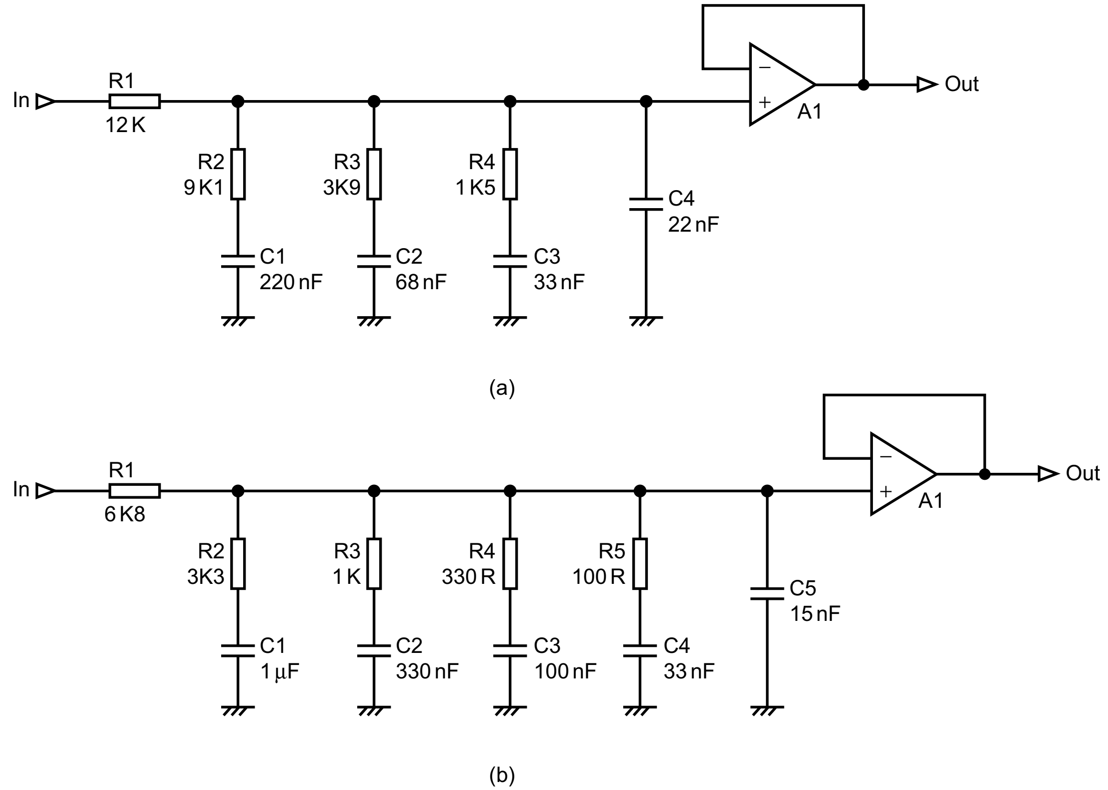


Figure 11.28: Two –3 dB/octave filters built from repeated RC networks. The version at (b) is more accurate over a wider frequency range.

been multiplied by ten times to make the deviations more visible. The worst errors in the 100 Hz–10 kHz range are +0.22 dB at 259 Hz, –0.09 dB at 2.5 kHz, and +0.36 dB at 8.2 kHz, which I hope you will agree is not bad for a simple circuit.

This 3x RC circuit was originally published in *Small Signal Audio Design* [11], but with R2 = 12K, R3 = 3K9, and R4 = 1K2. This had errors of up to ± 0.65 dB, because of a historical requirement to use only E12 resistor values. The 3x RC circuit is not the best option unless you are really closely counting the cost of every component, for reasons that will now appear.

Figure 11.30 shows the response of the version with four RC networks shown in Figure 11.28b. It has four pole-zero pairs, with a final unmatched pole. It gives rather better performance for two reasons, one obvious, the other less so. Clearly the more RC networks are used, the closer is the pole-zero spacing, and so there less the wobble on the frequency response. Less obvious is the fact that four RC networks allow the pole-zero frequencies required to match E6 capacitor values much better—you will note the tidy pattern of

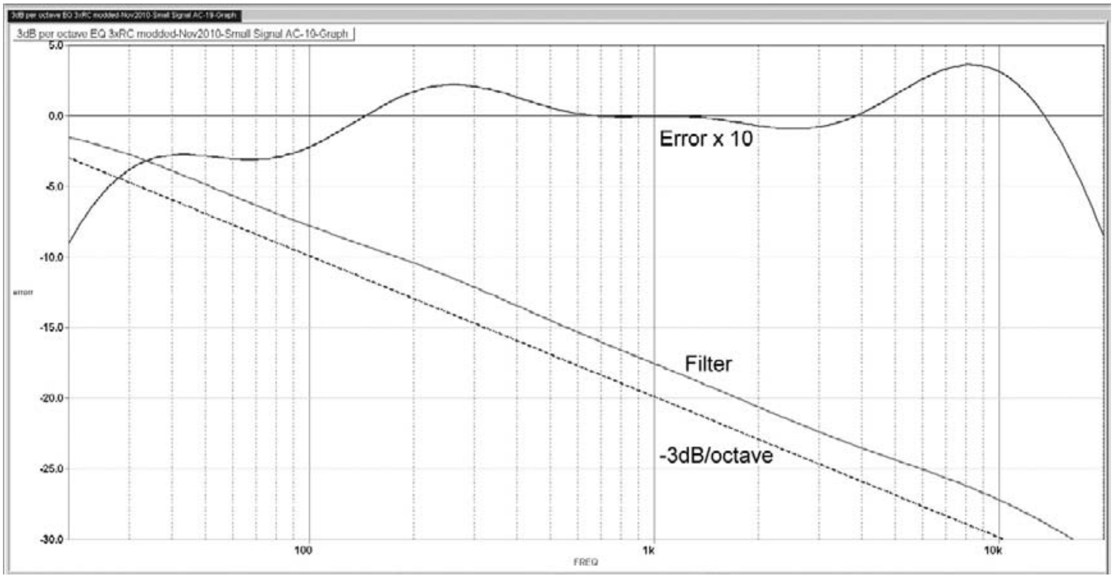


Figure 11.29: The response of the 3x RC –3 dB/octave filter in Figure 11.28a. The error trace at top (multiplied by ten) shows a maximum error of +0.36 dB in the 100 Hz–10 kHz range. The dotted line shows an exact –3 dB/octave slope.

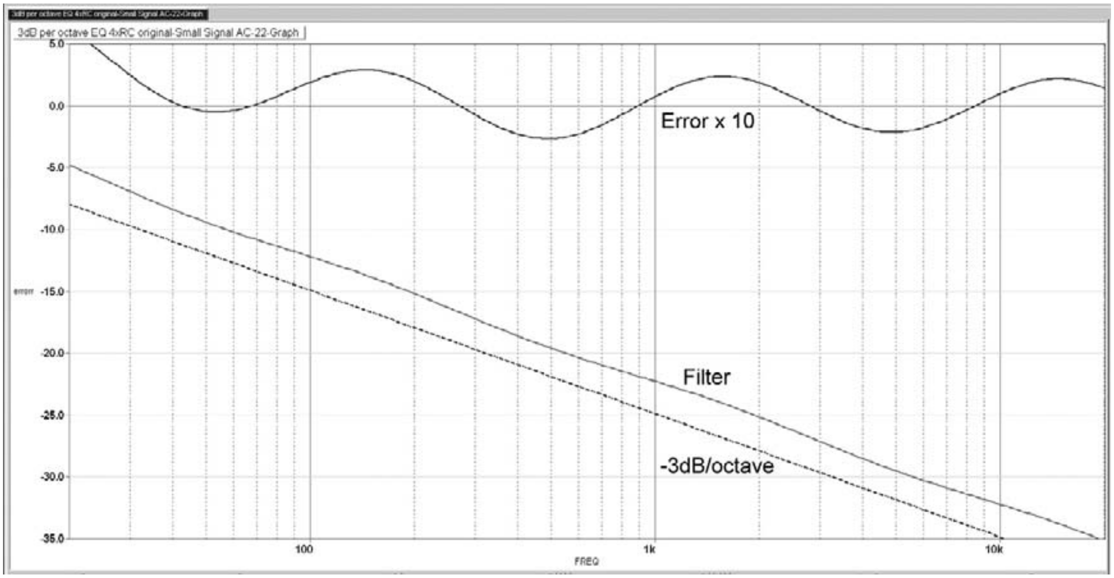


Figure 11.30: The response of the 4x RC –3 dB/octave filter in Figure 11.28b. The error trace at top (multiplied by ten) shows a maximum error of +0.28 dB over the 20 Hz–20 kHz range.

component values in Figure 11.28b, where 100/33 = 3.03 and 33/10 = 3.3, giving near-equal spacings on a log frequency scale.

The error plot now shows a pleasing sine-wave undulation around zero, with one cycle per decade of frequency, indicating that the errors (± 0.28 dB) are as low as they can be with this number of RC networks. The frequency range over which the errors are well-controlled is also much greater; a very useful 20 Hz–20 kHz span. This is a good return from adding just one resistor and one capacitor. Finally, you will note that our "premium" four-RC network has an overall lower impedance to reduce Johnson noise in the resistors. This means that C1 is relatively large at 1 uF, but I see no reason why the capacitance multiplication concept described for biquad equalisers should not work nicely in this application; I should however say that I have not yet tried it.

As we have just seen, it makes sense to go with the flow of the E6 capacitor series. The 4x RC network we have just looked at uses just two values from the series; 10 and 33. This principle can be extended by using three values instead of two; 10, 22, and 47. This gives ratios of 100/47 = 2.10, 47/22 = 2.14, and 22/10 = 2.20. The resulting circuit is seen in Figure 11.31, where exact resistor values derived from the capacitor ratios have are shown, rather than preferred values.

This gives considerably improved accuracy over the 4x RC network. The error is now within +0.1 dB over a frequency range extended to 10 Hz–20 kHz, though the error curve does not have the same symmetry; see Figure 11.32. It will of course be necessary to combine resistors to obtain those awkward values. With the E6 capacitor series the ultimate accuracy could be obtained by using all the values in a decade, with the ratios 100/68 = 1.47, 68/47 = 1.45, 47/33 = 1.42, 33/22 = 1.50, 22/15 = 1.47, and 15/10 = 1.50. For crossover design this would give much greater accuracy than that permitted by normal transducer variations.

This approach can be used to create a filter with any desired slope by altering the spacing of the poles and zeros. If a slope greater than 6 dB/octave but less than 12 dB/octave is required,

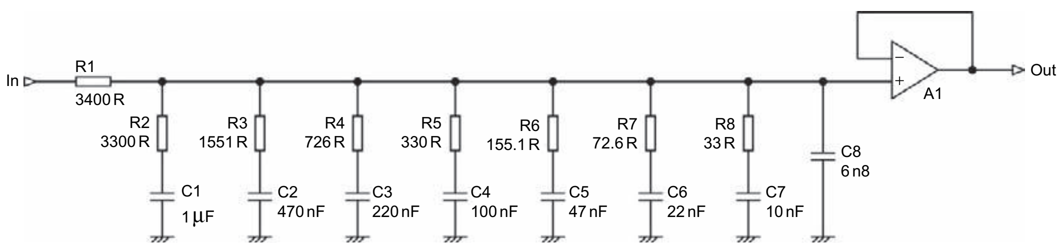


Figure 11.31: A –3 dB/octave filter with better accuracy using seven RC networks that fit in with the E6 capacitor series. Accuracy is within +0.1 dB over a 10 Hz–20 kHz frequency range.

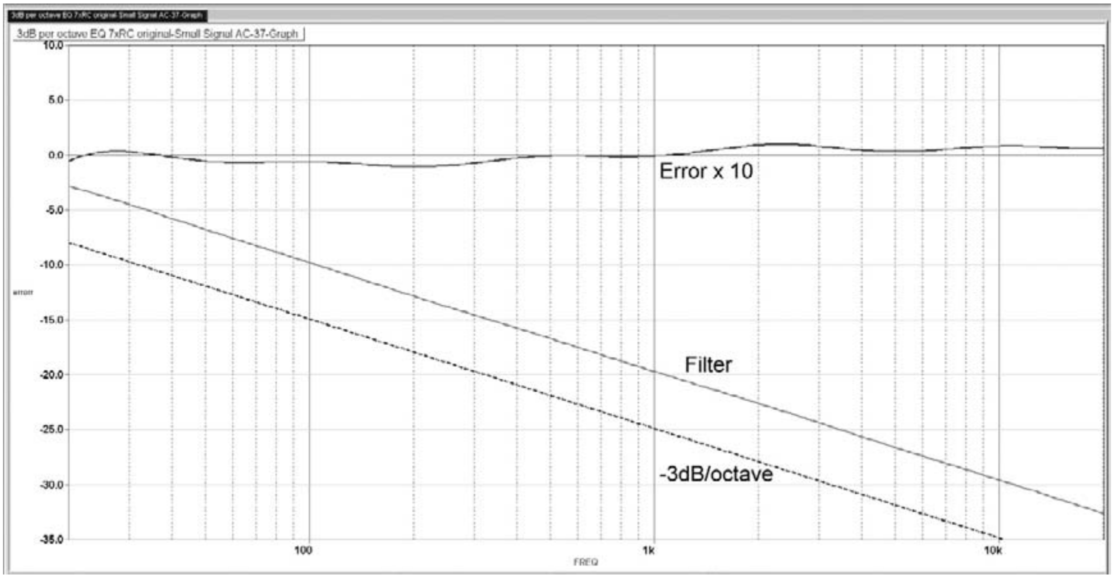


Figure 11.32: The response of the 7x RC –3 dB/octave filter in Figure 11.31. The error is now within +0.1 dB over a 10 Hz–20 kHz frequency range.

then a standard 6 dB/octave filter can be cascaded with one having a lesser slope. For example, cascading a –6 dB/octave filter with a –3 dB/octave filter gives a –9 dB/octave slope.

All the filters examined above give a response that falls at 3 dB/octave with frequency; to get a response that rises with frequency the networks can be inserted into the feedback network of an opamp. Stability will need to be checked.

11.15 Equalisation by Filter Frequency Offset

In Chapter 4 we saw that peaks and dips in the crossover region between a pair of filters can be manipulated and minimised by offsetting the frequencies of the two filters. If there is a peak at the crossover frequency then it can very often be reduced to a much smaller peak with two flanking dips by separating the filter frequencies. This technique can also be used for the purposes of response equalisation, though since there are only two variables (the highpass and lowpass filter cutoff frequencies) it is rather inflexible, and it is going to take a good deal of luck to get a response irregularity of just the right size and shape so it can be corrected simply by frequency offset. Even if this does happen, if you are manufacturing you might need to change one of the drive units for a different type, with different response irregularities, and you will then need to redesign the crossover to add extra equalisation stages that can cope with them.

11.16 Equalisation by Adjusting All Filter Parameters

Much greater flexibility is possible if other filter parameters apart from the cutoff frequencies are manipulated. If we assume we have a fourth-order Linkwitz–Riley crossover, then the lowpass path has two cascaded Butterworth second-order filters. Both the cutoff frequency and the Q of each Butterworth filter could be altered to meet specific requirements, giving us four variables. The highpass path of the crossover similarly has four variables, so there are many more degrees of freedom.

The problem is that all these variables have an interactive effect on the final response, and tweaking them to get a desired response is going to be a deeply tiresome business; some sort of automatic means of optimisation is highly desirable. The Linear-X Systems Filtershop CAD software [12] contains optimisation routines that can take measured drive unit amplitude response as input, and optimise the parameters of a chosen filter configuration to get the desired final response. It is a most impressive software package.

This approach has the advantage that it has much more ability to correct response irregularities than simple frequency-offsetting. It also requires no extra circuitry to perform the equalisation; on the other hand adding dedicated equalisation stages will only have a small impact on the price of a typical active crossover. It does have the serious drawback that the resulting circuitry is likely to be wholly opaque, with no indication of what parameters have been tweaked to compensate for what response irregularities. Good documentation is essential, and if specialised software is used to perform the optimisation, it will need to be kept available so that design changes can be made in necessary.

Passive crossovers must stringently minimise component count and power losses, so equalisation often has to be performed by manipulating filter cutoff frequencies and Q’s.

References

[1] D. Keele, What’s so sacred about exponential horns? AES Preprint No. 1038, 1975.
[2] H.F. Olson, Direct radiator loudspeaker enclosures, JAES 17 (1) (1969) 22–29.
[3] S. Linkwitz, Loudspeaker System Design, Electronics World, London, 1978.
[4] D. Self, Small Signal Audio Design, Chap. 10, Focal Press, Boston, MA, 2010, p. 271, ISBN 978-0-240-52177-0.
[5] D. Self, Small Signal Audio Design, Chap. 10, Focal Press, Boston, MA, 2010, p. 276, ISBN 978-0-240-52177-0.
[6] D. Self, Small Signal Audio Design, Chap. 10, Focal Press, Boston, MA, 2010, p. 286, ISBN 978-0-240-52177-0.
[7] N. Thiele, An active biquadratic filter for equalising overdamped loudspeakers, AES Convention Paper 6153, 116th Convention, 2004.
[8] S. Linkwitz, Loudspeaker System Design: Part 2, Wireless World, London, 1978, p. 71.
[9] R.A. Greiner, M. Schoessow, Electronic equalisation of closed-box loudspeakers, JAES 31 (3) (1983) 125–134.
[10] Heinzerling, Christhof, Adaptable Active Speaker System (subtractive), Electronics World, London, 2000, p. 105.
[11] D. Self, Small Signal Audio Design, Chap. 19, Focal Press, Boston, MA, 2010, p. 523, ISBN 978-0-240-52177-0.
[12] C. Strahm, Linear-X Systems Filtershop Application Manual, Application Note 5, p. 223.

CHAPTER 12

Passive Components for Active Crossovers

In this chapter we will look at some passive component properties that are especially important in the design of small-signal audio equipment in general and active crossovers in particular. All passive components differ from the ideal mathematical models—resistors have series inductance, capacitors have series resistance, and so on. There are also well-known issues with the accuracy of the component value, and the way that the value changes with temperature. What is less publicised is that some passive components can show significant non-linearity.

It is unwise to assume that all the distortion in an electronic circuit will arise from the active devices. This is pretty clearly not true if transformers or other inductors are in the audio path, but it is also a very unsafe assumption even if the only passive components you are using are resistors and capacitors. I recall that I was horrified when I first began designing active filters; the distortion from the capacitors completely obliterated the quite low THD from the discrete transistor unity-gain buffers I was using. More on this problem later.

Active filters are the building blocks of electronic crossovers, and their proper operation depends on having accurate ratios between resistors and capacitors. Setting up these ratios is complicated by the fact that capacitors are available in a much more limited range of values than resistors, often being restricted to the E6 series, which runs 10, 15, 22, 33, 47, 68. Fortunately resistors are widely available in the E24 series (twenty-four values per decade) and the E96 series (ninety-six values per decade). There is also the E192 series (you guessed it, 192 values per decade) but this is less freely available. Using the E96 or E192 series means you have to keep an awful lot of different resistor values in stock, so when non-standard values are required it is usually more convenient to use a series or parallel combination of two E24 resistors. More on this later, too.

12.1 Resistors: Values and Tolerances

Active filters very often require precise resistor values, or precise resistor ratios. When designing circuit blocks such as Sallen & Key highpass filters, where resistor ratios of exactly two are required, it is useful to keep in mind the options in the E24 series, as in Table 12.1, and the E96 series, as in Table 12.2. The E24 series offers six options for a ratio of two, while the E96 series offers twelve. In both cases the 1:2 pairs are closely spaced at the bottom of the decade, and this will have to be taken into account when choosing capacitor values.

The Design of Active Crossovers

Table 12.1: There Are Six E24 Resistor Values in a 1:2 Ratio

75	150
100	200
110	220
120	240
150	300
180	360

Table 12.2: There Are Twelve E96 Resistor Values in a 1:2 Ratio

100	200
105	210
113	226
140	280
147	294
158	316
162	324
174	348
187	374
196	392
221	442
232	464

In many cases you will aim to use a capacitor value or values in the relatively sparse E6 series, and this will almost invariably result in non-standard resistor values. These are easily obtained by putting two or more resistors in series or parallel, and multiple resistors will be cheaper than multiple capacitors. There are other advantages to this as I shall now explain.

12.2 Improving Accuracy with Multiple Components: Gaussian Distribution

Using two or more resistors to make up a desired value has a valuable hidden benefit. If it is done correctly it will actually increase the average accuracy of the total resistance value so it is *better* than the tolerance of the individual resistors; this may sound paradoxical but it is simply an expression of the fact that random errors tend to cancel out if you have a number of them. This works for any parameter that is subject to random variations, but for the time being we will focus on the concrete example of multiple resistors.

Resistor values are usually subject to a Gaussian distribution, also called a normal distribution. It has a familiar peaked shape, not unlike a resonance curve, showing that the majority of the values lie near the central mean, and that they get rarer the further away from the mean you look. This is a very common distribution in statistics, cropping up wherever there are many independent things going on that all affect the value of a given component. The distribution is defined by its mean and its standard deviation, which is the square-root of the sum of the squares of the distances from the mean—the RMS-sum, in other words. Sigma (σ) is the standard symbol for standard deviation. A Gaussian distribution will have 68.3% of its values within $\pm 1\ \sigma$, 95.4% within $\pm 2\ \sigma$, 99.7% within $\pm 3\ \sigma$, and 99.99% within $\pm 4\ \sigma$. This is illustrated in Figure 12.1, where the X-axis is calibrated in numbers of standard deviations on either side of the central mean value.

If we put two equal-value resistors in series, the total value has a narrower distribution than that of the original components. The standard deviation of summed components is the sum of the squares of the individual standard deviations, as shown in Equation 12.1. σ_{sum} is the overall standard deviation, and σ_1 and σ_2 are the standard deviations of the two resistors in series.

$$\sigma_{sum} = \sqrt{(\sigma_1)^2 + (\sigma_2)^2} \tag{12.1}$$

Thus if we have four 100 Ω 1% resistors in series, the standard deviation of the total resistance increases only by the square root of 4, that is 2 times, while the total resistance has increased by 4 times; thus we have inexpensively made an otherwise costly 0.5% close-tolerance 400 Ω resistor. There is a happy analogue here with the use of multiple amplifiers to reduce electrical noise; here we are using essentially the same technique to reduce

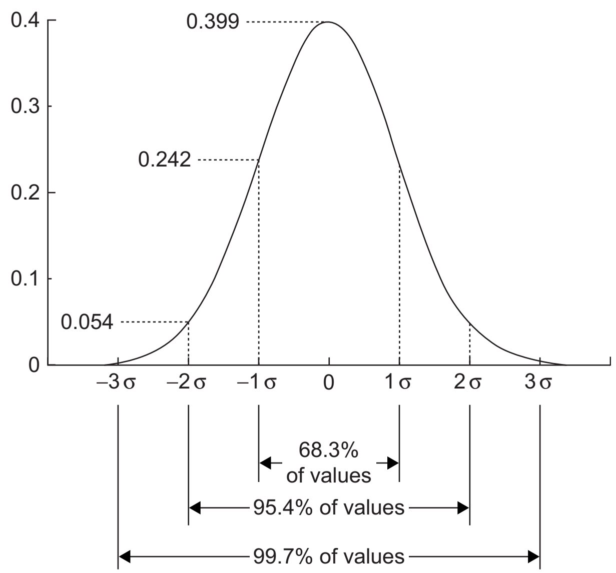


Figure 12.1: A Gaussian (normal) distribution with the X-axis marked in standard deviations on either side of the mean. The apparently strange value for the height of the peak is actually carefully chosen so the area under the curve is one.

"statistical noise." Note that this equation only applies to resistors in series, and cannot be used when resistors are connected in parallel to obtain a desired value. The improvement in accuracy nonetheless works in the same way.

You may object that putting four 1% resistors in series means that the worst-case errors can be four times as great. This is obviously true—if all the components are 1% low, or 1% high, the total error will be 4%. But the probability of this occurring is actually very, very small indeed. The more resistors you combine, the more the values cluster together in the centre of the range.

Perhaps you are not wholly satisfied that this apparently magical improvement in average accuracy is genuine. I could reproduce here the statistical mathematics, but it is not very exciting and can be easily found in the standard textbooks. I have found that showing the process actually at work on a spreadsheet makes a much more convincing demonstration.

In Excel, the usual random numbers have a uniform distribution and are generated by the function RAND(), but random numbers with a Gaussian distribution and specified mean and standard deviation can be generated by the function NORMINV(). Let us assume we want to make an accurate 20 kΩ resistance. We can simulate the use of a single 1% tolerance resistor by generating a column of Gaussian random numbers with a mean of 20 and a standard deviation of 0.2; we need to use a lot of numbers to smooth out the statistical fluctuations, so we will generate 400 of them. As a check we calculate the mean and standard deviation of our 400 random numbers using the AVERAGE() and STDEV() functions. The results will be very close to 20 and 0.2 but not identical, and will change every time we hit the F9 recalculate key as this generates a new set of random numbers. The results of five recalculations are shown in Table 12.3, demonstrating that 400 numbers are enough to get us quite close to our targets.

To simulate the series combination of two 10 kΩ resistors of 1% tolerance resistor we generate two columns of 400 Gaussian random numbers with a mean of 10 and a standard deviation of 0.1. We then set up a third column which is the sum of the two random numbers on the same row, and if we calculate the mean and standard deviation using AVERAGE() and STDEV() again, we find that the mean is still very close to 20 but the

Table 12.3: Mean and Standard Deviation of Five Batches of 400 Gaussian Random Resistor Values

Mean	Standard Deviation
20.0017	0.2125
19.9950	0.2083
19.9910	0.1971
19.9955	0.2084
20.0204	0.2040

standard deviation is reduced on average by the expected factor of √2. The result of five trials is shown in Table 12.4. Repeating this experiment with two 40 kΩ resistors in parallel gives the same results.

If we repeat this experiment by making our 20 kΩ resistance from a series combination of four 5 kΩ resistors of 1% tolerance we have to generate four columns of 400 Gaussian random numbers with a mean of 5 and a standard deviation of 0.05. We sum the four numbers on the same row to get a fifth column, and calculate the mean and standard deviation of that. The result of five trials is shown in Table 12.5. The mean is again very close to 20 but the standard deviation is now reduced on average by the a factor of √4, which is 2. Once again, repeating this for four parallel resistors gives the same improvement in accuracy.

I think this demonstrates quite convincingly that the spread of values is reduced by a factor equal to the square root of the number of the components used be they in series or parallel. The principle works equally well for capacitors or indeed any quantity with a Gaussian distribution of values. The downside is the fact that the improvement depends on the square of the number of equal-value components used, which means that big improvements require a lot of parts and the method quickly gets unwieldy. Table 12.6 demonstrates how this works; the rate of improvement slows down noticeably as the number of parts increases. Constructing a 0.1% resistance from 1% resistors would require a hundred of them, and would not normally be considered a practical proposition. The largest number of components I have ever used in this way for a production design is five. (Capacitors in a precision RIAA preamplifier)

Table 12.4: Mean and Standard Deviation of Five Batches of 400 Gaussian Resistors Made Up from Two in Series

Mean	Standard Deviation
19.9999	0.1434
20.0007	0.1297
19.9963	0.1350
20.0114	0.1439
20.0052	0.1332

Table 12.5: Mean and Standard Deviation of Five Batches of 400 Gaussian Resistors Made Up from Four in Series

Mean	Standard Deviation
20.0008	0.1005
19.9956	0.0995
19.9917	0.1015
20.0032	0.1037
20.0020	0.0930

Table 12.6: The Improvement in Value Tolerance with Number of Equal-Value Parts

Number of Equal-Value Parts	Tolerance Reduction Factor
1	1.000
2	0.707
3	0.577
4	0.500
5	0.447
6	0.408
7	0.378
8	0.354
9	0.333
10	0.316

The spreadsheet experiment is easy to set up, though it does involve a bit of copy-and-paste to set up 400 rows of calculation. It takes only about a second to recalculate on my PC, which is by no stretch of the imagination state-of-the-art.

You may be wondering what happens if the resistors used are *not* equal. If you are in search of a particular value the method that gives the best resolution is to use one large resistor value and one small one to make up the total, as this gives a very large number of possible combinations. However, the accuracy of the final value is essentially no better than that of the large resistor. Two equal resistors, as we have just demonstrated, give a $\sqrt{2}$ improvement in accuracy, and near-equal resistors give almost as much, but the number of combinations is very limited, and you may not be able to get very near the value you want. The question is, how much improvement in accuracy can we get with resistors that are some way from equal, such as one resistor being twice the size of the other?

The mathematical answer in the series case is very simple; even when the resistor values are not equal, the overall standard deviation is still the RMS-sum of the standard deviations of the two resistors, as shown in Equation 12.1 above; $\sigma 1$ and $\sigma 2$ are the standard deviations of the two resistors in series. Note that this equation is only correct if there is no correlation between the two values whose standard deviations we are adding; this is true for two separate resistors but would not hold for two film resistors fabricated on the same substrate.

Since both resistors have the same percentage tolerance, the larger of the two has the greater standard deviation, and dominates the total result. The minimum total deviation is thus achieved with equal resistor values. Table 12.7 shows how this works, and it is clear that using two resistors in the ratio 2:1 or 3:1 still gives a worthwhile improvement in average accuracy.

The entries for 19.5 K + 500 and 19.9 K + 100 demonstrate that when one large resistor value and one small are used to get a particular value, its accuracy is very little better than that of the large resistor alone.

Table 12.7: The Improvement in Value Tolerance with Unequal Resistors

Series Resistor Values Ω	Resistor Ratio	Standard Deviation
20 K single		0.2000
19.9 K + 100	199:1	0.1990
19.5 K + 500	39:1	0.1951
19 K + 1 K	19:1	0.1903
18 K + 2 K	9:1	0.1811
16.7 K + 3.3 K	5:1	0.1700
16 K + 4 K	4:1	0.1649
15 K + 5 K	3:1	0.1581
13.33 K + 6.67 K	2:1	0.1491
12 K + 8 K	1.5:1	0.1442
11 K + 9 K	1.22:1	0.1421
10 K + 10 K	1:1	0.1414

12.3 Resistor Value Distributions

At this point you may be complaining that this will only work if the resistor values have a Gaussian (also known as normal) distribution with the familiar peak around the mean (average) value. Actually, it is a happy fact this effect does *not* assume that the component values have a normal (Gaussian) distribution, as we shall see in a moment. An excellent account of how to handle statistical variations to enhance accuracy is in [1]. This deals with the addition of mechanical tolerances in optical instruments, but the principles are just the same.

You sometimes hear that this sort of thing is inherently flawed, because, for example, 1% resistors are selected from production runs of 5% resistors. If you were using the 5% resistors, then you would find there was a hole in the middle of the distribution; if you were trying to select 1% resistors from them, you would be in for a very frustrating time as they have already been selected out, and you wouldn't find a single one. If instead you were using the 1% components obtained by selection from the 5% population, then you would find that the distribution would be much flatter than Gaussian and the accuracy improvement obtained by combining them would be reduced, although there would still be a definite improvement.

However, don't worry. In general this is not the way that components are manufactured nowadays, though it may have been so in the past. A rare contemporary exception is the manufacture of carbon composition resistors [2] where making accurate values is difficult, and selection from production runs, typically with a 10% tolerance, is the only practical way to get more accurate values. Carbon composition resistors have no place in audio circuitry, because of their large temperature and voltage coefficients and high excess noise, but they live on in specialised applications such as switch-mode snubbing circuits, where their ability to absorb high peak power in bulk material rather than a thin film is useful, and in RF circuitry where the inductance of spiral-format film resistors is unacceptable.

So, having laid that fear to rest, what is the actual distribution of resistor values like? It is not easy to find out, as manufacturers are not very forthcoming with this sort of information, and measuring thousands of resistors with an accurate DVM is not a pastime that appeals to all of us. Any nugget of information in this area is therefore very welcome; Hugo Kroeze [3] reported the result of measuring 211 metal film resistors from the same batch with a nominal value of 10 kΩ and 1% tolerance. He concluded that:

1. The mean value was 9.995 kΩ
2. All the resistors were within the 1% tolerance range
3. The distribution appeared to be Gaussian, with no evidence that it was a subset from a larger distribution
4. The spread in value was surprisingly small, the standard deviation actually being about 10 Ω, ie only 0.1%

(Bear in mind here that the resistors were all from the same batch, and the spread in value across batches widely separated in manufacture date might have been less impressive.)

Now this is only one report, and it would be nice to have more confirmation, but there seems to be no reason to doubt that the distribution of resistance values is Gaussian, though the range of standard deviations you are likely to meet remains enigmatic. When ever I have attempted this kind of statistical improvement in accuracy, I have always found that the expected benefit really does appear in practice.

Up until now, we have just looked at the distribution of values around the mean, implicitly assuming that the mean is absolutely accurate. This is not as daft as it sounds because controlling the mean value emerging from a manufacturing process is usually relatively easy compared with controlling all the variables that lead to a spread in values. In the example given above, it appears that the mean is very well controlled indeed and the spread is very much under control as well.

12.4 Improving Accuracy with Multiple Components: Uniform Distribution

As I mentioned earlier, improving average accuracy by combining resistors does not depend on the resistance value having a Gaussian distribution. Even a batch of resistors with a uniform distribution gives better accuracy when two of them are combined. A uniform distribution of component values may not be likely but the result of combining two or more of them is highly instructive, so stick with me for a bit.

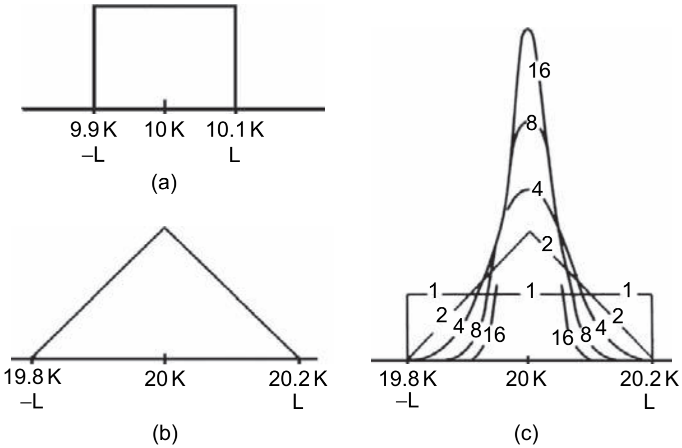


Figure 12.2: How a uniform distribution of values becomes a Gaussian (normal) distribution when more component values are summed (after Smith).

Figure 12.2a shows a uniform distribution that cuts off abruptly at the limits L and –L, and represents 10 kΩ resistors of 1% tolerance. We will assume again that we want to make a more accurate 20 kΩ resistance. If we put two of the uniform-distribution 10 kΩ resistors in series, we get not another uniform distribution, but the triangular distribution shown in Figure 12.2b. This shows that the total resistance values are already starting to cluster in the centre; it is possible to have the extreme values of 19.8 kΩ and 20.2 kΩ, but it is very unlikely.

Figure 12.2c shows what happens if we use more resistors to make the final value; when four are used the distribution is already beginning to look like a Gaussian distribution, and as we increase the number of components to 8 and the sixteen, the resemblance becomes very close.

Uniform distributions have a standard deviation just as Gaussian ones do. It is calculated from the limits L and –L as in Equation 12.2. Likewise, the standard deviation of a triangular distribution can be calculated from its limits L and –L as in Equation 12.3

$$\text{Standard deviation of uniform distribution } \sigma = \frac{1}{\sqrt{3}}L \qquad (12.2)$$

$$\text{Standard deviation of triangular distribution } \sigma = \frac{1}{\sqrt{6}}L \qquad (12.3)$$

Applying Equation 12.2 to the uniformly-distributed 10 kΩ 1% resistors in Figure 12.2a, we get a standard deviation of 0.0577. Applying Equation 12.3 to the triangular distribution of 20 kΩ resistance values in Figure 12.2b, we get 0.0816. The mean value has doubled, but the standard deviation has less than doubled, so we get an improvement in average accuracy; the ratio is √2, just as it was for two resistors with a Gaussian distribution. This is also easy to demonstrate with the spreadsheet method described earlier.

12.5 Obtaining Arbitrary Resistance Values

If you are making up a particular resistance value, the method that gives the best resolution is to use one large resistor value and one small one to make up the total, so for example 3100 Ω would be made up of 3000 Ω and 100 Ω; this is what you might call the asymmetrical solution. One the other hand, if it is possible to make up the required value with two resistors of roughly the same value, there is an advantage in terms of increased precision, for the statistical reasons described above. In the optimal case, where the two resistors are equal, the average accuracy is improved by √2, but resistors in the ratio 2:1 or 3:1 still give a useful improvement.

The best procedure is therefore to determine at the beginning how accurate the resistor value needs to be, start with two near-equal resistor values, and if no combination can be found within our error window, we try values that are more and more unequal until we get a satisfactory result. As an example, we will assume that we are using E24 values, and our calculations show that 2090 Ω is the exact value required. The closest we can get with near-equal resistors in series is 1000 Ω + 1100 Ω = 2100 Ω, an error of 0.48%, which you may well be able to live with. If not, we try again using near-equal resistors in parallel, and we come up with 4300 Ω in parallel with 3900 Ω = 2045 Ω; this is in error by –2.2%, which is rather less appealing. Clearly the series option comes up with the closer answer in this case, but if 2100 Ω is not close enough, one option is to abandon completely the "near-equal" constraint and go for one large value and one small value. The obvious answer is 2000 Ω + 91 Ω = 2091 Ω, which is an error of only +0.048%, small enough to be utterly lost in other component tolerances. However, there is no improvement in precision. As we saw in Table 12.7, intermediate conditions between "near-equal" and "one large and one small" will give intermediate improvements in precision.

The equivalent asymmetrical parallel-combination best approach is 2200 Ω in parallel with 43 k = 2092.9 Ω, which has an error of +0.14%, and so the series solution is in this case the better one.

A step-by step example of choosing parallel resistor combinations for the best possible increased precision while achieving a result within a specified percentage error window is given in Chapter 19, where a complete active crossover is designed.

Trying out the various resistor combinations on a hand calculator rapidly becomes life-threatingly tedious, even on my much-prized Casio FX-19. A spreadsheet is much quicker once you've set it up, but still involves a lot of work and the best answer is a clearly software that automatically comes up with the best answers. One example is a resistor combination calculator utility called "ResCalc" by Mark Lovell and Morgan Jones [4], which is superior to most as it allows mixed E24 and E96 series, but like all the other resistor-calculator applications I have seen, it does not appear to prioritise near-equal values to minimise tolerance errors. I have therefore written a Javascript program that explores the resistor combinations for a specified error window around the nominal value, keeping the resistors as nearly equal as possible. By the time this book appears I hope to have a fully debugged version on my website [5].

A closer approach to the desired value while keeping the values near-equal is possible by using combinations of three resistors rather than two. The cost is still low because resistors are cheap, but seeking out the best combination is an even more unwieldy process. I plan to write another Javascript app to deal with this problem.

All of the statistical features described here apply to capacitors as well, but are harder to apply because capacitors come in sparse value series, and it is also much more expensive to use multiple parts to obtain an arbitrary value. It is worth going to a good deal of trouble to come up with a circuit design that uses only standard capacitor values; the resistors will then almost certainly be all non-standard values, but this is easier and cheaper to deal with. A good example of the use of multiple parallel capacitors, both to improve accuracy and to make up values larger than those available, is the Signal Transfer RIAA preamplifier [6, 7]; see also the end of this book. This design is noted for giving very accurate RIAA equalisation at a reasonable cost. There are two capacitors required in the RIAA network, one being made up of four polystyrene capacitors in parallel and the other of five polystyrene capacitors in parallel. This also allows non-standard capacitance values to be used.

12.6 Resistor Noise: Johnson and Excess Noise

All resistors, no matter what their method of construction, generate Johnson noise. This is white noise, which has equal power in equal absolute bandwidth, ie with the bandwidth measured in Hz, not octaves. There is the same noise power between 100 and 200 Hz as there is between 1100 and 1200 Hz. The level of Johnson noise that a resistor generates is determined solely by its resistance value, the absolute temperature, (in degrees Kelvin) and the bandwidth over which the noise is being measured. For our purposes the temperature is 25°C and the bandwidth is 22 kHz, so the resistance is really the only variable. The level of Johnson noise is based on fundamental physics and is not subject to modification, negotiation or any sort of rule-bending. Sometimes it places the limit on how quiet a circuit can be, though often the noise from the active devices is dominant. It is a constant refrain in this book that resistor values should be kept as low as possible, without introducing distortion by overloading the circuitry, in order to minimise the Johnson noise contribution.

The rms amplitude of Johnson noise is calculated from the classic equation:

$$v_n = \sqrt{4\,kTRB} \tag{12.4}$$

Where:

v_n is the rms noise voltage
T is absolute temperature in °K
B is the bandwidth in Hz
k is Boltzmann's constant
R is the resistance in Ohms

The thing to be careful with here is to use Boltzmann's constant (1.380662×10^{-23}), and NOT the Stefan-Boltzmann constant (5.67×10^{-08}) which relates to black-body radiation, has nothing to do with resistors, and will give some impressively wrong answers. The voltage noise is often left in its squared form for ease of RMS-summing with other noise sources.

The noise voltage is inseparable from the resistance, so the equivalent circuit is of a voltage source in series with the resistance present. Johnson noise is usually represented as a voltage, but it can also be treated as a Johnson noise current, by means of the Thevenin-Norton transformation, which gives the alternative equivalent circuit of a current-source in shunt with the resistance. The equation for the noise current is simply the Johnson voltage divided by the value of the resistor it comes from:

$$i_n = v_n/R.$$

Excess resistor noise refers to the fact that some resistors, with a constant voltage drop across them, generate extra noise in addition to their inherent Johnson noise. This is a very variable quantity, but is essentially proportional to the DC voltage across the component; the specification is therefore in the form of a "Noise Index" such as "1 uV/V." The uV/V parameter increases with increasing resistor value and decreases with increasing resistor size or power dissipation capacity. Excess noise has a 1/f frequency distribution. It is usually only of interest if you are using carbon or thick film resistors—metal film and wirewound types should have little or no excess noise. A rough guide to the likely range for excess noise specs is given in Table 12.8.

Table 12.8: Resistor Excess Noise

Type	Noise Index uV/V
Metal film TH	0
Carbon film TH	0.2–3
Metal oxide TH	0.1–1
Thin film SM	0.05–0.4
Bulk metal foil TH	0.01
Wirewound TH	0

(Wirewound resistors are normally considered to be completely free of excess noise.)

One of the great benefits of opamp circuitry is that it is noticeably free of resistors with large DC voltages across them; the offset voltages and bias currents involved are much too low to cause measurable excess noise. If you are designing an active crossover on this basis then you can probably forget about the issue. If, however, you are using discrete transistor circuitry, it might possibly arise; specifying metal film resistors throughout, as you no doubt would anyway, will ensure you have no problems with excess noise.

To get a feel for the magnitude of excess resistor noise, consider a 100 kΩ 1/4 W carbon film resistor with a steady 10 V across it. The manufacturer's data gives a noise parameter of about 0.7 uV/V and so the excess noise will be of the order of 7 uV, which is −101 dBu. That could definitely be a problem in a low-noise preamplifier stage.

12.7 Resistor Non-Linearity

Ohm's Law is, strictly speaking, a statement about metallic conductors only. It is dangerous to assume that it invariably applies exactly to resistors simply because they have a fixed value of resistance marked on them; in fact resistors—whose main *raison d'etre* is packing a lot of controlled resistance in a small space—sometimes show significant deviation from Ohm's Law in that current is not exactly proportional to voltage. This is obviously unhelpful when you are trying to make low-distortion circuitry. Resistor non-linearity is normally quoted by manufacturers as a voltage coefficient, usually the number of parts per million (ppm) that the resistance changes when one volt is applied. The measurement standard for resistor non-linearity is IEC 6040.

The common through-hole metal film resistors show effectively perfect linearity, as do wirewound types, both having voltage coefficients of less than 1 ppm. Carbon film resistors, now almost totally obsolete, tend to be quoted at around 100 ppm; in many circumstances this is enough to generate more distortion than that produced by the active devices. Carbon composition resistors are of historical interest only so far as audio is concerned, and have rather variable voltage coefficients in the area of 350 ppm, something that might be pondered by connoisseurs of antique amplifying equipment. Today the real concern over resistor non-linearity is about thick-film surface-mount resistors, which have high and rather variable voltage coefficients; more on this below.

Table 12.9 (calculated with SPICE) gives the THD in the current flowing through the resistor for various voltage coefficients when a pure sine voltage is applied. If the voltage coefficient is significant this can be a serious source of non-linearity.

Distortion here is assumed to be second-order, and so varies proportionally with level. Third-order distortion, which will be dominant if a resistor has no steady voltage across it, rises as the square of level.

Table 12.9: Resistor Voltage Coefficients and the Resulting Distortion at +15 and +20 dBu

Voltage Coefficient	THD at +15 dBu	THD at +20 dBu
1 ppm	0.00011%	0.00019%
3 ppm	0.00032%	0.00056%
10 ppm	0.0016%	0.0019%
30 ppm	0.0032%	0.0056%
100 ppm	0.011%	0.019%
320 ppm	0.034%	0.060%
1000 ppm	0.11%	0.19%
3000 ppm	0.32%	0.58%

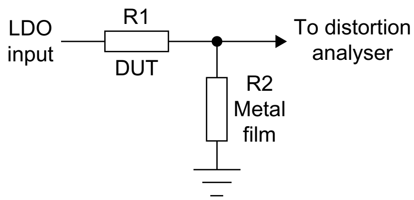


Figure 12.3: Test circuit for measuring resistor non-linearity. The not-under-test resistor R2 in the potential divider must be a metal-film type with negligible voltage coefficient.

My own test setup is shown in Figure 12.3. The resistors are usually of equal value, to give 6 dB attenuation. A very low-distortion oscillator that can give a large output voltage is necessary; the results in Figure 12.4 were taken at a 10 Vrms (+22 dBu) input level. Here thick-film surface-mount (SM) and through-hole (TH) resistors are compared. The gen-mon trace at the bottom is the record of the analyser reading the oscillator output and is the measurement floor of the AP System 1 used. The THD plot for the through-hole case is higher than this floor, but this is not due to distortion. It simply reflects the extra Johnson noise generated by two 10 kΩ resistors. Their parallel combination is 5 kΩ, and so this noise is at −115.2 dBu. The SM plot, however, is higher again, and the difference is the distortion generated by the thick-film component.

For both thin-film and thick-film SM resistors non-linearity increases with resistor value, and also increases as the physical size (and hence power rating) of the resistor shrinks. The thin-film versions are much more linear; see Figures 12.5 and 12.6.

Sometimes it is appropriate to reduce the non-linearity by using multiple resistors in series. If one resistor is replaced by two with the same voltage coefficient in series, the THD in the

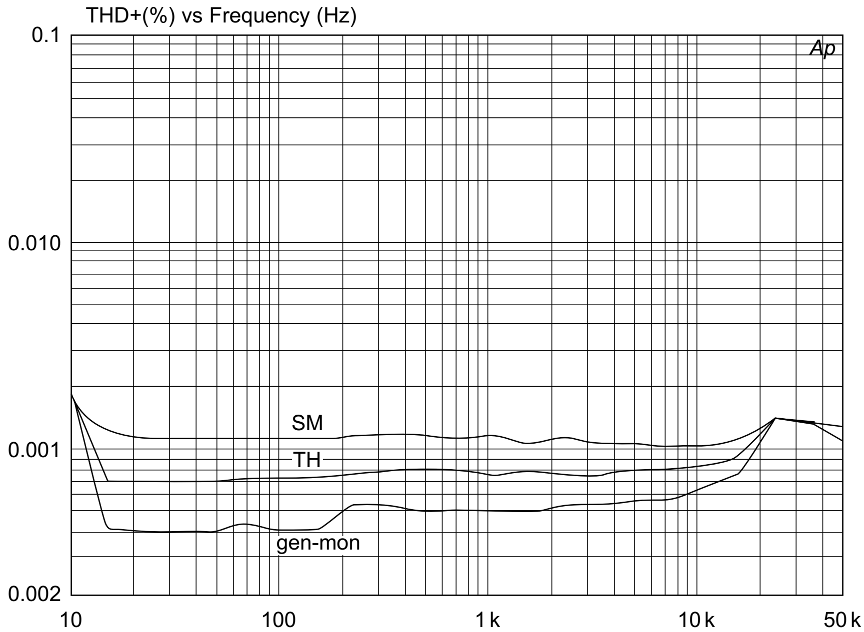


Figure 12.4: SM resistor distortion at 10 Vrms input, using 10 kΩ 0805 thick-film resistors.

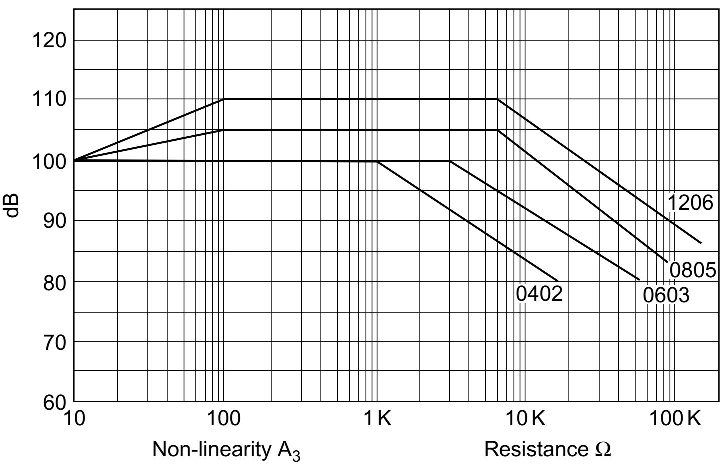


Figure 12.5: Non-linearity of thin-film surface-mount resistors of different sizes. THD is here in dB rather than percent.

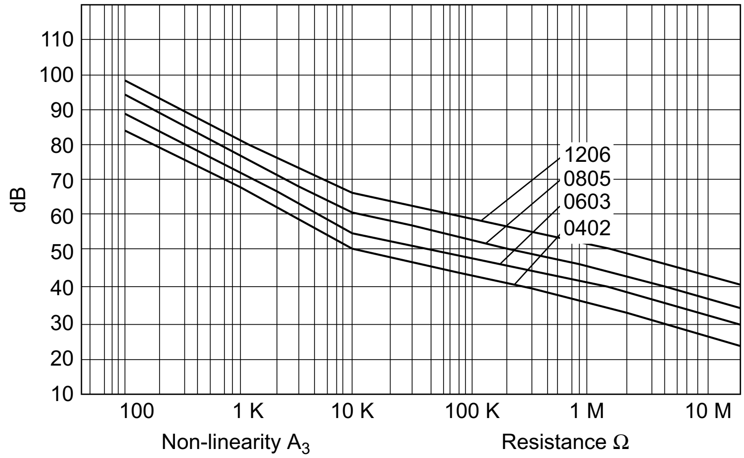


Figure 12.6: Non-linearity of thick-film surface-mount resistors of different sizes.

current flowing is halved. Similarly, three resistors reduces THD to a third of the original value. There are obvious economic limits to this sort of thing, and it takes up PCB area, but it can be useful in specific cases, especially where the voltage rating of the resistor is a limitation.

12.8 Capacitors: Values and Tolerances

The need for specific capacitor ratios creates problems as capacitors are available in a much more limited range of values than resistors, usually the E6 series, running 10, 15, 22, 33, 47, 68. If other values are needed then they often have to be made up of two capacitors in parallel; this puts up the cost and uses significantly more PCB area so it should be avoided if possible, but for applications like Sallen & Key lowpass filters where a capacitance ratio of two is required, it is often the only way. Sallen & Key filters with equal component values can be used (see Chapter 8), but they must be configured to give voltage gain, which is often unwanted and will force compromises over either noise performance or headroom. Selecting convenient capacitor values will almost invariably lead to a need for non-standard resistance values, but this is much less of a problem as combining two resistors to get the right value is much cheaper and uses less PCB area.

As for the case of resistors, when contriving a particular capacitor value, the best resolution is obtained by using one large value and one small one to make up the total, but two capacitors of approximately the same value give better average accuracy. When the two capacitors used are equal in nominal value, the accuracy is improved by $\sqrt{2}$. This effect is more important for capacitors because the cost premium for 1% parts is considerable, whereas for resistors it is very small, if it exists at all. When using multiple components like this to make up a value or improve precision, it is important to keep an eye on both the extra cost and the extra PCB area occupied.

The tolerance of non-electrolytic capacitors is usually in the range ±1% to ±10%; anything more accurate than this tends to be very expensive. Electrolytic capacitors used to have *much* wider tolerances, but things have recently improved and ±20% is now common. This is still wider than any other component you are likely to use in a crossover, but this is not a problem, for as described below, it is most unwise to try to define frequencies or time-constants with electrolytic capacitors.

12.9 Capacitor Shortcomings

Capacitors fall short of being an ideal circuit element in several ways, notably leakage, Equivalent Series Resistance (ESR), Equivalent Series Inductance (ESL), dielectric absorption, and non-linearity.

Capacitor leakage is equivalent to a high value resistance across the capacitor terminals, which allows a trickle of current to flow when a DC voltage is applied. Leakage is usually negligible for non-electrolytics, but is much greater for electrolytics. It is not normally a problem in audio design.

Equivalent Series Resistance (ESR) is a measure of how much the component deviates from a mathematically pure capacitance. The series resistance is partly due to the physical resistance of leads and foils, and partly due to losses in the dielectric. It can also be expressed as tan-δ, (tan-delta) which is the tangent of the phase angle between the voltage across and the current flowing through the capacitor. Once again it is rarely a problem in the audio field, the values being small fractions of an Ohm and very low compared with normal circuit resistances.

Equivalent Series Inductance (ESL) is always present. Even a straight piece of wire has inductance, and any capacitor has lead-out wires and internal connections. The values are normally measured in nano-Henries and have no effect in normal audio circuitry.

Dielectric absorption is a well known effect; take a large electrolytic, charge it up, and then fully discharge it. Over a few minutes the charge will partially reappear. This "memory effect" also occurs in non-electrolytics to a lesser degree; it is a property of the dielectric, and is minimised by using polystyrene, polypropylene, NPO ceramic, or PTFE dielectrics. Dielectric absorption is invariably modelled by adding extra resistors and capacitances to an ideal main capacitor, as shown in Figure 12.7 for a 1 uF polystyrene capacitor. Note there is no hint of any source of non-linearity [8]. However, the dielectric absorption mechanism does seem to have some connection with capacitor distortion, since the dielectrics that show the least dielectric absorption also show the lowest non-linearity.

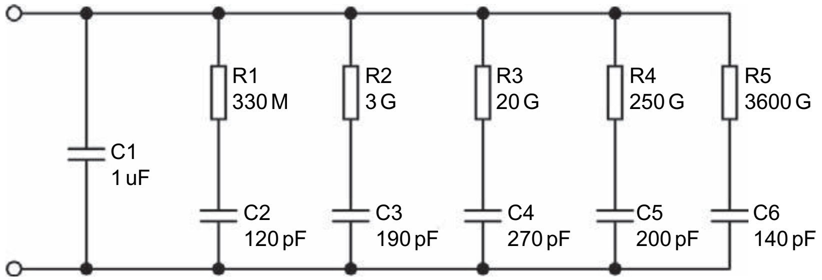


Figure 12.7: A model of dielectric absorption in a 1 uF polystyrene capacitor. All components are linear.

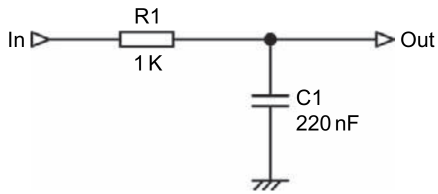


Figure 12.8: Simple lowpass test circuit for non-electrolytic capacitor distortion.

Dielectric absorption is a major consideration in sample-and-hold circuits and the like, but of no account in itself for normal linear audio circuitry; you can see from Figure 12.7 that the additional components are very small capacitors with very large resistances in series, and the effect these could have on the response of a filter is microscopic, and far smaller than the effects of component tolerances. Be aware that the model is just an approximation and is not meant to imply that each extra component directly represents some part of a physical process.

Capacitor non-linearity is the least known but by far the most troublesome of capacitor shortcomings. A typical RC lowpass filter can be made with a series resistor and a shunt capacitor, as in Figure 12.8, and if you examine the output with a distortion analyser, you will find to your consternation that the circuit is not linear. If the capacitor is a non-electrolytic type with a dielectric such as polyester, then the distortion is relatively pure third harmonic, showing that the effect is symmetrical. For a 10 Vrms input, the THD level may be 0.001% or more. This may not sound like much but it is substantially greater than the mid-band distortion of a good opamp. The definitive work on capacitor distortion is a magnificent series of articles by Cyril Bateman in Electronics World [9]. The authority of this is underpinned by Cyril's background in capacitor manufacturing.

Capacitors are used in audio circuitry for four main functions, where their possible non-linearity has varying consequences:

1. Coupling or DC blocking capacitors. These are usually electrolytics, and if properly sized have a negligible signal voltage across them at the lowest frequencies of interest. The non-linear properties of the capacitor are then unimportant unless current levels are high; power amplifier output capacitors can generate considerable mid-band distortion [10]. This makes you wonder what sort of non-linearity is happening in those big non-polarised electrolytics in passive crossovers.
   A great deal of futile nonsense has been talked about the mysterious properties of coupling capacitors, but it is all total twaddle. How could a component with negligible voltage across it put its imprint on a signal passing through it? For small-signal use, as long as the signal voltage across the coupling capacitor is kept low, non-linearity is not detectable by the best THD methods. The capacitance value is non-critical, as it has to be, given the wide tolerances of electrolytics.
2. Supply filtering or decoupling capacitors. Electrolytics are used for filtering out supply rail ripple, etc., and non-electrolytics, usually around 100 nF, are used to keep the supply impedance low at high frequencies and thus keep opamps stable. The capacitance value is again non-critical.
3. For active crossover purposes, by far the most important aspect of capacitors is their role in active filtering. This is a much more demanding application than coupling or decoupling, for firstly, the capacitor value is now crucially important as it defines the accuracy of the frequency response. Secondly, there is by definition a significant signal voltage across the capacitor and so its non-linearity can be a serious problem. Non-electrolytics are always used in active filters, though sometimes a time-constant involving an electrolytic is ill-advisedly used to define the lower end of the system bandwidth; this is a very bad practice because it is certain to introduce significant distortion at the bottom of the frequency range.
4. Small value ceramic capacitors are used for opamp compensation purposes, and sometimes in active filters in the HF path of a crossover. So long as they are NPO (C0G) ceramic types, their non-linearity should be negligible. Other kinds of ceramic capacitor, using the XR7 dielectric, will introduce copious distortion and must never be used in audio paths. They are intended for high-frequency decoupling where their linearity or otherwise is irrelevant.

12.10 Non-Electrolytic Capacitor Non-Linearity

It has often been assumed that non-electrolytic capacitors, which generally approach an ideal component more closely than electrolytics, and have dielectrics constructed in a totally different way, are free from distortion. It is not so. Some non-electrolytics show distortion at levels that is easily measured, and can exceed the distortion from the opamps in the circuit.

Non-electrolytic capacitor distortion is essentially third harmonic, because the non-polarised dielectric technology is basically symmetrical. The problem is serious, because non-electrolytic capacitors are commonly used to define time-constants and frequency responses (in RIAA equalisation networks, for example) rather than simply for DC-blocking.

Very small capacitances present no great problem. Simply make sure you are using the COG (NP0) type, and so long as you choose a reputable supplier, there will be no distortion. I say "reputable supplier" because I did once encounter some allegedly COG capacitors from China that showed significant non-linearity [11].

Middle-range capacitors, from 1 nF to 1 uF, present more of a problem. Capacitors with a variety of dielectrics are available, including polyester, polystyrene, polypropylene, polycarbonate and polyphenylene sulphide, of which the first three are the most common. (Note that what is commonly called "polyester" is actually polyethylene terephthalate, PET.)

Figure 12.8 shows a simple low-pass filter circuit which, with a good THD analyser, can be used to measure capacitor distortion. The values shown give a pole frequency, or –3 dB roll-off point, at 710 Hz. We will start off with polyester, the smallest, most economical, and therefore the most common type for capacitors of this size.

The THD results for a microbox 220 nF 100 V capacitor with a polyester dielectric are shown in Figure 12.9, for input voltages of 10, 15 and 20 Vrms. They are unsettling.

The distortion is all third harmonic. It peaks at around 300 to 400 Hz, well below the –3 dB frequency, and even with the input limited to 10 Vrms will exceed the non-linearity introduced by opamps such as the 5532 and the LM4562. Interestingly, the peak frequency

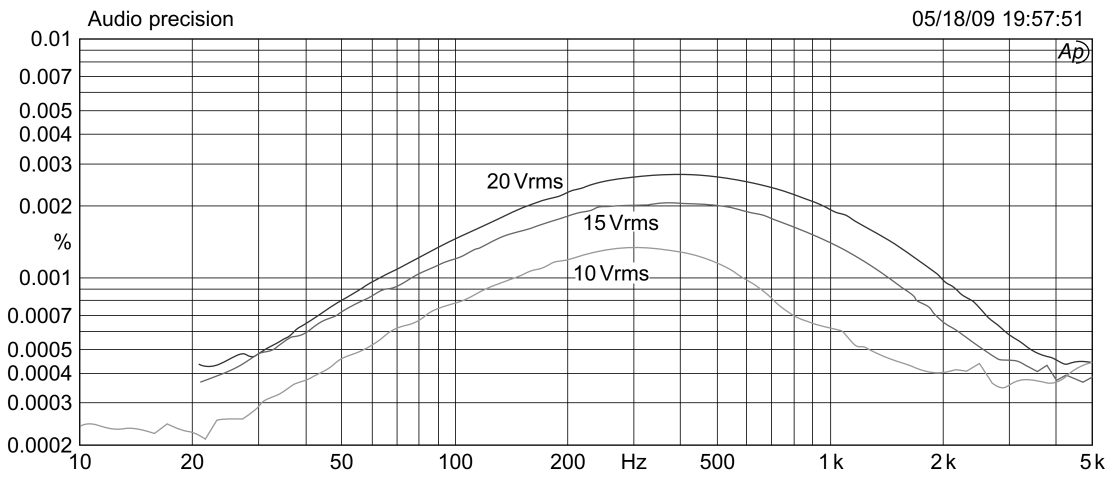


Figure 12.9: Third-harmonic distortion from a 220 nF 100 V polyester capacitor, at 10, 15, and 20 Vrms input level, showing peaking around 400 Hz.

changes with applied level. Below the peak, the voltage across the capacitor is constant but distortion falls as frequency is reduced, because the increasing impedance of the capacitor means it has less effect on a circuit node at a 1 kΩ impedance. Above the peak, distortion falls with increasing frequency because the lowpass circuit action causes the voltage across the capacitor to fall.

The level of distortion varies with different samples of the same type of capacitor; six of the above type were measured and the THD at 10 Vrms and 400 Hz varied from 0.00128% to 0.00206%. This puts paid to any plans for reducing the distortion by some sort of cancellation method.

The distortion can be seen in Figure 12.9 to be a strong function of level, roughly tripling as the input level doubles. Third harmonic distortion normally quadruples for doubled level, so there may well be an unanswered question here. It is however clear that reducing the voltage across the capacitor reduces the distortion. This suggests that if cost is not the primary consideration, it might be useful to put two capacitors in series to halve the voltage, and the capacitance, and then double up this series combination to restore the original capacitance, giving the series-parallel arrangement in Figure 12.10. The results are shown in Table 12.10, and once more it can be seen that halving the level has reduced distortion by a factor of three rather than four.

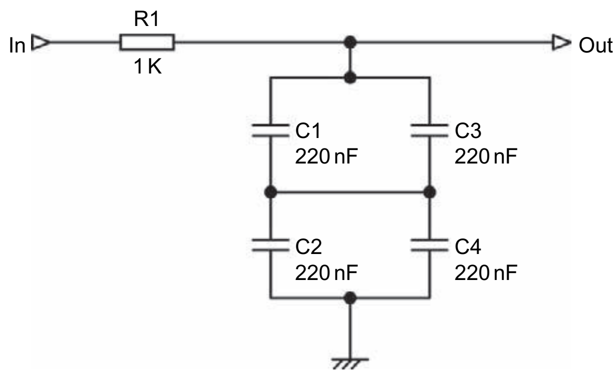


Figure 12.10: Reducing capacitor distortion by series-parallel connection.

Table 12.10: The Reduction of Polyester Capacitor Distortion by Series-Parallel Connection

Input Level Vrms	Single Capacitor	Series-Parallel Capacitors
10	0.0016%	0.00048%
15	0.0023%	0.00098%
20	0.0034%	0.0013%

The series-parallel arrangement has obvious limitations in terms of cost and PCB area occupied, but might be useful in some cases. It has the advantage that, as described earlier in the chapter, using multiple components improves the average accuracy of the total value.

Clearly polyester capacitors can generate significant distortion, despite their extensive use in audio circuitry of all kinds. The next dielectric we will try is polystyrene. Capacitors with a polystyrene dielectric are extremely useful for some filtering and RIAA-equalisation applications because they can be obtained at a 1% tolerance at up to 10 nF at a reasonable price. They can be obtained in larger sizes but at much higher prices.

The distortion test results are shown in Figure 12.11 for three samples of a 4n7 2.5% capacitor; the series resistor R1 has been increased to 4.7 kΩ to keep the −3 dB point inside the audio band, and it is now at 7200 Hz. Note that the THD scale has been extended down to a subterranean 0.0001%, because if it was plotted on the same scale as Figure 12.9 it would be bumping along the bottom of the graph. Figure 12.11 in fact shows no distortion at all, just the measurement noise floor, and the apparent rise at the HF end is simply due to the fact that the output level is decreasing, because of the lowpass action, and so the noise floor is relatively increasing. This is at an input level of 10 Vrms, which is about as high as might be expected to occur in normal opamp circuitry. The test was repeated at 20 Vrms, which might be encountered in discrete circuitry, and the results were the same, yielding no measurable distortion.

The tests were done with four samples of 10 nF 1% polystyrene from LCR at 10 Vrms and 20 Vrms, with the same results for each sample. This shows that polystyrene capacitors can be used with confidence; this finding is in complete agreement with Cyril Bateman's findings [12].

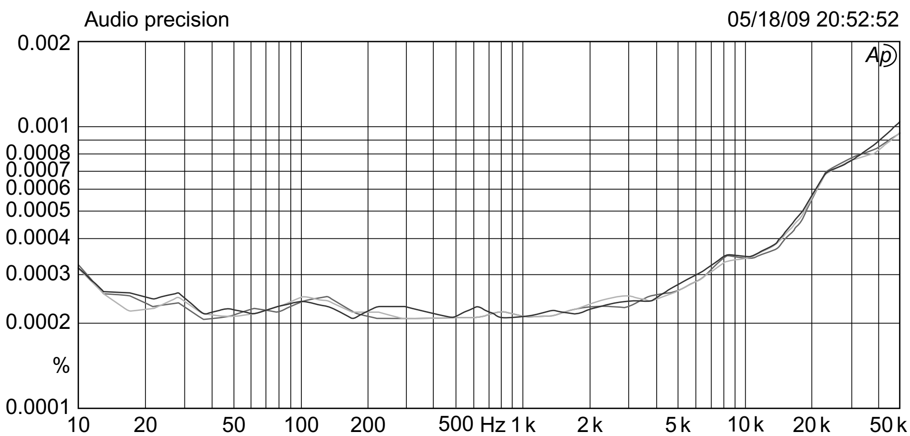


Figure 12.11: The THD plot with three samples of 4n7 2.5% polystyrene capacitors, at 10 Vrms input level. The reading is entirely noise.

Having resolved the problem of capacitor distortion below 10 nF, we need now to tackle it for larger capacitor values. Polyester having proven unsatisfactory, the next most common capacitor type is polypropylene, and I am glad to report that these are effectively distortion-free. Figure 12.12 shows the results for four samples of a 220 nF 250 V 5% polypropylene capacitor from RIFA. The plot shows no distortion at all, just the noise floor, with the apparent rise at the HF end being increasing relative noise due to the lowpass rolloff, as in Figure 12.11. This is also in agreement with Cyril Bateman's findings [13]. Rerunning the tests at 20 Vrms gave the same result—no distortion. This is very pleasing, but there is a downside. Polypropylene capacitors of this value and voltage rating are much larger than the commonly used 63 or 100 V polyester capacitor, and more expensive.

It was therefore important to find out if the good distortion performance was a result of the 250 V rating, and so I tested a series of polypropylene capacitors with lower voltage ratings from different manufacturers. Axial 47 nF 160 V 5% polypropylene capacitors from Vishay proved to be THD-free at both 10 Vrms and 20 Vrms. Likewise, microbox polypropylene capacitors from 10 nF to 47 nF, with ratings of 63 V and 160 V from Vishay and Wima proved to generate no measurable distortion, so the voltage rating appears not to be an issue. This finding is particularly important because the Vishay range has a 1% tolerance, making them very suitable for precision filters and equalisation networks. The 1% tolerance is naturally reflected in the price.

The higher values of polypropylene capacitors (above 100 nF) appear to be currently only available with 250 V or 400 V ratings, and that means a physically big component. For example, the EPCOS 330 nF 400 V 5% part has a footprint of 26 mm by 6.5 mm, with a height of 15 mm, and capacitors like that take up a lot of PCB area. One way of dealing with this is to use a smaller capacitor in a capacitance multiplication configuration, so, for example, a 100 nF 1% component could be made to emulate 330 nF. It has to be said that

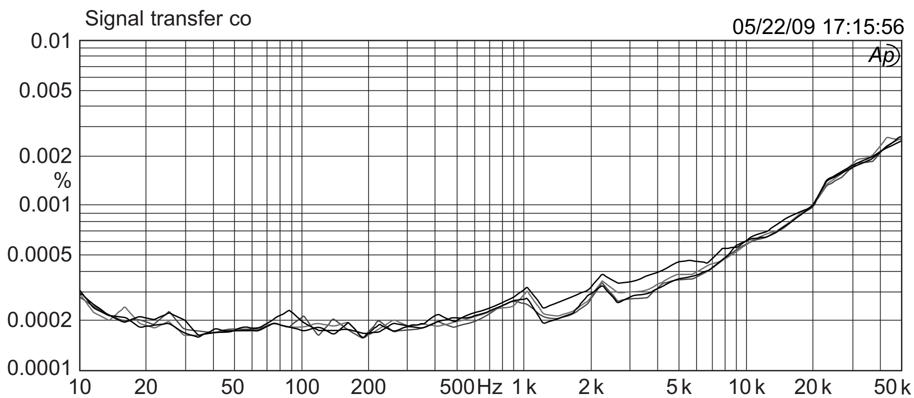


Figure 12.12: The THD plot with four samples of 220 nF 250 V 5% polypropylene capacitors, at 10 Vrms input level. The reading is again entirely noise.

this is only straightforward if one end of the capacitor is connected to ground; see Chapter 11 for an example of this technique applied to a biquad equaliser.

When I first started looking at capacitor distortion, I thought that the distortion would probably be lowest for the capacitors with the highest voltage rating. I therefore tested some RF-suppression X2 capacitors, rated at 275 Vrms, equivalent to a peak or DC rating of 389 V. The dielectric material is unknown. An immediate snag is that the tolerance is 10 or 20%, not exactly ideal for precision filtering or equalisation. A more serious problem, however, is that they are far from distortion-free. Four samples of a 470 nF X2 capacitor showed THD between 0.002% and 0.003% at 10 Vrms. A high voltage rating alone does not mean low distortion.

12.11 Electrolytic Capacitor Non-Linearity

Cyril Bateman's series in *Electronics World* [10] included two articles on electrolytic capacitor distortion. It proved to be a complex subject, and many long-held assumptions, such as "DC biasing always reduces distortion" were shown to be quite wrong. (My own results confirm this—DC biasing is at best pointless, and can increase distortion.) The distortion levels Cyril measured were in general a good deal higher than for non-electrolytic capacitors, and I can confirm that too.

My view is that electrolytics should never, ever, under any circumstances, be used to set time-constants in audio. There should be a time-constant early in the signal path, based on a non-electrolytic capacitor, that determines the lower limit of the system bandwidth; preferably proper bandwidth definition should be implemented with a subsonic filter. All the electrolytic-based time-constants should be much longer so that the electrolytic capacitors can never have significant signal voltages across them and so never generate detectable distortion. Electrolytics have large tolerances, and cannot be used to set accurate time-constants anyway.

However, even if you think you are following this plan, you can still get into trouble. Figure 12.13 shows a simple highpass test circuit representing an electrolytic capacitor in

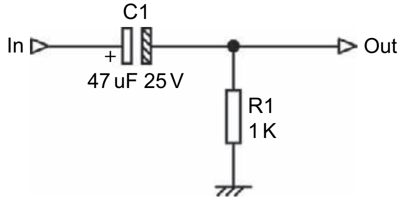


Figure 12.13: Highpass test circuit for examining electrolytic capacitor distortion.

use for coupling or DC-blocking. The load of 1 KΩ is the sort of value that can easily be encountered if you are using low-impedance design principles. The calculated –3 dB rolloff point is 3.38 Hz, so the attenuation at 10 Hz, at the very bottom of the audio band, will be only 0.47 dB; at 20 Hz it will be only 0.12 dB, which is surely a negligible loss. As far as frequency response goes, we are doing fine. But… examine Figure 12.14, which shows the measured distortion of this arrangement. Even if we limit ourselves to a 10 Vrms level, the distortion at 50 Hz is 0.001%, already above that of a good opamp. At 20 Hz it has risen to 0.01%, and at 10 Hz is a most unwelcome 0.05%. The THD is increasing by a ratio of 4.8 times for each octave fall in frequency, in other words increasing faster than a square law. The distortion residual is visually a mixture of second and third harmonic, and the levels proved surprisingly consistent for a large number of 47 uF 25 V capacitors of different ages and from different manufacturers.

Figure 12.14 also shows that the distortion rises rapidly with level; at 50 Hz going from an input of 10 Vrms to 15 Vrms almost doubles the THD reading. To underline the point, consider Figure 12.15, which shows the measured frequency response of the circuit with 47 uF and 1 KΩ; note the effect of the capacitor tolerance on the real versus calculated response. The rolloff that does the damage, by allowing an AC voltage to exist across the capacitor, is very modest indeed, less than 0.2 dB at 20 Hz.

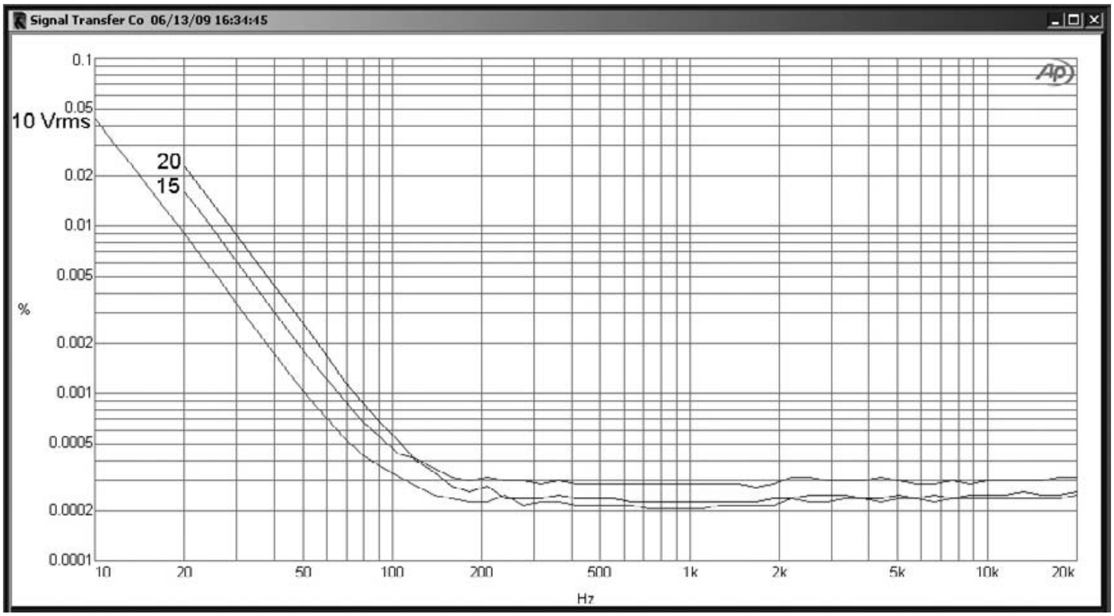


Figure 12.14: Electrolytic capacitor distortion from the circuit in Figure 12.13. Input level 10, 15, and 20 Vrms.

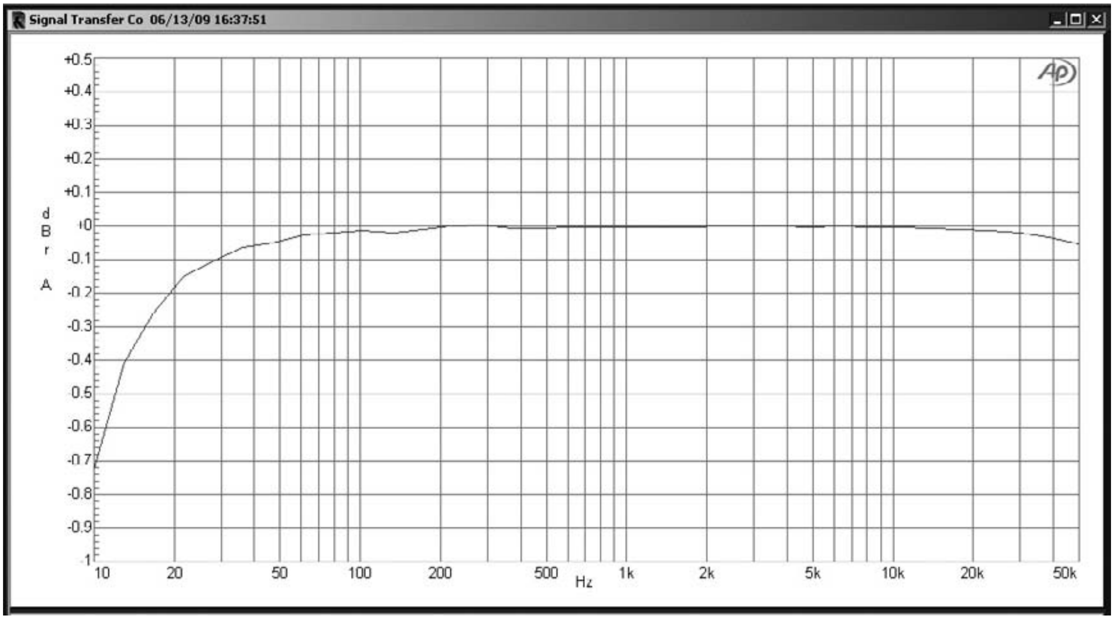


Figure 12.15: The measured rolloff of the highpass test circuit for examining electrolytic capacitor distortion.

Having demonstrated how insidious this problem is, how do we fix it? As we have seen, changing capacitor manufacturer is no help. Using 47 uF capacitors of higher voltage does not work—tests showed there is very little difference in the amount of distortion generated. An exception was the sub-miniature style of electrolytic, which was markedly worse.

The answer is simple—just make the capacitor bigger in value. This reduces the voltage across it in the audio band, and since we have shown that the distortion is a strong function of the voltage across the capacitor, the amount produced drops more than proportionally. The result is seen in Figure 12.16, for increasing capacitor values with a 10 Vrms input.

Replacing C1 with a 100 uF 25 V capacitor drops the distortion at 20 Hz from 0.0080% to 0.0017%, an improvement of 4.7 times; the voltage across the capacitor at 20 Hz has been reduced from 1.66 Vrms to 790 mV rms. A 220 uF 25 V capacitor reduces the voltage across itself to 360 mV, and gives another very welcome reduction to 0.0005% at 20 Hz, but it is necessary to go to 1000 uF 25 V to obtain the bottom trace, which is the only one indistinguishable from the noise floor of the AP-2702 test system. The voltage across the capacitor at 20 Hz is now only 80 mV. From this data, it appears that the AC voltage across an electrolytic capacitor should be limited to below 80 mV rms if you want to avoid distortion. I would emphasise that these are ordinary 85°C rated electrolytic capacitors, and in no sense special or premium types.

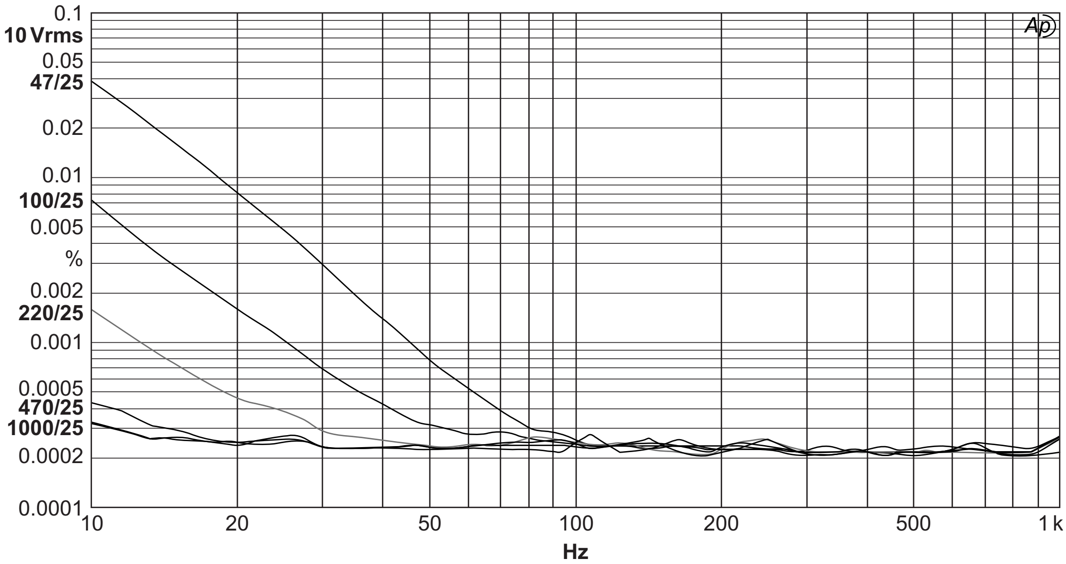


Figure 12.16: Reducing electrolytic capacitor distortion by increasing the capacitor value. Input 10 Vrms.

This technique can be seen to be highly effective, but it naturally calls for larger and somewhat more expensive capacitors, and larger footprints on a PCB. This can be to some extent countered by using capacitors of lower voltage, which helps to bring back down the CV product and hence the can size. I tested 1000 uF 16 V and 1000 uF 6V3 capacitors, and both types gave exactly the same results as the 1000 uF 25 V part in Figure 12.16, which seems to indicate that the maximum allowable signal voltage across the capacitor is an absolute value and not relative to the voltage rating. 1000 uF 16 V and 1000 uF 6V3 capacitors naturally gave very useful reductions in CV product, can size, and PCB area occupied. This does of course assume that the capacitor is, as is usual, being used to block small voltages from opamp offsets to prevent switch clicks and pot noises rather than for stopping a substantial DC voltage.

The use of large coupling capacitors in this way does require a little care, because we are introducing a long time-constant into the circuit. Most opamp circuitry is pretty much free of big DC voltages, but if there are any, the settling time after switch-on may become undesirably long.

References

[1] W.J. Smith, Modern Optical Engineering, McGraw-Hill, London, 1990, p. 484, ISBN 0-07-059174-1.
[2] H. Johnson, http://www.edn.com/article/509250-7_solution.php, 2010.
[3] H. Kroeze, http://www.rfglobalnet.com/forums/Default.aspx?g=posts&m=61096, 2002.

[4] M. Lovell, M. Jones, http://www.pmillett.com/rescalc.htm.
[5] D. Self, http://www.douglas-self.com/.
[6] D. Self, Small Signal Audio Design Handbook, Chap. 7, Newnes, Boston, MA, 2009, p. 189, ISBN 978-0-240-52177-0.
[7] The Signal Transfer Company, http://www.signaltransfer.freeuk.com/.
[8] K. Kundert, Modelling Dielectric Absorption in Capacitors, http://www.designers-guide.org/Modeling/da.pdf, 2008 (accessed 12.10).
[9] C. Bateman, Capacitor Sound? Parts 1–6, Electronics World, London, 2002–2003.
[10] D. Self, Audio Power Amplifier Handbook, fifth ed., Newnes, Boston, MA, p. 43 (amp output cap).
[11] D. Self, Audio Power Amplifier Design Handbook, fifth ed., Newnes, Boston, MA, p. 205 (COG cap).
[12] C. Bateman, Capacitor Sound? Part 3, Electronics World, London, 2002, pp. 16–18.
[13] C. Bateman, Capacitor Sound? Part 4, Electronics World, London, 2002, p. 47.

CHAPTER 13

Opamps for Active Crossovers

13.1 Active Devices for Active Crossovers

It is a truth universally acknowledged, that if you are designing an active crossover, you will need active devices. In this day and age that almost always means opamps. Active crossovers can be built with discrete transistor circuitry if this is desirable for performance benefits or marketing reasons. This is relatively straightforward for Sallen & Key filters that require only a unity-gain buffer or voltage-follower, which can be implemented as some form of emitter-follower. It is a bit more complex for MFB filters and some equaliser circuits, which require a high-open-loop-gain inverting amplifier, and more complex again for configurations such as state-variable filters, which require multiple differential amplifiers. (The first active crossover I designed for production was in fact a discrete-transistor system, solely on the grounds of performance, for the affordable opamps of the time were really not very good.) However, even if the crossover architecture is confined to Sallen & Key filters, matching the distortion performance of the newer opamps is going to be no easy matter. The noise performance of discrete circuitry should be slightly better, but in a typical application the differences will be marginal. Discrete transistor circuitry undoubtedly scores on the issue of headroom, because you can use supply rails that are pretty much as high as you like; the downside is that the rail voltages have to be increased considerably to get a meaningful increase in headroom when it is expressed in dB, and equipment capable of producing very high output voltages can be dangerous to other parts of the system if the output levels are mismanaged. While the study of suitable discrete circuitry for active crossovers would be fascinating, its doubtful utility means that space cannot be given to it here, and this chapter concentrates solely on opamps.

It is of course also possible to make active crossovers using valves. Given the large number of active elements required in a high-quality crossover, and the necessity of working at high impedance levels with accompanying higher noise, this is for most of us a truly unattractive proposition. If you are one of those who insist on using directly-heated triodes left over from WW1, then the prospect becomes quite surreal.

An active crossover is an optional part of an audio system, in the sense that you could use passive crossovers instead. If you are inserting an extra piece of equipment into the audio path, especially one that requires a good deal of expenditure on additional power amplifiers and so on, it really makes sense that it should be of high quality. The cost of a mediocre

The Design of Active Crossovers

active crossover is not going to be radically different from that of a really good one, the main differences being the cost of the opamps, the cost of the capacitors (there is a great deal on this issue in Chapters 8 and 12), and perhaps the cost of balanced outputs with their associated connectors. In active crossovers there are likely to be four or more filter stages in succession, so the quality of each one must be high. For these reasons this chapter focuses on achieving the best possible performance rather than cutting the last penny off the costing sheet.

13.2 Opamp Types

You might be questioning how a discourse on opamps for active crossovers will differ from one on opamps for general use in audio circuitry. One issue is that active crossover design tends to make great use of Sallen & Key filters, which rely on voltage-followers as the active part of the circuitry. Opamps in voltage-followers work under the most demanding conditions possible as regards common-mode (CM) distortion, and this topic, which gets little attention in most textbooks, is therefore examined in detail here.

General audio design has for a great many years relied on a very small number of opamp types; the dual-opamp TL072, the single-opamp 5534, and the dual-opamp 5532 dominated the audio small-signal scene for a long time. The TL072, with its JFET inputs, was used wherever its negligible input bias currents, lower power consumption, and lower cost were important. For a long time the 5534 and 5532 were much more expensive than the TL072, so the latter was used wherever feasible in an audio system, despite its markedly inferior noise, distortion, and load-driving capabilities. The 5534 or 5532 was reserved for critical parts of the circuitry. Although it took many years, the price of the 5534 or 5532 is now down to the point where it is usually the cheapest opamp you can buy, and its cost/performance ratio is outstanding. You need a very good reason to choose any other type of opamp for audio work.

The TL072 and the 5532 are dual opamps; the single equivalents are TL071 and 5534. Dual opamps are used almost universally, as the package containing two is usually cheaper than the package containing one, simply because it is more popular. The 5534 also requires an external compensation capacitor for closed-loop gains of less than three, which adds to the cost.

It took a long time for better opamps for audio to come along. Some were marketed specifically for audio applications, such as the idiosyncratic OP275, which had both BJT and JFET input devices. Unfortunately it also had higher noise and higher distortion than a 5532, and cost six times as much, so it made little headway It is only in the last few years that opamps have appeared that have better performance than the 5532. Notable examples are the LM4562 and the LME49990, and these are examined in this chapter.

There are many opamps on the market which could be applied to audio purposes, and to deal with them all in detail would fill this book, so only a selected range is covered here. Samuel Groner [1] has measured a wide range of opamp types and his published measurements should be your first recourse if your favourite opamp is not included here.

13.2.1 Opamp Properties: Noise

Table 13.1 ranks the opamps most commonly used for audio in order of voltage noise. The great divide is between JFET input opamps and BJT (bipolar junction transistor) input opamps. The JFET opamps have more voltage noise but less current noise than bipolar input opamps, the TL072 being particularly noisy. The BJT opamps have the lowest voltage noise, but the current noise is much higher. The voltage noise of a modern JFET-input opamp such as the OPA2134 is 4 dB greater than that of the old faithful 5532; and the JFET part is a good deal more costly.

It is important to realise that the voltage noise, like the poor, is always with you, but the effect of the current noise is negotiable in that it only becomes measurable and audible voltage noise when it flows through an impedance. Current noise can therefore be reduced by using suitably low circuit impedances, and here the ability of an opamp to drive heavy loads without increased distortion becomes very important. This noise advantage is

Table 13.1: Opamps Ranked by Typical Voltage Noise Density

Opamp	e_n nV/rtHz	i_n pA/rtHz	Input Device Type	Bias Cancel?
TL072	18	0.01	JFET	No
OPA604	11	0.004	JFET	No
NJM4556	8	Not spec'd	BJT	No
OPA2134	8	0.003	JFET	No
LME49880	7	0.006	JFET	No
OP275	6	1.5	BJT + FET	No
OPA627	5.2	0.0025	DIFET	No
5532A	5	0.7	BJT	No
LM833	4.5	0.7	BJT	No
MC33078	4.5	0.5	BJT	No
5534A	3.5	0.4	BJT	No
OP270	3.2	0.6	BJT	No
OP27	3	0.4	BJT	**YES**
LM4562	2.7	1.6	BJT	No
LME49710	2.5	1.6	BJT	No*
LME49990	0.9	2.8	BJT	No*
AD797	0.9	2	BJT	No
LT1115	0.85	1	BJT	No*
LT1028	0.85	1	BJT	**YES**

*Not directly stated on datasheet but inferred from specifications.
NB: A DiFET is a dielectrically-isolated FET, i.e., a MOSFET rather than a JFET.

one reason why BJT opamps are a better choice for high-quality circuitry where the impedance levels can be chosen for best results.

Looking at the range of BJT opamps in Table 13.1, the current noise increases as the voltage noise decreases, as both depend on the standing current in the input devices, so parts like the AD797 and LME49990 are definitely best suited to low impedance circuitry. The LM4562 has almost 6 dB less voltage noise than the 5532, while both the LME49990 and the AD797 are almost 15 dB quieter, given sufficiently low impedance levels that their high current noise does not intrude.

A further complication in this noise business is that the OP27, and LT1028 devices have bias cancellation systems which can cause unexpectedly high noise, for reasons described later. Opamps with bias cancellation circuitry are often unsuitable for audio use due to the extra noise this creates. The amount depends on circuit impedances, and is not taken into account in Table 13.1.

13.2.2 Opamp Properties: Slew Rate

Slew rates vary more than most parameters; a range of 100:1 is shown here in Table 13.2. A maximum slew rate greatly in excess of what is required appears to confer no benefits whatever.

Table 13.2: Opamps Ranked by Typical Slew Rate

Opamp	V/us
OP270	2.4
OP27	2.8
NJM4556	3
MC33078	7
LM833	7
5532A	9
LT1028	11
TL072	13
5534A	13
LT1115	15
LME49880	17
LME49710	20
OPA2134	20
LM4562	20
AD797	20
OP275	22
LME49990	22
OPA604	25
OPA627	55

The 5532 slew rate is typically ±9 V/us. This opamp is internally compensated for unity-gain stability, not least because there are no spare pins for compensation when you put two opamps in an 8-pin dual package. The single-amp version, the 5534, can afford a couple of compensation pins, and so is made to be stable only for gains of 3x or more. The basic slew rate is therefore higher at ±13 V/us.

Compared with power-amplifier specs, which often quote 100 V/us or more, these speeds may appear rather sluggish. In fact they are not; even ±9 V/us is more than fast enough. Assume you are running your opamp from ±18 V rails, and that it can give a ±17 V swing on its output. For most opamps this is distinctly optimistic, but never mind. To produce a full-amplitude 20 kHz sine wave you only need 2.1 V/us, so even in the worst case there is a safety-margin of at least four times. Such signals do not of course occur in actual use, as opposed to testing. More information on slew-limiting is given in the section on opamp slew-limiting distortion.

13.2.3 Opamp Properties: Common-Mode Range

This is simply the range over which the inputs can be expected to work as proper differential inputs. It usually covers most of the range between the rail voltages, with one notable exception. The data sheet for the TL072 shows a common-mode (CM) range that looks a bit curtailed at −12 V. This bland figure hides the deadly trap this IC contains for the unwary. Most opamps, when they hit their CM limits, simply show some sort of clipping. The TL072, however, when it hits its negative limit, promptly inverts its phase, so your circuit either latches up, or shows nightmare clipping behaviour with the output bouncing between the two supply rails. The positive CM limit is in contrast trouble-free. This behaviour can be especially troublesome when TL072s are used in high-pass Sallen & Key filters; they are definitely not recommended for this role.

13.2.4 Opamp Properties: Input Offset Voltage

A perfect opamp would have its output at 0 V when the two inputs were exactly at the same voltage. Real opamps are not perfect and a small voltage difference usually a few milliVolts is required to zero the output. These voltages are large enough to cause switches to click and pots to rustle, and DC blocking capacitors are very often required to keep them in their place.

The typical offset voltage for the 5532A is ±0.5 mV typical, ±4 mV maximum at 25 °C; the 5534A has the same typical spec but a lower maximum at ±2 mV. The input offset voltage of the new LM4562 is only ±0.1 mV typical, ±4 mV maximum at 25 °C.

13.2.5 Opamp Properties: Bias Current

Bipolar-input opamps not only have larger noise currents than their JFET equivalents, they also have much larger bias currents. These are the base currents taken by the input transistors. This current is much larger than the input offset current, which is the difference between the bias current for the two inputs. For example, the 5532A has a typical bias current of 200 nA, compared with a much smaller input offset current of 10 nA. The LM4562 has a much lower bias current of 10 nA typical, 72 nA maximum. In the case of the 5532/4 the bias current flows into the input pins as the input transistors are NPN.

Bias currents are a considerable nuisance when they flow through variable resistors they make them noisy when moved. They will also cause significant DC offsets when they flow through high-value resistors.

It is often recommended that the effect of bias currents can be cancelled out by making the resistance seen by each opamp input equal. Figure 13.1a shows a shunt-feedback stage with a 22 kΩ feedback resistor. When 200 nA flows through this it will generate a DC offset of 4.4 mV, which is a good deal more than we would expect from the input offset voltage error.

If an extra resistance Rcompen, of the same value as the feedback resistor, is inserted into the non-inverting input circuit then the offset will be cancelled. This strategy works well and appears to be done almost automatically by some designers. However, there is a snag. The resistance Rcompen generates extra Johnson noise, and to prevent this it is necessary to shunt the resistance with a capacitor, as in Figure 13.1b. This extra component costs money and takes up PCB space, so it is questionable if this technique is actually very useful for audio work. It is usually more economical to allow offsets to accumulate in a chain of opamps, and then remove the DC voltage with a single output blocking capacitor. This assumes that there are no stages with a large DC gain, and that the offsets are not large

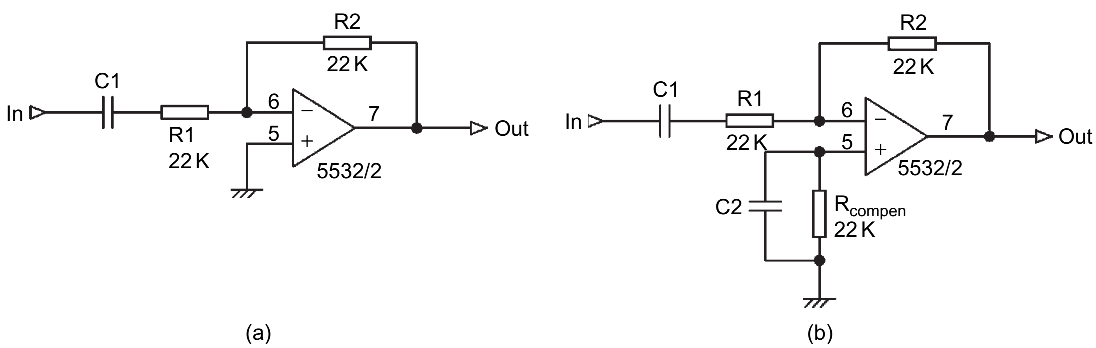


Figure 13.1: Compensating for bias current errors in a shunt-feedback stage. The compensating resistor must be bypassed by a capacitor C2 to prevent it adding Johnson noise to the stage.

enough to significantly reduce the available voltage swing. Care must also be taken if controls are involved, because even a small DC voltage across a potentiometer will cause it become crackly, especially as it wears.

FET input opamps have very low bias current at room temperature; however it doubles for every 10° Centigrade rise. This is pretty unlikely to cause trouble in most audio applications, but a combination of high internal temperatures and high-value pots could lead to some unexpected and unwelcome crackling noises.

13.2.6 Opamp Properties: Cost

While it may not appear on the datasheet, the price of an opamp is obviously a major factor in deciding whether or not to use it. Table 13.3 lists the opamps commonly used in audio, and was derived from the averaged prices for 25+ quantities across a number of UK distributors. At the time of writing (2010) the 5532 was not only by far the most popular audio opamp but also the cheapest, so its price was taken as unity and used as the basis for the price ratios given.

The table is ranked by cost per package, but of course some are dual and some are single packages. The column on the right shows cost per actual opamp, with the singles bolded. This shows that some parts are relatively very expensive indeed, costing almost 100 times as much per opamp as a 5532.

Table 13.3: Opamps Ranked by Price per Package (2010) Relative to the 5532

Opamp	Format	Price Ratio per Package 25+	Price Ratio per Opamp 25+
5532	Dual	**1.00**	**1.00**
TL072	Dual	1.00	1.00
LM833	Dual	1.12	1.12
MC33078	Dual	1.27	1.27
5534A	Single	1.52	**3.04**
TL052	Dual	2.55	2.55
OP275GP	Dual	3.42	3.42
OPA2134PA	Dual	4.45	4.45
OPA604	Dual	5.03	5.03
OP27	Single	6.76	**13.52**
LM4562	Dual	9.06	9.06
LME49990	Single	9.58	19.15
LT1115	Single	12.73	**25.45**
AD797	Single	13.09	**26.18**
LT1028	Dual	17.88	17.88
OPA270	Single	24.42	**48.85**
LME49710	Single	30.58	**61.15**
OPA627	Single	48.42	**96.85**

(No representative UK distributor was found for the LME49880 at time of writing.)

Table 13.3 was compiled using prices for DIL packaging and the cheapest variant of each type. It is obviously only a rough guide. Purchasing in large quantities or in different countries may change the rankings somewhat, but the basic look of things will not alter too much. One thing is obvious—the 5532 is one of the great opamp bargains of all time.

Active crossovers are not going to be used in low-cost systems where every penny counts; they are found in high-end hi-fi systems and professional PA rigs, and so they are expected to have appropriately good performance. There is usually no point in compromising this to make minor cost savings. However, as we shall see, there is in fact very little in the way of agonising dilemmas about cost/performance tradeoffs. The 5532/5534 may be the cheapest opamp, but it is also an exceptionally good one. When first introduced it was very much an expensive premium part, but its popularity has driven prices down. When used with a little care (and I shall be going into that in detail in this chapter) it is capable of a performance that can only be beaten by using parts that cost ten times as much, such as the LM4562.

13.2.7 Opamp Properties: Internal Distortion

This is what might be called the basic distortion produced by the opamp you have selected. Sam Groner calls it "transfer distortion" [1]. Even if you scrupulously avoid clipping, slew-limiting, common-mode issues, and excessive output loading, opamps are not completely distortion free, though some types such as the 5532, the LM4562, and the LME49990 do have very low levels indeed. If distortion appears when the opamp is run with shunt feedback, to prevent common-mode voltages on the inputs, and with very light output loading, then it is probably wholly internal and all you can do is a) run the opamp at the maximum safe supply rails (this will improve linearity but not usually by very much) or b) pick a better opamp.

If the distortion is higher than expected, the cause may be internal instability provoked by putting a capacitative load directly on the output, or neglecting the supply decoupling. The classic example of the latter effect is the 5532, which shows high distortion if there is not a capacitor across the supply rails close to the package; 100 nF is usually adequate. No actual HF oscillation is visible on the output with a general-purpose oscilloscope, so the problem is presumably instability in one of the intermediate gain stages.

13.2.8 Opamp Properties: Slew-Rate Limiting Distortion

This is essentially an overload condition, and it is the designer's responsibility to make sure it never happens. If users crank up the gain until the signal is within a hair of clipping, they should still be able to assume that slew-limiting will never occur, even with aggressive material full of high frequencies.

Arranging this is not too much of a problem. If the rails are set at the usual maximum voltage, i.e., ±18 V, then the maximum possible signal amplitude is 12.7 Vrms, ignoring the saturation voltages of the output stage. To reproduce this level cleanly at 20 kHz requires a minimum slew rate of only 2.3 V/usec. Most opamps can do much better than this, though with the OP27 (2.8 V/usec) you are sailing rather close to the wind. This calculation obviously assumes that the incoming signal is free of ultrasonic signals at any significant level; these may not be audible, but if they are large enough to provoke slew-limiting, there will be severe intermodulation distortion in the audio band. Bandwidth definition filters at the start of the audio chain will stop such unwanted signals.

Horrific as it may now appear, audio paths full of LM741s were quite common in the early 1970's. Entire mixers were built with no other active devices, and what complaints there were tended to be about noise rather than distortion. (This is not surprising as the voltage noise of a 741 has been measured at 22nV/√Hz; higher than any opamp in Table 13.1. No wonder it never appeared on the spec sheets) The reason for this is that full-level signals at 20 kHz simply do not occur in reality; the energy at the HF end of the audio spectrum is well-known to be lower than that at the bass end. See Chapter 14.

This assumes that slew-limiting has an abrupt onset as level increases, rather like clipping. This is in general the case with opamps. As the input frequency rises and an opamp gets closer to slew-limiting, the input stage is working harder to supply the demands of the compensation capacitance. There is an absolute limit to the amount of current this stage can supply, and when you hit it the distortion shoots up, much as it does when you hit the supply rails and induce voltage clipping. Before you reach this point, the linearity may be degraded, but usually only slightly until you get close to the limit. It is not normally necessary to keep big margins of safety when dealing with slew-limiting. If you are employing the Usual Suspects in the audio opamp world—the 5532 and TL072, with maximal slew rates of 9 and 13 V/usec respectively, you are most unlikely to suffer any slew-rate non-linearity.

13.2.9 Opamp Properties: Distortion Due to Loading

Output stage distortion is always worse with heavier output loading because the increased currents flowing exacerbate the gain changes in the Class-B output stage. These output stages are not in general individually trimmed for optimal quiescent conditions (as are audio power amplifiers) and so the crossover distortion produced by opamps tends to be both higher and can be more variable between different specimens of the same chip. On the other hand, the intimate contact between biasing circuits and the output transistors means that it is possible to make the quiescent conditions very stable against temperature, especially when compared with discrete-component power amplifiers where the thermal losses and lags are much greater. Distortion increases with loading in different ways for different opamps. It may rise only at the high-frequency end, or there may be a general rise at all frequencies. Often both effects occur, as in the TL072 and the 5532.

The lowest load that a given opamp can be allowed to drive is an important design decision. It will typically be a compromise between the distortion performance required and opposing factors such as number of opamps in the circuit, cost of load-capable opamps, and so on. It even affects noise performance, for the lower the load resistance an amplifier can drive, the lower the resistance values in the negative feedback can be, and hence the lower the Johnson noise they generate. There are limits to what can be done in noise-reduction by this method, because Johnson noise is proportional to the square-root of circuit resistance, and so improves only slowly as opamp loading is increased. Voltage noise from the opamps is not reduced at all, but the effect of current noise falls proportionally to the value of the circuit resistances. Overall the sum of these contributions decreases at a fairly gentle rate. Opamp distortion, however, at least in the case of the ubiquitous 5532, tends to rise rapidly when the loading exceeds a certain amount, and a careful eye needs to be kept on this issue. More modern devices such as the AD797 and the LM4562 handle heavy loading better, with less increase in distortion, and the very recent LME49990 barely reacts at all to loads down to 500 Ω. (See Figure 13.19 below)

13.2.10 Opamp Properties: Common-Mode Distortion

This is the general term for extra distortion that appears when there is a large signal voltage on both the opamp inputs. The voltage difference between these two inputs will be very small, assuming the opamp is in its linear region, but the common-mode (CM) voltage can be a large proportion of the available swing between the rails, and in the case of the voltage-follower will equal it.

Common-mode distortion appears to be the least understood distortion mechanism, and it gets little or no attention in opamp books, but it is actually one of the most important influences on opamp non-linearity. It is simple to separate this effect from the basic forward-path distortion by comparing THD performance in series and shunt-feedback modes; this should be done at the same noise gain. The distortion is often a good deal lower for the shunt-feedback case where there is no common-mode voltage. BJT and JFET input opamps show rather different behaviour as regards common-mode distortion, and this is an important difference between the two types. This fact seems to be very little appreciated.

A BJT opamp requires both a CM voltage and a significant source resistance driving an input for it to generate CM distortion. While this is somewhat speculative, my hypothesis is that this is due to Early effect occurring in the input stage when there is a large CM voltage, modulating the high input bias currents, and this is the cause of the distortion. (Early effect occurs in a bipolar transistor when changes in its collector-emitter voltage cause changes in the collector current, even though the base-emitter voltage is constant.) The signal input currents, which are in general non-linear, are much smaller due to the high open-loop gain of the opamp, and appear to have a negligible effect. There is more on this in the section below on the CM distortion behaviour of the 5532.

With JFET inputs the problem is not the input bias currents of the input devices themselves, which are quite negligible, but the currents drawn by the non-linear junction capacitances inherent in field-effect devices. These capacitances are effectively connected to one of the supply rails. For P-channel JFETs, as used in the input stages of most JFET opamps, the important capacitances are between the input JFETs and the substrate, which is normally connected to the V- rail. See, for example, Jung [2]. According to the Burr-Brown data sheet for the OPA2134, "The P-channel JFETs in the input stage exhibit a varying input capacitance with applied CM voltage." It goes on to recommend that the input impedances should be matched if they are above 2 kΩ, to cancel out the non-linearity.

The amount of CM distortion generated by a given type of opamp is very important for our purposes here because active crossover circuitry very often uses voltage-followers, primarily in Sallen and Key filters, but also for general buffering purposes. This configuration is the worst case for CM distortion because the full output voltage appears as a CM signal on the inputs. With BJT opamps CM distortion can be rendered negligible by keeping the source impedances low, which is also a very good idea as it reduces noise and susceptibility to electrostatic interference.

13.3 Opamps Surveyed

As we have seen, opamps with JFET inputs tend to have higher voltage noise but lower current noise than BJT-input types, and therefore give a better noise performance with high source resistances. Their very low bias currents often allow circuitry to be simplified by omitting DC blocking. These advantages are, however, of little use in active crossover circuitry. Their CM distortion performance also tends to be much worse, and for this reason, and to economise on space, I shall only look at one JFET opamp before examining some BJT types in detail.

Figure 13.2 shows some of the test circuits used in this chapter.

13.4 The TL072 Opamp

The TL072 is, or perhaps was, one of the most popular opamps, having very high-impedance inputs, with effectively zero bias and offset currents. The JFET input devices give their best noise performance at medium impedances, in the range 1 kΩ–10 kΩ. It has a modest power consumption at typically 1.4 mA per opamp section, which is significantly less than the 5532. The slew rate is higher than for the 5532, at 13 V/us against 9 V/us. The TL072 is a dual opamp; there is a single version called the TL071 that has offset null pins. Nowadays, the TL072 is regarded as somewhat obsolescent, as a result of its high voltage noise and mediocre load-driving capabilities, but it still gives a good illustration of JFET opamp issues.

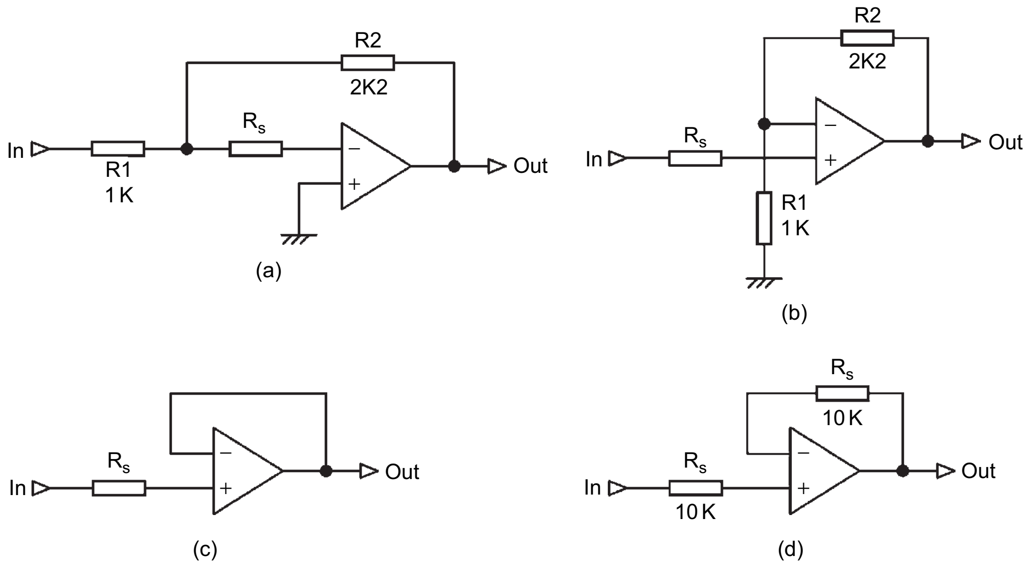


Figure 13.2: Opamp test circuits with added source resistance Rs: (a) shunt; (b) series; (c) voltage-follower; (d) voltage-follower with cancellation resistor in feedback path.

Figure 13.3 shows the distortion performance of the TL072 in voltage-follower mode as in Figure 13.2c, with varying extra source resistance in the input path. With a low driving impedance the THD at 10 kHz is 0.0025%, but with a 10 kΩ source resistance inserted, this rises to an alarming 0.035%. Note that the flat parts of the traces to the left are not the noise floor, as would be the case with more modern opamps; it is real distortion at about 0.0006%. The noise floor is equivalent to 0.00035%. Be aware that the test level is almost as high as possible at 10 Vrms, to get the low-frequency distortion clear of the noise floor; practical internal levels such as 3 Vrms will give much lower levels of distortion. This applies to all the distortion tests in this chapter, and indeed to the whole book.

Figure 13.3 should be compared with the same test carried out on BJT opamps; see Figure 13.10 (5532) where for Rs = 10 KΩ at 10 kHz the THD is 0.0015% against 0.035% for the TL072. Figure 13.19 (LM4562) shows 0.0037%, Figure 13.24 (LME49990) shows 0.024%, and Figure 13.27 (AD797) shows 0.0025%. Clearly the TL072 has far worse CM distortion than all except the LME49990, and that opamp is redeemed by its very low voltage noise.

Since the inputs of an opamp are nominally identical, it is possible to greatly reduce the effects of CM distortion by making both input see the same source impedance. In the voltage-follower case, this can be done by inserting an equal resistance in the feedback path, as in Figure 13.2d above. Figure 13.4 shows the how the high level of CM distortion can

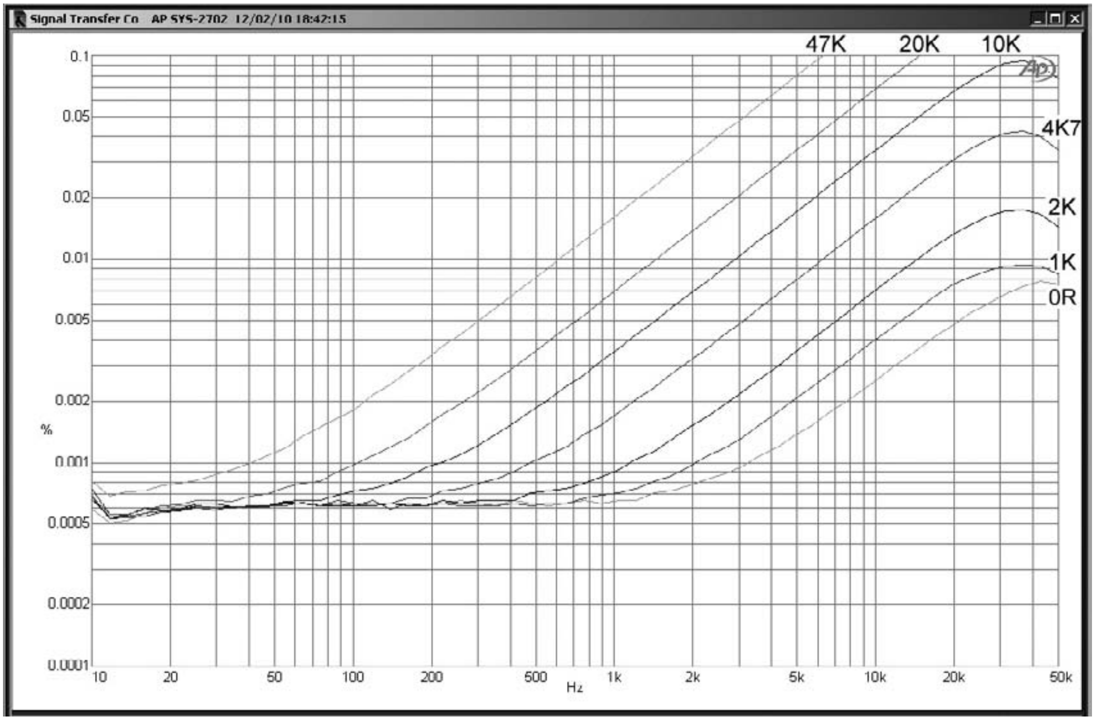


Figure 13.3: The TL072 in voltage-follower mode, with varying extra source resistance in the input path. CM distortion is much higher than for the BJT opamps. 10 Vrms out, ±18 V supply rails.

be radically reduced—observe the "Cancel" trace. However, adding resistances for distortion cancellation in this way has the obvious disadvantage that they introduce extra Johnson noise into the circuit. Taking that with the higher voltage noise of JFET opamps, this approach is definitely going to be noisy.

This cancellation technique is not pursued further in the sections on BJT opamps because there the effects of extra resistance noise are relatively more serious due to the lower opamp voltage noise.

It would clearly be desirable to examine more JFET opamps, but for reasons of space we need to cut to the chase and look in detail at the far more promising BJT opamps.

13.5 The NE5532 and NE5534 Opamps

Since the 5534/5532 is by far the most popular audio opamp I make no apology for describing its behaviour in quite a bit of detail. This will also illuminate issues that apply to all the other opamps examined here.

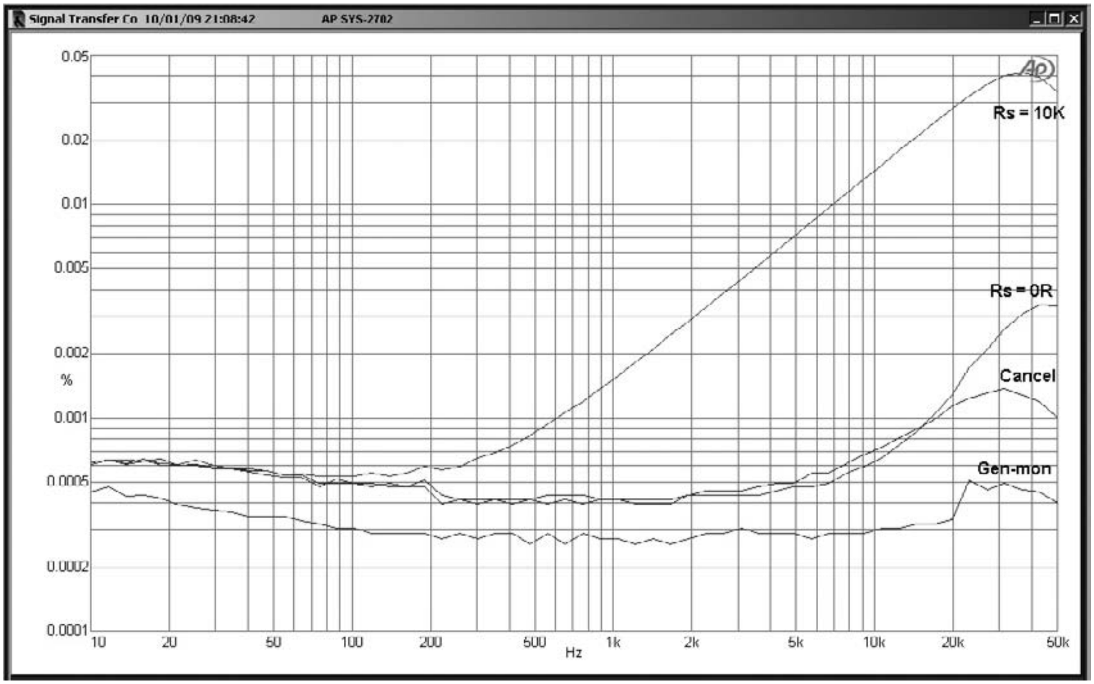


Figure 13.4: A TL072 voltage-follower driven from a low source resistance produces reasonably low distortion (Rs = 0R), but adding a 10 kΩ source resistance makes things much worse (Rs = 10K). Putting a 10 kΩ resistance in the feedback path as well gives complete cancellation of this extra distortion (Cancel). Output 5 Vrms, supply ±18 V. 'Gen-mon' means generator monitor i.e., the testgear output.

The 5532 is a low-noise, low distortion bipolar dual opamp, with internal compensation for unity-gain stability. The 5534 is a single version internally compensated for gains down to three, and an external compensation capacitor can be added for unity-gain stability; 22 pF is the usual value. The dual 5532 is used much more than the single 5534 as it is cheaper per opamp and does not require an external compensation capacitor when used at unity gain; the 5534 is however significantly quieter. The common-mode range of the inputs is a healthy ±13 V, with no phase inversion problems if this is exceeded. It has a distinctly higher power consumption than the TL072, drawing approx 4 mA per opamp section when quiescent. The DIL version runs perceptibly warm when quiescent on ±17 V rails.

The internal circuitry of the 5532 has never been publicly explained, but appears to consist of nested Miller loops that permit high levels of internal negative feedback.

The 5534/5532 has BJT input devices. It therefore gives low noise with low source resistances, but the downside is that it draws relatively high bias currents through the input pins. The input transistors are NPN, so the bias currents flow into the chip from the positive rail.

If an input is fed through a significant resistance then the input pin will be more negative than ground due to the voltage-drop caused by the bias current. The inputs are connected together with back-to-back diodes for reverse-voltage protection, and therefore should not be forcibly pulled to different voltages.

The 5532 and 5534 type opamps require adequate supply-decoupling if they are to remain stable; otherwise they appear to be subject to some sort of internal instability that seriously degrades linearity without being visible as oscillation on a normal oscilloscope. The essential requirement is that the +ve and −ve supply rails should be decoupled with a 100 nF capacitor between them, not more than a few millimetres from the opamp; normally one such capacitor is fitted per package as close to it as possible. It is *not* necessary, and often not desirable to have two capacitors going to ground; every capacitor between a supply rail and ground carries the risk of injecting rail noise into the ground.

The 5532 is a robust opamp but it is possible to permanently damage it so that it keeps working but shows high distortion. This seems to be associated with faults where one supply rail fails; see Chapter 18 for ways of guarding against this. Obviously such specimens should be disposed of at once to prevent confusion in the future.

13.6 The 5532 with Shunt Feedback

Figure 13.5 shows the distortion from a 5532 working in shunt mode with low-value feedback resistors of 1 kΩ and 2k2 setting a gain of 2.2 times, at an output level of 5 Vrms. This is the circuit of Figure 13.2a with Rs set to zero. This is the simplest situation as the use of shunt mode means there is no CM voltage, and hence no extra CM distortion. The THD is well below 0.0005% up to 20 kHz; this underlines what a superlative bargain the 5532 is.

Figure 13.6 shows the same situation but with the output increased to 10 Vrms (the clipping level on ±18 V rails is about 12 Vrms), and there is now some significant distortion above 10 kHz, though it only exceeds 0.001% when the frequency reaches 18 kHz. This remains the case when Rs in Figure 13.2a is increased to 10 kΩ and 47 kΩ- the noise floor is higher but there is no real change in the distortion behaviour. The significance of this will be seen later when we look at CM distortion.

13.7 5532 Output Loading in Shunt-Feedback Mode

When a significant load is placed on a 5532 output, the distortion performance deteriorates in a predictable way. Figure 13.7 shows the effect on a shunt-feedback amplifier (to eliminate the possibility of input common-mode distortion). The output of the opamp is of

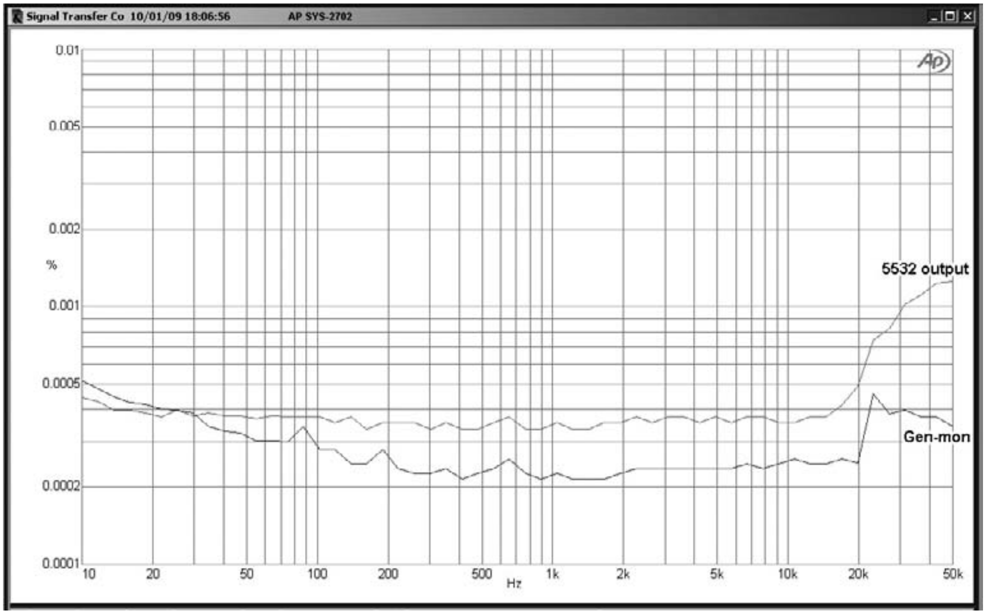


Figure 13.5: 5532 distortion in a shunt-feedback circuit (as in Figure 13.2a at 5 Vrms out. This shows the AP SYS-2702 output (lower trace) and the opamp output (upper trace). Supply ±18 V.

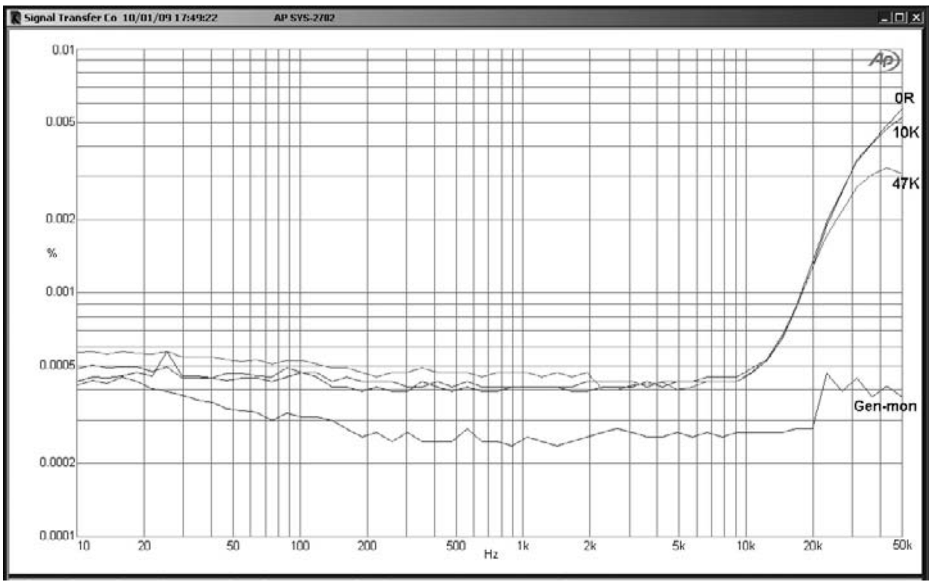


Figure 13.6: 5532 distortion in the shunt-feedback circuit of Figure 13.2a. Adding extra resistances of 10 kΩ and 47 kΩ in series with the inverting input does not degrade the distortion at all, but does bring up the noise floor a bit. Test level is now 10 Vrms out, supply ±18 V.

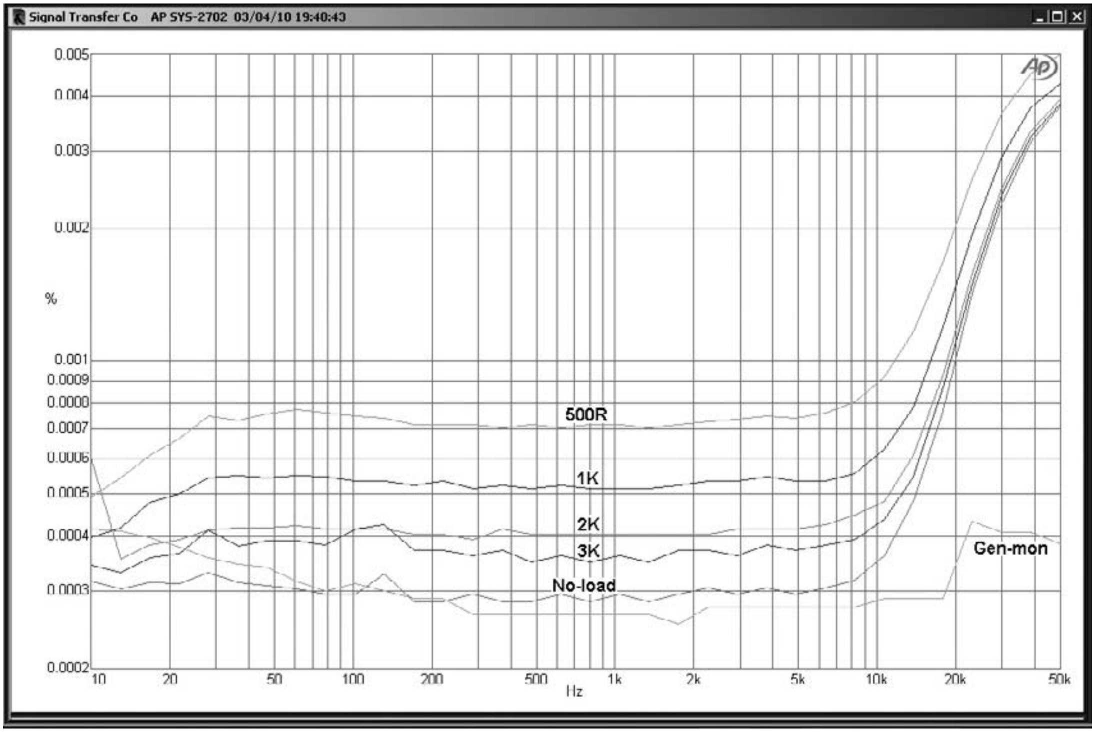


Figure 13.7: The effect of output loading on a shunt-feedback 5532 stage; 500 Ω, 1 KΩ, 2 KΩ, 3 KΩ, and No load, plus gen-mon (bottom trace). Feedback resistors 1 KΩ and 2K2, noise gain 3.2 times. Output 9 Vrms, supply ±18 V.

Table 13.4: THD at 1 kHz for Varying External Loads on a Shunt Feedback 5532

External Load Resistance Ω	THD at 1 kHz
No external load	0.00030%
3 KΩ	0.00036%
2 KΩ	0.00040%
1 KΩ	0.00052%
500 Ω	0.00072%

course always loaded by the feedback resistor, which is effectively connected to ground at its other end. The circuit is as shown in Figure 13.2b.

There are two different distortion regimes visible. At low frequencies (below 10 kHz) the distortion is flat with frequency. Above this, the THD rises rapidly, at approximately 12 dB/octave. This is very different from Blameless power amplifier behaviour, where the distortion normally rises at only 6 dB/octave when it emerges from the noise floor, typically around 2 kHz. The increase in THD in the flat region (at 1 kHz) is summarised in Table 13.4.

With no external load, the THD trace is barely distinguishable from the Audio Precision output until we reach 10 kHz, where the THD rises steeply. In this condition the opamp is of course still loaded by the 2K2 feedback resistor.

13.8 The 5532 with Series Feedback

Figure 13.8 shows the distortion from a 5532 working in series-feedback mode, as in Figure 13.2b. Note that the stage gain is greater at 3.2 times but the opamp is working at the same noise gain and so has the same amount of negative feedback available to reduce distortion. The working conditions are however less favourable, as we shall see in the next section on common-mode problems, and the distortion now begins to rise from 5 kHz and reaches 0.0025% at 20 kHz, as opposed to 0.0014% at 20 kHz for the shunt version, as in Figure 13.7.

Figure 13.8 also shows the effect of loading the output. As for the shunt case, the increased distortion is mostly apparent in the flat LF section of the traces.

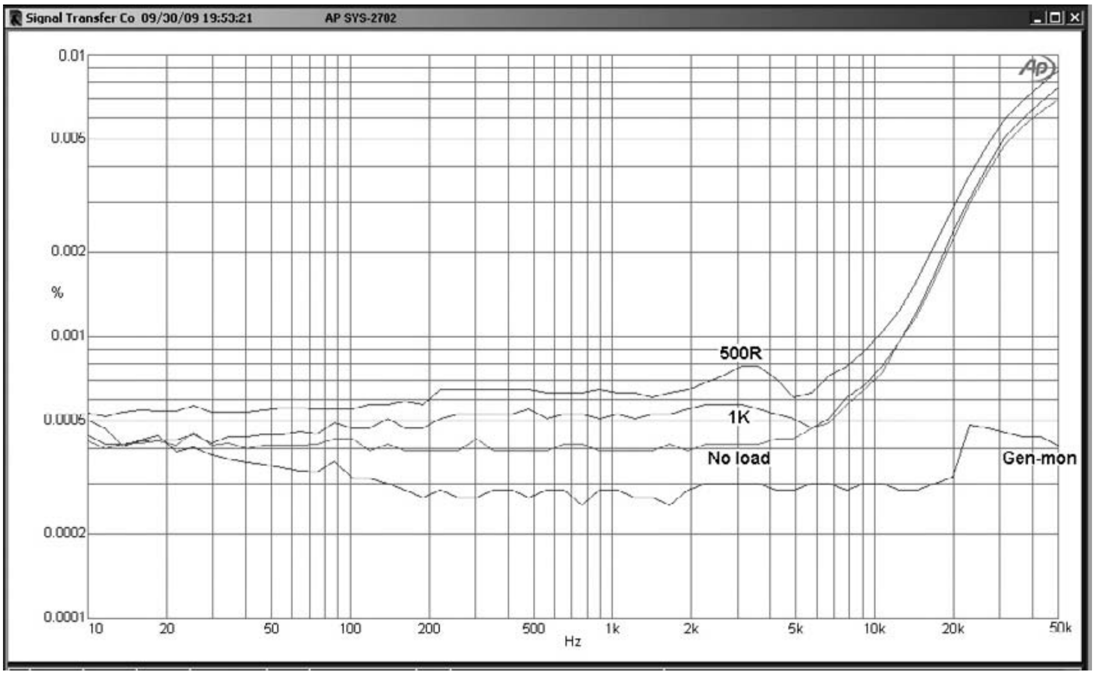


Figure 13.8: 5532 distortion in a series feedback stage with 2k2 & 1 kΩ feedback resistors, (gain 3.2 times) and zero source resistance. The output level is 10 Vrms with 500 Ω, 1 kΩ loads, and no load. The gen-mon trace is the output of the distortion analyser measured directly. Supply ±18 V.

13.9 Common-Mode Distortion in the 5532

In a series-feedback amplifier with a gain of 3.2 times the CM voltage is 3.1 Vrms for a 10 Vrms output. See Figure 13.9. The trace labelled "0R" (i.e., zero source resistance) is the same as the "No load" trace of Figure 13.8. The source resistance seen by the inverting input is not zero, because of the impedance of the feedback network, but this is only 2K2 in parallel with 1 KΩ; in other words 687 Ω. Figure 13.9 implies that this will have only a very small effect, but more on that later. When we add some source resistance Rs, the picture is radically worse, with serious midband distortion rising at 6 dB/octave, and roughly proportional to the amount of resistance added. We will note it is 0.0015% at 10 kHz with Rs = 10 kΩ. The horizontal low-frequency parts of the traces are raised by the Johnson noise from the source resistances and also by the opamp current noise flowing in those resistances.

The worst case for CM distortion is the voltage-follower configuration, as in Figure 13.2c, where the CM voltage is equal to the output voltage. Voltage-followers typically give low distortion because they have the maximum possible amount of negative feedback, and Figure 13.8 shows that even with a CM voltage of 10 Vrms, the distortion when driven from a low impedance is no greater than for the shunt mode. However, when source resistance is inserted in series with the input, the distortion mixture of second, third and other low-order harmonics increases markedly; in the 3.2 times stage 10 KΩ of source

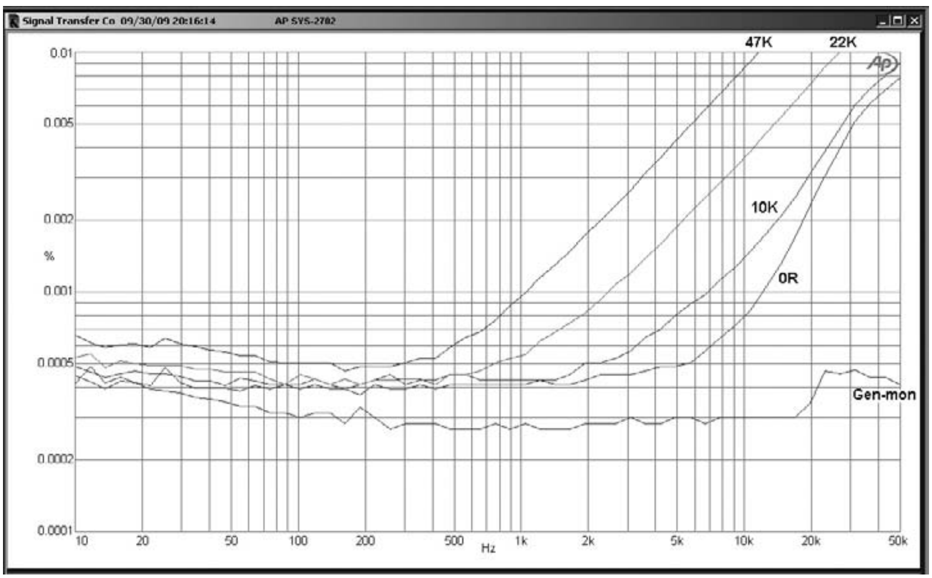


Figure 13.9: 5532 distortion in a series feedback stage with 2k2 & 1 kΩ feedback resistors, (gain 3.2 times) and varying source resistances. 10 Vrms output.

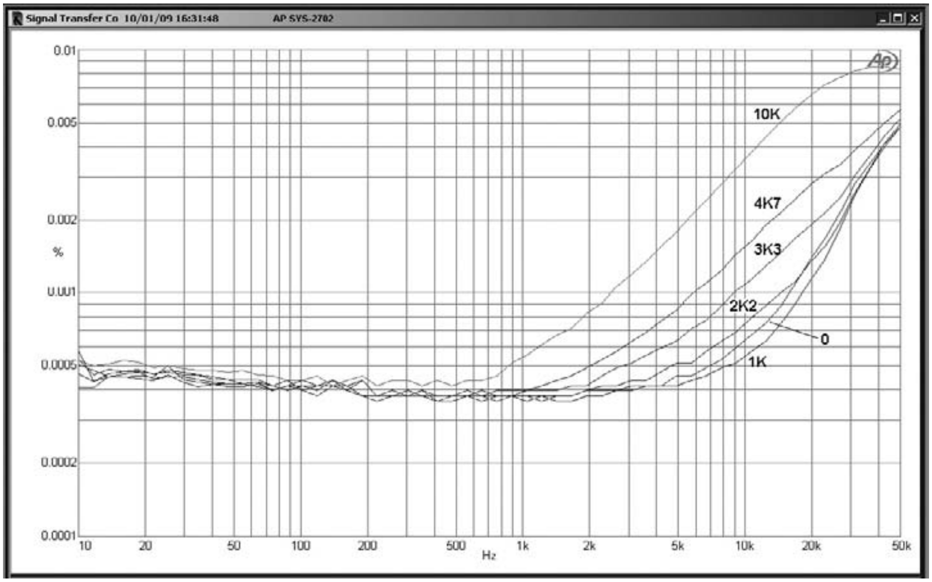


Figure 13.10: 5532 distortion in a voltage-follower circuit with a selection of source resistances; a 1 kΩ source resistance actually gives less distortion than no resistance, due to cancellation. The gen-mon THD was as before. Test level 10 Vrms, supply ±18 V.

resistance caused 0.0015% THD at 10 kHz, but with a voltage-follower we get 0.0035%. The distortion increases with output level, approximately quadrupling as level doubles. Figure 13.10 shows what happens with source resistances of 10 kΩ and below; when the source resistance is below 2k2, the distortion is close to the zero source resistance trace.

Close examination reveals the intriguing fact that a 1 kΩ source actually gives *less* distortion than no source resistance at all, reducing THD from 0.00065% to 0.00055% at 10 kHz. Minor variations around 1 kΩ make no measurable difference. This effect is due to the 1 kΩ source resistance cancelling the effect of the source resistance of the feedback network, which is 1 kΩ in parallel with 2k2, i.e., 688 Ω. The improvement is small, but if you are striving for the very best linearity, the result of deliberately adding a small amount of source resistance appears to be repeatable enough to be exploited in practice. A large amount will compromise the noise performance, as seen in Figure 13.9, and this is another argument for keeping feedback network impedances as low as practicable without impairing distortion.

This CM distortion behaviour is unfortunate because voltage-followers are very frequently required in active crossover design, primarily in Sallen and Key filters, but also for general buffering purposes. The only cure is to keep the source impedances low, which is obviously a good idea from the noise point of view also. A total of 2 kΩ is about as high as you want to go, and this is why the lowpass Sallen & Key filters in Chapter 8 have been given series resistors that do not exceed this figure.

So, what exactly is going on here with common-mode distortion? Is it simply due to non-linear currents drawn being drawn by the opamp inputs? Audio power amplifiers have discrete input stages which are very simple compared with those of most opamps, and draw relatively large input currents. These currents show appreciable non-linearity even when the output voltage of the amplifier is virtually distortion-free, and if they flow through significant source resistances will introduce added distortion [3].

If this was the case with the 5532 then the extra distortion would manifest itself whenever the opamp was fed from a significant source resistance, no matter what the circuit configuration. But as we saw earlier, it only occurs in series-feedback situations; increasing the source resistance in a shunt-feedback amplifier does not perceptibly increase distortion.

The only difference is that the series circuit has a CM voltage of about 3 Vrms, while the shunt circuit does not, and the conclusion is that with a bipolar input opamp, you must have *both* a CM voltage and a significant source resistance to see extra distortion. The input stage of a 5532 is a straightforward long-tailed pair, with a simple tail current source and no fancy cascoding, and I suspect that Early Effect is significant when there is a large CM voltage, modulating the quite high input bias currents, and this is what causes the distortion. The signal input currents are much smaller, due to the high open-loop gain of the opamp, and as we have seen appear to have a negligible effect.

13.10 Reducing 5532 Distortion by Output Biasing

There is an extremely useful, though relatively little-known (and where it is known it is almost universally misunderstood) technique for reducing the distortion of the 5532 opamp. While the general method may be applicable to some other opamp types, here I concentrate on the 5532, as probably the most popular opamp in the world, and it must not be assumed that the relationships or results will be emulated by any other opamp type.

The principle is that if a biasing current of the right polarity is injected into the opamp output, then the output stage distortion can be significantly reduced. This technique is sometimes called "output biasing" though it must be understood that the DC voltage conditions are not significantly altered; because of the high level of voltage feedback the actual DC potential at the output is shifted by only a tenth of a millivolt or so.

You may have recognised that this scheme is very similar to the Crossover Displacement (Class XD) system I introduced for power amplifiers, which also injects an extra current, either steady or signal-modulated, into the amplifier output [4]. It is not however quite the same in operation. In power amplifiers the main aim of Crossover Displacement is to prevent the output stage from traversing the crossover region at low powers. In the 5532 at least the crossover region is not easy to spot on the distortion residual, the general effect being of second and third harmonic distortion rather than spikes or edges; it appears that the

5532 output stage is more linear when it is pulling down rather than pulling up, and the biasing current is compensating for this.

For the 5532, the current *must* be injected from the positive rail; currents from the negative rail make the distortion emphatically worse. This confirms that the output stage of the 5532 is in some way asymmetrical in operation, for if it was simply a question of suppressing crossover distortion by Crossover Displacement, a bias current of either polarity would be equally effective. The continued presence of the crossover region, albeit displaced, would mean that the voltage range of reduced distortion would be quite small, and centred on 0 V. It is rather the case that there is a general reduction in distortion across the whole of the 5532 output range, which seems to indicate that the 5532 output stage is better at sinking current than sourcing it, and therefore injecting a positive current is effective at helping out.

Figure 13.11a shows a 5532 running in shunt-feedback mode with a moderate output load of 1 KΩ; the use of shunt feedback makes it easier to see what's going on by eliminating the possibility of common-mode distortion. With normal operation we get the upper trace in

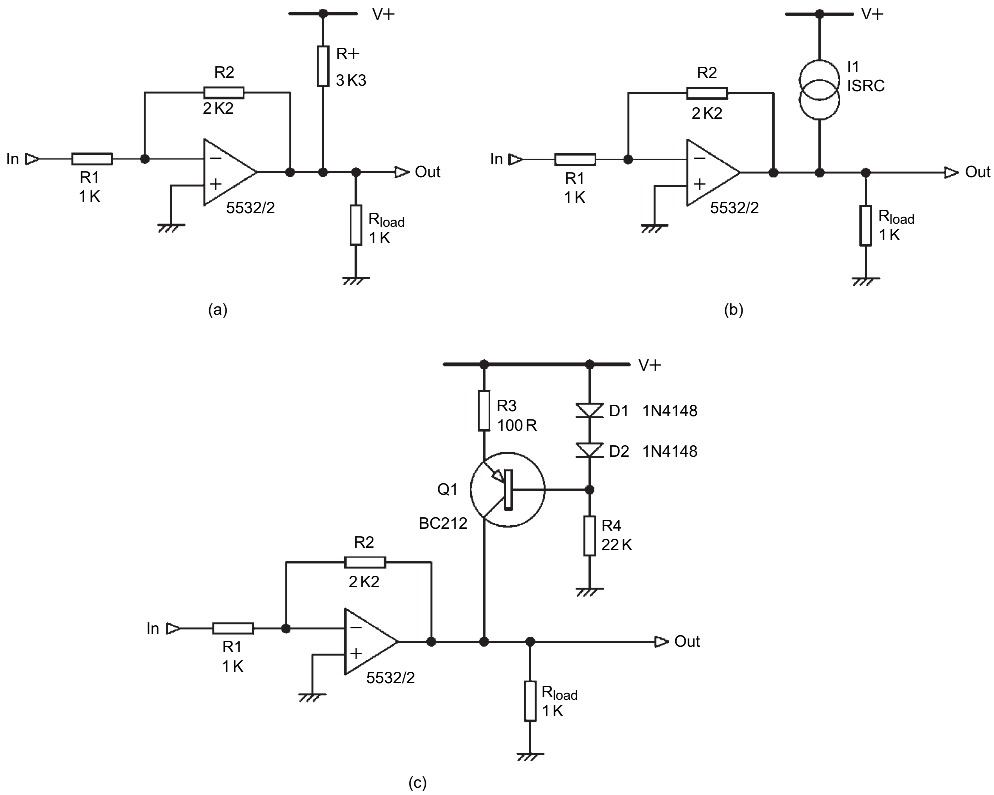


Figure 13.11: Reducing 5532 distortion in the shunt-feedback mode by biasing the output stage with a current injected through a resistor R+ or a current source.

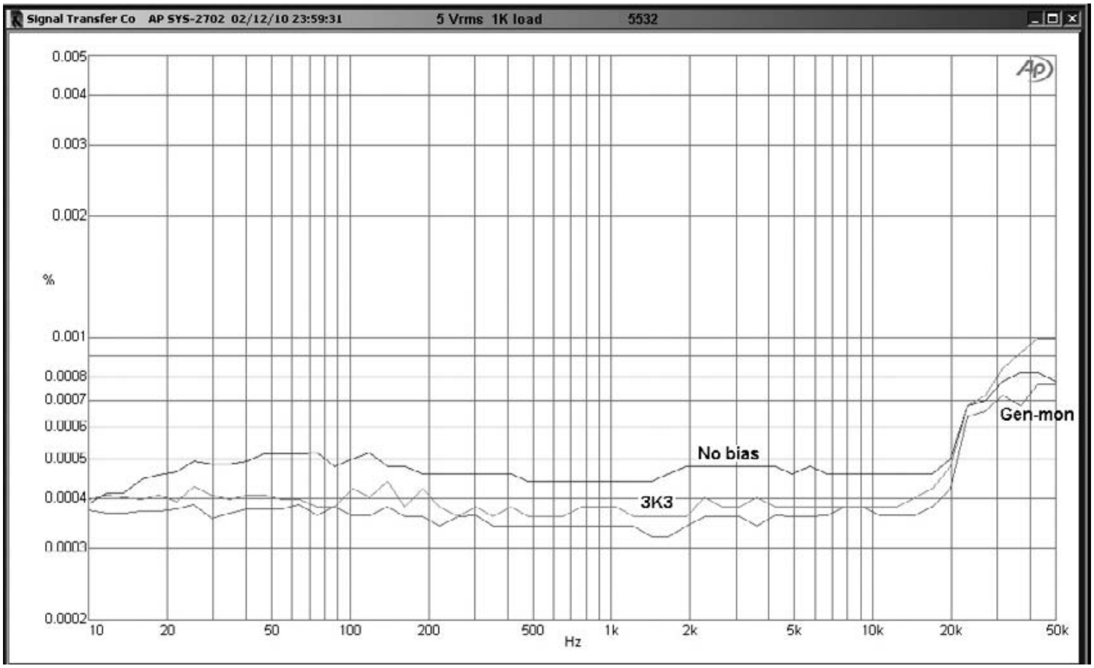


Figure 13.12: The effect of output biasing, with a 3K3 resistor to V+, on a unity-gain shunt-feedback 5532 stage. Output load 1 KΩ, input and feedback resistors are 2K2, noise gain 2.0 times. Output 5 Vrms, supply ±18 V.

Figure 13.12, labelled "No bias." If we then connect a current-injection resistor between the output and to the V+ rail, we find that the LF distortion (the flat bit) drops almost magically, giving the trace labelled "3K3", which is only just above the gen-mon trace. Since noise makes a significant contribution to the THD residual at these levels, the actual reduction in distortion is greater than it appears.

The optimum resistor value for the conditions shown (5 Vrms and 1 KΩ load) is about 3K3, which injects a 5.4 mA current into the output pin. A 2K2 resistor gives greater distortion than 3K3, due to the extra loading it imposes on the output; in AC terms the injection resistor is effectively in parallel with the output load. In fact, 3K3 seems to be close to the optimal value for a wide range of output levels and output loadings.

The extra loading that is put on the opamp output by the injection resistor is a disadvantage, limiting the improvement in distortion performance that can be obtained. By analogy with the canonical series of Class-A power amplifier outputs [5], a more efficient and elegant way to inject the required biasing current is by using a current source connected to the V+ rail, as in Figure 13.11b. Since this has a very high output impedance the loading on the opamp output is not increased. Figure 13.11c shows a simple but effective way to do this; the current source is set to the same current as the 3K3 resistor injects when the output is at 0 V, (5.4 mA) but the improvement in distortion is greater. There is nothing magical about this figure; however, increasing the injection current to say, 8 mA, gives only a small further improvement in the THD figure, and in some cases may make it worse; also the circuit dissipation is considerably increased, and in general I would not recommend using a current-source value of greater than 6 mA. Here in Table 13.5 are typical figures for a unity-gain shunt amplifier as before, with the loading increased to 680 Ω to underline that the loading is not critical; output biasing is effective with a wide range of loads.

As mentioned before, at such low THD levels the reading is largely noise, and the reduction of the distortion part of the residual is actually greater than it looks from the raw figures. Viewing the THD residual shows a dramatic difference.

You might be concerned about the Cbc of the transistor, which is directly connected to the opamp output. The 5532/5534 is actually pretty resistant to HF instability caused by load capacitance, and in the many versions of this configuration I tested I have had no HF stability problems whatever. The presence of the transistor does not reduce the opamp output swing.

Output biasing is also effective with series feedback amplifier stages in some circumstances. Table 13.6 shows it working with a higher output level of 9.6 Vrms, and a 1 KΩ load. The feedback resistors were 2K2 and 1 KΩ to keep the source resistance to the inverting input low.

Table 13.5: Output Biasing Improvements with Unity-Gain Shunt Feedback, 5 Vrms Out, Load 680 Ω, Supply ±18 V

Injection Method	THD at 1 kHz (22 kHz Bandwidth)
None	0.00034%
3K3 resistor	0.00026%
5.4 mA current-source	0.00023%
8.1 mA current-source	0.00021%

Table 13.6: Output Biasing Improvements with 3.2 Times Gain, Series Feedback, 9.6 Vrms Out, Load 1 KΩ, Supply ±18 V

Injection Method	THD at 1 kHz (22 kHz Bandwidth)
None	0.00037%
3K3 resistor	0.00033%
5.4 mA current-source	0.00027%
8.1 mA current-source	0.00022%

The output biasing technique is in my experience only marginally useful with voltage followers, as the increased feedback factor with respect to a series amplifier with gain reduces the output distortion below the measurement threshold. Table 13.7 demonstrates this.

As a final example, Figure 13.13 shows that the output biasing technique is particularly effective with higher gains, here 14 times. The distortion with the 5.4 mA source is barely distinguishable from the testgear output up to 2 kHz. The series-feedback stage had its gain

Table 13.7: Output Biasing Improvements for Voltage-Follower, 9.6 Vrms out, Load 680 Ω, Supply ±18 V

Injection Method	THD at 1 kHz (22 kHz Bandwidth)
None	0.00018% (almost all noise)
3K3 resistor	0.00015% (all noise)
5.4 mA current-source	0.00015% (all noise)
8.1 mA current-source	0.00015% (all noise)

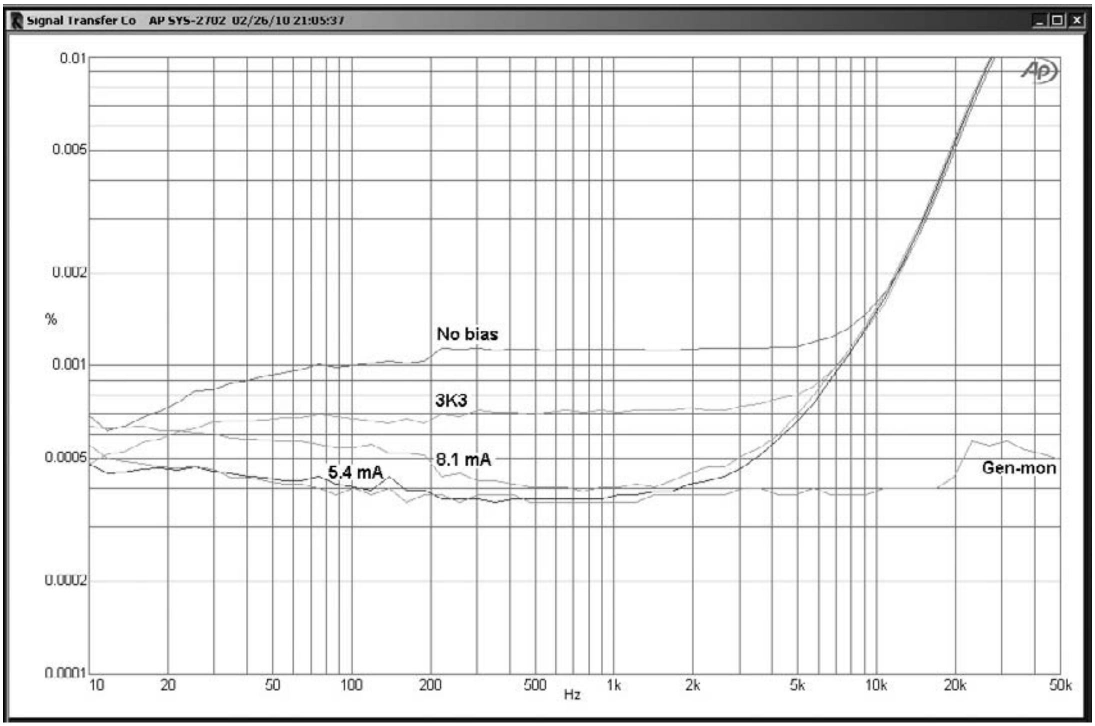


Figure 13.13: Reducing 5532 distortion with series feedback by biasing the output stage with a 3K3 resistor, or 5.2 or 8.1 mA current sources. Gain 14 times, no external load. Test level 5 Vrms out, supply ±18 V.

set by 1K3 and 100 Ω feedback resistors, their values being kept low to minimise common-mode distortion. It also underlines the point that in some circumstances an 8.1 mA current source gives worse results than the 5.4 mA version.

When extra common-mode distortion is introduced by the presence of a significant source resistance, this extra distortion is likely to swamp the improvement due to output biasing. In a 5532 amplifier stage with a gain of 3.2 times and a substantial source resistance, the basic output distortion with a 1 KΩ load at 9.6 Vrms, 1 kHz out was 0.0064%. A 3K3 output biasing resistor to V+ reduced this to 0.0062%, a marginal improvement at best, and an 8.1 mA current source could only reduce it to 0.0059%.

Earlier I said that the practice of output stage biasing appears to be pretty much universally misunderstood, judging by how it is discussed on the Internet. The evidence is that every application of it that my research has exposed shows a resistor (or current source) connected between the opamp output and the *negative* supply rail. This is very likely based on the assumption that displacing the crossover region in either direction is a good idea, coupled with a vague feeling that a resistor to the negative rail is somehow more "natural" and looks more like the familiar drawing of a load to ground. However, the assumption that the output stage is symmetrical is usually incorrect; as we have seen, it is certainly not true for the 5532/5534. For the 5532—which surely must be the most popular audio opamp by a long way—a pulldown resistor would be completely inappropriate as it *increases* rather than decreases the output stage distortion.

You may be thinking that this is an ingenious method of reducing distortion, but rather clumsy compared with simply using a more linear opamp like the LM4562. This is true, but on the other hand, if the improvement from output biasing is adequate, it will be much cheaper than switching to a more advanced opamp that costs ten times as much.

13.11 Which 5532?

It is an unsettling fact that not all 5532s are created equal. The part is made by a number of manufacturers, and there are definite performance differences. While the noise characteristics appear to show little variation in my experience, the distortion performance does vary noticeably from one manufacturer to another. Although, to the best of my knowledge, all versions of the 5532 have the same internal circuitry, they are not necessarily made from the same masks, and even if they were there would inevitably be process variations between manufacturers.

The main 5532 sources at present are Texas Instruments, Fairchild Semiconductor, ON Semiconductor (was Motorola), NJR (New Japan Radio), and JRC (Japan Radio Company). I took as wide a range of samples as I could, ranging from brand-new devices to parts over twenty years old, and it was reassuring to find that without exception, every part tested gave the good linearity we expect from a 5532. But there were differences. I did THD tests on six samples from Fairchild, JRC, and Texas, plus one old Signetics 5532 for historical interest.

All tests were done using a shunt-feedback stage with a gain of unity, both input and feedback resistors being 2 KΩ; the supply rails were ±18 V. The output level was high at 10.6 Vrms, only slightly below clipping. No external load was applied, so the load on the output was solely the 2 KΩ feedback resistor; applying extra loading will make the THD figures worse. The test instrumentation was an Audio Precision SYS-2702.

For most of the manufacturers distortion only starts to rise very slowly above the measurement floor at about 5 kHz, and remains well below 0.0007% at 20 kHz. The exceptions were the six Texas 5532s, which consistently showed somewhat higher distortion at high frequencies. Distortion at 20 kHz ranged from 0.0008% to 0.0012%, showing more variation than the other samples as well as being generally higher in level. The low-frequency section of the plot, below 10 kHz, was at the measurement floor, as for all the other devices, and distortion is only just visible in the noise in the THD residual.

Compared with other maker's parts, the THD above 20 kHz is much higher—at least three times greater at 30 kHz. Fortunately this should have no effect unless you have very high levels of ultrasonic signals that could cause intermodulation. If you do, then you have bigger problems than picking the best opamp manufacturer.

All of the measurements given in this book were performed using the Texas version of the 5532, to ensure worst-case results. If you use 5532s from one of the other manufacturers then your high-frequency distortion results should be somewhat better.

13.12 The 5534 Opamp

The single-opamp 5534 is somewhat neglected compared with the 5532, often being regarded as the same thing but inconveniently packaged with only one opamp per 8-pin DIL and in many cases requiring the expense of an external compensation capacitor. However, it can in fact be extremely useful when a somewhat better performance than the 5532 can give needs to be achieved economically, as it is between two and three times as expensive per opamp as the 5532 but about five times cheaper than the still somewhat exotic LM4562. The price ratio is likely to be nearer two when purchasing large quantities. (There was once a 5533 which was basically a dual 5534, with external compensation and offset null facilities in a 14-pin DIL package, but it seems to have achieved very little market penetration—the only place I have ever seen one is on the equaliser board of the famous EMT turntable. However, Philips were still putting out spec sheets for it in 1994.)

The 5534 has a very significant distortion advantage when gains of greater than unity are required, because the lighter internal compensation means that a greater NFB factor can be applied, with distortion reduced proportionally.

The 5534 also has lower noise than the 5532, its input voltage noise density being 3.5 uV/√Hz rather than 5 uV/√Hz. This means that in situations where the voltage noise dominates over the current noise- the sort of situation where you would want to use a BJT input opamp—the noise performance is potentially improved by 3.0 dB. Johnson noise from the circuit resistances is likely to reduce this difference in some cases.

Samuel Groner points out [1] that the higher voltage noise of the 5532 is almost certainly because the input pair has emitter degeneration resistors added to make unity-gain compensation easier. These are not shown on any internal schematic I have ever seen, but their presence is suggested by the fact that the 5532 has more voltage noise but much the same current noise, and has a faster slew rate than a unity-gain compensated 5534 because its compensation capacitor can be smaller.

Figure 13.14 demonstrates the excellent linearity of an uncompensated 5534 with shunt feedback for 3.2 times gain and no external loading; distortion is 0.0005% at 20 kHz and 10 Vrms out. Compare this with Figure 13.6, which shows 0.0015% at 20 kHz for a 5532

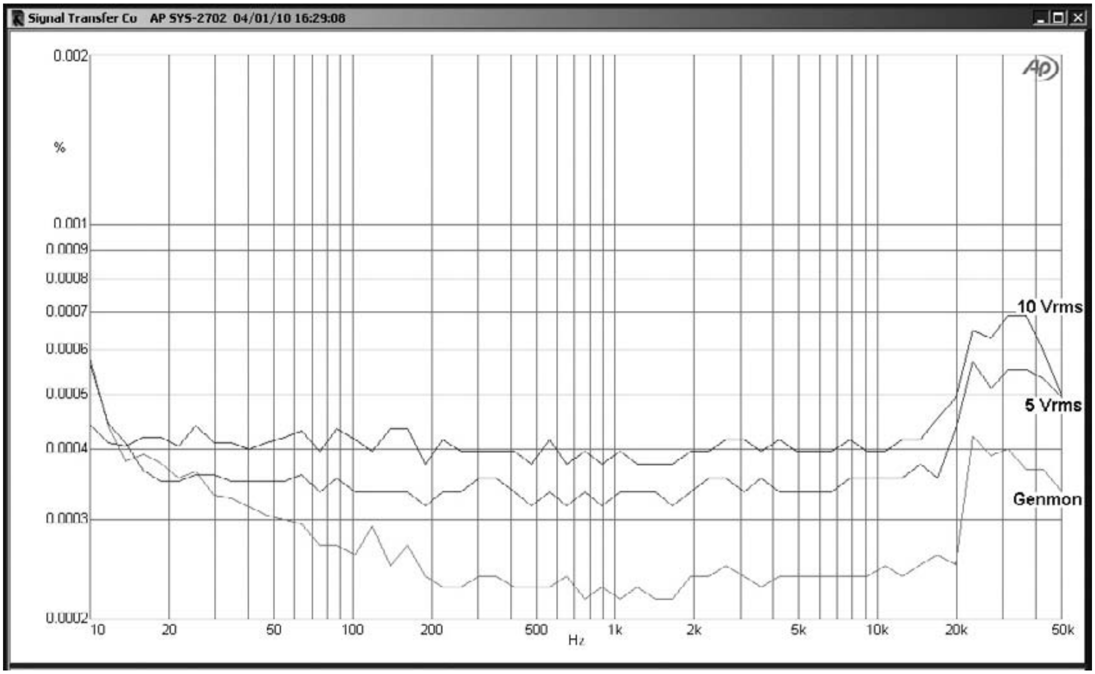


Figure 13.14: Distortion from an uncompensated 5534 with shunt feedback with 3.2 times gain and no external loading, at 5 Vrms and 10 V rms out. The gen-mon (testgear output) is also shown. Supply ±18 V.

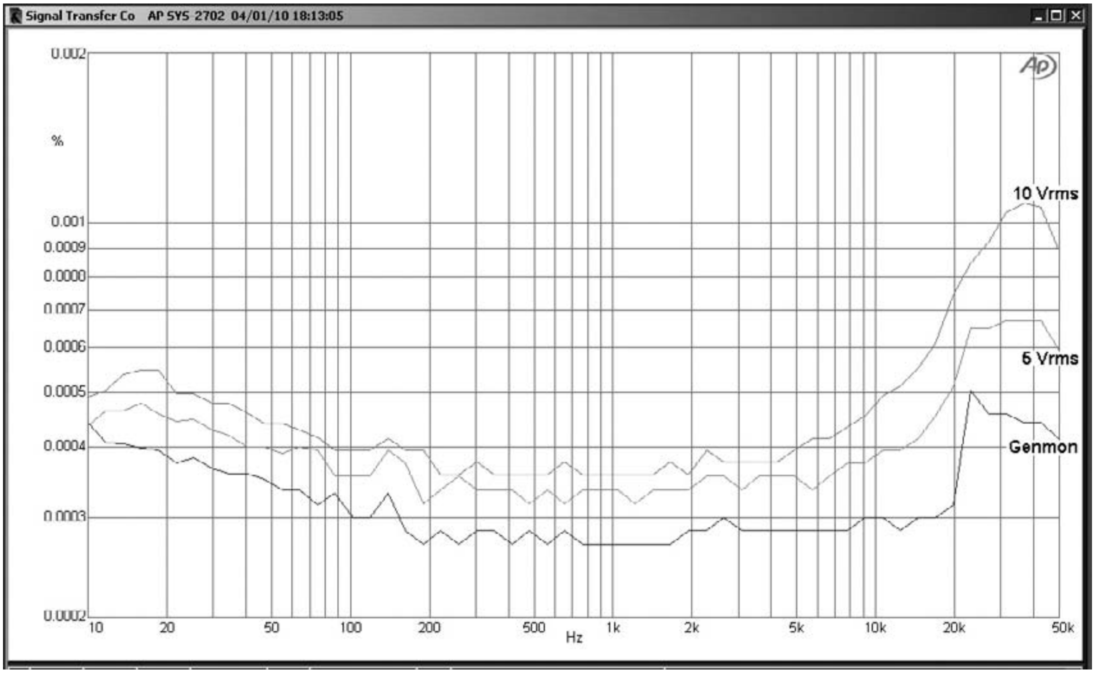


Figure 13.15: Distortion from an uncompensated 5534 with series feedback and 3.2 times gain, no external loading, at 5 Vrms and 10 V rms out. Supply ±18 V.

section in the same situation. The 5534 distortion is reduced proportionally to the increased NFB factor, being three times less.

Figure 13.15 shows an uncompensated 5534 with series feedback for 3.2 times gain and no external loading; distortion is 0.00077% at 20 kHz and 10 Vrms out. Compare this with the "No-load" trace in Figure 13.8, which shows 0.0022% at 20 kHz for a 5532 section in the same situation. The reduction is again proportional to the increased NFB factor, being three times lower once more. Note that there is no source resistance added so the 5534 is fed from a low impedance and will not exhibit CM distortion.

If the 5534 is used with external compensation to allow unity-gain stability, then the distortion advantage is lost but the better noise performance remains. The generally accepted value is 22 pF for unity-gain stability, though this does not as far as I know have any official seal of approval from a manufacturer. In fact I have often found that uncompensated 5534s were stable and apparently quite happy at unity gain, but this is not something to rely on.

For a unity-gain shunt-feedback stage the opamp noise gain is two times. There appears to be no generally accepted value for the external compensation capacitor required in this situation, or other circuits requiring a 5534 to work at a noise gain of two, but I have always found that 10 pF does the job.

13.13 The LM4562 Opamp

The LM4562 is a relatively new opamp, which first become freely available at the beginning of 2007. It is a National Semiconductor product. It is a dual opamp—there is no single or quad version. It costs about ten times as much as a 5532.

The input noise voltage is typically 2.7 nV/√Hz, which is substantially lower than the 5 nV/√Hz of the 5532. For suitable applications with low source impedances this translates into a useful noise advantage of 5.3 dB. The bias current is 10 nA typical, which is very low and would normally imply that bias-cancellation, with its attendant noise problems, was being used. However in my testing I have seen no sign of excess noise, and the data sheet is silent on the subject. No details of the internal circuitry have been released so far, and quite probably never will be. The LM4562 is not fussy about decoupling, and as with the 5532, 100 nF across the supply rails close to the package seems to ensure HF stability. The slew rate is typically ±20 V/us, more than twice as fast as the 5532.

The first THD plot in Figure 13.16 shows the LM4562 working at a closed-loop gain of 2.2 times in shunt-feedback mode, at a high level of 10 Vrms. The top of the THD scale is now

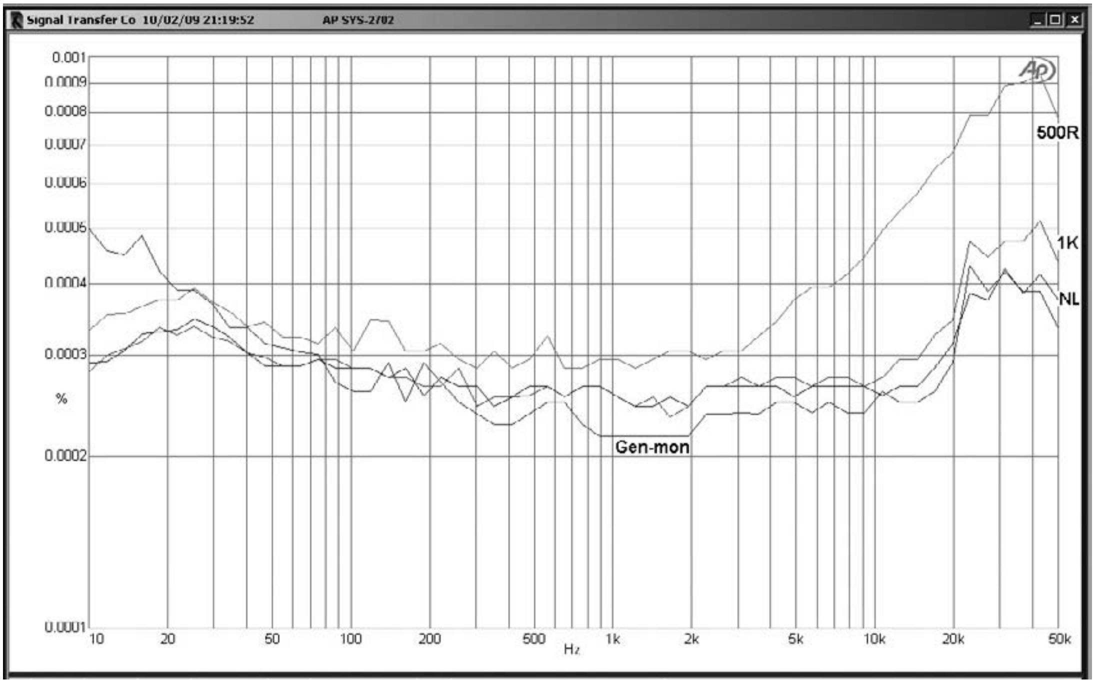


Figure 13.16: The LM4562 in shunt-feedback mode, with 1 kΩ, 2k2 feedback resistors giving a gain of 2.2x. Shown for no load (NL) and 1 kΩ, 500 Ω loads. Note the vertical scale ends at 0.001% this time. Output level is 10 Vrms. ±18 V supply rails.

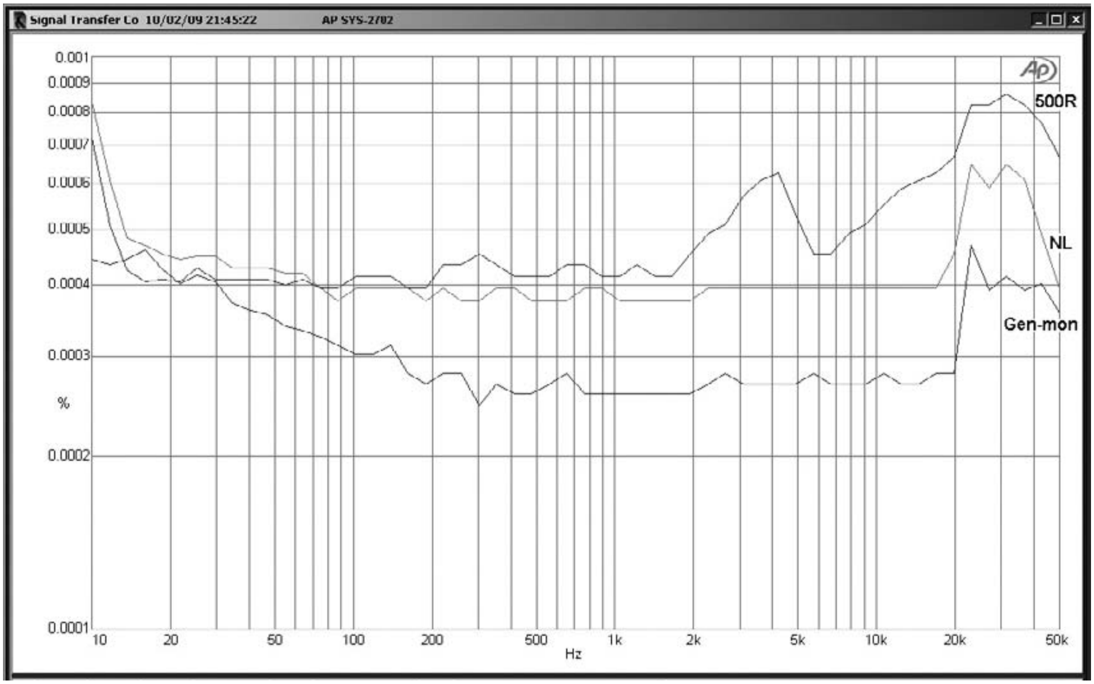


Figure 13.17: The LM4562 in series feedback mode, with 1 kΩ, 2k2 feedback resistors giving a gain of 3.2x. No load (NL) and 500 Ω load. 10 Vrms output. ±18 V supply rails.

a very low 0.001%, and the plots look a bit jagged because of noise. The no-load trace is barely distinguishable from the AP SYS-2702 output, and even with a heavy 500 Ω load driven at 10 Vrms there is only a very small amount of extra THD, reaching 0.0007% at 20 kHz. Compare this with the 5532 performance in Figure 13.7, where loading brings up the distortion in the flat region below 10 kHz to 0.0008%, and gives a THD of 0.0020% at 20 kHz with a 500 Ω load, as opposed to only 0.0007% for the LM4562.

Figure 13.17 shows the LM4562 working at a gain of 3.2x in series feedback mode, both modes having a noise gain of 3.2 times. The extra distortion from the 500 Ω loading is very low.

13.14 Common-Mode Distortion in the LM4562

For Figures 13.16 and 13.17 the feedback resistances were 2k2 and 1 kΩ, so the minimum source resistance presented to the inverting input is 688 Ω. In Figure 13.18 extra source resistances were then put in series with the input path (as was done with the 5532 in the section above on common-mode distortion), and this revealed a remarkable property of the LM4562—with moderate levels of CM voltage it is much more resistant to common-mode distortion than the 5532. At 10 Vrms and 10 kHz, with a 10 kΩ source resistance the 5532

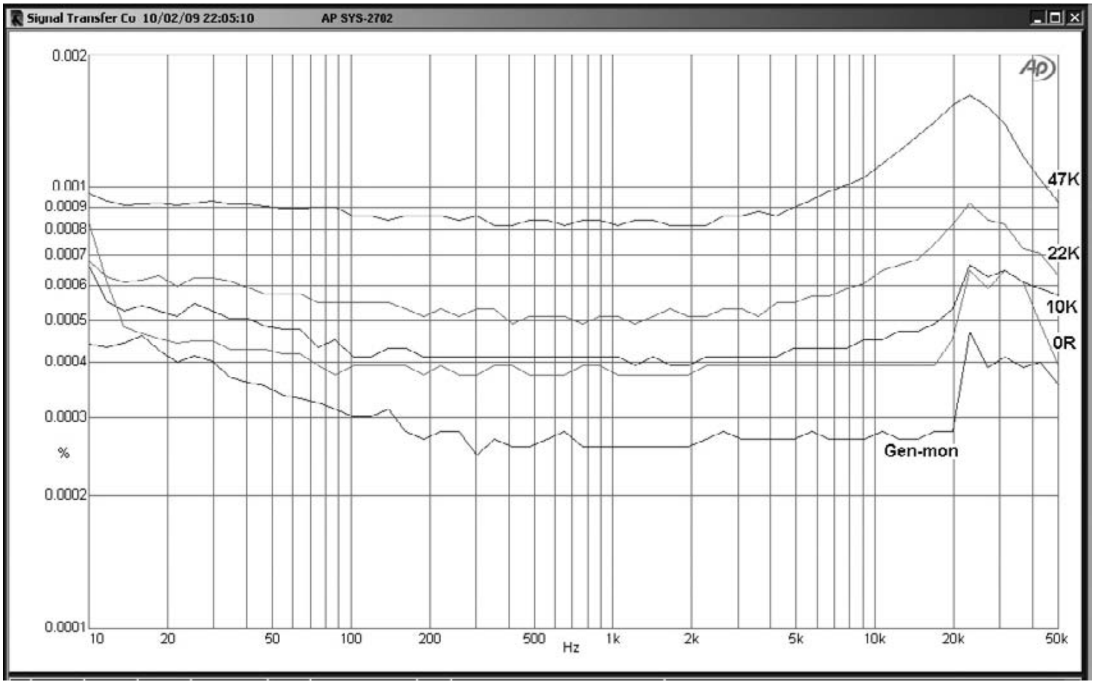


Figure 13.18: The LM4562 in series feedback mode, gain 3.2x, with varying extra source resistance in the input path. The extra distortion is much lower than for the 5532. 10 Vrms out, ±18 V supply rails.

generates 0.0014% THD (see Figure 13.9), but the LM4562 gives only 0.00046% under the same conditions. I strongly suspect that the LM4562 has a more sophisticated input stage than the 5532, probably incorporating cascoding to minimise the effects of common-mode voltages. Note that only the rising curves to the right represent actual distortion. The raised levels of the horizontal traces at the LF end are due to Johnson noise from the added series resistances, plus opamp current noise flowing in them.

As we saw with the 5532, the voltage-follower configuration is the most demanding test for CM distortion, because the CM voltage is at a maximum. Here the LM4562 does not work quite so well. At 10 kHz the distortion with a 10 KΩ source resistance has leapt up from 0.00046% in Figure 13.18 to 0.0037% in Figure 13.19, which allowing for the noise component in the former must be an increase of about ten times. This is a surprising and unwelcome result, because it means that despite its much greater cost the performance of the LM4562 in the voltage-follower configuration is no better than that of the 5532; compare Figure 13.10 above.

The reason for this rapid increase in distortion with CM voltage appear to be a non-linearity mechanism that is activated when the CM voltage exceeds 4 Vrms. This is illustrated in

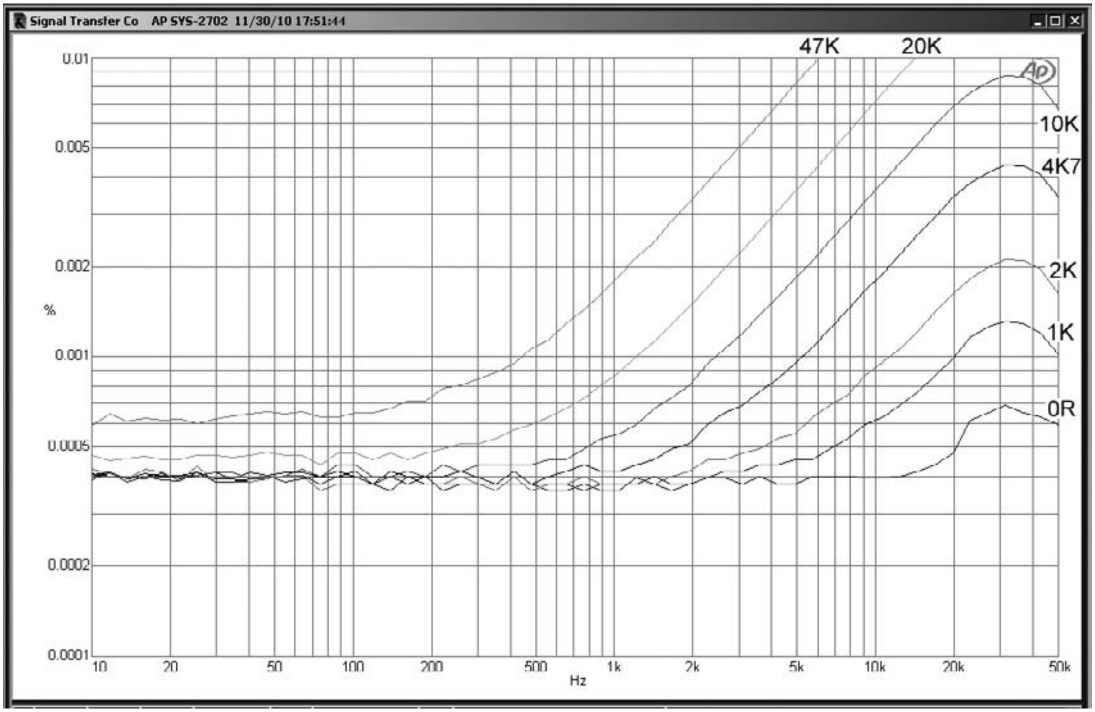


Figure 13.19: The LM4562 in voltage-follower mode, with varying extra source resistance in the input path. CM distortion is much higher than for the series-feedback amplifier in Figure 13.16. 10 Vrms out, ±18 V supply rails.

Figure 13.20, which shows distortion against level for the voltage-follower at 10 kHz, with different source resistances. The left side of the plot shows only noise decreasing relatively as the test level increases; the steps in the lowest trace are measurement artefacts. Since the CM voltage for the series feedback configuration is only about 3 Vrms, the non-linearity mechanism is not activated and CM distortion in that case is very low.

The conclusion has to be once more that if you are using voltage-followers and want low distortion it is well worthwhile to expend some time and trouble on getting the source impedances as low as possible.

It has taken an unbelievably long time—nearly thirty years—for a better audio opamp than the 5532 to come along, but at last it has happened. The LM4562 is superior in just about every parameter, apart from having more than twice as much current noise and no better CM distortion in voltage-follower use. These issues should not be serious problems if the low-impedance design philosophy is followed. At present the LM4562 has a much higher price; hopefully that will change, but it may take a very long time, based on the

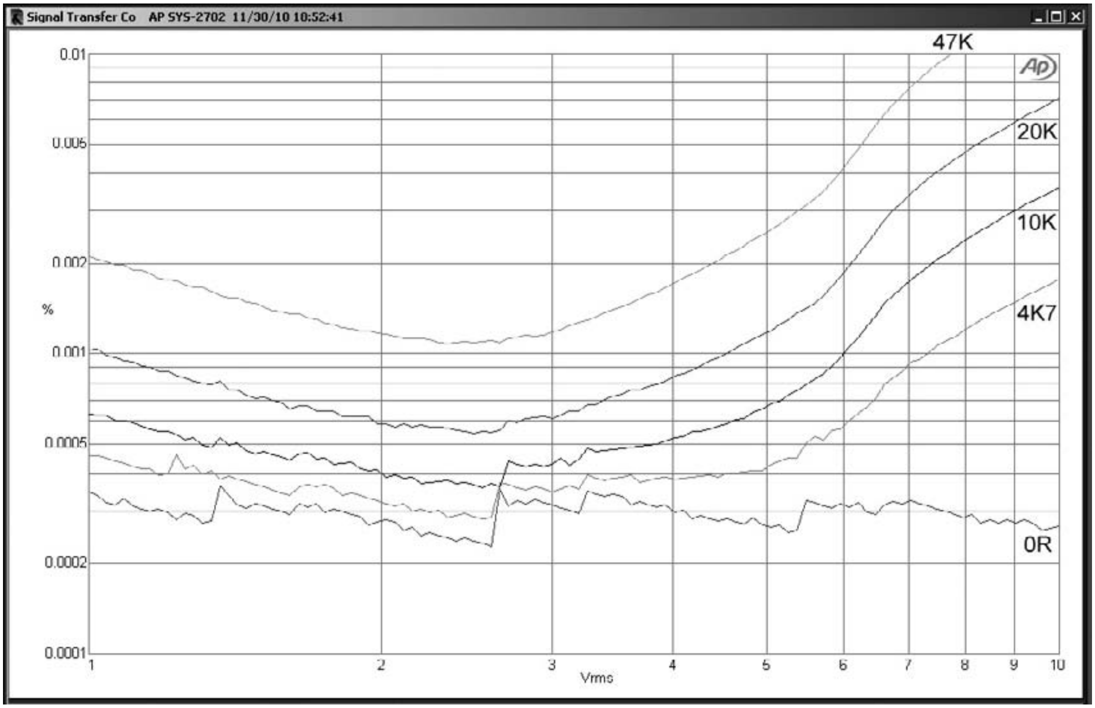


Figure 13.20: The LM4562 in voltage-follower mode, showing CM distortion versus signal voltage with varying extra source resistance. The frequency is 10 kHz. A non-linearity is kicking in at about 4 Vrms. 1 to 10 Vrms out, ±18 V supply rails.

price history of other opamps. It has already begun to appear in hi-fi equipment, one example being the Benchmark DAC1 HDR, a combined DAC and preamplifier.

13.15 The LME49990 Opamp

The LME49990 from National Semiconductor is a single opamp available only as an 8-lead narrow body SOIC surface-mount package. It was released in early 2010. It is part of their "Overture" series, which the data sheet describes as an "ultra-low distortion, low noise, high slew rate operational amplifier series optimized and fully specified for high performance, high fidelity applications," and from my measurements on the LME49990 I'll go along with that. The Overture series also includes the LME49880, which is a dual JFET-input opamp. The LM49710 is another BJT opamp with very low noise and distortion specs but for unknowable reasons it does not appear to be part of the Overture series.

Figure 13.21 shows the distortion performance in the shunt-feedback configuration, to prevent any common-mode distortion issues. The input and feedback resistors are 1 kΩ and 2k2, giving a gain of 2.2 times (and a noise gain of 3.2 times, as for the series version of this test). The

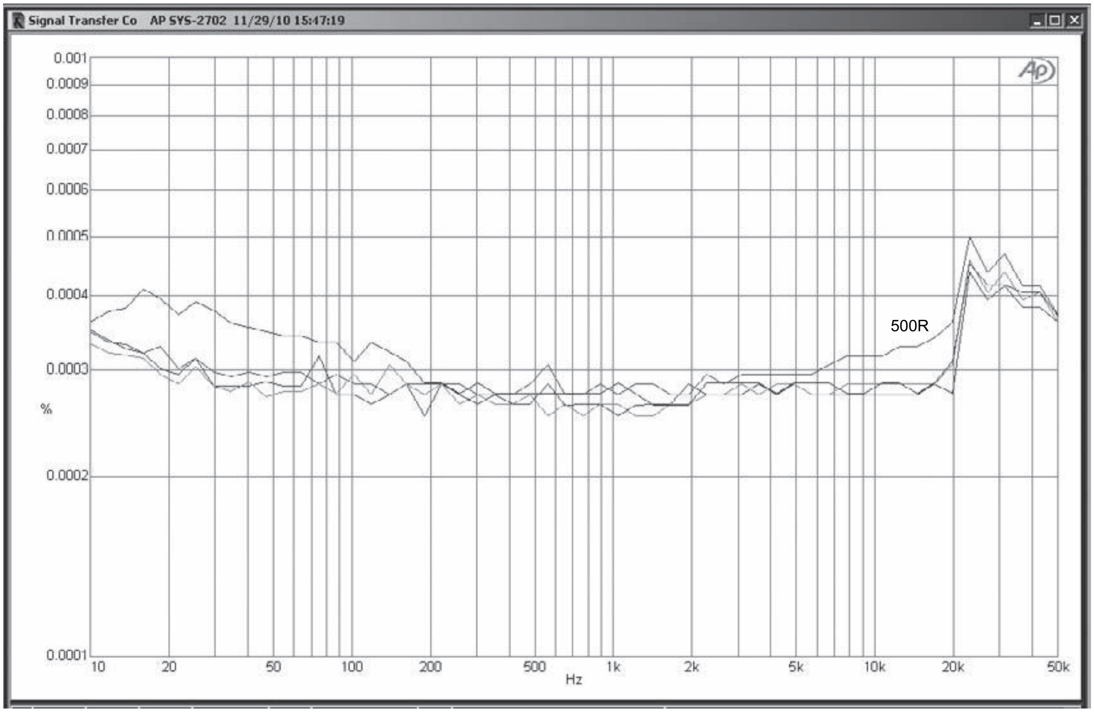


Figure 13.21: The LME49990 in shunt-feedback mode, with a 1 kΩ input resistor and a 2k2 feedback resistor giving a gain of 2.2x. Shown for no load (NL) and 1 kΩ, 500 Ω loads. The generator output is also plotted. Note the top of the vertical scale is at only 0.001%. The output level is 9 Vrms. with ±17 V supply rails.

traces are for no load (apart from the feedback resistor) and 1 kΩ, and 500 Ω loads at an output of 9 Vrms, and also the AP 2722 output for reference. As you can see, these traces are pretty much piled up on top of each other, with no distortion visible on the residual except for a small amount between 10 kHz and 20 kHz with the 500 Ω load; clearly the LME49990 is very good at driving 500 Ω loads. The step at 20 kHz is an artefact of the Audio Precision SYS-2702 measuring system.

If we compare this plot with Figure 13.5 above, we can see that the LME49990 is somewhat superior to the 5532, but since we are down in the noise floor most of the time the differences are not great.

In the series configuration, with 1 kΩ and 2k2 feedback resistors giving a gain of 3.2 times, and a significant common-mode voltage of 3 Vrms, things are not quite so linear; there is now clearly detectable distortion at high frequencies, as shown in Figure 13.22. Even so, the distortion is less than half that of a 5532 in the same situation—compare Figure 13.6 above.

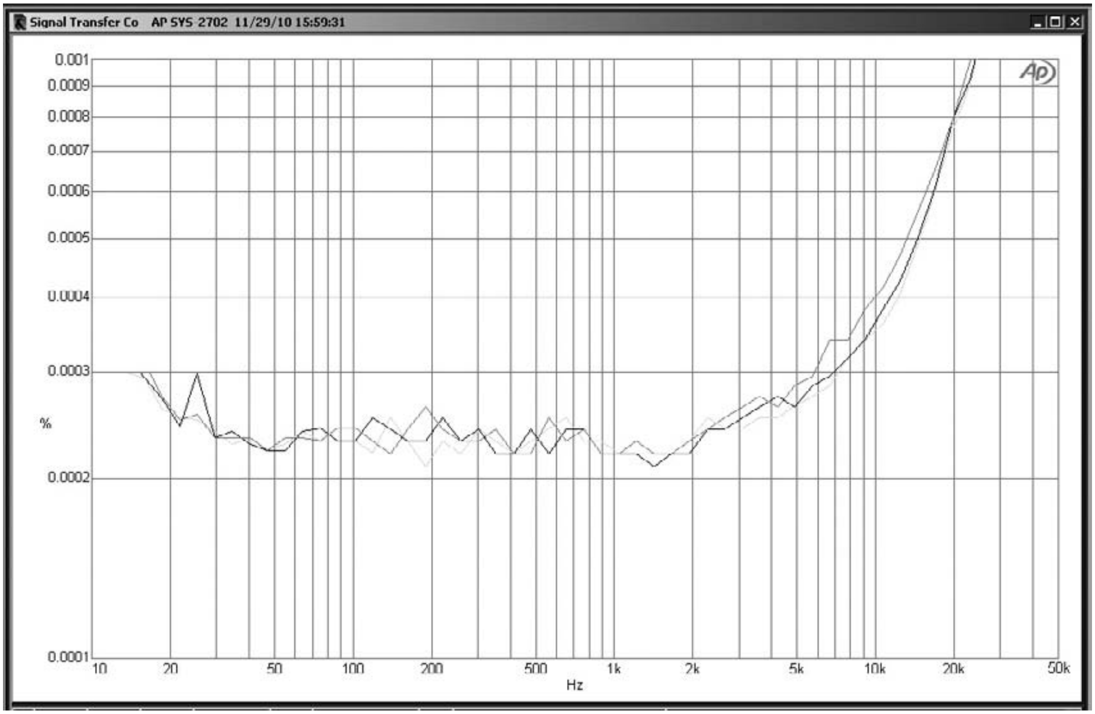


Figure 13.22: The LME49990 in series feedback mode, with 1 kΩ, 2k2 feedback resistors giving a gain of 3.2x. No load, 1 kΩ, and 500 Ω loads. 9 Vrms output. ±17 V supply rails.

13.16 Common-Mode Distortion in the LME49990

It looks as though common-mode distortion may be more of an issue with the LME49990 than it was with the LM4562. As we saw earlier, BJT input opamps do not show common-mode distortion unless the configuration has both a significant common-mode voltage *and* a significant source impedance. If we repeat the series feedback test with no external load, but increasing source resistance, we get Figure 13.23, where as usual more source resistance means more high-frequency distortion. The exception is for a 1 kΩ source resistance, where the distortion actually decreases; this is because the 1 kΩ is partially cancelling the 688 Ω source resistance of the feedback network. We saw exactly the same effect with the 5532; see Figure 13.10 above. The horizontal low-frequency parts of the traces are raised by the Johnson noise from the added source resistances and also by the opamp current noise flowing in those resistances. There is no distortion visible in this region.

The voltage-follower configuration has the worst working conditions for CM distortion because since there is no amplification, the CM voltage on the inputs is as large as the output

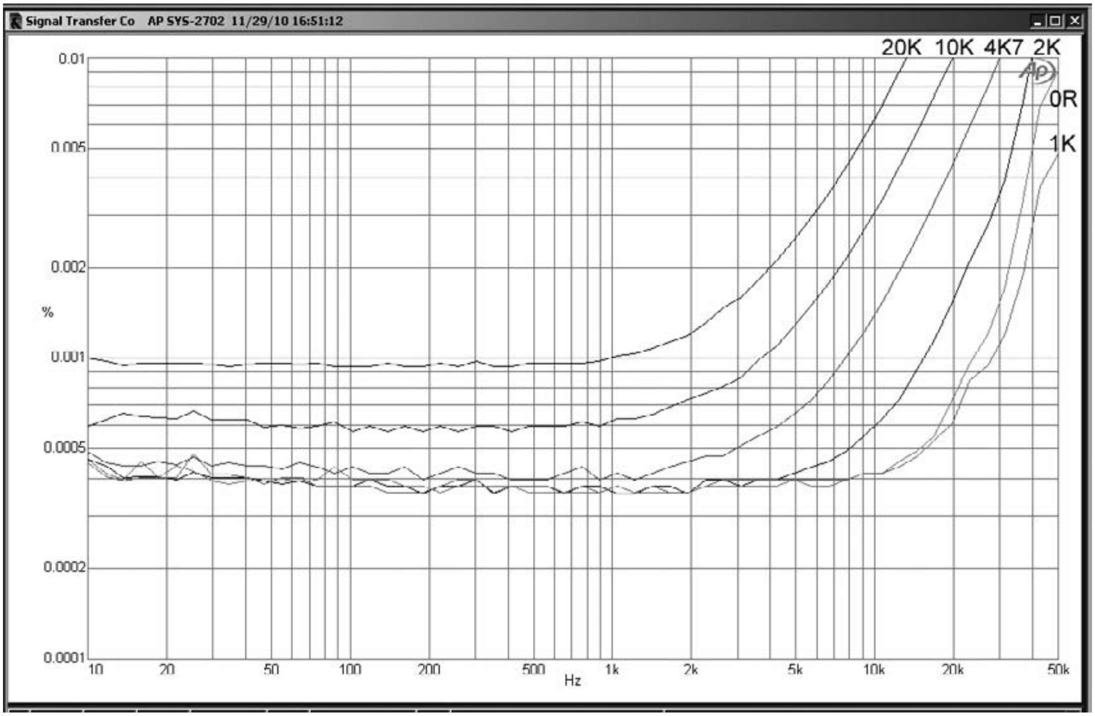


Figure 13.23: The LME49990 in series feedback mode, gain 3.2x, with varying extra source resistance in the input path; note that a 1 kΩ source resistance actually gives less distortion than none. The CM distortion is lower than for the 5532, but higher than for the LM4562. 10 Vrms out, ±18 V supply rails.

voltage. Figure 13.24 shows that in this case the CM distortion is much worse. With a 2 KΩ source resistance the THD at 10 kHz has increased from 0.0015% to 0.0042%, and all the other figures show a similar increase. The conclusion has to be that if you are working with a large CM voltage and a significant source resistance, the LME49990 is not the best choice, and either the 5532, the LM4562, or the AD797 will give considerably lower distortion.

The SOIC-only format of the LME4990 is not helpful for experimentation or home construction. I soldered my samples to an 8-pin DIL adaptor and plugged that into a prototype board without any problems. The LME49990 samples were kindly supplied by Don Morrison.

13.17 The AD797 Opamp

The AD797 (Analog Devices) is a single opamp with very low voltage noise and distortion. It appears to have been developed primarily for the cost-no-object application of submarine sonar, but it works very effectively with normal audio- if you can afford to use it. The cost is something like twenty times that of a 5532. No dual version is available, so the cost ratio per opamp section

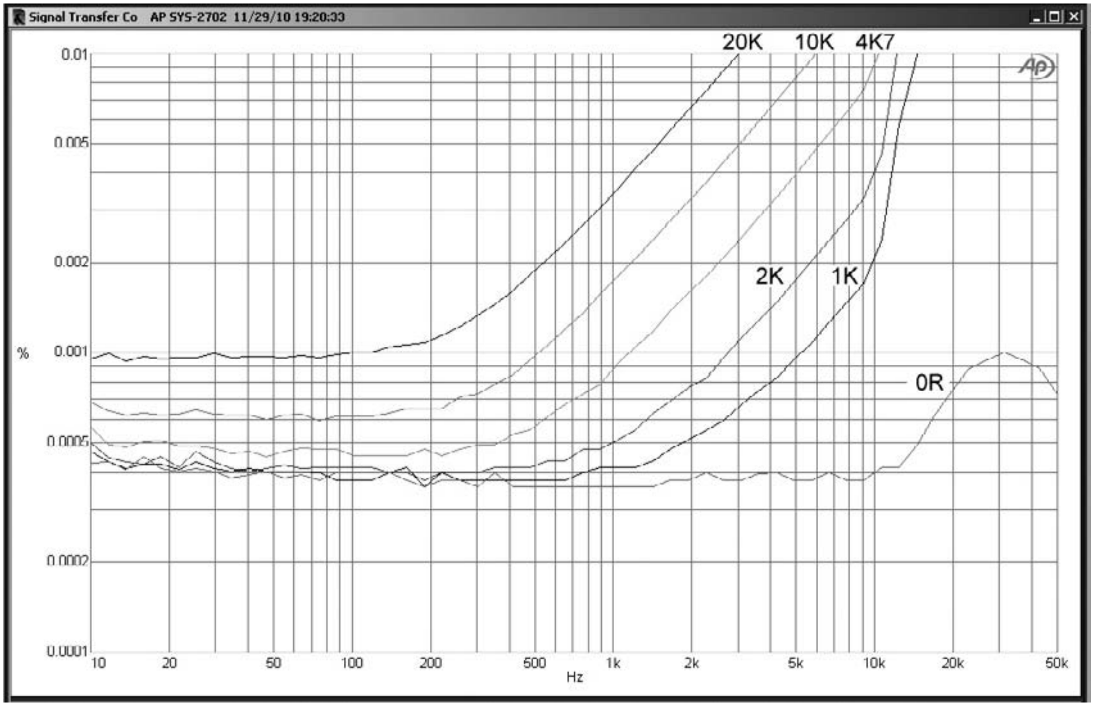


Figure 13.24: LME49990 distortion in a voltage-follower circuit with a selection of source resistances; test level 10 Vrms, supply ±18 V.

is a hefty forty times. The AD797 incorporates an ingenious error-correction feature for internal distortion cancellation, the operation of which is described on the manufacturer's data sheet. The measurements presented below seem to show that it works effectively.

The AD797 is a remarkably quiet device in terms of voltage noise, but current noise is correspondingly high due to the large standing currents in the input devices. Early versions appeared to be quite difficult to stabilise at HF; the current product seems to be easier to handle but still a bit harder to stabilise than the 5532 or the LM4562. Possibly there has been a design tweak, or on the other hand my impression may be wholly mistaken.

It has taken quite a long time for the AD797 to make its way into audio circuitry, possibly because of the discouraging early reports on instability, but more likely because of the high cost. At the time of writing (2010) they are showing up in commercial audio equipment, a contemporary example being the Morrison E.L.A.D. line-level preamplifier, and they are also appearing in DIY designs. Audio Precision use this opamp in their 2700 series state-of-the-art distortion analysers.

Figure 13.25 shows the AD797 in shunt-feedback mode, to assess how it copes with output loading; as you can see, the effect is very small, with a very modest rise above 10 kHz with

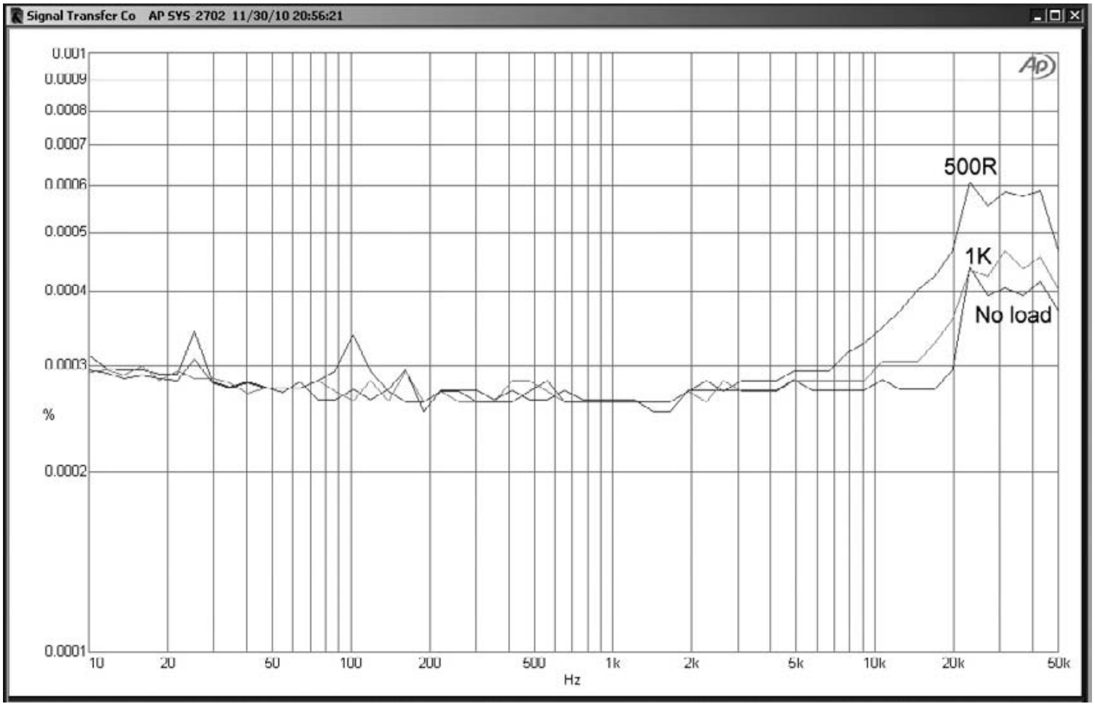


Figure 13.25: The AD797 in shunt-feedback mode, with 1 kΩ, 2k2 feedback resistors giving a gain of 2.2x. No load and 1 kΩ, 500 Ω loads. Note top of vertical scale is at 0.001%. Output level is 10 Vrms. ±18 V supply rails.

the 500 Ω load. Below 10 kHz the loading has no detectable effect at all, which makes it a better opamp than the 5532—see Figure 13.5. The step at 20 kHz is a measurement artefact. When the loads were applied a 33 pF capacitor had to be put across the 2k2 feedback resistor to obtain stability—this was not required with either the 5532, the LM49990, or the LM4562, and shows that the AD797 does require a little more care to get dependable HF stability.

Figure 13.26 shows the AD797 in series-feedback mode, with no output loading apart from its own 2k2 feedback resistor, to test its CM distortion with moderate CM voltages, here 3.1 Vrms. The series feedback test also needed 33 pF across the feedback resistor to ensure HF stability.

13.18 Common-Mode Distortion in the AD797

The AD797 is next tested in voltage-follower mode with a 10 Vrms signal level and various source impedances. The traces in Figure 13.27 do not fall out very neatly because there appear to be some sort of cancellation effect going on, but from the difference between the

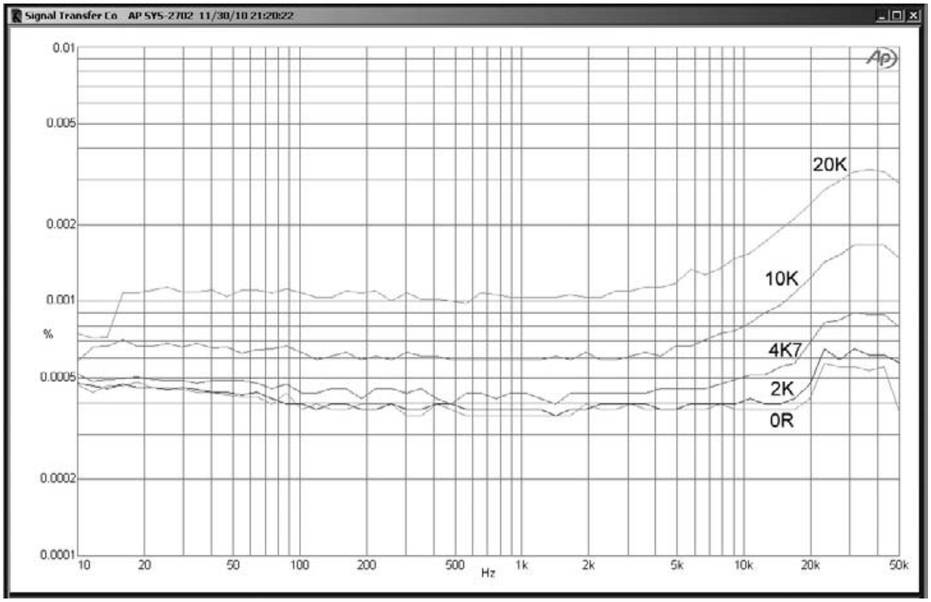


Figure 13.26: AD797 with series feedback THD with no external load, at 10 Vrms. Gain = 3.2x, ±18 V supply rails.

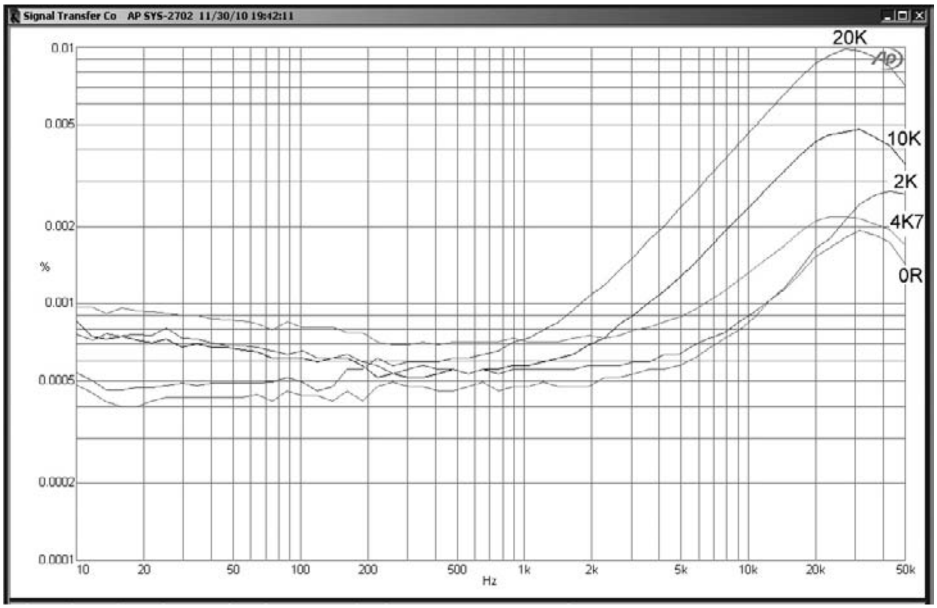


Figure 13.27: AD797 voltage-follower THD, at 10 Vrms. Output is virtually indistinguishable from input.

Rs = 4k7 and Rs = 10 kΩ traces it is abundantly clear that, as usual, source impedances need to be kept low when there is a large CM signal. The results are significantly better than the 5532 in voltage-follower mode–compare above.

13.19 The OP27 Opamp

The OP27 from Analog Devices is a bipolar-input single opamp primarily designed for low noise and DC precision. It was not intended for audio use, but in spite of this it is frequently recommended for such applications as RIAA and tape head preamps. This is unfortunate, because while at first sight it appears that the OP27 is quieter than the 5534/5532, as the e_n is 3.2 nV/rtHz compared with 4 nV/rtHz for the 5532, in practice it is usually slightly noisier. This is because the OP27 is in fact optimised for DC characteristics, and so has input bias-current cancellation circuitry that generates common-mode noise. When the impedances on the two inputs are very different the CM noise does not cancel, and this can degrade the overall noise performance significantly, and certainly to the point where the OP27 is noisier than a 5532.

For a bipolar input opamp, there appears to be a high level of common-mode input distortion, enough to bury the output distortion caused by loading; see Figures 13.28 and 13.29. It is likely that this too is related to the bias-cancellation circuitry, as it does not occur in the 5532.

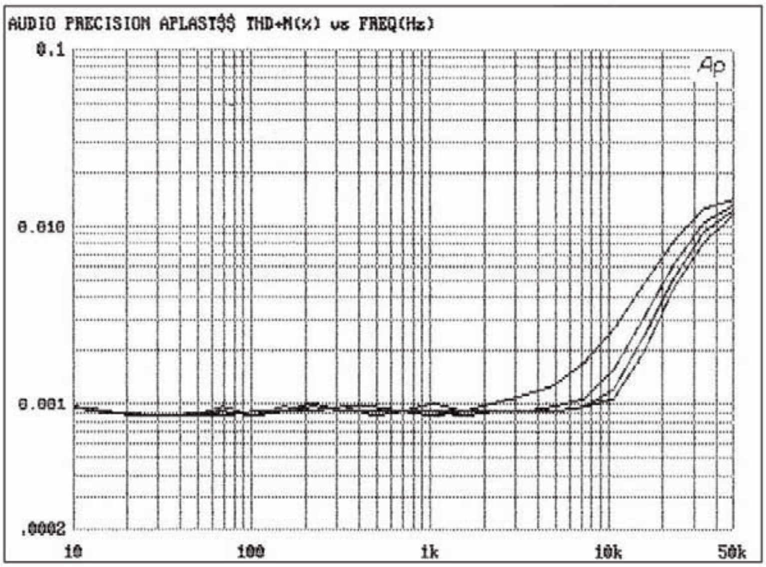


Figure 13.28: OP27 THD in shunt-feedback mode with varying loads. This opamp accepts even heavy (1 kΩ) loading relatively gracefully.

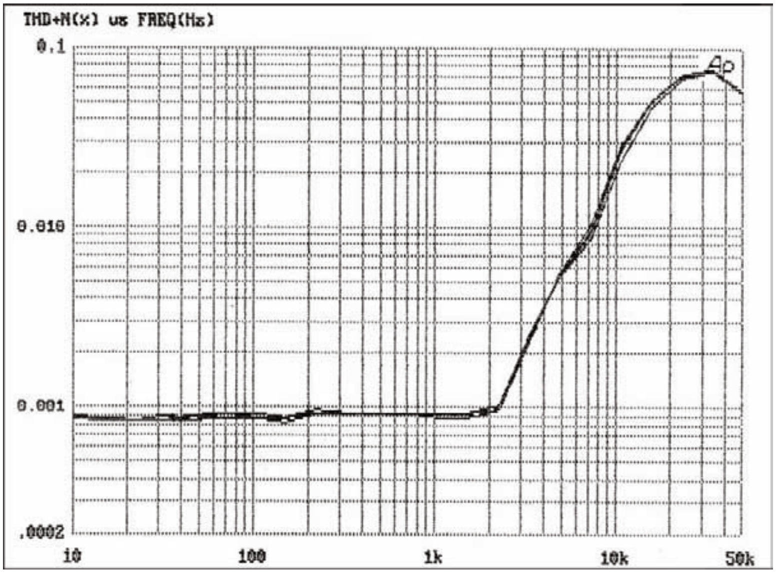


Figure 13.29: OP27 THD in series feedback mode. The common-mode input distortion completely obscures the output distortion.

The maximum slew rate is low compared with other opamps, being typically 2.8V/us. However, this is not the problem it may appear. This slew rate would allow a maximum amplitude at 20 kHz of 16 Vrms, if the supply rails permitted it. I have never encountered any particular difficulties with decoupling or stability of the OP27.

13.20 Opamp Selection

In audio work, the 5532 is pre-eminent. It is found in almost every mixing console, and in a large number of preamplifiers. Distortion is very low, even when driving 600 Ω loads. Noise is very low, and the balance of voltage and current noise in the input stage is well-matched to low-impedance audio circuitry. Large-quantity production has brought the price down to a point where a powerful reason is required to pick any other device.

The 5532 is not, however, perfect. It suffers common-mode distortion. It has high bias and offset currents at the inputs, as an inevitable result of using a bipolar input stage (for low noise) without any sort of bias-cancellation circuitry.

With tiresome inevitability, the very popularity and excellent technical performance of the 5532 has led to it being savagely criticised by Subjectivists who have contrived to convince themselves that they can tell opamps apart by listening to music played through them. This always draws a hollow laugh from me, as there is probably no music on the planet that has not passed through a hundred or more 5532s on its way to the consumer.

In some cases, such as variable frequency crossover filters, bipolar-style bias currents are a nuisance because keeping them out of pots to prevent scratching noises requires the addition of blocking capacitors. JFET-input opamps have negligible bias currents and so do not need these extra components, but there is no obviously superior device that is the equivalent of the 5532 or 5534. The TL072 has been used in the EQ sections of low-cost preamplifiers and mixers application for this purpose for many years, but its HF linearity is poor by modern standards and distortion across the band deteriorates badly as output loading increases. It is also more subject to common-mode distortion than bipolar types. There are of course more modern JFET opamps such as the OPA2134, but their general linearity is not much better and they also suffer from common-mode distortion.

If you are looking for something better than the 5532, the newer opamps (LM4562, LME49990, AD797) have definite advantages in noise performance. The LM4562 has almost 6 dB less voltage noise, while both the LME49990 and the AD797 are nearly 15 dB quieter, given sufficiently low impedance levels. Both have better load-driving characteristics than the 5532. The LME49990 has really excellent load-driving capabilities but its CM distortion is disappointingly high. The 3 dB noise advantage of the 5534A over the 5532 should not be forgotten.

To summarise some related design points:

- The trade-off between noise and distortion-reducing circuit impedances will reduce noise relatively slowly. The effect of opamp current noise is proportional to impedance, but Johnson noise is proportional to square-root of impedance.
- Keep circuit impedances as low as possible to minimise the effect of current noise flowing through them, and their Johnson noise. This will also reduce common-mode distortion and capacitive crosstalk.
- If distortion is more important than noise, and you don't mind the phase inversion, use the shunt configuration.
- If noise is more important than distortion, or if a phase inversion is undesirable, use the series configuration. In this case keeping circuit impedances low has an extra importance as it minimises common-mode distortion arising in the feedback network.
- To minimise distortion, keep the output loading as light as possible. This of course runs counter to the need to keep feedback resistances as low as possible.
- If the loading on a 5532/5534 output is heavy, consider using output biasing to reduce distortion. If the improvement is sufficient, this will be cheaper than switching to a more advanced opamp.

References

[1] S. Groner, Operational Amplifier Distortion. http://www.sg-acoustics.ch/analogue_audio/index.html, 2009 (accessed 12.10).
[2] W. Jung (Ed.), Op-amp Applications Handbook, Chap. 5, Newnes, Boston, MA, p. 399.
[3] D. Self, Audio Power Amplifier Design Handbook, fifth ed., Focal Press, Boston, MA, 2009, pp. 96–104, ISBN 978-0-240-52162-6.
[4] D. Self, Audio Power Amplifier Design Handbook, fifth ed., Focal Press, Boston, MA, 2009, pp. 328–343, ISBN 978-0-240-52162-6.
[5] D. Self, Audio Power Amplifier Design Handbook, fifth ed., Focal Press, Boston, MA, 2009, p. 300, ISBN 978-0-240-52162-6.

CHAPTER 14

Active Crossover System Design

So far we have looked at the design of the functional blocks that go together to make an active crossover. It is now time to look at putting them together into a complete system. This involves not only selecting the optimal types of circuit block to perform the required crossover functions, but also considerations about what order in which to place them in the signal path, and what nominal signal levels to operate them at. This chapter also covers minor crossover functions such as output level controls and channel muting switches.

14.1 Crossover Features

The actual crossover functions of filtering and equalisation make up the main part of a crossover's circuitry, but other functions are required to make a device that is practically useful. Not all crossovers have all of them. They are as follows.

14.1.1 Input Level Controls

These allow varying incoming levels to be normalised to the nominal internal level of the crossover. They frequently have a deliberately restricted gain range, typically not going down to zero; they are not intended to be used as volume controls. As described later in this chapter, better signal-to-noise performance can be obtained by using active gain controls that actually vary the gain of an amplifier stage, as opposed to combinations of fixed-gain amplifiers and variable passive attenuation. This often means designing a variable-gain balanced input stage, which requires some subtlety if a good CMRR is to be obtained at all gain settings. This issue is thoroughly explored in Chapter 16 on line inputs.

14.1.2 Subsonic Filters

High-pass filters used to stop subsonic signals and so protect loudspeakers. They are typically of the second-order or third-order Butterworth (maximally flat) configuration, rolling off at rates of 12 dB/octave and 18 dB/octave respectively. Fourth order 24 dB/octave filters are less common, presumably due to worries about the possible audibility of rapid phase changes at the very bottom of the audio spectrum; when they are used the cutoff frequency is lower. There seems to be a general consensus that a third-order Butterworth filter with a cutoff frequency around 20 to 25 Hz is the correct approach, though at least one

The Design of Active Crossovers

manufacturer has used a fourth-order filter starting at 15 Hz. A third-order Butterworth filter with a −3 dB cutoff at 20 Hz will be −18.6 dB down at 10 Hz, and −36.0 dB down at 5 Hz. The 30 Hz response is down by −0.37 dB, and if you feel that is too much the filter cutoff frequency can be moved down a little.

Subsonic filters are placed as early in the signal path as possible, typically immediately after a balanced input amplifier. This is not so much to avoid the generation of intermodulation distortion (which should be negligible at low frequencies in a decent opamp), but to prevent headroom being eroded by large subsonic signals. Standard Sallen & Key filters are normally used.

14.1.3 Ultrasonic Filters

These are also intended for speaker protection, in the event of HF instability somewhere upstream in the audio system A typical ultrasonic filter would be a second-order Butterworth with a cutoff frequency around 40 kHz; this will be only −0.24 dB at 20 kHz. If you feel that is too much of an intrusion on the audio spectrum, moving the cutoff frequency up to 50 kHz gives a 20 kHz loss of only −0.09 dB. Ultrasonic filters are rarely steeper than second-order, presumably because of a fear that phase changes are more audible at the very top of the audio spectrum than the very bottom. The ultrasonic filter may be placed just after the balanced input amplifier to minimise HF intermodulation distortion in the following circuitry, or alternatively placed only in the HF crossover path, the assumption being that the lowpass filters in the LF and MID paths will very effectively remove any ultrasonic signals, and that it is better to put as little of the audio spectrum as possible through as few stages as possible, to minimise signal degradation. Standard Sallen & Key filters are commonly used.

The combination of a subsonic filter and an ultrasonic filter is often called a bandwidth definition filter. There is much information on how to make them economically in Chapter 8.

14.1.4 Output Level Trims

Output level controls tend also to have a limited range; once again they are not intended to be used as volume controls. Allowing for power amplifier gain variations in a nominally identical set of power amplifiers may require as little as ±1 dB, but it is advisable to provide at least ±3 dB to allow for drive unit variations and ±6 dB is more usual. If the crossover is intended to work with a wide range of power amplifiers of differing sensitivities the range may need to be considerably greater than this. There is more on output level controls later in this chapter.

14.1.5 Output Mute Switches, Output Phase-Reverse Switches

These are also dealt with later in this chapter.

14.1.6 Control Protection

Active crossovers are not some sort of glorified tone-control—they do a quite different job and are meant to be carefully set up to match the power amplifiers and loudspeakers and then left alone. They are not meant to be tweaked, twiddled, or frobbed by every casual passer-by. It is possible to cause serious damage by maladjustment—if the highpass cutoff frequency for the HF loudspeakers is adjusted radically downwards then expensive and gig-cancelling damage is virtually a certainty.

It is therefore common to cover up crossover controls with a panel to prevent tampering; this is sometimes called a "security cover." This might be a substantial piece of clear plastic (the look-but-don't-touch approach) or a solid piece of metal, which is more robust against impact. It is usually fixed with so-called "security screws" but since you can buy sets of drivers for those very easily we're not exactly talking Fort Knox here.

14.2 Features Usually Absent

There are also some features that, while appearing in many kinds of audio equipment, are rarely if ever found in active crossovers. These include the following.

14.2.1 Metering

Active crossovers are not usually fitted with comprehensive level metering, as the assumption is that they will be installed in some secluded spot where a visual display will not be seen. Peak detect or clip-detect indicators can however be useful if there is a possibility of internal clipping. These are dealt with later in this chapter. Signal-present indicators, that illuminate some way below the nominal signal level, (often at −20 dB) can be useful for fault-finding.

14.2.2 Relay Output Muting

Active crossovers are not normally expected to have relay output muting, the function of which is to avoid sending out unpleasant transients at power-up and power-down. Since there are likely to be six outputs, probably balanced, the extra cost of six good-quality two-changeover relays is significant. In sound-reinforcement work putting mute relays on every piece of equipment is very much not favoured as they present one more place for things to go wrong and stop the signal. Thump suppression is normally considered to be the job of the power amplifier output muting relays, which must always be fitted as one of their most important functions is protection of the loudspeakers from DC fault conditions.

Manual mute switches for each output are however often fitted to sound reinforcement crossovers to simplify checking and fault finding.

14.2.3 Switchable Crossover Modes

Crossovers intended for sound reinforcement are often constructed so they can be used in several different modes. Figure 14.1 shows the block diagram of a stereo 2-way crossover that can be switched to act as a mono 3-way crossover, or as a 2-way crossover with a single mono LF output for a subwoofer. Alternatively, a crossover might be switchable between stereo 2-way and mono 3-way crossover, with a wholly separate mono subwoofer output always available.

Figure 14.1 has the mode switches in the normal stereo 2-way crossover position, and 2-way outputs are obtained as shown in the left column of the text in Figure 14.1. If the 2-way/3-way switch is operated then the Right input is not used, and the input to the Right filter block now comes from the highpass output of the Left filter block. If the Left filter block is set to the LF/MID crossover frequency, and the Right filter block set to the MID/HF crossover frequency, 3-way outputs are obtained as shown in the middle column of the text in Figure 14.1. This kind of mode switching requires a very wide range of filter frequency variation, as the filter block must be able to cover both LF/MID and MID/HF crossover points.

For mono-sub operation the 2-way/3-way switch is left in its normal position, and the mono-sub switch operated instead. This causes the Left LF output to be fed with the mono sum of the two LF outputs from the filter blocks. The Right LF output is not used. In this case the Left and Right filter blocks are set to the same frequency—the crossover point

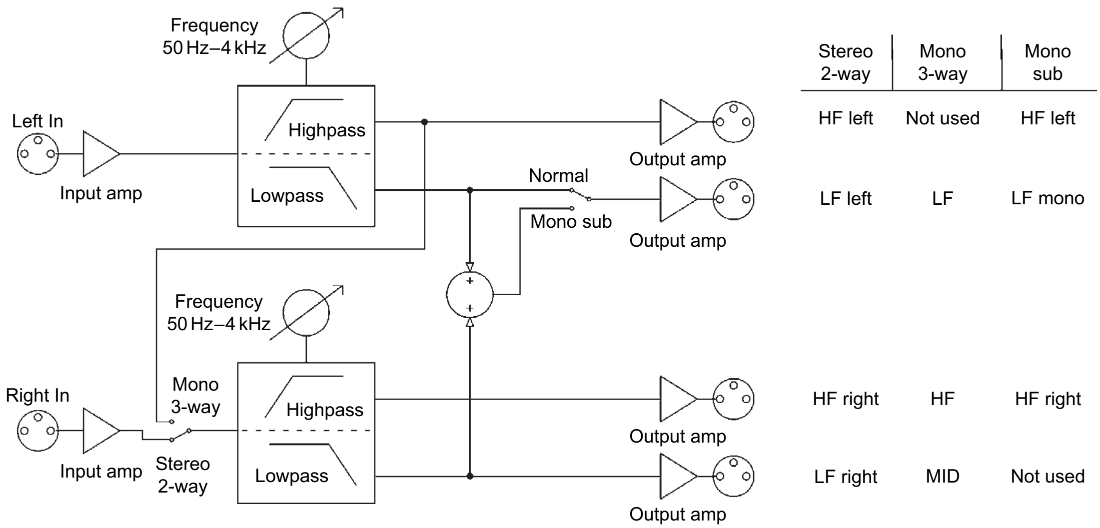


Figure 14.1: Mode-switching in an electronic crossover: it can be used as a stereo 2-way crossover, a mono 3-way crossover, or a 2-way crossover with a mono subwoofer output.

between the two HF outputs and the single LF output. The summing function has to be implemented carefully if crosstalk between the two channels feeding it is to be avoided.

Each filter block is shown with a single frequency control to emphasise that the cutoff frequencies of the highpass and lowpass sections move together; this is usually implemented with a state-variable filter (SVF) that simultaneously gives highpass and lowpass outputs.

More complex mode switching schemes are possible. A stereo 3-way crossover could be switchable to work as a mono 4-way or 5-way crossover. This is all very ingenious, but it does require a lot of complicated switching and a very clear head when you are setting up all those crossover frequencies.

Manufacturers often warn that mode switches should not be operated while the whole system is active, stating that this can lead to damaging transients; presumably they are worried that you might get an excessive level you don't expect, rather than concerned about minor DC clicks.

14.3 Noise, Headroom, and Internal Levels

The choice of the internal signal level in a piece of audio equipment is a serious matter, as it controls both the signal-to-noise ratio and the headroom available before clipping occurs. A vital step in any design is the determination of the optimal signal level at each point in the circuit; there is no reason why you have to stick to the same level in every section. Obviously a real signal, as opposed to a test sinewave, continuously varies in amplitude, and the signal level chosen is purely a nominal level. One must steer a course between two evils:

1. If the signal level is too low, it will be contaminated unduly by noise. The absolute level of noise in a circuit is not of great significance in itself—what counts is how much greater the signal is than the noise—in other words the signal to noise ratio.
2. If the signal level is too high there is a risk it will clip and introduce severe distortion.

You will note that the first evil is a certainty, while the second is a statistical risk.

The wider the gap between the noise level and the clipping level, the greater is the dynamic range. If the best possible signal-to-noise is required for hi-fi use, then the internal level should be high, and if there is an unexpected overload it's not the end of the world. In sound-reinforcement applications it will often be preferable to use a lower internal level, sacrificing some noise performance to reduce the risk of clipping. Heavy clipping, which in an active-crossover system can be surprisingly hard to detect by ear, is likely to imperil HF speaker units, though not for the frequently quoted but quite untrue reason that extra harmonics are generated; the real problem is the general rise in level [1]. Later in this chapter we will look at some ways of detecting and indicating clipping.

The internal levels chosen depends on the purpose of the equipment. For example, suppose you are designing a mixing console. If it is intended for studio recording you only have to get the performance right once, but you do have to get it exactly right, that is, with the best possible signal-to-noise ratio, so the internal level is relatively high, very often −2 dBu (615 mV rms), which gives a headroom of about 24 dB. If it is intended for broadcast work to air you only have one chance to get it right, and a mildly impaired signal-to-noise ratio is much acceptable than a crunching overload, so the internal levels need to be significantly lower. The Neve 51 Series broadcast consoles used −16 dBu (123 mV rms), which gives a much increased headroom of 38 dB. Apart from this specialised application, general audio equipment might be expected to have a nominal internal level in the range −6 dBu (388 mV rms) to 0 dBu (775 mV rms) with −2 dBu probably being the most popular choice.

If you have a given dynamic range and you're not happy with it, you can either increase the maximum signal level or lower the noise floor. The maximum signal levels in opamp-based equipment are set by the voltage capabilities of the opamps used, and this usually means a maximum signal level of about 10 Vrms or +22 dBu. Discrete transistor technology removes this absolute limit on supply voltage, and allows the voltage swing to be at least doubled before the supply rail voltages get inconveniently high. For example, +/−40 V rails are quite practical for small-signal transistors and permit a theoretical voltage swing of 28 Vrms or +31 dBu. However, in view of the complications of designing your own discrete circuitry, and the greater space and power it requires, the extra 9 dB of headroom is bought at a high price. You will need a lot more PCB area, and of course the knowledge of how to design discrete transistor stages. My book, *Small Signal Audio Design*, will be of help with the latter [2]. If you are using high signal levels like 28 Vrms you may need to consider what will happen if they are applied to opamp circuitry.

A current example of a crossover with all discrete circuitry in the signal path is the Bryston 10 B [3].

14.4 Circuit Noise and Low-Impedance Design

Increasing the dynamic range by reducing the noise levels in the circuitry is more practical (and in general a good deal cheaper) than increasing the nominal level, but there are some quite restrictive limits on how much you can do this. Adopting Low-Impedance Design—in other words using the lowest resistor values you can without creating extra distortion by overloading the opamps—will reduce the Johnson noise the resistors generate. It also makes the circuit more immune to capacitive crosstalk and interference pick-up. However, Johnson noise is proportional to the square-root of the resistance, and so moving from 10 kΩ to 1 kΩ will only reduce the noise by 10 dB (√10 times) rather than 20 dB (10 times). A reduction of 10 dB is nevertheless very well worth having. Things get more difficult if you want to reduce the impedance levels further, as opamp distortion will start to increase due to the heavier loading.

This can be countered by using opamps in parallel to increase the drive capability, assuming you're not designing to the absolute minimum cost. Two opamps working together allow the circuit impedances to be halved, giving us another 3 dB improvement, while four opamps allow them to be halved again, giving 6 dB less noise. This is going to be about as far as it is economical to go unless you're designing really gold-plated gear, so we have a possible Johnson noise improvement from Low-Impedance Design of 16 dB.

Johnson noise is, however, only one component of the circuit noise, the other two important contributions coming from the voltage noise and the current noise of the active devices. Reducing the circuit impedances reduces the effect of current noise—proportionally this time, as the current noise only manifests itself when it causes a voltage drop across an impedance. Voltage noise is a tougher proposition to reduce, the options being a) shell out for quieter and more expensive active devices; or b) make use of opamps in parallel again. If two opamp stages of the same gain are connected together by low-value resistors (say 10 Ω) then at their junction you get the average of the two outputs, so the signal level is unchanged, but the noise drops by 3 dB ($1/\sqrt{2}$) as the two noise components are uncorrelated and so partially cancel. Four opamp stages give a 6 dB improvement. This technique obviously goes extremely well with using opamps in parallel to allow circuit impedance reduction, and can make for some very neat and effective circuitry.

14.5 Using Raised Internal Levels

When setting the internal levels of an active crossover a great deal depends on the way that it is going to built into the overall system. If the crossover is running directly into a power amplifier with no intermediate level control, it can be guaranteed that the output of the power amplifier will clip long before the crossover outputs, as even the most insensitive amplifiers are unlikely to need more than +8 dBu (2 Vrms) to drive them fully. It therefore occurred to me that it would be quite safe to raise the crossover internal level to 2 Vrms, which would theoretically give a 10 dB better signal-to-noise than the often-used −2 dBu level. That is a significant improvement.

As an example, look at Figure 14.2a, where the crossover is operating with an internal level of 0 dBu throughout, which is also the input required by the power amp. Only one of the paths through the crossover is shown. There is a unity-gain input amplifier A1 which has an equivalent input noise level of −100 dBu. (The input signal is assumed to be completely noise-free; keeping down the noise level from the preamplifier is someone else's problem.) The filters, etc. of the crossover have a noise level of −85 dBu, and when this is summed with the −100 dBu from A1 we get −84.9 dBu at the crossover output; the noise from A1 makes only a tiny contribution. The signal is then passed directly to the power amplifier, and our signal-to-noise ratio is 84.9 dB.

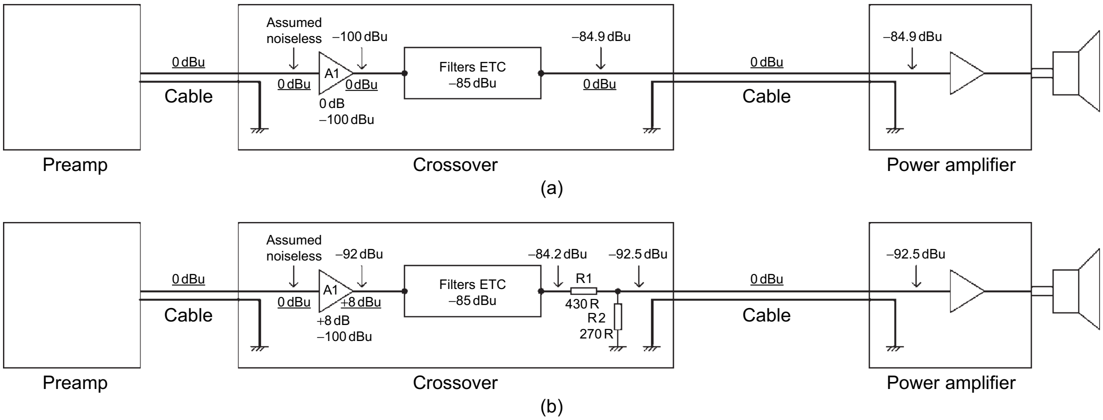


Figure 14.2: Gain structure for the preamplifier-crossover-power amplifier chain: (a) internal crossover level of 0 dBu gives an S/N ratio of 84.9 dB; (b) raising the internal level to +8 dBu gives an S/N ratio of 92.5 dB, an improvement of 7.6 dB.

An elevated internal level +8 dBu (2 Vrms) is used in Figure 14.2b, input amplifier A1 having a gain of +8 dB. The noise out from A1 is now −92 dBu, and with the −85 dBu of noise from the crossover filters added the total is −84.2 dBu. A passive 8 dB output attenuator R1, R2 then reduces the signal level back to the 0 dBu required by the power amplifier, and the noise is also reduced by 8 dB, giving us −92.5 dBu at the amplifier input. The signal-to-noise ratio has therefore increased from 84.9 dB to 92.5 dBu, an improvement of 7.6 dBu. We do not get the whole 8 dB because of the increased noise from input amplifier A1.

An elevated internal level not only makes the signal more proof against noise as it passes through the signal chain, but also against hum and other interference, though a good design should have negligible levels of these anyway.

You will note that I specified a passive output attenuator, so that the very low noise output is not compromised by an opamp stage after it. The attenuation can be made variable to give an output level trim control, working over perhaps a ±3 dB range. Given opamps with good load-driving capability, it is possible to make the passive attenuator with low resistance values so the output impedance is still acceptably low for driving long cables. The attenuator values shown in Figure 14.2b give an output impedance of 166 Ω, which is not perhaps to the highest professional standards but quite good enough for a domestic installation with limited cable runs. The load on the last opamp in the crossover is 700 Ω, which is high enough to prevent significantly increased distortion from a 5532 stage. There is more on such output networks, and how they can be combined with balanced outputs, later in this chapter.

There is an assumption here that the crossover is mostly composed of unity-gain circuitry such as the standard Sallen & Key filters. This is not always true—if you are using equal-C

Sallen and Key Butterworth filters then each second-order stage has a gain of +4.0 dB, and this will have to be dealt with somehow. Equalisation circuitry may have gains greater than this at some frequencies, with a corresponding reduction in headroom; this applies particularly to equalisers intended to extend the LF speaker unit response, which may have gains of +6 dB or more.

However, let's take it a little further. If our power amplifier clips with an input of 0 dBu, which corresponds to a crossover internal level of +8 dBu (2 Vrms), then as far as the crossover is concerned there is a range of signal levels from 2 Vrms to 10 Vrms (opamp maximum output) that is unusable. That is a 14 dB range. The effective signal-to-noise ratio could be further improved if that the crossover ran at a still higher level, say 5 Vrms (we want to keep a little safety margin). With the signal more heavily attenuated at the output to reduce it to 0 dBu; the internal crossover signals would then be another 8 dB higher, and we should now get some 16 dB more signal-to-noise ratio than with the 0 dBu internal level.

If there are full-range level controls between the crossover outputs and the power amplifiers, to allow level trimming, then the operator may unwisely set them for a considerable degree of attenuation; they will probably then crank up the input into the crossover to compensate, so there is now a higher nominal level in the crossover to obtain the same final output level. This means that the headroom in the crossover is reduced and the option of running it at a deliberately elevated internal level to improve the signal-noise ratio now looks less attractive. The question here is whether the noise performance should be compromised by increasing the headroom to allow for maladjustment.

14.6 Placing the Output Attenuator

Let's stick for the moment with the situation that there are no full-range level controls on the power amplifiers, just an output level trim on the crossover as described above. We therefore have our passive attenuator at the crossover outputs. The signal has therefore been brought back down to something of the order of 1 Vrms before it passes along the interconnection between crossover and power amplifier. However, if the attenuator is placed at the destination end of the interconnect cable, as in Figure 14.3a, any hum and interference picked up because of currents flowing through the cable ground will also be attenuated.

Ideally the output attenuator should be actually inside the power amplifier, as in Figure 14.3b, as this would also deal with any voltages caused by ground current flowing through the connector earth pins, but this requires a specialised amplifier design which would have a low input impedance, and low overall gain. Its input parameters would have to be defined by a Domestic Electronic Crossover Standard and unless there was some sort of input switching it would not be usable as an ordinary power amplifier. A better idea is to put the attenuator inside the connector that plugs into the power amp.

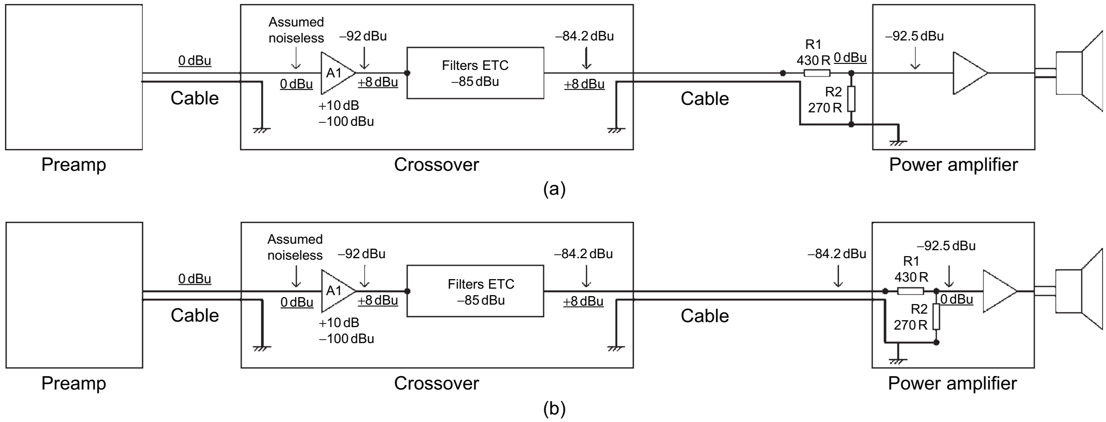


Figure 14.3: The preamplifier-crossover-power amplifier chain: (a) attenuator placed in the connector at the far end of the crossover-power amplifier; (b) attenuator placed in the power amplifier itself.

By now you are probably thinking: "Why not use a balanced interconnection? They are designed to discriminate against ground voltages," which is of course absolutely correct. However, if you peruse the chapter on line inputs and outputs, you will note that a practical real-life balanced input has a fairly modest Common-Mode Rejection Ratio (CMRR) of the order of 40 to 50 dB, and the most straightforward way of improving that is to have a preset CMRR trim. On a six-channel amplifier that is six adjustments to make, and while they are set-and-forget adjustments, this won't appeal to all manufacturers. Putting the passive attenuator at the power-amplifier end of the cable gives another 10 to 16 dB of immunity against ground noise on top of the basic CMRR. This plan obviously makes for a more complicated cable but it may be worth it, because ground-loop problems are a never-ending cause of distress in the audio business.

14.7 The Amplitude/Frequency Distribution of Musical Signals and Internal Levels

In considering the headroom requirements, it is important to remember that the signal levels in the HF, MID, and LF bands of the typical crossover are very different. The energy in the HF band is much lower, and this prompts me to suggest the possibility of running the HF channel at a higher nominal level than the others so noise performance can be improved without significantly compromising headroom. This is particularly appropriate for the HF channel because it contains no low-pass filters and therefore may be expected to have a higher noise output.

So, how much should the HF channel level be raised? Statistics on the distribution of amplitude with frequency are surprisingly hard to find, given how fundamental this information is to audio design; some information is given in [4] but it is not presented in an very convenient form for our purposes. A good source is Greiner and Eggars [5] who derived a lot of statistical data from 30 CDs of widely varying musical genres. The data is presented as the level in each of ten octave bands which is exceeded 90% of the time, 50% of the time, 10% of the time, and so on. This is a great deal of very useful information, but it does not directly tell us what we want to know, which might be phrased as: "If I have a 3-way system with crossover frequencies at 500 Hz and 3 kHz, how much level can I expect in each of the three crossover bands?" Greiner and Eggars do however at the end of their paper summarise the spectral energy levels in each octave band, which is more useful. See Table 14.1, which gives the levels in dB with respect to full CD level, for eight different musical genres, and the average of them all. The frequencies are the centres of the octave bands. Note that the results are from 1989, before the so-called 'Loudness Wars' in CD mastering broke out.

Figure 14.4 shows how the average levels are distributed across the audio band. The maximal levels are the region 100−2 kHz, with roll-offs of about 10 dB per octave at each end of the audio spectrum.

What we have to do now is decide which of the octave bands fit into our crossover bands, and sum the levels in those bands to arrive at a composite figure for each crossover band. A puzzling difficulty is to decide how to sum them—should they be treated as correlated (so two −6 dB levels sum to 0 dB) or uncorrelated (two −3 dB levels sum to 0 dB)? At first it seems unlikely that there would be much correlation between the octave bands, but on the other hand harmonics from a given musical instrument are likely to spread over several of them, giving some degree of correlation. I tried both ways, and for our purposes here the results are not very different.

Let's assume that our LF crossover band includes the bottom four octave bands (31.5 to 250 Hz), the MID crossover band includes the middle three octave bands (500 Hz to 2 kHz) and the HF crossover band includes the top three octave bands (4 kHz to 16 kHz).

Table 14.1: Spectral Energy Levels in Octave Bands for Different Genres, in dB (After Greiner and Eggars)

Octave Centre Hz	31.5	63	125	250	500	1 K	2 K	4 K	8 K	16 K
Piano B	−63	−47	−34	−28	−27	−33	−38	−46	−58	−63
Organ A	−32	−30	−29	−31	−30	−29	−32	−37	−50	−69
Orchestra B	−34	−33	−29	−29	−28	−30	−32	−39	−48	−58
Orchestra C	−26	−26	−30	−32	−32	−33	−35	−38	−45	−54
Chamber music	−78	−62	−45	−39	−41	−46	−49	−51	−65	−78
Jazz A	−48	−36	−33	−31	−29	−27	−29	−32	−39	−49
Rock A	−39	−35	−34	−32	−31	−32	−30	−36	−48	−57
Heavy metal	−45	−31	−27	−27	−33	−37	−31	−29	−36	−46
AVERAGE	−45.6	−37.5	−32.6	−31.1	−31.4	−33.4	−34.5	−38.5	−48.6	−59.3

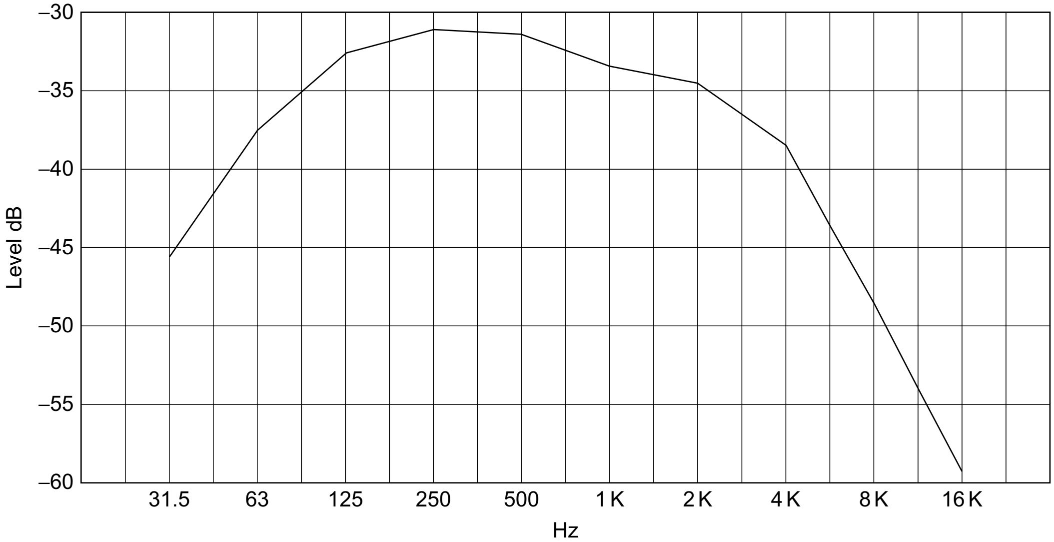


Figure 14.4: Average spectral levels in musical signals from CDs (after Greiner and Eggars).

Table 14.2: Uncorrelated Sums of Levels in the Three Crossover Bands, HF Band = Top Three Octaves. The Two Columns on the Right Are Levels Relative to the 31.5–250 Hz LF Band

Octave Centre Hz	LF dB 31.5–250	MID dB 500–2 K	HF dB 4 K–16 K	MID dB 500–2 K Relative to LF	HF dB 4 K–16 K Relative to LF
Piano B	−27.0	−25.8	−45.7	1.2	−18.7
Organ A	−24.3	−25.4	−36.8	−1.1	−12.4
Orchestra B	−24.7	−24.9	−38.4	−0.3	−13.8
Orchestra C	−21.8	−28.4	−37.1	−6.6	−15.4
Chamber music	−38.0	−39.3	−50.8	−1.3	−12.8
Jazz A	−28.1	−23.5	−31.1	4.6	−3.1
Rock A	−28.3	−26.2	−35.7	2.2	−7.4
Heavy Metal	−23.2	−28.3	−28.1	−5.1	−5.0
AVERAGE	−26.9	−27.7	−38.0	−0.8	−11.1

Uncorrelated summing gives us Table 14.2, where the actual levels in each crossover band are on the left, and the relative levels of the MID and HF bands compared with the LF band are in the two columns on the right.

Looking at the bottom line. the average of all the genres, we see that in general the MID crossover band will have similar levels in it to the LF crossover band. This is pretty much as expected, though it's always good to have confirmation from hard facts. We also see that the average HF level is a heartening 11 dB below the other two crossover bands, so it looks as if we could run the HF channel at an increased level, say 10 dB, and get a corresponding improvement in signal-to-noise ratio.

However… look a little deeper than the average. Table 14.2 shows that the HF level is significantly lower in all cases except for "Heavy Metal" where the MID and HF levels are the same. This is so unlike all the other data that I am not convinced it is correct. It looks as though we could indeed run the HF channel a good 10 dB hotter if it were not for the existence of Heavy Metal—a sobering thought.

If we change our assumptions, moving the upper crossover frequency higher so that the MID crossover band now includes the middle four octave bands (500 Hz to 4 kHz) and the HF crossover band includes only the top two octave bands, (8 kHz and 16 kHz) then things look rather better. Uncorrelated summing now gives us Table 14.3, where the HF levels are now down by some 20 dB with respect to the LF and MID channels. Now even Heavy Metal can have the level in the HF path pumped up by 10 dB.

Table 14.3: Uncorrelated Sums of Levels in the Three Crossover Bands, HF Band = Top Two Octaves. The Two Columns on Right Are Levels Relative to the 31.5–250 Hz LF Band

Octave Centre Hz	LF dB 31.5–250	MID dB 500–4 K	HF dB 8 K–16 K	MID dB 500–4 K Relative to LF	HF dB 8 K–16 K Relative to LF
Piano B	−27.0	−25.7	−56.8	1.3	−29.8
Organ A	−24.3	−25.1	−49.9	−0.8	−25.6
Orchestra B	−24.7	−24.8	−47.6	−0.1	−22.9
Orchestra C	−21.8	−27.9	−44.5	−6.2	−22.7
Chamber music	−38.0	−39.0	−64.8	−1.0	−26.8
Jazz A	−28.1	−22.9	−38.6	5.2	−10.5
Rock A	−28.3	−25.7	−47.5	2.6	−19.2
Heavy metal	−23.2	−25.6	−35.6	−2.4	−12.4
AVERAGE	−26.9	−27.1	−48.2	−0.2	−21.2

Nevertheless, what we can only call The Heavy Metal Problem is messing up the nice clear results from this study. The HM track used by Greiner and Eggars is only identified as "Metallica" but a little detective work revealed it to be the track "Fight Fire with Fire" from the CD "Ride The Lightning." I am listening to it as I write this, and I cannot honestly say that the HF content is obviously greater than that of the normal run of rock music. Are we justified in regarding this Heavy Metal data as an outlier that can be neglected? Hard to say.

At this point the only conclusion can be that you can run the HF channel hotter, but how much depends on where the upper crossover point is. If it is around 3 kHz then elevating the HF channel level by 6 dB looks pretty safe, and this is the value adopted in my crossover design example in Chapter 19.

Another area where the distribution of amplitude with frequency is important is that of transformer-balanced input and outputs. Transformers are notorious for generating levels of third-harmonic distortion that rise rapidly as frequency falls. It is fortunate that Table 14.1 above shows that in almost all cases (church organ music being a notable exception) the signal levels in the bottom audio octave are roughly 10–12 dB lower than the maximum amplitudes, which occur in the middle octaves. This reduces the transformer distortion considerably.

14.8 Gain Structures

There are some very basic rules for putting together an effective gain structure in a piece of audio equipment. Breaking them reduces the dynamic range of the circuitry, either by raising the noise floor or lowering the headroom. Three simple rules are:

1. DON'T AMPLIFY THEN ATTENUATE. It is all too easy to thoughtlessly add a bit of gain to make up for a loss later in the signal path, and immediately a few dB of precious headroom are gone for good. This assumes that each stage has the same power rails and hence the same clipping point, which is usually the case. Figure 14.5a shows a fragment of a system with a gain control designed to have +10 dB of gain at maximum. There is assumed to be no noise at the input. A and B are unity gain buffers and each contributes −100 dBu of its own noise; Amplifier 1 has a gain of +10 dB and an EIN of −100 dBu.

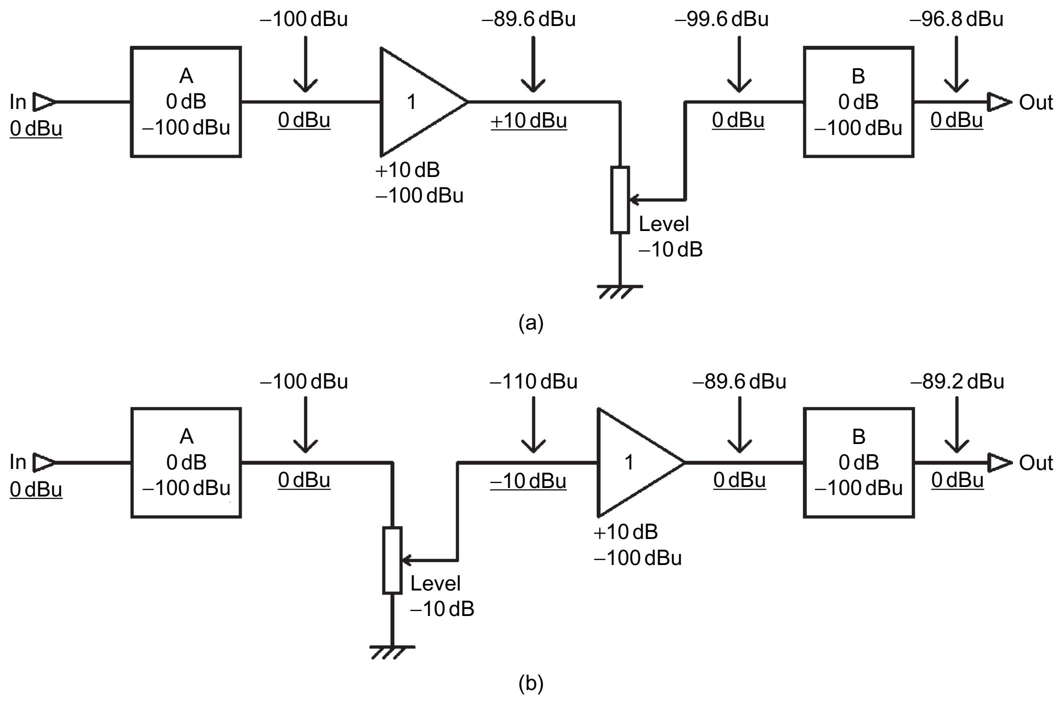


Figure 14.5: Gain structures: (a) Amplification then attenuation. Amplifier 1 always clips first, reducing headroom; (b) Attenuation then amplification. Noise from Amplifier 1 degrades the S/N ratio at low gain settings. Noise levels along the signal path indicated by arrows; signal levels are underlined.

The expectation is that the level control will spend most of its time set somewhere near the "0 dB" position where it introduces 10 dB of attenuation. To keep the nominal signal level at 0 dBu we need 10 dB of gain, and Amplifier 1 has been put before the gain control. This is a bad decision, as this amplifier will clip 10 dB before any other stage before it in the system, and this introduces what one might call a headroom bottleneck. On the positive side, the noise output is only –96.8 dBu, because the signal level never falls below 0 dBu and so is relatively robust against the noise introduced by the stages.

2. DON'T ATTENUATE THEN AMPLIFY. Since putting our 10 dB of amplification before the gain control has disastrous effects on headroom, it is more usual to put it afterwards, as shown in Figure 14.5b. Now noise performance rather than headroom suffers; the amount of degradation depends on the control setting, but as a rule it is much more acceptable than a permanent 10 dB reduction in headroom. The signal-to-noise ratio is impaired for all gain control settings except maximum; if we dial in 10 dB of attenuation as shown then the signal reaching Amplifier 1 is 10 dB lower, at −10 dBu. The noise generated by Amplifier 1 is unchanged, and almost all the noise at its output is its own EIN amplified by 10 dB. This is slightly degraded by the noise of block B to give a final noise output of −89.2 dBu, worse than Figure 14.5a by 7.6 dB. If there are options for the amplifier stages in terms of a noise/cost trade-off and you can only afford one low-noise stage then it should clearly be Amplifier 1.
3. AMPLIFY AS SOON AS POSSIBLE. Get the signal up to the nominal internal level as soon as it can be done, preferably in the first stage, to minimise its contamination with noise from later stages. Consider the signal path in Figure 14.6a, which has a nominal input level of –10 dBu and a nominal internal level of 0 dBu. It has an input amplifier with 10 dB of gain followed by two unity-gain buffers A and B. As before, all circuit stages are assumed to have an Equivalent Input Noise level of −100 dBu, and the incoming signal is assumed to be entirely noise free. The noise output from the first amplifier is therefore −100 dBu + 10 dB = −90 dBu. The second stage A adds in another −100 dBu, but this is well below −90 dBu and its contribution is very small, giving us −89.6 dBu at its output. Block B adds another −100 dBu, and the final noise output is −89.2 dBu.

Compare that with a second version of the signal path in Figure 14.6b, which has an input amplifier with 5 dB of gain, followed by block A, a second amplifier with another 5 dB of gain, then block B. The signal has received exactly the same amount of gain, but the noise output is now 1.7 dB higher at −87.5 dB, because the signal passed through block A at −5 dBu rather than 0 dBu. There is also an extra amplifier stage to pay for, and the second version is clearly an inferior design.

An even worse architecture is shown in Figure 14.6c, where both the two unity-gain blocks, A and B are now ahead of a +10 dB amplifier. The low-level signal at −10 dBu is now more vulnerable to the noise from block B, and the output noise is now −85.2 dBu, 4.0 dB

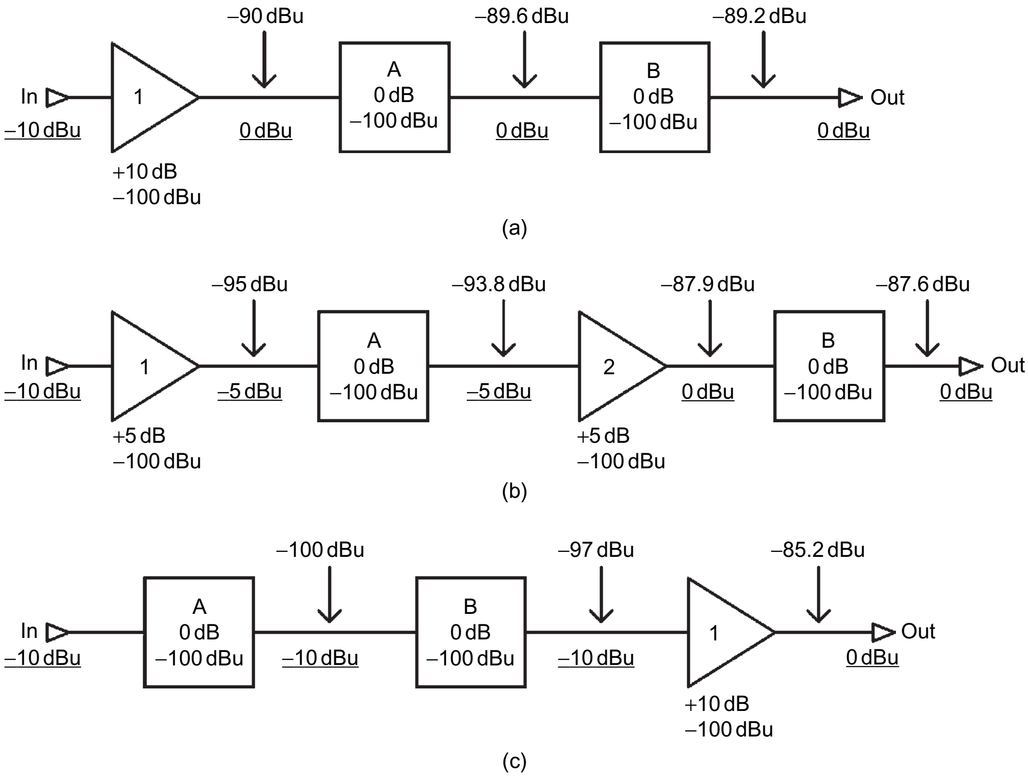


Figure 14.6: Why you should amplify as soon as possible: (a) All amplification in first stage gives best noise performance of –89.2 dBu; (b) Amplification split over two stages. Noise from Stage A degrades the S/N ratio; (c) Amplification late in chain. Now both Stages A and B degrade the S/N ratio. Noise levels indicated by arrows; signal levels are underlined.

worse than the optimal arrangement in Figure 14.2a where all the amplification is placed at the start of the signal path.

In active crossover design the problem is not usually quite so marked as in the examples above, because any level controls are usually trims rather than full-range volume controls, but the principles still stand. Let us examine a typical scenario; assume that you want to make a fourth-order Linkwitz–Riley filter, and you are planning to use two cascaded second-order Sallen and Key Butterworth filters of the equal-C type to make component procurement simpler. For a second-order Sallen and Key filter to be equal-C the stage gain has to be 1.586 times (+4 dB) and when you cascade two to make the fourth-order Linkwitz–Riley filter you have a total gain of 2.5 times (+8 dB) to deal with. There are three possibilities:

Firstly, you can attenuate the signal by 8 dB *before* the filters, but the low level in the first filter will almost certainly degrade the noise performance.

Secondly, you can attenuate the signal by 8 dB *after* both filters, but now there are higher signal levels, in particular 8 dB higher at the output of the second filter, and headroom problems may occur at this point.

Thirdly, and usually best, is some compromise between these two extremes. For example, putting 4 dB of attenuation after the first filter, and then another 4 dB of attenuation after the second filter will reduce the headroom problems by 4 dB. Alternatively, putting 4 dB of attenuation before the first filter, followed by 4 dB of attenuation after it, will reduce the noise contribution of the first stage.

The aforegoing assumes you do not need to add buffers between the attenuators and the filters to give the latter a low source impedance. There is some useful information on avoiding buffers and economically combining attenuators with filters in Chapter 8.

If you are seeking the best possible performance, then probably your best option is to not use Sallen and Key equal-C filters in the first place, but stick to the more usual unity-gain sort.

14.9 Noise Gain

Before we dig deeper into the details of getting the cleanest possible signal path, we need to look at the concept of noise gain. Figure 14.7 shows two inverting stages, both with unity gain. The only difference is the presence of an innocent-looking resistor R3 in the (b) version of the circuit. Since the inverting input of the opamp is at virtual ground (i.e., it has a negligible signal voltage on it but is not actually connected to ground) because of the very high negative feedback factor, the gain will be very nearly exactly equal to –1 for both circuits. So they are functionally identical? No indeed. The (b) version will be 3.5 dB noisier, simply because R3 is lurking in there.

This is nothing to do with Johnson noise from R3, which I have ignored; even if R3 was some sort of magical resistor with no Johnson noise (which is not possible in this universe),

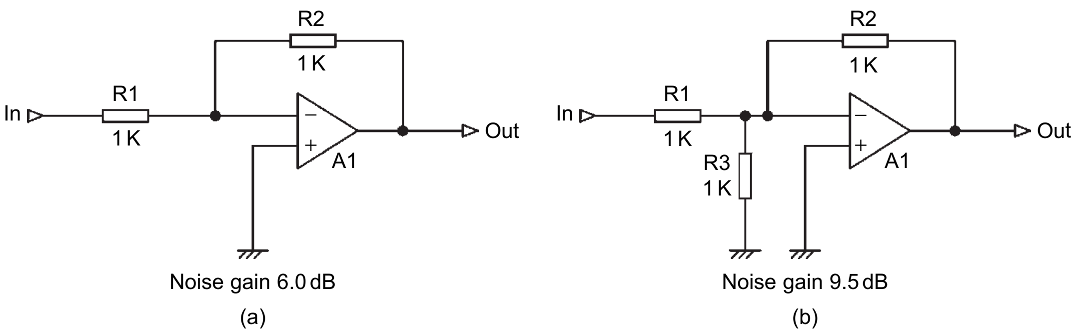


Figure 14.7: Demonstration of the concept of noise gain: (a) Unity-gain inverting stage with noise gain of 6.0 dB; (b) Unity-gain inverting stage with noise gain of 9.5 dB.

the (b) version would be 3.5 dB noisier. This is because although it has the same signal gain, it has a higher *noise gain*.

Opamp voltage noise is accurately modelled by assuming a noiseless opamp and putting a voltage noise generator between the two inputs. (I am also ignoring current noise as it has negligible effects with low resistances like those shown.) One side of this generator is grounded via the non-inverting input, and as far as the opamp is concerned the stage input is also effectively at ground because it is fed from a low source impedance. Therefore, as far as noise is concerned, we have not an inverting amplifier, but a non-inverting amplifier, with the noise generator as the input and R1, R2 in Figure 14.7a as the feedback network. R1 equals R2 so the gain from noise generator to output is 2 times; (+6 dB) thus Figure 14.7a has unity gain for signal but 6 dB of gain for its own noise.

For Figure 14.7b the story is different; so far as noise is concerned the upper arm of the feedback network is unchanged, but the lower arm now consists of two equal resistors so their combined resistance is half the value, and the gain from noise generator to output is 3 times (+9.5 dB).

When concocting amplifier circuits you should always be aware of the possibility of accidentally building in R3 or some less obvious equivalent. Noise gain is a really major issue in mixing consoles where the large number of mix resistors feeding the virtual earth summing bus [6] are effectively in parallel and represent a very low value for R3; a 32-input console mix system has a noise gain of +30.4 dB, and you can see why summing amplifiers with the lowest possible noise are required. In active crossovers noise gain issues are most likely to arise in equaliser circuits such as the bridged-T and the biquad equaliser; see Chapter 11.

14.10 Active Gain-Controls

So far we have dealt only with combinations of fixed-gain amplifiers and passive attenuation controls, and we have seen that either noise performance or headroom must be compromised when a gain-control function is included in the signal path. If, however, we move beyond the idea of a fixed-gain amplifier this compromise can be to a large extent avoided. If we have an amplifier stage with variable gain, we can set it to give just the gain we want and no more. Using less gain when the maximum is not required will reduce the noise generated compared with a fixed-gain amplifier placed after a level control, because the fixed-gain amplifier has to have the maximum gain required, and is hissing away all the time with no following level control to turn it down.

Simply making a stage with variable gain is straightforward; you just vary the amount of negative feedback. Achieving a given gain law requires a little more thought (see Chapter 16 on line inputs). In a crossover it will not normally be necessary for the gain to be variable right down to zero, as it would be for a preamplifier volume-control [7].

Another advantage of active gain controls is that they can give much better channel balance by eliminating the uncertainties of log pots. The Baxandall active gain configuration gives excellent channel balance as it depends solely on the mechanical alignment of a dual linear pot- all mismatches of its electrical characteristics are cancelled out, and there are no quasi-log dual slopes to cause heartache [8].

To demonstrate the advantages of active gain controls, take a look at Figure 14.8. The circuit at (a) is a conventional passive level control followed by a +10 dB amplifier stage. At (b) is an active gain control also giving a maximum of +10 dB gain. In this case we have a shunt-feedback circuit that can be adjusted all the way down to zero gain but that is not an essential feature. If we assume both opamps have the same voltage noise (the actual level does not matter) and neglect current noise and Johnson noise, which make little contribution with such low impedances, then it is straightforward to calculate the signal-to-noise (S/N) ratio for each circuit. This is expressed as the difference between the two circuits in Table 14.4, since this makes the actual signal level irrelevant. The level control setting is recorded as dB down from fully up, whatever the maximum gain.

The S/N advantage of the active gain control is greatest at intermediate and low level settings, and this advantage increases as the maximum gain increases. However, with the circuits we have selected to look at here, when the control is actually set near or at maximum gain, the S/N is actually marginally worse. This is because the noise gain of Figure 14.8b is always one times more than the signal gain, because of its shunt feedback configuration, whereas for Figure 14.8a the signal gain and noise gain are always equal. The higher the maximum gain required, the less significant this effect is; for 20 dB max gain with the control fully up the active version is 0.8 dB noisier, but backing it off by only 1 dB reverses the situation and the passive version is now 0.1 dB noisier. At lower control settings the advantage of the active version increases rapidly, being a very useful 7.6 dB at the −10 dB setting, and asymptotically increasing to 20 dB.

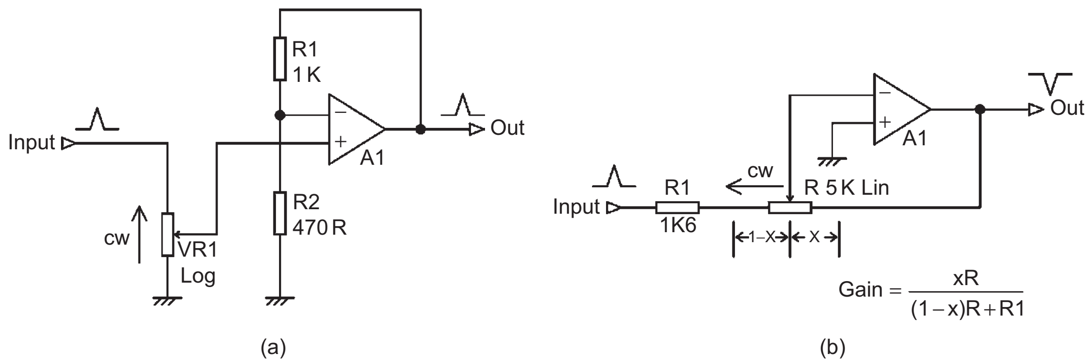


Figure 14.8: Passive and active gain controls, both with a maximum gain of +10 dB.

Table 14.4: The S/N Improvement with an Active Gain Control for Various Maximum Gains

	S/N Ratio Compared with Passive Circuit in Fig 14.8a		
Max Gain =	10 dB	15 dB	20 dB
Level Control dB	dB	dB	dB
0	−2.4	−1.4	−0.8
−1	−1.6	−0.6	0.1
−2	−0.9	0.2	1.0
−5	1.1	2.6	3.6
−7	2.4	4.1	5.2
−10	4.0	6.1	7.6
−20	7.6	11.1	14.0
−30	9.2	13.6	17.6
−40	9.7	14.5	19.2
−50	9.9	14.8	19.7
−60	10.0	15.0	19.9
−70	10.0	15.0	20.0
−80	10.0	15.0	20.0

If we compare our active stage not with the "passive control then amplifier" structure of Figure 14.8a, but with the rather impractical "amplifier first then passive gain control" of Figure 14.5a, the active noise performance is inferior at low gain settings because the noise gain of the shunt feedback stage cannot fall below unity, whereas a final passive control can attenuate the noise fed to it indefinitely. However, as we have seen, "amplifier first then passive gain control" is rarely if ever a good choice in system design because of the crippling loss of headroom. There is never any issue of reduction of headroom with the active gain control as only the amount of gain required is actually in use.

Whenever you need a level control you should carefully consider the possibility of using active gain control circuitry.

14.11 Filter Order in the Signal Path

The audio performance is affected by the order in which filters and other circuit blocks are placed. If you have lowpass filters in a signal path then put them last, then they will attenuate noise from the other circuit blocks ahead of them.

A an example of this, Figure 14.9 shows a typical crossover MID path composed of fourth-order Linkwitz–Riley lowpass (400 Hz) and highpass (3 kHz) filters. Each Linkwitz–Riley filter is made up of two identical second-order Butterworth filters; note that "2 × 47 nF" means two 47 nF capacitors in parallel. The lowpass filters are here placed *before* the highpass filters in the signal path. The noise levels in the signal path were measured and the results are shown in Figure 14.10.

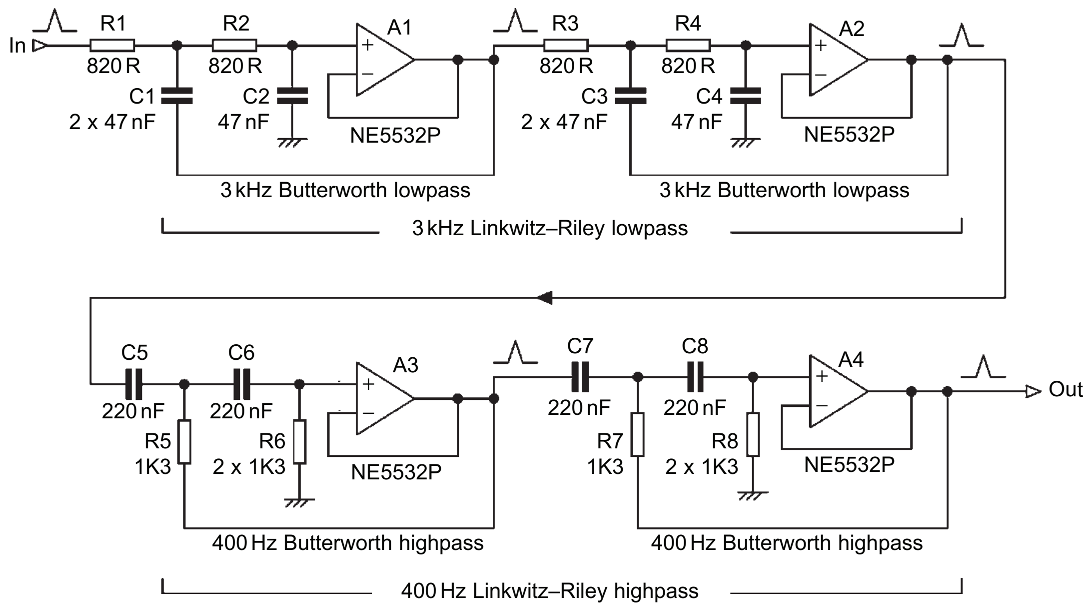


Figure 14.9: A typical Linkwitz–Riley MID path with the lowpass filters placed before the highpass filters. Reversing the order reduces the noise output.

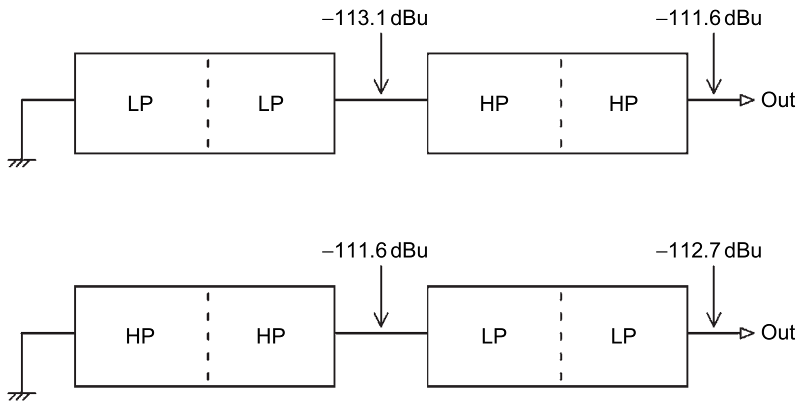


Figure 14.10: How changing the filter order so the lowpass filters come after the highpass filters reduces the noise output.

The order of the filters was then reversed so the lowpass filters came at the end of the chain, and Figure 14.10 shows that we reduce the noise by 1.1 dB by this simple change. I do not pretend this is a radical improvement, but it has the advantage that it costs *absolutely nothing*. If there is more circuitry in the signal path than just the crossover filters, such as frequency/ amplitude equalisers or delay-compensation allpass filters, these should also be placed upstream of the lowpass filters, and the improvement in noise performance will be greater as these circuits tend to be noisier than simple highpass filters. Bridged-T equalisers in particular have a reputation for being noisy due to their high internal noise gain.

You might wonder if the distortion performance is in some way impaired. It isn't. It is only possible to make meaningful measurements between about 200 Hz and 5 kHz, as the rapid fourth-order amplitude roll-offs mean that you are very quickly just measuring noise rather than distortion, but in this region the distortion is in fact slightly lower with the lowpass filters at the end of the path. This is because they tend to filter out distortion harmonics from the preceding highpass filters.

One possibility to consider there might be increased HF intermodulation in the highpass filters as they are now handling a more extended range of the audio spectrum. Since the amount of negative feedback is reduced at HF because of limited opamp bandwidth, this would of greater concern than at LF. I have seen no sign that this could be a significant problem in practice.

When settling the arrangement of the filter blocks in the crossover, you should be aware of the amplitude response irregularities that can be caused by filter phase-shift when crossover frequencies are not widely spaced. This very important issue is dealt with at the end of Chapter 4.

14.12 Output Level Controls

The first question is how much gain control should be available on the output of a crossover. We are not trying to make a volume control that can be adjusted from full up to zero, firstly because the place for the volume control is on the preamplifier where you can get at it—the crossover will usually be tucked away out of sight and possibly hidden in an equipment cabinet. Secondly, if we did want to put a volume control in this position, a three-way crossover would require a six-gang pot, which would be expensive, probably have some unsettling channel-balance errors, and an unpleasant feel.

If we are simply intending to allow for gain variations in a nominally identical set of power amplifiers then a vernier control of as little as ±1 dB may be sufficient. If in addition the gain controls must allow for production variations in transducers then a wider range of ±3 dB or ±6 dB may be appropriate, but it is essential to realise that simply altering the gain of each channel will probably not give anything like enough control; making really accurate allowances for transducer tolerances will probably require tuning of crossover frequencies and several equalisation parameters.

This sort of calibration is quite feasible if it is computer-based, and each crossover example is permanently allocated to its matched loudspeaker. There is no guarantee that loudspeakers will have identical units in a stereo pair, and so it is essential not to swap left and right channels as it is unlikely that the crossover settings will be identical.

As we saw earlier in this chapter, the use of raised internal levels in the crossover circuitry followed by passive attenuation can give significant improvements in signal-to-noise ratio. The passive attenuator can be conveniently combined with an output trim network as shown in Figure 14.11, where the component values are chosen to give a ±6 dB adjustment range. This is a fragment from one of my active crossover designs where the internal signal path was running at an elevated level of 3 Vrms, and the nominal output was 880 mV rms.

The design considerations are:

1. That the trim network resistances should be high enough to not excessively load the last opamp in the signal path, to ensure no excess distortion is generated. The load imposed by Figure 14.11 is 625 Ω, which is about as heavy as you would want to make it.
2. That the network resistances should be low enough to give a suitably low output impedance to drive reasonable lengths of cable without HF losses. The network in Figure 14.11 has a maximum output impedance of 156 Ω, which occurs when the control is fully up. The output impedances with the control central, and fully down, are 130 Ω and 66 Ω respectively. In the worst case (fully up) this arrangement could drive up to 53 metres of 150 pF/metre cable before the loss at 20 kHz reached 0.1 dB.

You can of course only have a ±6 dB trim range if the internal signal level is at least 6 dB above the nominal output level. It is very likely that the preset pot track resistance will have a wider tolerance than the fixed resistors, and this can cause channel balance errors. Note that no separate drain resistor is required to remove any charges left on C1 by external DC voltages.

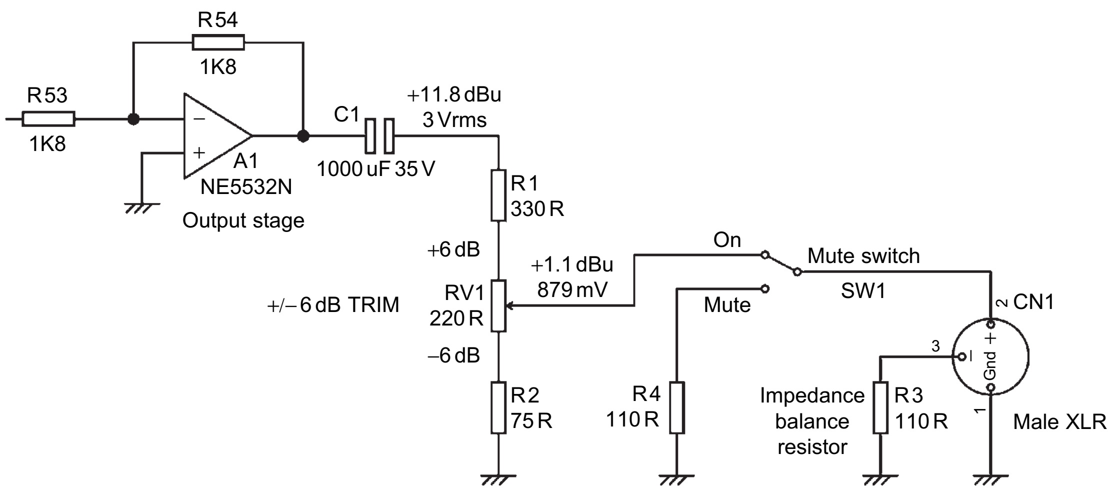


Figure 14.11: Low-impedance passive output attenuator with gain trim, mute switch, and impedance balancing resistor.

The unbalanced output uses an XLR to allow impedance-balanced operation; R3 is connected between the cold output pin and ground to balance the impedances seen on the output pins and so maximise the CMRR when driving a balanced input. The value of 110 Ω is the average of the maximum and minimum output impedances as the setting of the gain trim is varied. This variation inevitably compromises the balancing of the impedances, and picking an average figure for R3 is the best we can do. There is more on impedance balancing in Chapter 17 on line outputs, and in Chapter 19. Note that C1 is rather big at 1000 uF; this is not to extend the frequency response, but to make sure it does not generate capacitor distortion as it drives the relatively low impedance of the attenuator network. It is a non-polarised component so it is immune to external fault voltages of either polarity.

14.13 Mute Switches

Crossovers intended for sound reinforcement are often fitted with mute switches for each output to simplify system checking and debugging. The straightforward method shown in Figure 14.11 connects the output to ground and gives very good attenuation indeed, as almost all the crossover circuitry is disconnected. The “impedance-balanced” feature is retained because R4 = R3. Slightly lower output noise would be obtained by replacing R4 with a short-circuit, but is more important to retain the best possible CMRR, even in the test-only mute mode. Illuminated switches are often used to make operation easier in shadowy backstage areas.

14.14 Phase-Invert Switches

This is another feature which is pretty much restricted to crossovers for sound reinforcement. The ability to invert the phase of just one output can be useful for checking and for the rapid correction of a phase error somewhere else in the system. Some crossovers (e.g., Behringer Super-XPRO CX2310) have separate phase-invert switches on all the main outputs and also the sub-woofer output.

A phase-invert switch is easy to arrange if you have a balanced output; the outputs are just swapped over, as shown in Figure 14.12. With an unbalanced output it will be necessary to switch in a unity-gain inverting stage. If you are doing that then you might as well make use of the added opamp section all the time, and press it into service to implement a balanced output. Once again, illuminated switches are useful.

14.15 Distributed Peak Detection

An electronic crossover is not normally fitted with level meters in the way that mixing consoles or power amplifiers are, it being assumed that it will be carrying out its function in some out-of-the-way corner where no one will see it. Nevertheless some indication of grossly wrong signal levels is useful.

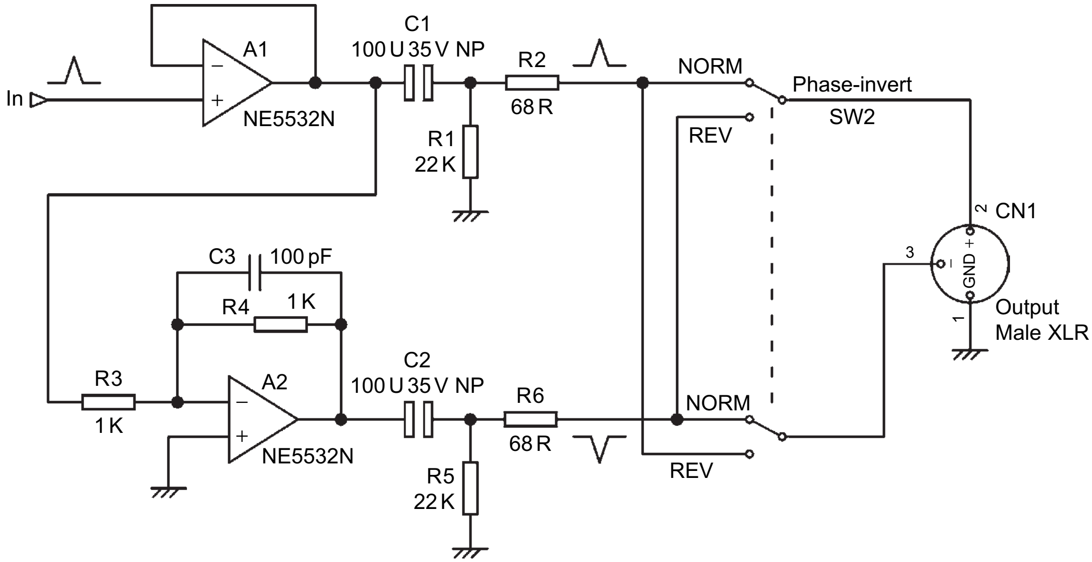


Figure 14.12: A phase-invert switch added to a balanced output.

When an audio signal path consists of a series of circuit blocks, each of which may give either gain or attenuation, it is something of a challenge to make sure that excessive levels do not occur anywhere along the chain. Simply monitoring the level at the end of the chain is no use because a circuit block that gives gain, leading to clipping, may be followed by one that attenuates the clipped signal back to a lower level that does not trip a final peak-detect circuit. It is not, however, necessary to add a bipolar peak detection circuit at the output of every opamp to be sure that no clipping is happening anywhere along the path.

If a stage is followed by another stage with a flat gain, say of +4 dB, then there is of course no need to monitor the first stage as the one following it will always clip first. If, however, the gain of the second stage is not flat, as in an electronic crossover it will very often not be, it may be necessary to monitor both stages. For example, if the second stage is a lowpass filter with a cutoff frequency of 400 Hz, then a signal at 10 kHz may cause the first stage to clip heavily without the second stage getting anywhere near its clipping point, even if it has gain in its passband.

A multi-point or distributed peak detection circuit that I have made extensive use of is shown in Figure 14.13, where it is shown monitoring a circuit block in each path of a 3-way crossover. It can detect when either a positive or negative threshold is exceeded, at any number of points desired; to add another stage to its responsibilities you need only add another pair of diodes, so it is very economical. If a single peak detector is triggered by too

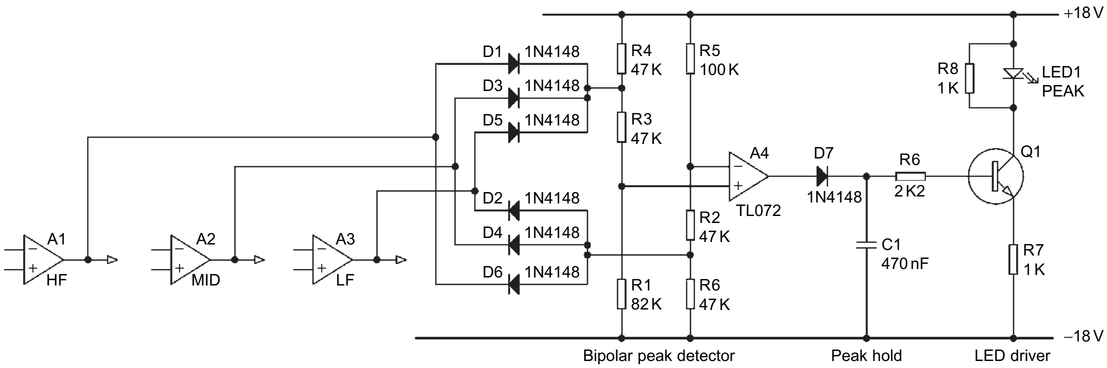


Figure 14.13: A multipoint bipolar peak detector, monitoring three circuit blocks.

many points in the signal path, it can be hard to determine at which of them the excessive level exists, but the basic message is still clear—turn it down a bit.

The operation is as follows. Because R5 is greater than R1, normally the non-inverting input of the opamp is held below the inverting input and the opamp output is low. If any of the inputs to the peak system exceed the positive threshold set at the junction of R4, R3, one of D1, D3, D5 conducts and pulls up the non-inverting input, causing the output to go high. Similarly, if any of the inputs to the peak system exceed the negative threshold set at the junction of R2, R6, one of D2, D4, D6 conducts and pulls down the inverting input, once more causing the opamp output to go high. When this occurs C1 is rapidly charged via D7. The output-current limiting of the opamp discriminates against very narrow noise pulses. When C1 charges Q1 turns on, and illuminates D8 with a current set by the value of R7. R8 ensures that the LED stays off when A4 output is low, as it does not get close enough to the negative supply rail for Q1 to be completely turned off.

Each input to this circuit has a non-linear input impedance, and so for this system to work without introducing distortion into the signal path, it is essential that the diodes D1–D6 are driven directly from the output of an opamp or an equivalently low impedance. Do not try to drive them through a coupling capacitor as asymmetrical conduction of the diodes can create unwanted DC-shifts.

The peak-detect opamp A4 must be a FET-input type to avoid errors due to its bias currents flowing in the relatively high value resistors R1–R6, and a cheap TL072 works very nicely here; in fact the resistor values could probably be raised significantly without any problems.

You will note all the circuitry operates between the two supply rails, with no connection to ground, preventing unwanted currents from finding their way into the ground system and causing distortion.

14.16 Power Amplifier Considerations

When selecting the power amplifiers that are to work with active crossovers, the first consideration is how much power is required in each band. Earlier in this chapter we looked at the amplitude/frequency distribution of musical signals, and found that while the LF and MID bands may have roughly the same amplitude, the HF band may be between 5 and 20 dB lower, depending on where the MID/HF crossover frequency is set. Converting this into Watts (with their square-law relationship with Volts) tells us that if the LF and MID bands require 100 W amplifiers, that for the HF channel need be capable of no more than 31 Watts in the worst case. There is some useful information on amplifier power requirements in Penkov [4].

Another important factor is the efficiency of the speaker drive units for each band. Getting good low frequency extension from an enclosure of limited size means low efficiency and for this reason more power is generally required for LF channels than for MID channels.

In general there is little to be gained by designing different amplifiers for the frequency bands, apart from the power differences described above. You might be able to cut down on the size of the power supply reservoir capacitors in the MID amplifier compared with the LF amplifier, because of the absence of sustained transients, but the cost saving is hardly worth the trouble.

The exception to that statement is the HF power amplifier. A DC fault in this amplifier can vaporise a tweeter in the twinkling of a voice-coil, as unlike a tweeter connected to a passive crossover, it is not protected by a series capacitor. This would seem to demand that DC-offset protection circuitry for HF amplifiers would need to be especially fast-acting to be effective; since such protection would inevitably be triggered by large bass signals, we have created an amplifier that cannot be used for anything other than tweeter-driving duties. This runs completely counter to the desirability of the user being able to connect up any amplifiers he feels are the best for the job. The vulnerability of HF drive units is an excellent reason to not use HF amplifiers that are any more powerful than is necessary to do the job.

The DC offset protection could be much improved by adding a capacitor in series with the tweeter, but this re-introduces a large and expensive component, of the sort we thought we had left behind when we adopted an active crossover; there should however be no need for it to maintain a precise value as would be required if it was part of a passive crossover. The series capacitor is likely to introduce its own shortcomings in the form of non-linearity, especially if it is an electrolytic, when it may introduce distortion even in the midband where it is not implementing any kind of roll-off [9]. There is also the point that in the world of hi-fi, where almost any technology that can make some sort of noise has its adherents, it seems that no one is prepared to advocate capacitor-coupled power amplifiers.

References

[1] M. Ross, Power Amplifier Clipping and Its Effects On Loudspeaker Reliability, RaneNote 128, Rane Corporation, USA.

[2] D. Self, Small Signal Audio Design Handbook, Chap. 3, Focal Press, Boston, MA, 2009, ISBN 978-0-240-52177-0.

[3] Bryston, http://bryston.com/10b_m.html, 2010 (accessed 12.10).

[4] G. Penkov, Power & real signals in an audio system, JAES 31 (6) (1983) 423–429.

[5] R.A. Greiner, J. Eggars, The spectral amplitude distribution of selected compact discs, J. Audio Eng. Soc. 37 (1989) 246–275.

[6] D. Self, Small Signal Audio Design Handbook, Focal Press, Boston, MA, 2009, pp. 443–445, ISBN 978-0-240-52177-0.

[7] D. Self, Small Signal Audio Design Handbook, Focal Press, Boston, MA, 2009, p. 219, et seq (Volume controls), ISBN 978-0-240-52177-0.

[8] D. Self, Small Signal Audio Design Handbook, Focal Press, Boston, MA, 2009, pp. 235–237, (Baxandall vol control), ISBN 978-0-240-52177-0.

[9] D. Self, Audio Power Amplifier Design Handbook, fifth ed., Chap. 2, Focal Press, Boston, MA, 2009, p. 43, ISBN 978-0-240-52162-6.

CHAPTER 15

Subwoofer Crossovers

15.1 Subwoofer Applications

A subwoofer loudspeaker is a separate enclosure from the main loudspeakers that is specifically designed to reproduce low bass. They are normally used singly rather than in stereo pairs, the assumption being that low frequencies can be regarded as non-directional. Another justification for mono subwoofers is that bass information is commonly pan-potted to the centre of the stereo stage. This is essential if the recorded material is likely to be used to cut vinyl discs, as big differences in low frequency information between left and right channels creates large up-and-down, as opposed to lateral, contours in the groove that increase the possibility of mistracking and can in bad cases throw the stylus out of the groove altogether. Another good reason for centralising the bass is that it needs large amounts of amplifier power to reproduce it, and it is therefore desirable to make full use of both channels.

In home entertainment the sub-woofer is normally of modest size, and is often installed under the television screen, in front of the listeners, though other placements are possible, exploiting the non-directional characteristics of its output. In automotive use subwoofers are installed in the boot (the trunk, to some of you) or the rear cabin space.

The typical frequency range for subwoofer operation is between 20 and 200 Hz. Sound reinforcement subwoofers typically operate below 100 Hz, and THX-approved systems in cinemas operate below 80 Hz.

15.2 Subwoofer Technologies

A wide variety of types of loudspeaker enclosure are used in subwoofers. The relatively small frequency range that has to be covered allows techniques such as bandpass enclosures to be used when they would not be usable with LF or MID drive units. The main types are:

Sealed (Infinite baffle)
Reflex (ported)
Transmission line
Bandpass

Isobaric
Dipole
Horn-loaded

The first two of these are by far the most popular. While I have not done a comprehensive survey, it seems clear that sealed enclosures are more popular than ported types, which is perhaps surprising given the greater bass extension possible with ported enclosures. Brief descriptions of the operation of each type are given below; I have kept them fairly short as many aspects are covered in Chapter 2 on loudspeakers.

15.2.1 Sealed-Box (Infinite Baffle) Subwoofers

The general characteristics of a sealed subwoofer enclosure, as shown in Figure 15.1a, are much the same as for normal loudspeakers, as examined in Chapter 2. For subwoofer use the driver may be mounted on one of the sides of the box, or on the top or bottom; sometimes there are multiple drivers mounted on different sides of the box. It has a good transient response, good LF power handling because the drive unit is always loaded by the box air, and lower sensitivity to parameter misalignment than other approaches. Sealed-box enclosures have higher low-frequency cutoff points and lower sensitivity than other subwoofer types for the same box volume. Despite this they appear to be the most popular configuration for subwoofers.

The efficiency of a drive unit in a sealed box is proportional to the cube of the drive unit resonance frequency (the Thiele-Small parameter F_S) [1] so quite small improvements to low-frequency extension with the same drive unit and box volume lead to big reductions in the efficiency of converting electrical energy into sound energy. Subwoofers of reasonable size are therefore typically very inefficient by the standards of normal loudspeakers and require both powerful amplifiers, and drivers with considerable power handling capability. For this reason enclosures other than a sealed box (e.g., bass reflex designs) are sometimes used for subwoofers to increase the efficiency of the driver/enclosure system and reduce the amplifier power requirements.

15.2.2 Reflex (Ported) Subwoofers

Reflex or ported subwoofer enclosures work exactly as described for normal loudspeakers in Chapter 2. At low frequencies the port output is in phase with the forward radiation of the driver and allows the bass response to be extended without the lower efficiency that occurs with a sealed box.

With everything else being equal, a ported subwoofer can have lower drive unit distortion, higher power handling, and a lower cutoff frequency than a sealed box system with the same drive unit. Distortion rapidly increases below the cutoff frequency

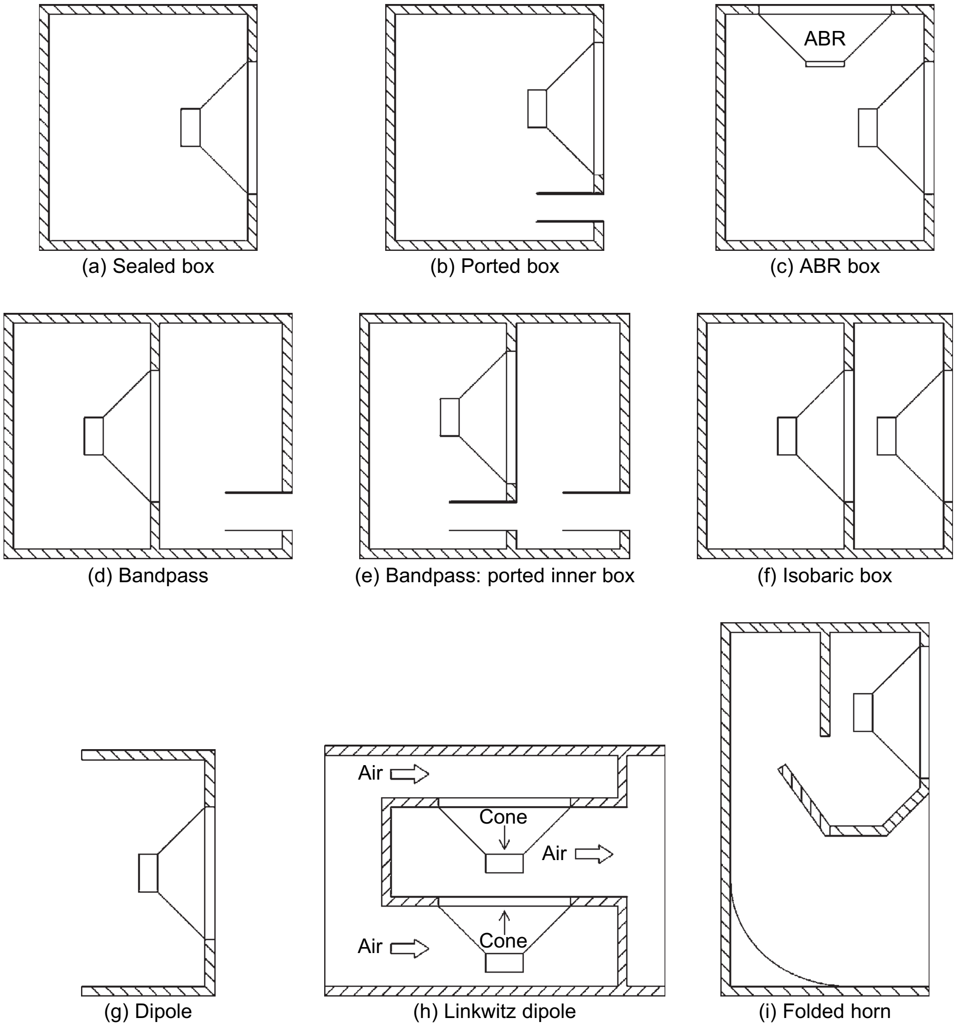


Figure 15.1: The common types of subwoofer enclosure.

however as the driver becomes unloaded and its excursion increases. The transient response of a ported subwoofer is usually worse than that of a sealed box system using the same driver. The drivers used in ported subwoofers usually have a Q_{ts} value in the range 0.2 to 0.5.

Because of the low frequencies involved, large amounts of air are in motion, and the port should have the largest practicable diameter and be flared at both ends to minimise chuffing noises from turbulent air-flow. It is often recommended that air velocity should be kept below 25 metre/sec, though maxima between 15 and 34 metre/sec have also been quoted.

15.2.3 Auxiliary Bass Radiator (ABR) Subwoofers

In auxiliary bass radiator subwoofers (also called passive radiator subwoofers) the mass of air in a port is replaced by the mass of the ABR cone. The response is therefore very similar to that of a ported subwoofer using the same driver, but with a notch in the response corresponding to the resonance frequency of the ABR. The larger the ABR, the more mass its cone will have, and the lower its resonance frequency will be for the same target F_b, (the resonance frequency of speaker and box combined) pushing the notch further out of the subwoofer passband. The large size of suitable ABR subwoofer units means that they are commonly mounted on a different face of the box from the main drive unit, as shown in Figure 15.1c, where the ABR is mounted on the top. Two ABR units are sometimes combined with a single active driver, occupying three sides of the box.

15.2.4 Transmission Line Subwoofers

The transmission line approach is not a very practical technology for subwoofers. To get worthwhile reinforcement the transmission line needs to be a quarter of a wavelength long at the frequency of interest, and at subwoofer frequencies that means an impractically long duct. The wavelength of sound at 25 Hz is 13.8 metres, and a quarter of that is 3.4 metres; that is obviously impractical as a straight duct, unless you live in a cathedral, and even if folded, as is usual in transmission line speakers, it will take up a lot of space. This means that transmission line subwoofers are relatively rare, but it seems that no audio technology is without its supporters. The Wisdom Audio transmission line subwoofer [2] has a duct folded once and is 90 inches high (2.3 metres). Drivers used with transmission line loudspeakers normally have a low Q_{ts} in the region of 0.25 to 0.4.

15.2.5 Bandpass Subwoofers

This kind of enclosure is specific to subwoofers because of its very limited bandwidth. The bandpass enclosure in Figure 15.1d consists of a sealed chamber mounting the drive unit with a second ported chamber in front of it. This has a second-order lowpass roll-off (12 dB/octave) at low frequencies and a second-order highpass roll-off at its upper frequency limit, and is sometimes called a fourth-order bandpass enclosure. I think this is unnecessarily confusing as there is no actual fourth-order filtering or response, and it would be far better to call it a second-order + second-order system, or something similar. Here I will use 2nd + 2nd order to describe it. The second-order lowpass roll-off gives a more gently falling bass response than that shown by ported or ABR subwoofers, but gives a lower cutoff frequency for the same driver and volume than the sealed-box approach; it is often claimed that the low-frequency response can be extended by about an octave compared with the sealed box alignment. The box volume required for a given cutoff frequency can be further reduced by using two drivers in an isobaric configuration. The bandpass principle is not a new idea; it was first patented in 1934. Variations on the concept can be made by replacing the ports with auxiliary bass radiators (ABRs).

2nd + 2nd order bandpass subwoofers have good power handling characteristics at low frequencies because the driver cone never becomes unloaded. It is generally considered that the transient response is second only to that of sealed box subwoofers. In a bandpass subwoofer of this sort, all of the output is via the port, which therefore needs to have the largest practicable diameter to minimise air-flow noise. The port ends are often flared to reduce turbulence; the larger the flare radius the better the results.

Figure 15.1e shows a 3rd + 3rd order bandpass subwoofer. Now both the front and rear chambers are tuned by ports. As for a straightforward ported subwoofer, power handling is poor at frequencies below the passband due to lack of loading on the rear of the drive unit cone. The transient performance of a 3rd + 3rd order bandpass subwoofer is usually inferior to either sealed, ported or 2nd + 2nd order bandpass systems, so this type of bandpass subwoofer is more appropriate to sound reinforcement applications than to hi-fi. Once again the entire acoustical output is by way of the outer port, which must have the largest practicable diameter. The inner port also needs to be designed with the issue of port noise in mind, but it is less critical because its noise output will be attenuated by the lowpass action of the front chamber.

It is also possible to design a 4th + 4th order bandpass subwoofer, but there seems to be a consensus that the transient response is so poor that any improvements, such as in bass extension, are irrelevant.

15.2.6 Isobaric Subwoofers

In an isobaric subwoofer there are two drive units acoustically in series, driven by the same electrical signal. One version is shown in Figure 15.1f. At low frequencies the two cones move together, and are equivalent to a single drive unit with twice the cone mass. This reduces the resonant frequency to about 70% of what it would have been with one drive unit of the same type working in a sealed box with the same volume. The same cutoff frequency can be achieved with half the box volume, though the volume of the chamber between the two drive units must be added; the air volume in this chamber has no acoustic function beyond coupling the drive units so it is desirable to make it as small as the physical constraints allow.

At higher frequencies cone break-up will occur, and the drive units no longer act as one unit. Other configurations are possible with the drive units mounted to front-to-front instead of front-to-back; this requires the phase of one drive unit to be reversed. There seems to be a general feeling that isobaric loudspeakers are a new idea, but this is not so. Once again, the technique is older than you might think. It was introduced by Harry F. Olson in the early 1950s.

15.2.7 Dipole Subwoofers

The concept of the dipole subwoofer is quite different from the other subwoofer configurations we have looked at. They all aspire to be monopole radiators—ideally a single point source of sound that has the same polar response in all directions. A large area source, or a multiple source, causes reinforcements and cancellations in the sound pressure level that causes major amplitude response irregularities. The rear radiation from the drive unit must be dealt with in some way, either by suppressing it (sealed boxes) or getting it into phase with the front radiation to reinforce the output (ported boxes and transmission lines).

A dipole subwoofer does neither of these things; instead the rear radiation is allowed out into the room without modification. Typically one or more large-diameter drive units are mounted on a relatively small baffle, as in Figure 15.1g; this means that the bass will begin a 6 dB/octave roll-off relatively early, because of cancellation around the edges of the baffle. To counteract this, drivers with a high Q_{ts} are sometimes used to give a peaky underdamped response that lifts the overall response before roll-off; adding dipole equalisation to an active crossover is a far better approach as it gives complete controllability and eliminates the need for special drivers that may be hard to get. This usually consists of a low-frequency shelving characteristic, where the amount of boost plateaus at low frequencies to avoid excessive cone excursions; more details are given in Chapter 11 on equalisation.

When the path around the baffle edge from front to rear of the drive unit equals a wavelength, front and back radiation are in phase and will reinforce, giving a 6 dB peak in the response; the frequency at which this happens can be simply calculated from the baffle dimensions. However, reinforcement only happens at a single frequency when the baffle is circular, so all paths are of equal length, and this is not likely to be a practical shape. The open-backed folded baffle shown in Figure 15.1g gives a more complex response but is far more compact and structurally rigid.

Figure 15.1h shows a dipole subwoofer design put forward by Siegfried Linkwitz [3] which gives greater baffle path lengths and is intended for stereo subwoofer operation. It uses two drive units working in anti-phase, which should offer at least the possibility of cancelling some even-order distortion products.

Since a dipole subwoofer has effectively two point sources, these interact to give a highly non-uniform polar response with a "figure of eight" shape. The output is greatest directly in front and behind the baffle, decreasing to a zero null at each side, where the front and rear waveforms cancel each other out. The presence of these polar response nulls is held to be the major advantages of a dipole bass system, it being claimed that the much reduced sideways radiation excites fewer room modes and so leads to a smoother overall amplitude response in the room.

15.2.8 Horn-Loaded Subwoofers

As with horn-loaded loudspeakers for higher frequency ranges, this type of subwoofer uses a horn as an acoustic transformer to match the impedance of the drive unit cone to that of the air; it makes the driver cone appear to have a much greater surface area than it actually does. This increases the efficiency of electric-acoustic considerably; the efficiency of a sealed-box design may only be 1%, but a well-designed horn loudspeaker can give 10% or more. Obviously this makes a radical difference to the number of power amplifiers you have to haul to the gig.

The size of horn required depends on the wavelength of sound being handled, and so subwoofer horns are very big- impractically so for most home entertainment purposes. The low-frequency cutoff of a horn depends on both its mouth diameter and the rate at which it expands along its length, known as the "flare-rate." A cutoff frequency of 40 Hz requires something like a mouth diameter of 2.5 metres and a length of some 4 metres. Even for sound reinforcement applications such a straight horn would be infeasibly huge, and so the horn is normally folded, as shown in Figure 15.1i (there are many ways to fold a horn and this is just one example) This can lead to amplitude response irregularites at the upper end of the operating range due to resonances and reflections as the sound output makes its way round the corners in the horn.

15.3 Subwoofer Drive Units

As we have noted, subwoofers are typically very inefficient and require driver units with considerable power handling ability. Driver cone excursion increases at 12 dB/octave with decreasing frequency for a constant sound pressure (SPL) because four times as much air has to be moved, so a large X_{max} (maximum linear excursion of the cone) and X_{mech} (maximum physical excursion before physical damage) are required. The voice coil must also be designed to withstand considerable thermal stress.

15.4 Hi-fi Subwoofers

When the main format for music delivery was vinyl, the first problem to be overcome in reproducing loud and deep bass was to get the information off the wretched disc. The lower the frequency, the greater the amplitude of the groove deviations for a constant level, and the greater the chance that mistracking of the stylus would occur. From subsonic up to about 1 kHz, a limit on groove amplitude is the constraint on the maximum level that can be cut on the disc. The welcome appearance of the CD format meant that much greater levels of clean, low bass could be accessed, and this gave a great stimulus to the development of subwoofers and the pursuit of an extended bass response in general.

If the subwoofer approach is applied to upscale music-listening rather than an audio-visual experience, it is normal not to take chances with the possibility of losing low-frequency stereo information, and two subwoofers are used, for left and right in the usual way. The subwoofers are often placed under the main speakers, or very close to them, to preserve what stereo cues can be extracted from their output. They are not placed almost at random in the listening room in the way that mono subwoofers often appear to be. A classic application of stereo subwoofers is the extension of the bass response of electrostatic loudspeakers, notably those by Quad, such as the ESL-57 introduced in 1955 and the later ESL-63.

Since the technology of the hi-fi and the home entertainment subwoofer are similar, they are dealt with together in the next section.

15.5 Home Entertainment Subwoofers

When the emphasis is watching television rather than listening to music, it is more common to use a single subwoofer. In multi-channel formats the extra directional information from rear and centre channels means that any lack of stereo in the deep bass is more likely to go unnoticed, and a single subwoofer takes up less space and is easier to fit into a room. In this application the drive units are typically between 4 and 15 inches in diameter.

Table 15.1 below gives the vital statistics of a handful of home entertainment subwoofers picked pretty much at random from those on the market now (2010). This does not in any way claim to be a representative selection, but it does give some feel for the basic subwoofer format. Note that ABR stands for Auxiliary Bass Radiator.

Domestic considerations require the subwoofer to use as small a box as possible, while at the time being capable of reproducing deep bass. This means that efficiency is inevitably low, and powerful amplifiers are needed to generate the desired sound levels—considerably more powerful than those driving the main loudspeakers. It is common for the subwoofer amplifier to have ten times the power capability in Watts compared with the main amplifiers.

Figure 15.2 shows the block diagram of a typical hi-fi subwoofer with its electronics. The facilities offered are subject to some variation, but typical features you might expect to find are:

Low-level inputs (unbalanced)
Low-level inputs (balanced)
High-level inputs
High-level outputs
Mono summing

Table 15.1: Specs for Some Current Subwoofer Designs on the Market

Model	Driver Diameter cm	Driver Orientation	Box Size H × W × D cm	Box Type	Amplifier Power W rms
Monitor Audio Vector VW-8	20	Forward	32 × 28 × 28	Ported	100
Velodyne Impact-Mini	16.5	Forward	25 × 25 × 30	Sealed	180
B&W ASW610	25	Forward	31 × 31 × 31	Sealed	200
Wilson Benesch Torus	36	Upward	45 × 90 × 30	Sealed	200
Energy ESW-M6	1 × 16.5 active 2 × 16.5 passive	Forward & sides	20 × 20 × 20	ABR	200
Audio Pro B1.36	25	Forward	45 × 35 × 38	Ported	200
Wharfedale Diamond SW250	25	Downward	42 × 42 × 38	Sealed	250
Mordaunt-Short Mezzo 9	2 × 20	Forward	32 × 34 × 35	Sealed	375
Velodyne SPL-1500R	38	Forward	47 × 46 × 44	Sealed	1000

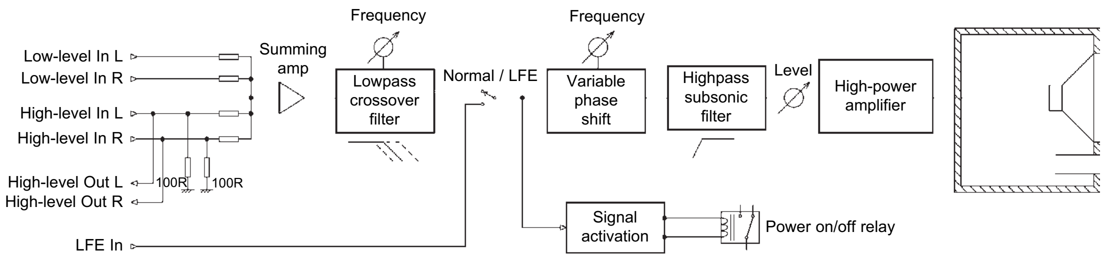


Figure 15.2: Block diagram of a typical home-entertainment subwoofer.

LFE input
Level control
Crossover in/out switch (LFE/normal)
Crossover frequency control (lowpass filter)
Highpass subsonic filter
Phase switch (normal/inverted)
Variable phase (delay) control
Signal activation out of standby

15.5.1 Low-Level Inputs (Unbalanced)

The low-level inputs are intended to be driven from a preamplifier or AV processor. In many cases they are phono (RCA) connectors and so are inherently unbalanced. The input impedance should not be less than 10 kΩ.

15.5.2 Low-Level Inputs (Balanced)

More upscale subwoofers are likely to have balanced inputs which will reject ground noise caused by ground loops, etc. This means paying for an XLR connector, but the cost of the electronics to implement the balanced function is small. Much more information on balanced inputs can be found in Chapter 16 on line inputs and outputs.

15.5.3 High-Level Inputs

The high-level inputs are designed to be connected directly to the amplifier outputs that feed the main loudspeakers. They drive a resistive attenuator, usually with high resistor values to reduce power dissipation, which reduces the incoming level to that of the low-level inputs. Input protection clamping diodes are often fitted to prevent damage from excessive input levels. The circuitry is often arranged to sum the low-level and high-level inputs so that a switch between low-level and high-level inputs is not necessary; this can be done by the same circuitry that sums the Left and Right halves of the incoming signal. This approach is made practical by making the input impedance of the high-level inputs quite low, so they are not liable to pickup external noise. A typical loading resistance value is 100 Ω; the downside is that such a resistor will wastefully dissipate a lot of power when connected to a powerful amplifier, and it needs to be a substantial wire-wound component.

An important objection to this method is that the signals reaching the high-level inputs have passed through the main power amplifiers, and will be degraded by whatever noise, hum and distortion those power amplifiers introduce. The signal has then to go through the subwoofer amplifier, so it is degraded twice instead of once. For these reasons, the use of high-level inputs should be avoided if possible, and manufacturers that provide them state in their instruction manuals that the use of the low-level inputs is preferred.

15.5.4 High-Level Outputs

High-level outputs are sometimes also provided, so the subwoofer and main loudspeakers can be daisy-chained via high-level inputs instead of being connected in parallel. This means that the power amplifier signal has to go first to the subwoofer, and then out again to main loudspeakers, passing through twice as many connectors and extra lengths of cable. If the subwoofer is installed away from the rest of the system, then the extra length of speaker cable may be considerable, and this will increase the impedance seen by the loudspeakers and reduce the so-called "damping factor" of the system. As is now well established, the actual effect of speaker cable resistance on loudspeaker damping is very small, because most of the resistance is in the voice-coils, but a real and more worrying effect is irregularities in the frequency response caused by variations in loudspeaker impedance interacting with cable resistance.

15.5.5 Mono Summing

Both the low-level and high-level inputs are in stereo, and so must be summed to mono before they are presented to the crossover and power amplifier. As mentioned above, the same summing circuit is often used to sum the low-level and high-level inputs, to save on a selector switch.

15.5.6 LFE Input

LFE stands for Low Frequency Effects. There will only be one LFE input connector as the LFE channel (which is generated by an AV processor) is already in mono. The input is typically unbalanced; the input impedance should not be less than 10 kΩ. An LFE channel is already lowpass filtered and does not pass through the subwoofer crossover.

15.5.7 Level Control

This adjusts the volume of the subwoofer relative to the rest of the system.

15.5.8 Crossover In/Out Switch

This is sometimes labelled "Normal/LFE" as when the input is the LFE channel the internal subwoofer crossover is not used.

15.5.9 Crossover Frequency Control (Lowpass Filter)

The filtering in a subwoofer is not strictly speaking a crossover because instead of splitting the input between two or more outputs, it simply rejects all high frequencies; it is more accurate to simply call it a lowpass filter. The cutoff frequency is always adjustable, and a typical range is 50 Hz–150 Hz. The filter is normally a simple second-order Butterworth lowpass using the Sallen & Key configuration. See Chapter 8 for more information on variable-frequency filters.

15.5.10 Highpass Subsonic Filter

Many subwoofers use reflex (ported) enclosures to obtain more bass extension. At very low frequencies these enclosures put no restraint on cone movement, and when this factor is combined with the high output capability of subwoofer power amplifiers, you have a recipe for disaster; large subsonic signals (oops, dropped the needle) will almost certainly cause excess cone displacement and serious damage. For this reason most subwoofers—and especially those of the reflex type—include subsonic filters in the signal path. These are usually second, third, or fourth order, and usually based on the Sallen & Key configuration.

15.5.11 Phase Switch (Normal/Inverted)

This control puts the subwoofer output in phase or 180 degrees out of phase with the incoming signal material. The intention is to try to cope with the fact that the difference between the distance from the main loudspeakers to the listener, and the distance from the subwoofer to the listener, is unpredictable, and the resulting time delay must be compensated to give good subwoofer integration. A phase-invert switch can only do this very crudely, as what is really required is a continuously variable control. A passage like this appears in most subwoofer user manuals:

"There is no correct or incorrect setting of the phase switch. The proper setting depends on many variables such as subwoofer placement, room acoustics, and listener position. Set the phase switch to maximize bass output at the listening position."

This sort of thing is a bit disingenuous. There certainly will be a correct phase-shift which gives the flattest and best bass response, but it is unlikely to coincide with either of the two arbitrary settings provided by a phase-invert switch.

15.5.12 Variable Phase Control

As just explained above, a phase-invert switch is very much a gesture at achieving correct phase-compensation. What is really required is a variable amount of phase-shift. See Chapter 10 on variable delays in time-domain filtering to see how this can be accomplished.

15.5.13 Signal Activation Out of Standby

Increasing attention is being paid to economy in the use of energy, and it is now common for power amplifiers to have a standby mode where the main transformer is disconnected from the supply when the unit is not in use, but a small transformer remains energised to power circuitry that will bring the unit out of standby when either an incoming signal is detected or a +12 V trigger voltage is applied to a control input. This also applies to the powerful amplifiers found in subwoofers, and they commonly include a signal activation facility. It is convenient to have the unit wake-up automatically when a signal is applied. It is now only necessary to pop in a CD or whatever, and start it playing; there is no need to push a button on the subwoofer. This has particular force as subwoofers are likely to be hidden away out of sight where they cannot easily be got at.

A full discussion of the technical challenges presented by signal activation, for example the need to avoid switch-on in response to isolated noise clicks, is given in my book on power amplifiers [4].

15.6 Home Entertainment Crossovers

So far as the crossover is concerned there are two modes of subwoofer operation. In 5.1 systems there is a dedicated channel for the lowest frequencies called the LFE (Low Frequency Effects) channel. This is the "0.1" in 5.1 and 7.1 discrete surround systems; it is important to realise that it contains low frequency bass material not included in the other channels, and so cannot be derived from the other channels by filtering. AV receivers and processors therefore have a mono output labelled "Subwoofer Out" or "LFE Out." The LFE channel has a digital brickwall lowpass filter at 120 Hz, and to keep clear of this producers normally use their own gentler lowpass roll-off at around 80 Hz.

AV processors have so-called "bass management" facilities which can redirect bass from any channel to a subwoofer output. (This output is *not* the same as an LFE output, because, as I said, the LFE channel contains unique information.) This is useful if no LFE channel exists, as in a normal stereo signal in Format 2.0.

AV processors classify the main loudspeakers into "Large" and "Small" with Large being considered capable of handling the full audio bandwidth, while Small speakers cannot cope with the lowest frequencies so their signal is subjected to a highpass rolloff, as part of the LFE crossover function. The unused low frequencies are mixed and sent to the LFE output; the phase-shifts between the low frequency signals in the satellite channels are generally small, so simple summation of the signals is all that is required. Processors differ in the amount of control they offer over the crossover frequency. The possibilities are Fixed, Variable, and Multiple Variable, which work as follows.

15.6.1 Fixed Frequency

The only choice is between Large or Small main loudspeakers. When "Small" is selected, the crossover frequency is fixed, usually in the range 80 to 90 Hz. This restricted approach is considered obsolescent.

15.6.2 Variable Frequency

There is the choice of Large or Small for the main loudspeakers, plus ability to alter a single crossover frequency for optimal subwoofer and main loudspeaker integration. This facility is available on most modern AV receivers and processors. A typical range of crossover frequency is 80 to 160 Hz.

15.6.3 Multiple Variable

There is the choice of Large or Small for the main loudspeakers, plus the extra flexibility of being able to set different crossover frequencies for multiple speakers, such as the front, centre, and rear channels.

15.7 Power Amplifiers for Home Entertainment Subwoofers

Because of the powerful amplifiers required, more efficient operating modes than Class-B are commonly used. When quality is the first consideration Class-G is preferred [5]. When keeping down heatsink size and overall weight are the priorities, Class-D amplifiers are used. Despite considerable development effort, Class-D amplifiers remain much inferior in their distortion performance to other types of power amplifier, and it is often stated that they are only acceptable for subwoofer applications because their high-order distortions are not reproduced by the drive unit or units. This does overlook the possibility of intermodulation generating non-linearity products in the lower frequencies.

Class-G amplifiers, on the other hand, are capable of linearity competitive with all but the very best Class-B power amplifiers, so long as modern design approaches are used minimise distortion [5].

15.8 Subwoofer Integration

The output of a subwoofer must combine properly with the bass output of the main loudspeakers. This issue is often ignored simply because it is complex and the problems are not easy to solve [3]. There are many factors to consider in integrating the LF and MID outputs in a conventional loudspeaker, where the drive units are in a fixed physical relationship to each other. Subwoofer integration adds the complications of summing stereo to mono, and the variable placing of the subwoofer relative to the main loudspeakers. The very presence of a phase-invert switch, and often a variable-phase control, indicates the uncertainty in the situation.

Subwoofers can integrate with the main loudspeakers in two ways—Augmentation and Crossover. In Augmentation mode the main loudspeakers handle the full frequency range, down to the bottom of the bass end, but are limited in their ability to reproduce deep bass at high levels without distortion. The subwoofer supplements their output, typically below 80 Hz. There is a body of opinion that some rooms can benefit from smoother bass by getting deep bass from three room locations, that is, the subwoofer plus the main left and right speakers, rather than from the subwoofer alone.

In Crossover mode, the subwoofer replaces the deep bass output from the main loudspeakers. There is a crossover from the main loudspeaker LF drive units to the subwoofer, just as there is from the MID to the LF drive units within the main loudspeakers. A highpass filter is therefore applied to the main loudspeaker signal, removing the low bass.

A very important consideration is the choice of the crossover frequency between the main loudspeakers and the subwoofer. In current variable crossovers the frequencies available range from 10 Hz to 4.2 kHz, though it is not easy to see how the latter figure counts as subwoofer territory. It is usually best to make the crossover frequency between the subwoofer and the main loudspeakers as low as possible; certainly below 90 Hz, and preferably below 60 Hz if this can be done without overtaxing the main loudspeakers, as this reduces the chance of the subwoofer becoming audible as a separate sound source. This is less important if the subwoofer is at the front of the room between the main loudspeakers, as that is usually where the bass is placed in the stereo stage. Let us examine the various factors affecting the choice of crossover frequency.

The advantages of higher crossover frequencies for subwoofer integration are:

1. Less output is required from the main loudspeakers, which should give lower distortion, and less thermal compression where it is more audible in the audio spectrum; this does of course assume that the subwoofer is itself capable of reproducing low frequencies cleanly, which it certainly should be.
2. Since a subwoofer has a greater flexibility of placement than the main loudspeakers, it may be possible to position it so that room standing waves are excited less. The higher the crossover frequency, the more likely this is to be successful.
3. For a given crossover alignment the group delay of the crossover filters will be inversely proportional to the crossover frequency. Thus the group delay for a 50 Hz crossover frequency will be twice that of a 100 Hz crossover frequency. It is however highly unlikely that this could be audible.
4. When the main channels have speaker size set to "Small" a typical AV processor sums all these channels with the LFE channel and this combined signal goes through a lowpass crossover filter. If this crossover frequency is lower than 80 Hz (the practical upper limit of the LFE channel), then information at the upper end of the LFE bandwidth will be lost.
5. In the right conditions, the use of a subwoofer can give a more consistent low frequency response. When the bass from the main loudspeaker channels is routed to a subwoofer, the frequency response over the subwoofer operating range will be identical between channels (though not necessarily good in itself). As the frequency range of the subwoofer becomes wider, this advantage grows in significance. In contrast, the bass signals radiated by the main loudspeakers will have differing amplitude responses because they are in different physical locations in the listening space.

The disadvantages of higher crossover frequencies for subwoofer integration are:

1. As higher frequencies are more audible than lower frequencies, a higher crossover frequency will place the crossover transitions in a more sensitive part of the audio spectrum. Any response irregularities induced by the crossover will therefore be more obvious.
2. It is normal to use a single subwoofer, and so the higher the crossover frequency, the greater the chance there is that stereo bass information will be lost. Stereo low frequency information will most probably be in the form of phase differences between channels as opposed to level differences.
3. It is clearly important not to operate the subwoofer above its intended frequency range. This will almost certainly lead to increased distortion and greater frequency response anomalies.
4. The crossover needs to have a suitably steep lowpass slope so that the higher bass frequencies that are easier to localise (say above 100 Hz), are not rendered in mono. A fourth-order Linkwitz–Riley configuration is recommended.

The greatest problems with subwoofer integration arise when a mono subwoofer is placed some distance from the main speakers. As we have seen in Chapter 10, time-alignment of the drivers is crucial for obtaining the desired amplitude/frequency response. It is difficult enough when all the drivers are mounted in the same enclosure, but separate stereo subwoofers may be metres away, requiring considerable time-delay compensation, and a mono subwoofer makes things worse because it may not be symmetrically placed between the two main speakers and therefore different time compensation delays would have to be applied to each main channel.

With a metre or more spacing the delays required are much longer, but implementing them is not too hard because the allpass filters only need to maintain a constant delay up to a relatively low frequency—that at which the subwoofer output has become negligible compared with the output of the LF drive units of the main loudspeakers.

15.9 Sound Reinforcement Subwoofers

Subwoofers are now pretty much standard in quality sound reinforcement systems. The requirements are very different from those for home entertainment; the output levels are much higher, and the need to have the smallest possible box volume is less pressing (though not jettisoned altogether) and so they are usually much bigger. Sound reinforcement subwoofers are usually of the sealed-box, ported-box, or horn-loaded types. The drive units are typically between 8 and 21 inches in diameter.

It is normal to use multiple subwoofers (up to a hundred at a time have been used) to get the sound output required, and this multiplicity can be exploited to beam the sound in the desired direction. This not only increases efficiency but reduces the likelihood of bitter complaints from outside the venue. Several techniques for steering the sound energy are used; most of them require signal delays to do this, and this is where the crossover setup comes in. For reasons of space only the two most important methods are described here; it is a big and a most fascinating subject.

15.9.1 Line or Area Arrays

Installing multiple subwoofers in a vertical line array focuses the sound energy into a narrow beam so that a relatively small amount is sent up into the air or down towards the floor; most of it is focused at the audience. The reduction in the amount of low frequency sound reflected from the ceiling (if working indoors) reduces frequency response and feedback problems. The longer the array, the greater the directional effect; note that this technique is implemented simply by the physical placement of the subwoofers and signal delays are not required.

Giving the speaker array greater numbers horizontally as well gives an area array, which also focuses the sound radiation into a beam in the horizontal plane; this effect can become a problem rather than an advantage if a long line of subwoofers is require to get enough output, as the beam becomes too narrow to cover the whole audience. If however the signals to the outer subwoofers are delayed by a few milliseconds prior relative to those at the centre, considerable control can be exercised over the beam width. This technique is sometimes called a “delay-shaded array.”

15.9.2 Cardioid Subwoofer Arrays

The various types of cardioid subwoofer array (CSA) are, as the name implies, more concerned with altering the front/back ratio of output power rather than controlling beam width. The polar response is very similar to that of a cardioid microphone. One method of creating a cardioid pattern is applicable to a horizontal subwoofer array across the front of the stage. The polarity of every third subwoofer is reversed (this can be done simply by turning the cabinet around so it’s firing backwards) and the signals to these are delayed. The result is that the radiation pattern is no longer quasi-omnidirectional; instead the sound energy being sent backward to the stage is much reduced, reducing feedback problems and making the lives of the performers more tolerable. The amount of delay can be manipulated to maximise the cancellation of the most troublesome frequencies in stage area. The technique is only effective over a limited frequency range, but the delays can usually be adjusted to give a significant improvement over slightly more than an octave.

15.9.3 Aux-Fed Subwoofers

Subwoofers are sometimes used in sound reinforcement quite separately from the main loudspeaker system with its active crossovers. In what is called “aux-fed” operation, the subwoofer is fed via a lowpass filter (which may or may not be part of the main crossover system, but is at any rate handling only the subwoofer signal) from a dedicated auxiliary send on the mixing console. This send is used to make a mix specifically of those instruments with the greatest amount of bass output, such as bass guitar, kick drum, and keyboards. Aux-fed operation is claimed to give a cleaner sound, because microphones tend to pick up bass information that is not intended for them, with unpredictable phase delays. This low frequency rubbish goes into a conventional stereo mix and is fed to the subwoofers, producing what is usually described as a "muddy" effect; the bass-cut filters normally present on mixer channels are generally not considered effective at controlling this because of their limited slope. (Usually 12 dB/octave) if the subwoofer is aux-fed then the unwanted low frequencies are filtered out by the main crossover system and never reach the subwoofer.

A disadvantage is that in most mixing consoles, fading down the entire audio system will require an aux master knob to be turned down to fade the subwoofers at the same time as the main faders are pulled back. This is obviously undesirable and consoles specially designed for aux-feeding have an extra fader placed next to the main output faders so easy simultaneous control is possible.

15.10 Automotive Audio Subwoofers

The loudspeakers fitted as standard to cars have a very limited low-frequency capability because of their small size; they are installed in car doors or dashboards so very little space is available. If some serious bass is required, one or more subwoofers are installed in the boot or back seat space.

Getting decent sound in a car by any means is a serious challenge. The cabin volume is much smaller than the average listening room, and consequently the effects of resonances and reflections are much more severe. Vance Dickason [6] describes the listening space as a "lossy pressure field." A true pressure field would have perfectly rigid walls, but the thin metal panels of a car are a long way from rigid and Vance points out that this leads to unpredictable variation in the low-frequency response of the space over a 3 to 6 dB range. Simply opening a window (more optional in these days of wide-spread air-conditioning) has a radical effect on the response of the space.

The resonances and reflections in the listening space can give a considerable lift to low frequencies; this is sometimes called "cabin gain." Vance reports one test that yielded a boost of 7 to 8 dB between 40 and 50 Hz, and a frightening 20 dB boost at 20 Hz. This sort of thing has obvious implications for crossover design; an unexpected hefty bass note from a CD could pop your windows out, so some effective subsonic filtering is an extremely good idea. If you are implementing variable equalisation, bear in mind that while it very often comes only in the form of bass boost, in this case the ability to cut the bass is very necessary. Circuitry that can give 20 dB of attenuation at 20 Hz, 8 dB at 40 Hz, and very little at, say, 60 Hz is going to need at least 12 dB/octave slopes and will need something more sophisticated than the conventional Baxandall configuration. See Chapter 11 on equalisation.

Automotive subwoofer enclosures are either sealed boxes or ported designs. Transmission-line and horn-loaded subwoofers are not generally used, mostly because the physical size required for low-frequency operation is quite impossible to fit into any normal-sized car.

One of the main problems of car audio is that only a nominal 12 V is available to power amplifiers. This, even assuming a lossless amplifier, only allows 2.25 W into an 8 Ω loudspeaker. Speaker impedances of 4 Ω give only 4.5 W and lower impedances than this do not give worthwhile improvements because the losses in amplifiers and wiring resistances become large. Using a bridged pair of power amplifiers, driving each side of the speaker in anti-phase, doubles the voltage swing available and so theoretically quadruples the power, giving 9 W into 8 Ω and 18 W into 4 Ω, though because of amplifier losses the real increase will be significantly less than 4 times. When higher powers than this are required, the normal practice is to use a switch-mode power supply to convert 12 V into whatever higher voltage is required.

Many after-market automotive amplifiers are of high power, to drive relatively small subwoofer enclosures that are inevitably inefficient. A typical model might be in two-channel format, giving 2 × 350 W into 4 Ω and 2 × 700 W into 2 Ω, with a bridging facility to give 1 × 1400 W into 4 Ω. Built-in variable-frequency crossovers are often provided, providing an HF output to the main loudspeakers as well as the lowpass feed to the subwoofer. These are usually based on second-order Butterworth filters. Subsonic filters are usually included for drive unit protection; variable equalisation (invariably usually in the form of bass boost) is sometimes also provided.

Power amplification of this sort obviously makes heavy demands on a car's a electrical system. The alternator fitted as standard will typically be capable of generating 80 Amps maximum, and this is not adequate for high-power audio systems; it will commonly be replaced with a special high-output alternator that can give up to 200 Amps. These are not normally direct physical replacements and some mechanical engineering is required to fit them. An alternative approach is the "split" system where the original vehicle electrical system is left alone, and a second alternator of high-output is installed that charges a separate set of batteries dedicated to the audio system. Split systems are commonly used on emergency vehicles, though in that case the extra batteries run lighting, defibrillators, etc.

Capacitor banks, which are large numbers of high value capacitors connected in parallel, are sometimes wired across the power supply when it is feared that the batteries, with their associated wiring, will not be able to respond quickly enough to a sudden demand for current. The enormous capacitances used are measured in Farads rather than microFarads, and the values used range from 1 F to at least 50 F.

There is a distinctive genre of competition in car audio, where constructors contend simply to create the most awesome installations. Car audio competitions started in the early 1980s, with the first known event in 1981 at Bakersfield in California. While some competitions focus on sound quality and neat installation, the majority appear to be held simply to find the highest sound pressure levels; this is sometimes called "dB drag racing." In recent years the two aims, sound quality vs SPL, appear to have become almost mutually exclusive.

Cars built for dB-drag-racing are nearly undriveable as the interior is almost completely filled with extra batteries, capacitor banks, amplifiers, and loudspeakers; but they must be driven 20 feet to prove that some movement is possible. Sound pressure levels, with all cabin openings sealed, of 155 or 160 dB above threshold are commonly reached. The current world record of 180.5 dB is held by Alan Dante. This was achieved with a single 18-inch subwoofer driven by four amplifiers totalling 26 kW, powered by fifteen 16 Volt batteries. This equipment was installed in a Volvo that appears to have been weighted down with concrete. It should perhaps be mentioned that people do not sit in the cars during testing. It would not be a survivable experience.

References

[1] http://en.wikipedia.org/wiki/Thiele/Small, 2010 (accessed 12.10).

[2] Anon, RegenerativeTransmission Line™ Subwoofers. http://www.wisdomaudio.com/pdfs/RTL.pdf, 2010 (accessed 12.10).

[3] S. Linkwitz, http://www.linkwitzlab.com/frontiers_4.htm#S, 2010 (accessed 12.10).

[4] D. Self, Audio Power Amplifier Design Handbook, fifth ed., Newnes, Boston, MA, p. 573 (signal activation).

[5] D. Self, Audio Power Amplifier Design Handbook, fifth ed., Newnes, Boston, MA, pp. 344–365 (Class-G).

[6] V. Dickason, Loudspeaker Design Cookbook, fifth ed., Audio Amateur Press, Peterborough, NH, p. 159, ISBN 1-882580-10-9.

CHAPTER 16

Line Inputs and Outputs

16.1 External Signal Levels

There are several standards for line signal levels. The −10 dBv standard is used for a lot of semi-professional recording equipment as it gives more headroom with unbalanced connections—the professional levels of +4 dBu and +6 dBu assume balanced outputs which inherently give twice the output level for the same supply rails as it is measured between two pins with signals of opposite phase on them. See Table 16.1.

Signal levels in dBu are expressed with reference to 0 dBu = 775 mV rms; the origin of this odd value is that it gives a power of 1 mW in a purely historical 600 Ω load. Signals in dBv (or dBV) are expressed with reference to 0 dB = 1.000 V rms.

These standards are well established, but that does not mean that all equipment uses them as a nominal level. Many power amplifiers require more than 0 dBV for full output; the Yamaha P7000S power amplifier requires +8 dBu (1.95 Vrms) to give its full output of 750 W into 8 Ω. The "0 VU" on VU meters is nominally 1.23 Volts.

16.2 Internal Signal Levels

In any audio system it is necessary to select a suitable nominal level for the signal passing through it. This level is always a compromise—the signal level should be high so it suffers minimal degradation by the addition of circuit noise as it passes through the system, but not so high that it is likely to suffer clipping before it reaches a gain control, or generate undue distortion below the clipping level. (This last constraint is not normally a problem with modern circuitry, which gives very low distortion right up to the clipping point.)

The choice of internal levels for active crossovers is complicated by the fact that there are two or three parallel signal paths carrying signals with completely different spectral distributions. It is well known that signal levels at the upper end of the audio band are much lower than those at low and middle frequencies, which raises the question of whether an HF signal path will require more gain so that it can be run at a higher nominal internal level. This would give a better signal-to-noise ratio, which is important as noise is much more obvious in the HF path. The LF and MID paths both contain low-pass filters, which if they are appropriately placed late in the processing chain, discriminate heavily against the most audible part of the noise spectrum.

The Design of Active Crossovers

Table 16.1: Nominal Signal Levels

	V rms	dBu	dBv
Semi-professional	0.316	−7.78	−10
Professional	1.228	+4.0	+1.78
German ARD	1.55	+6.0	+3.78

This is a complicated issue and is dealt with in much more detail in Chapter 14 on crossover system design.

If the incoming signal does have to be amplified, this should be done as early as you can in the signal path, to get the signal well above the noise floor as soon as possible. If the gain is in the input amplifier, balanced or otherwise, the signal will pass through later stages at a high level and so their noise contribution will be less. However, if the input stage is configured with a fixed gain, this must be kept low as it is not possible to turn it down to avoid clipping. Ideally any input stage should have variable gain. It is not straightforward to combine this feature with a good balanced input, but several ways of doing it are shown later in this chapter.

16.3 Input Amplifier Functions

It is important that the very first thing an incoming signal meets is some form of RF filtering, to prevent RF breakthrough and other EMC problems. It must be done before the incoming signal encounters any semiconductors where RF demodulation could occur, and can be regarded as a “roofing filter.” Next, the low-end frequency response is given an early limit by DC-blocking capacitors, and in some cases overvoltage spikes are clamped by diodes. An input amplifier should have a reasonably high impedance; certainly not less than 10 kΩ, and preferably more. It must have a suitable gain—possibly switched or variable—to scale the incoming signal to the nominal internal level. Balanced input amplifiers also accurately perform the subtraction process that converts differential signals to single-ended ones, so noise produced by ground loops and the like is rejected. That’s quite a lot of functionality for one stage.

16.4 Unbalanced Inputs

The simplest unbalanced input would feed the incoming signal directly to the first stages of the crossover. This is not practical because these stages will very likely be active filters that require a low source impedance to give the expected response. In addition, the input impedance would vary with frequency and could fall to rather low values. Some sort of input amplifier which can be fed from a significant impedance without ill effect is needed.

A typical unbalanced input amplifier is shown in Figure 16.1. The opamp U1:A is a unity-gain voltage follower; it could be altered to give a fixed gain by adding two series feedback resistors. A 5532 bipolar opamp is used here for its low distortion and low noise; with the low source impedances that are likely, an FET-input opamp would be noisier by 10 dB or more. R1 and C1 are a first-order low pass filter to stop incoming RF before reaches the opamp where it would be likely to be demodulated into the audio band; once this has happened any further attempts at RF filtering are pointless. R1 and C1 must be as physically close to the input socket as possible to prevent RF being radiated inside the equipment enclosure before it is shunted to ground, and this is why they should always be the first components in the signal path.

Selecting component values for input RF filters of like this is always a compromise, because the output impedance of the source equipment is not known. If the source is an active preamplifier, then the output impedance ought to be around 50 Ω, but it could be 200 Ω or as high as 1 kΩ. If the source is one of those oxymoronic "passive preamplifiers." In other words, just an input selector switch and a volume potentiometer, and an improbably large price-tag, then the output impedance will be much higher in some circumstances. If you really must use a piece of equipment that blazons forth its internal contradictions in its very name, you will find that by far the most popular potentiometer value is 10 kΩ, with a maximum output impedance (when set for 6 dB of attenuation) of 2.5 kΩ, very much higher than the 50 Ω we might expect from a good active preamplifier. This is in series with R1 and affects the turnover frequency of the RF filter. Effective RF filtering is very desirable, but it is also important to avoid a frequency response that sags significantly at 20 kHz. Valve equipment is also likely to have a high output impedance.

Taking 2.5 kΩ as the worst-case source impedance and summing it with R1 we get 2.6 kΩ. With a 100 pF capacitor we would get −3 dB at 612 kHz; the loss at 20 kHz is a wholly negligible 0.005 dB, so we might decide that C1 could be usefully increased. If we make it

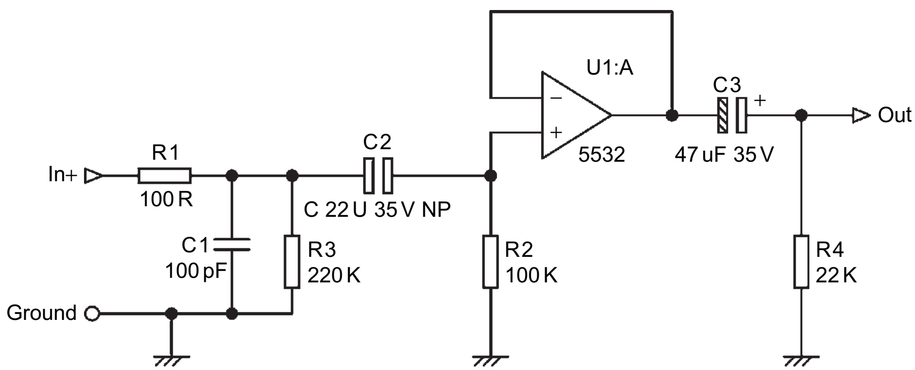


Figure 16.1: An unbalanced input amplifier with RF filter, DC drain, and input and output DC blocking.

220 pF then the 20 kHz loss is still a tiny 0.022 dB, but the −3 dB point is now 278 kHz, much improving the rejection of what used to be called The Medium Wave. If we take C1 as 220 pF and assume an active output with a 50 Ω impedance in the source equipment, then together with the 100 Ω of R1 we have 150 Ω, which in conjunction with 100 pF gives us −3 dB at 4.82 MHz. This is a bit higher than is really desirable, but it is not easy to see what to do about it. If there was a consensus that the output impedance of a respectable piece of audio equipment should not exceed 100 Ω, then things would be much easier in this area.

Another important consideration is that the series resistance R1 must be kept as low as practicable to minimise Johnson noise; but lowering this resistance means increasing the value of shunt capacitor C1, and if it becomes too big then its impedance at high audio frequencies will become too low. Not only will there be too low a roll-off frequency if the source has a high output impedance, but there might be an increase in distortion at high audio frequencies because of excessive loading on the source output stage.

Replacing R1 with a small inductor to make an LC lowpass filter will give much better RF rejection at increased cost. This is justifiable in professional audio equipment, but it is much less common in hi-fi, one reason being that the unpredictable source impedance makes the filter design difficult, as we have just seen. In the professional world one *can* assume that the source impedance will be low. Adding more capacitors and inductors allows a 3 or 4-pole LC filter to be made. If you do use inductors then it is important to check the frequency response to make sure it is as intended and there is no peaking at the turnover frequency, as you could have made a filter that does more harm than good.

C2 is a DC-blocking capacitor to prevent voltages from ill-conceived source equipment getting into the input amplifier. It is a non-polarised type as voltages from outside are of unpredictable polarity, and it is rated at not less than 35 V so that even if it gets connected to defective equipment with an opamp output jammed hard against one of its supply rails, the capacitor will not be damaged. It will give no protection against faulty valve equipment that may put out a couple of hundred volts. R3 is a DC drain resistor that prevents any charge put on C2 by external equipment from remaining there for a long time and causing a thud when connections are replugged; as with all input drain resistors, its value is a compromise between discharging the capacitor quickly and keeping the input impedance high. The input impedance here is R3 in parallel with R2, that is, 220 kΩ in parallel with 100 kΩ, giving 68 kΩ. This is a suitably high value and should work well with just about any source equipment, including that valve-based stuff.

R2 provides the biasing for the opamp input; it must be a high value to keep the input impedance up, but bipolar input opamps draw significant input bias current. The Fairchild 5532 data sheet quotes 200 nA typical, and 800 nA maximum, and these currents would cause a voltage drop across R2 of 20 mV and 80 mV respectively. This offset voltage will be reproduced at the output of the opamp, with the input offset voltage added; this is only

4 mV maximum, much less than the offset due to the bias current. The 5532 has NPN input transistors, so the bias current flows into the input pins, and the voltage at Pin 3 and hence the opamp output may be negative with respect to ground by anything up to 84 mV.

Such offset voltages do not significantly affect the output voltage swing, but they will generate unpleasant clicks and pops if the input stage is followed by any sort of switching, and they are big enough to make potentiometers crackly; the DC voltage (you know perfectly well what I mean) is therefore blocked by C3. R4 is another DC drain to keep the output at zero volts. It can be made lower in value than the input drain R3 as the only requirement is that it should not significantly load the opamp output; 22 kΩ or 47 kΩ resistors are commonly used.

FET-input opamps have much lower input bias currents, so that the offsets they generate as they flow through biasing resistors are usually negligible, but they still have input offsets of a few milliVolts, so DC blocking will still be needed if switches downstream are to work silently.

This input stage, with its input terminated by 50 Ω to ground, has a noise output of only −119.0 dBu over the usual 22–22 kHz bandwidth. This is very quiet indeed, and is a reflection of the low voltage noise of the 5532, and the fact that R1, the only resistor in the signal path, has the low value of 100 Ω and so generates very little Johnson noise; −132.6 dBu, to be precise. This noise is wholly swamped by the voltage noise of the opamp, which is basically all we see; its current noise has negligible effect because of the low circuit impedances.

16.5 Balanced Interconnections

Balanced inputs are used to prevent noise and crosstalk from affecting the input signal, especially in applications where long interconnections are used. They are standard on professional audio equipment, and are quite quickly becoming more common in the world of hi-fi. Their importance is that they can render ground loops and other connection imperfections harmless. Since there is no point in making a wonderful piece of equipment and then feeding it with an impaired signal, making sure you have an effective balanced input really is of the first importance, and I will go into it in some detail.

The basic principle of balanced interconnection is to get the signal you want by subtraction, using a three-wire connection. In some cases a balanced input is driven by a balanced output, with two anti-phase output signals; one signal wire (the hot or in-phase) sensing the in-phase output of the sending unit, while the other senses the anti-phase output.

In other cases, when a balanced input is driven by an unbalanced output, as shown in Figure 16.2, one signal wire (the hot or in-phase) senses the single output of the sending unit,

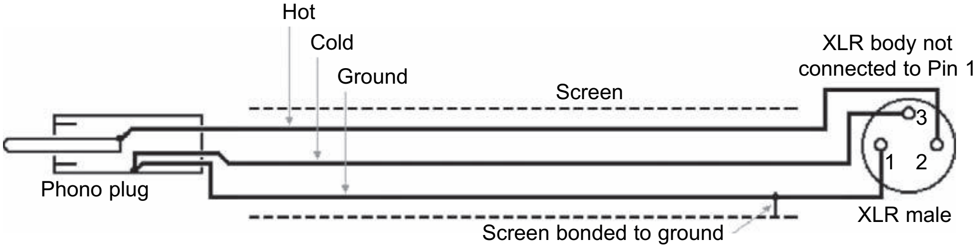


Figure 16.2: Unbalanced output to balanced input interconnection, with cold joined to ground at the unbalanced end.

while the other (the cold or phase-inverted) senses the unit's output-socket ground, and once again the difference between them gives the wanted signal. In either of these two cases, any noise voltages that appear identically on both lines (i.e., common-mode signals) are in theory completely cancelled by the subtraction. In real life the subtraction falls short of perfection, as the gains via the hot and cold inputs will not be precisely the same, and the degree of discrimination actually achieved is called the Common-Mode Rejection Ratio, (CMRR) of which more later. I should also say of Figure 16.2 that the CMRR is often much improved by putting a resistance in series with the Cold line, where it goes into the phono plug, of a value equal to the output impedance of the unbalanced output; more on that later, too.

It is deeply tedious to keep referring to non-inverting and inverting inputs, and so these are usually abbreviated to "Hot" and "Cold" respectively. This does *not* necessarily mean that the hot terminal carries more signal voltage than the cold one. For a true balanced connection, the voltages will be equal. The "hot" and "cold" terminals are also often referred to as IN+ and IN–, and this latter convention has been followed in the diagrams here.

The subject of balanced interconnections is a large one, and a big book could be written on this topic alone; two of the classic papers on the subject are by Neil Muncy [1] and Bill Whitlock, [2] both of which are very well worth reading.

To make a start, let us look at the pros and cons of balanced connections.

16.6 The Advantages of Balanced Interconnections

- Balanced interconnections discriminate against noise and crosstalk, whether they result from ground currents, or electrostatic or magnetic coupling to signal conductors.
- Balanced connections make ground-loops much less intrusive, and usually inaudible, so people are less tempted to start "lifting the ground" to break the loop, with possibly fatal consequences. In the absence of a dedicated ground-lift switch that leaves the external metalwork firmly connected to mains safety earth, the foolhardy and the optimistic will break the mains earth (not quite so easy now that moulded mains plugs

are standard) and this is highly dangerous, as a short-circuit from mains to the equipment chassis will then result in live metalwork but dead people.
- A balanced interconnection incorporating a true balanced output gives 6 dB more signal level on the line, potentially giving 6 dB more dynamic range. This is true not only with respect to *external* noise but also the noise generated by a balanced input amplifier at the receiving end of the link. As is described later in this chapter, a standard balanced input using 10 KΩ resistors is about 14 dB noisier than the unbalanced input shown in Figure 16.1 above.
- Balanced connections are usually made with XLR connectors. These are a professional 3-pin format, and are far superior to the phono (RCA) type normally used for unbalanced connections. More on this below.

16.7 The Disadvantages of Balanced Interconnections

- Balanced inputs are inherently noisier than unbalanced inputs by a large margin, in terms of the noise generated by the input circuitry itself rather than external noise. This may appear paradoxical but it is all too true, and the reasons will be fully explained in this chapter.
- More hardware means more cost. Small-signal electronics is relatively cheap; unless you are using a sophisticated low-noise input stage, of which more later, most of the extra cost is likely to be in the balanced input connectors.
- Balanced connections do not of themselves provide any greater RF immunity than an unbalanced input. For this to happen both legs of the balanced input would have to demodulate the RF in equal measure for common-mode cancellation to occur. The chances of this happening over any sort of frequency range are effectively zero.
  It remains vital to provide the passive RF filtering before the first active electronics.
- There is the possibility of introducing a phase error. It is all too easy to create an unwanted phase inversion by confusing hot and cold when wiring up a connector, and this can go undiscovered for some time. The same mistake on an unbalanced system interrupts the audio completely and leaves little room for doubt that something is amiss.
- Balanced connectors (usually XLRs) are inevitably more expensive. However, their security of connection and general quality make it money well spent in my view.

16.8 Balanced Cables and Interference

In a balanced interconnection two wires carry the signal, and the third connection is the ground wire which has two functions.

Firstly it joins the grounds of the interconnected equipment together. This is not always desirable, and if galvanic isolation is required a transformer balancing system will be necessary because the large common-mode voltages are likely to exceed the range of an electronic balanced input. A good transformer will also have a very high CMRR, which will be needed to get a clean signal in the face of large CM voltages.

Secondly, the presence of the ground allows electrostatic screening to shield the two signal wires, preventing both the emission and pick-up of unwanted signals. In cheap cables this will mean a "lapped screen," with wires laid parallel to the central signal conductors. The screening coverage is not total, and can be badly degraded as the screen tends to open up on the outside of cable bends. A braided screen around the signal wires gives better coverage but still not 100%. This is much more expensive, as it is harder to make.

The best solution is an overlapping foil screen, with the ground wire (often called the drain wire in this context) running down the inside of the foil and in electrical contact with it. This is usually the most effective as the foil is a solid sheet and cannot open up on bends. It should give perfect electrostatic screening, and it is much easier to work with than either lap screen or braided cable.

There are three main ways in which an interconnection is susceptible to hum and noise.

16.8.1 Electrostatic Coupling

An interfering signal at significant voltage couples directly to the inner signal line, through stray capacitance. The stray capacitance between imperfectly-screened conductors will be a fraction of a pF in most circumstances, as electrostatic coupling falls off with the square of distance. This form of coupling can be serious in studio installations with unrelated signals running down the same ducting.

The three main lines of defense against electrostatic coupling are effective screening, low impedance drive, and a good CMRR maintained up to the top of the audio spectrum. As regards screening, an overlapped foil screen provides complete protection.

Driving the line from a low impedance, of the order of 100 Ω or less, is also helpful because the interfering signal, having passed through a very small stray capacitance, is a very small current and cannot develop much voltage across such a low impedance. This is convenient because there are other reasons for using a low output impedance, such as optimising the interconnection CMRR, minimising HF losses due to cable capacitance, and driving multiple inputs without introducing gain errors. For the best immunity to crosstalk the output impedance must remain low up to as high a frequency as possible. This is definitely an issue as opamps invariably have a feedback factor that begins to fall from a low, and quite possibly sub-audio frequency, and this makes the output impedance rise with frequency as the negative feedback factor falls, as if an inductor were in series. Some line outputs have physical series inductors to improve stability or EMC immunity, and these should not be so large that they significantly increase the output impedance at 20 kHz. From the point of view of electrostatic screening alone, the screen does not need to be grounded at both ends, or form part of a circuit [3]. It must of course be grounded at some point.

If the screening is imperfect, and the line impedance non-zero, some of the interfering signal will get into the hot and cold conductors, and now the CMRR must be relied upon to make the immunity acceptable. If it is possible, rearranging the cable-run away from the source of interference and getting some properly screened cable is more practical and more cost-effective than relying on very good common-mode rejection.

Stereo hi-fi balanced interconnections almost invariably use XLR connectors. Since an XLR can only handle one balanced channel, two separate cables are almost invariably used and interchannel capacitive crosstalk is not an issue. Professional systems, on the other hand, use multi-way connectors that do not have screening between the pins and there is an opportunity for capacitive crosstalk here, but the use of low source impedances should reduce it to below the noise floor.

16.8.2 Magnetic Coupling

If a cable runs through an AC magnetic field, an EMF is induced in both signal conductors and the screen, and according to some writers, the screen current must be allowed to flow freely or its magnetic field will not cancel out the field acting on the signal conductors, and therefore the screen should be grounded at both ends, to form a circuit [4]. In practice the magnetic field cancellation will be very imperfect and reliance is better placed on the CMRR of the balanced system to cancel out the hopefully equal voltages induced in the two signal wires. The need to ground both ends to possibly optimise the magnetic rejection is not usually a restriction, as it is rare that galvanic isolation is required between two pieces of audio equipment.

The equality of the induced voltages can be maximised by minimising the loop area between the hot and cold signal wires, for example by twisting them tightly together in manufacture. In practice most audio foil-screen cables have parallel rather than twisted signal conductors, but this seems adequate almost all of the time. Magnetic coupling falls off with the square of distance, so rearranging the cable-run away from the source of magnetic field is usually all that is required. It is unusual for it to present serious difficulties in a hi-fi application.

16.8.3 Ground Voltages

These are the result of current flowing through the ground connection, and is often called "common-impedance coupling" in the literature [1]. This is the root of most ground-loop problems. The existence of a loop in itself does no harm, but it is invariably immersed in a 50 Hz magnetic field that induces mains-frequency currents plus harmonics into it.

This current produces a voltage drop down non-negligible ground-wire resistances, and this effectively appears as a voltage source in each of the two signal lines. Since the CMRR is finite a proportion of this voltage will appear to be a differential signal, and will be reproduced as such.

16.9 Balanced Connectors

Balanced connections are most commonly made with XLR connectors, though it can be done with stereo (tip-ring-sleeve) jack plugs. XLRs are a professional 3-pin format, and are a much better connector in every way than the usual phono (RCA) connectors used for unbalanced interconnections. Phono connectors have the great disadvantage that if you are connecting them with the system active (inadvisable, but then people are always doing inadvisable things) the signal contacts meet before the grounds and thunderous noises result. The XLR standard has Pin 2 as hot, Pin 3 as cold, and Pin 1 as ground. As described in Chapter 1, in domestic crossover use there is a good case for using multiway connectors that carry several 3-wire balanced connections, in order to cut down the amount of visible cabling.

16.10 Balanced Signal Levels

Many pieces of equipment, including preamplifiers and power amplifiers designed to work together, have both unbalanced and balanced inputs and outputs. The general consensus in the hi-fi world is that if the unbalanced output is say 1 Vrms, then the balanced output will be created by feeding the in-phase output to the hot output pin, and also to a unity-gain inverting stage, which drives the cold output pin with 1 Vrms phase-inverted. The total balanced output voltage between hot and cold pins is therefore 2 Vrms, and so the balanced input must have a gain of ½ or −6 dB relative to the unbalanced input to maintain consistent internal signal levels.

16.11 Electronic versus Transformer Balanced Inputs

Balanced interconnections can be made using either transformer or electronic balancing. Electronic balancing has many advantages, such as low cost, low size and weight, superior frequency and transient response, and no low-frequency linearity problems. Transformer balancing has advantages of its own, particularly for work in very hostile RF/EMC environments, but serious drawbacks. The advantages are that transformers are electrically bullet-proof, (and quite possibly physically bullet-proof) retain their high CMRR performance forever, and consume no power even at high signal levels. They are essential if galvanic isolation between ground is required. Unfortunately transformers can generate LF distortion, particularly if they have been made with minimal core sizes to save weight and cost.

They are liable to have HF response problems due to leakage reactance and distributed capacitance, and compensating for this requires a carefully designed Zobel network across the secondary. Inevitably they are heavy and expensive compared with an opamp and a few R's and C's. Transformer balancing is therefore relatively rare, even in professional audio applications, and the greater part of this chapter deals with electronically balanced inputs.

16.12 Common Mode Rejection Ratio (CMRR)

Figure 16.3 shows a balanced interconnection reduced to its bare essentials; hot and cold line outputs with source resistances Rout+, Rout– and a standard differential amplifier at the input end. The output resistances are assumed to be exactly equal, and the balanced input in the receiving equipment has two exactly equal input resistances to ground R1, R2. The ideal balanced input amplifier senses the voltage difference between the points marked IN+ (hot) and IN- (cold) and ignores any common-mode voltage which are present on both. The amount by which it discriminates is called the Common-Mode Rejection Ratio or CMRR, and is usually measured in dB. Suppose a differential voltage input between IN+ and IN– gives an output voltage of 0 dB; now reconnect the input so that IN+ and IN– are joined together and the same voltage is applied between them and ground. Ideally the result would be zero output, but in this imperfect world it won't be, and the output could be anywhere between −20 dB (for a bad balanced interconnection, which probably has something wrong with it) and −140 dB (for an extremely good one). The CMRR when plotted may have a flat section at low frequencies, but it very commonly degrades at high audio frequencies, and may also deteriorate at very low frequencies. More on that later.

In one respect balanced audio connections have it easy. The common-mode signal is normally well below the level of the wanted signal, and so the common-mode range of the input is not an issue. In other area of technology, such as electrocardiogram amplifiers, the common-mode signal may be many times greater than the wanted signal.

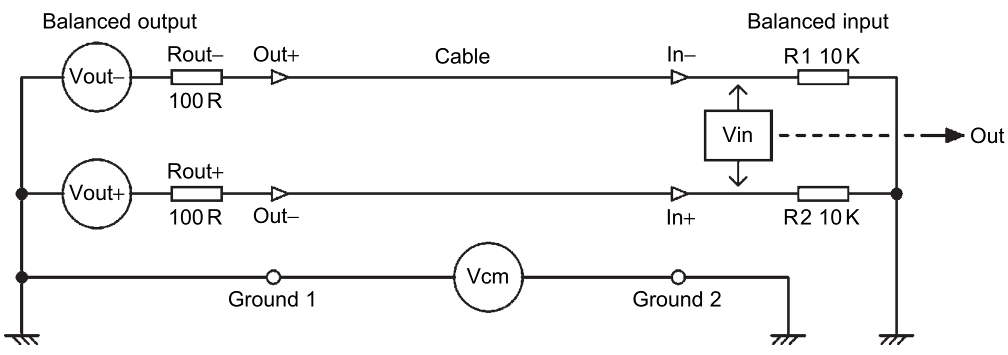


Figure 16.3: A theoretical balanced interconnection showing how the output and input impedances influence CMRR.

The simplified conceptual circuit of Figure 16.3, under SPICE simulation, demonstrates the need to get the resistor values right for a good CMRR, before you even begin to consider the rest of the circuitry. The differential voltage sources Vout+, Vout− which represent the actual balanced output are set to zero, and Vcm, which represents the common-mode voltage drop down the cable ground, is set to 1 Volt to give a convenient result in dBV. The output resulting from the presence of this voltage source is measured by a mathematical subtraction of the voltages at IN+ and IN– so there is no actual input amplifier to confuse the results with its non-ideal performance.

Let us begin with Rout+ and Rout– set to 100 Ω and input resistors R1, R2 set to 10 kΩ. These are typical real-life values as well as being nice round figures. When all four resistances are exactly at their nominal value, the CMRR is in theory infinite, but if just one of the output resistors or one of the input resistors is then altered in value by 1%, then the CMRR drops sharply to −80 dB. If the deviation is 10%, things are predictably worse and the CMRR degrades to −60 dB. The CMRR is here flat with frequency because our simple model has no frequency-dependent components.

Each line of the connection is a potential divider with Rout at the top and R1 at the bottom. With these values the attenuation on each line is the very small figure of 0.0087 dB. If we can reduce this attenuation further then the gain on each line will be closer to unity, and minor changes in resistor values will have less effect. As an example, if we increase the input resistors R1, R2 to 100 kΩ with Rout+, Rout– kept at 100 Ω then a 1% resistor deviation only degrades the CMRR to −100 dB. This change is quite practical, so long as you buffer the inputs to the actual balanced amplifier—the technique is explained in more detail below. An even higher value for R1, R2 of 1 MΩ, which is still feasible but slightly more difficult, gives −120 dB for a single 1% resistance deviation, and −100 dB for a single 10% deviation.

We could also improve CMRR by reducing the output impedances Rout+, Rout–. Dropping them to 10 Ω, and using the conventional 10 kΩ input resistors gives −100 dB for a single 1% resistor deviation. The problem is that these resistors are in the source equipment, and usually outside our control. A 10 Ω output resistor is also too low to prevent HF instability caused by cable capacitance, and would need to be supplemented by expensive output chokes. Alternatively, and much more economically, the output stage could be configured as a “zero-impedance output” as described in the section of the chapter on line outputs; an output impedance of a fraction of an Ohm at 1 kHz is very easy to achieve.

If as usual, however, we have to take the source equipment as it comes, we cannot assume the output resistors will be less than 100 Ω, and they may be a good deal higher. If the output is unbalanced then we effectively have one output resistor at, say, 100 Ω, while the other is zero as there is a direct connection between the Cold line and the output ground. This imbalance messes things up dramatically, and the CMRR collapses to 43 dB.

Time for a reality check. You will have noticed that the CMRR figures we are dealing with, of 80 or 100 dB, are much better than we measure in reality. This is because we are only altering one of four resistances—in real life all four will be subject to a statistical distribution and the CMRR results likewise come out as a statistical distribution.

There is no point in going into that level of detail here because there are other and more important influences on CMRR, which we will look at shortly. The lesson to take away here is that we need the lowest possible output impedances (if we have any say in their value) and the highest possible input impedances to get the maximum common-mode rejection. This is highly convenient because low output impedances are already needed to drive multiple inputs and cable capacitance, and high input impedances are needed to minimise loading and consequent signal losses. However… it will soon emerge that influences on CMRR are such that what we should really conclude is that balanced line impedances should be as high as possible *without compromising anything else*.

16.13 The Basic Electronic Balanced Input

Figure 16.4 shows the basic balanced input amplifier. To achieve balance R1 must be equal to R3 and R2 equal to R4. The amplifier in Figure 16.4 has a gain of R2/R1 (=R4/R3). The standard one-opamp balanced input or differential amplifier is a very familiar circuit block, but its operation often appears somewhat mysterious. Its input impedances are *not* equal when it is driven from a balanced output; this has often been commented on [5]. Some confusion has resulted.

The source of the confusion is that a simple differential amplifier has interaction between the two inputs, so that the input impedance seen on the cold input depends on the signal applied to the hot input. Input impedance is measured by applying a signal and see how much current flows into the input, so it follows that the apparent input impedance on each

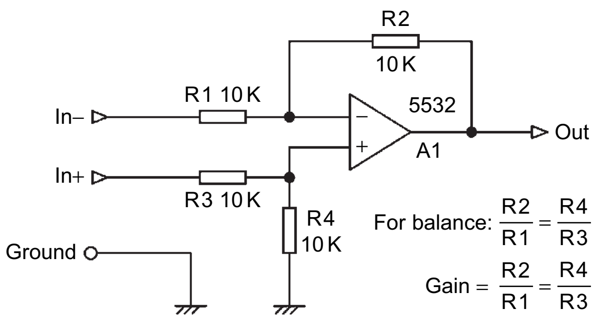


Figure 16.4: The basic balanced input amplifier, with standard 10 kΩ resistors.

leg varies according to how the cold input is driven. If the amplifier is made with four 10 kΩ resistors, then the input impedances on hot and cold are:

Some of these impedances are not exactly what you might expect, and require some explanation. They are summarised in Table 16.2.

Table 16.2: The Input Impedances for Different Input Drive Conditions

Case	Pins Driven	Hot Input Res Ω	Cold Input Res Ω
1	Hot only	20 k	Grounded
2	Cold only	Grounded	10 k
3	Both (balanced)	20 k	6.66 k
4	Both common-mode	20 k	20 k
5	Both floating	10 k	10 k

16.13.1 Case 1

The balanced input is being used as an unbalanced input by grounding the cold input and driving the hot input only. The input impedance is therefore simply R3 + R4. Resistors R3 and R4 reduce the signal by a factor of a half, but this loss is undone as R1 and R2 set the amplifier gain to two times, and the overall gain is unity. If the cold input is not grounded then the gain is 0.5 times. The attenuate-then-amplify architecture, plus the Johnson noise from the resistors, makes this configuration much noisier than the dedicated unbalanced input of Figure 16.1, which has only a single 100 Ω resistor in the signal path.

16.13.2 Case 2

The balanced input is again being used as an unbalanced input, but this time by grounding the hot input, and driving the cold input only. This gives a phase inversion and it is unlikely you would want to do it except as an emergency measure to correct a phase error somewhere else. The important point here is that the input impedance is now only 10 kΩ, the value of R1, because shunt negative feedback through R2 creates a virtual earth at Pin 2 of the opamp. Clearly this simple circuit is not as symmetrical as it looks. The gain is unity, whether or not the hot input is grounded; grounding it is desirable because it not only prevents interference being picked up on the hot input pin, but also puts R3 and R4 in parallel, reducing the resistance from opamp Pin 3 to ground and so reducing Johnson noise.

16.13.3 Case 3

This is the standard balanced interconnection. The input is driven from a balanced output with the same signal levels on hot and cold, as if from a transformer with its centre-tap grounded, or an electronically balanced output using a simple inverter to drive the cold pin. The input impedance on the hot input is what you would expect; R3 + R4 add up to 20 kΩ. However, on the cold input there is a much lower input impedance of 6.66 kΩ. This at first sounds impossible as the first thing the signal encounters is a 10 kΩ series resistor, but the crucial point is that the hot input is being driven simultaneously with a signal of the opposite phase, so the inverting opamp input is moving in the opposite direction to the cold input due to negative feedback, and what you might call anti-bootstrapping reduces the effective value of the 10 kΩ resistor to 6.66 kΩ. These are the differential input impedances we are examining, the impedances seen by the balanced output driving them. Common-mode signals see a common-mode impedance of 20 kΩ, as in Case 4 below. You will sometimes see the statement that these unequal differential input impedances "unbalance the line." From the point of view of CMRR, this is not the case, as it is the CM input impedance that counts. The line is, however, unbalanced in the sense that the cold input draws three times the current from the output that the hot one does. This current imbalance might conceivably lead to inductive crosstalk in some multi-way cable situations, but I have never encountered it. The differential input impedances can be made equal by increasing the R1 and R2 resistor values by a factor of three, but this degrades the noise performance markedly and makes the common-mode impedances to ground unequal, which is a much worse situation as it compromises the rejection of ground voltages, and these are almost always the main problem in real life.

16.13.4 Case 4

Here both inputs are driven by the same signal, representing the existence of a common-mode voltage. Now both inputs shown an impedance of 20 kΩ. It is the symmetry of the common-mode input impedances that determines how effectively the balanced input rejects the common-mode signal. This configuration is of course only used for CMRR testing.

16.13.5 Case 5

Now the input is driven as from a floating transformer with the centre-tap (if any) unconnected, and the impedances can be regarded as equal; they must be, because with a floating winding the same current must flow into each input. However, in this connection the line voltages are *not* equal and opposite: with a true floating transformer winding the hot input has all the signal voltage on it while the cold has none at all, due to the negative feedback action of the balanced input amplifier. This seemed very strange when it emerged in SPICE simulation, but a sanity-check with real components proves it true. The line has been completely unbalanced as regards crosstalk to other lines, although its own common-mode rejection remains good.

Even if absolutely accurate resistors are assumed, the CMRR of the stage in Figure 16.4 is not infinite; with a TL072 it is about −90 dB, degrading from 100 Hz upwards, due to the limited open-loop gain of the opamp. We will now examine this effect.

16.14 Common-Mode Rejection Ratio: Opamp Gain

In the earlier section on CMRR we saw that in a theoretical balanced line, choosing low output impedances and high input impedances would give very good CM rejection even if the resistors were not perfectly matched. Things are a bit more complex (i.e., worse) if we replace the mathematical subtraction with a real opamp. We quickly find that even if perfectly matched resistors everywhere are assumed, the CMRR of the stage is not infinite, because the two opamp inputs are not at exactly the same voltage. The negative feedback error-voltage between the inputs depends on the open-loop gain of the opamp, and that is neither infinite nor flat with frequency into the far ultra-violet. Far from it. There is also the fact that opamps themselves have a common-mode rejection ratio; it is high, but once more it is not infinite.

As usual, SPICE simulation is instructive, and Figure 16.5 shows a simple balanced interconnection, with the balanced output represented simply by two 100 Ω output resistances connected to the source equipment ground, here called Ground 1, and the usual differential opamp configuration at the input end, where we have Ground 2.

A common-mode voltage Vcm is now injected between Ground 1 and Ground 2, and the signal between the opamp output and Ground 2 measured. The balanced input amplifier has all four of its resistances set to precisely 10 kΩ, and the opamp is represented by a very simple model that has only two parameters; a low-frequency open-loop gain, and a single pole frequency that says where that gain begins to roll-off at 6 dB per octave. The opamp input impedances and the opamp's own CMRR are assumed infinite, as in the world of simulation they so easily can be. Its output impedance is set at zero.

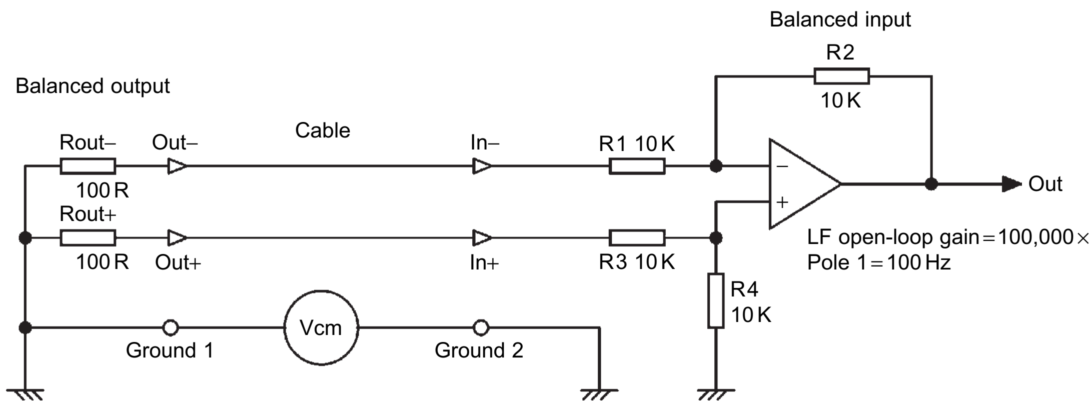


Figure 16.5: A simple balanced interconnection for SPICE simulation to show the effect that opamp properties have on the CMRR.

Table 16.3: The Effect of Finite Opamp Gain on CMRR for the Circuit of Figure 16.5

Open-Loop Gain	CMRR dB	CMRR Ratio
10,000	−74.0	19.9×10^{-5}
30,000	−83.6	66.4×10^{-6}
100,000	−94.0	19.9×10^{-6}
300,000	−103.6	6.64×10^{-6}
1,000,000	−114.1	1.97×10^{-6}

For the first experiments, even the pole frequency is made infinite, so now the only contact with harsh reality is that the opamp open-loop gain is finite. That is however enough to give distinctly non-ideal CMRR figures, as Table 16.3 shows.

With a low-frequency open-loop gain of 100,000, which happens to be the typical figure for a 5532 opamp, even perfect components everywhere will never yield a better CMRR than −94 dB. The CMRR is shown as a raw ratio in the third column so you can see that the CMRR is inversely proportional to the gain, and so we want as much gain as possible.

16.15 Common-Mode Rejection Ratio: Opamp Frequency Response

To examine these we will set the low-frequency gain to 100,000 which gives a CMRR "floor" of −94 dB, and then introduce the pole frequency that determines where it rolls-off. The CMRR now worsens at 6 dB/octave, starting at a frequency set by the interaction of the low-frequency gain and the pole frequency. The results are summarised in Table 16.4 which shows that as you might expect, the lower the open-loop bandwidth of the opamp, the lower the frequency at which the CMRR begins to fall off. Figure 16.6 shows the situation diagrammatically.

Table 16.5 gives the open-loop gain and pole parameters for a few opamps of interest. Both parameters, but especially the gain, are subject to considerable variation; the typical values from the manufacturers' data sheets are given here.

Some of these opamps have very high open-loop gains, but only at very low frequencies. This may be good for DC applications, but in audio line input applications, where the lowest frequency of CMRR interest is 50 Hz, they will be operating above the pole frequency and so the gain available will be less—possibly considerably so, in the case of opamps like the OPA2134. This is not however a real limitation, for even if a humble TL072 is used the perfect-resistor CMRR is about −90 dB, degrading from 100 Hz upwards. This sort of performance is not attainable in practice. We will shortly see why not.

Table 16.4: The Effect of Opamp Open-Loop Pole Frequency on CMRR for the Circuit of Figure 16.5

Pole Frequency	CMRR Breakpoint Freq
10 kHz	10.2 kHz
1 kHz	1.02 kHz
100 Hz	102 Hz
10 Hz	10.2 Hz

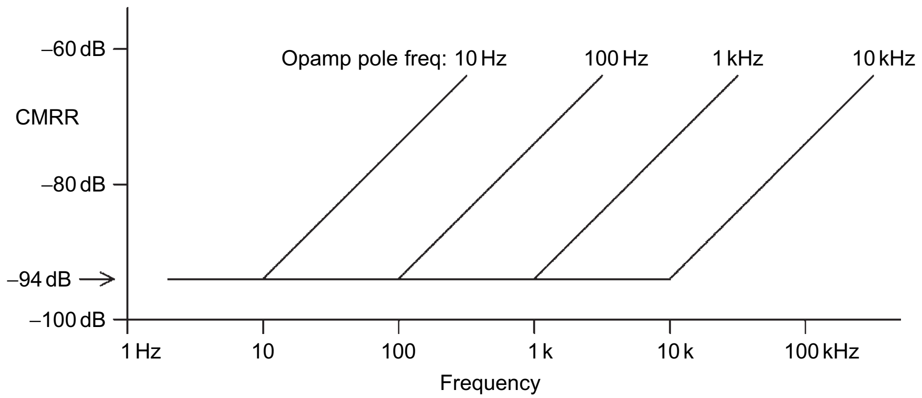


Figure 16.6: How the CMRR degrades with frequency for different opamp pole frequencies. All resistors are assumed to be perfectly matched.

Table 16.5: Typical LF Gain and Open-Loop Pole Frequency for Some Opamps Commonly Used in Audio

Name	Input Device Type	LF Gain	Pole Freq	Opamp LF CMRR dB
NE5532	Bipolar	100,000	100 Hz	100
LM4562	Bipolar	10,000,000	below 10 Hz	120
LT1028	Bipolar	20,000,000	3 Hz	120
TL072	FET	200,000	20 Hz	86
OP27	FET	1,800,000	3 Hz	120
OPA2134	FET	1,000,000	3 Hz	100
OPA627	FET	1,000,000	20 Hz	116

16.16 Common-Mode Rejection Ratio: Opamp CMRR

Opamps have their own common-mode rejection ratio, and we need to know how much this will affect the final CMRR of the balanced interconnection. The answer is that if all resistors are exactly correct, the overall CMRR is equal to the CMRR of the opamp [6].

Since opamp CMRR is typically very high (see the examples in Table 16.5) it is most unlikely to be the limiting factor.

The CMRR of an opamp begins to degrade above a certain frequency, typically at 6 dB per octave. This is (fortunately) at a higher frequency than the open-loop pole, and is frequently around 1 kHz. For example the OP27 has a pole frequency at about 3 Hz, but the CMRR remains flat at 120 dB until 2 kHz, and it is still greater than 100 dB at 20 kHz.

16.17 Common-Mode Rejection Ratio: Amplifier Component Mismatches

We saw earlier in this chapter that when the output and input impedances on a balanced line have a high ratio between them and are accurately matched we got a very good CMRR; this was compromised by the imperfections of opamps, but the overall results were still very good—and much higher than the real CMRRs that we measure in practice. There remains one place where we are still away in theory-land; we have so far assumed the resistances around the opamp were all exactly accurate. We must now face reality, admit that these resistors will not be perfect, and see how much damage to the CMRR they will do.

SPICE simulation gives us Table 16.6. The situation with LF opamp gains of both 100,000 and 1,000,000 is examined, but the effects of finite opamp bandwidth or opamp CMRR are not included. R1 in Figure 16.5 is varied while R2, R3 and R4 are all kept at precisely 10 kΩ, and the balanced output source impedances are set to exactly 100 Ω.

Table 16.6 shows with glaring clarity that our previous investigations, which took only output and input impedances into account, and determined that 100 Ω output resistors and 10 kΩ input impedances gave a CMRR of −80 dB for a 1% deviation in either, were

Table 16.6: How Resistor Tolerances Affect the CMRR for Two Realistic Opamp Open-Loop Gains

R1 Ω	R1 Deviation	Gain x	CMRR dB
10 k	0%	100,000	−94.0
10.001 k	0.01%	100,000	−90.6
10.01 k	0.1%	100,000	−66.5
10.1 k	1%	100,000	−46.2
11 k	10%	100,000	−26.6
10 k	0%	1,000,000	−114.1
10.001 k	0.01%	1,000,000	−86.5
10.01 k	0.1%	1,000,000	−66.2
10.1 k	1%	1,000,000	−46.2
11 k	10%	1,000,000	−26.6

definitely optimistic, and even adding in opamp imperfections left us with implausibly good results. Looking at component imbalances in the amplifier itself brings us down to earth; when a 1% tolerance resistor is used for R1 (and nowadays there is no financial incentive to use anything less accurate), the CMRR plummets to −46 dB; the same figure results from varying any other one of the four resistances by itself. If you are prepared to shell out for 0.1% tolerance resistors, the CMRR is a rather better −66 dB.

Table 16.6 also shows that there really is no point in getting anxious about the gain of the opamp you use in balanced inputs; unless you're planning to use 0.01% resistors (and I'm sure you're not) the effect of the opamp gain is negligible.

The results in Table 16.6 give an illustration of how resistor accuracy affects CMRR, but it is only an illustration, because in real life—a phrase that seems to keep cropping up, showing how many factors affect a practical balanced interconnection—all four resistors will of course be subject to a tolerance, and a more realistic calculation would produce a statistical distribution of CMRR rather than a single figure. One method is to use the Monte Carlo function in SPICE, which runs multiple simulations with random component variations and collates the results. However you do it you must know (or assume) how the resistor values are distributed within their tolerance window. Usually you don't know, and finding out by measuring hundreds of resistors is not a task that appeals to all of us.

It is straightforward to assess the worst-case CMRR, which occurs when all resistors are at the limit of the tolerance in the most unfavourable direction. The CMRR in dB is then:

$$\mathrm{CMRR} = 20\log\left(\frac{1 + \mathrm{R2/R1}}{4\mathrm{T}/100}\right) \tag{14.1}$$

Where R1 and R2 are as in Figure 16.5, and T is the tolerance in %.

This rather pessimistic equation tells us that 1% resistors give a worst-case CMRR of only 34.0 dB, that 0.5% parts give only 40.0 dB and expensive 0.1% parts yield but 54.0 dB. Things are not however quite that bad in actuality, as the chance of everything being as wrong as possible is actually very small indeed. I have measured the CMRR of more of these balanced inputs, built with 1% resistors, than I care to contemplate, but I do not ever recall that I ever saw one with an LF CMRR worse than 40 dB.

There are 8-pin SIL packages that offer four resistors that ought to have good matching, if not accurate absolute values; be very, very wary of these as they usually contain thick-film resistive elements that are not perfectly linear. In a test I did a 10 kΩ SIL resistor with 10 Vrms across it generated 0.0010% distortion. Not a huge amount perhaps, but in the quest for perfect audio, resistors that do not stick to Ohm's Law are not a good start.

To conclude this section, it is clear that in practical use it is the errors in the balanced amplifier resistors that determine the CMRR, though both unbalanced capacitances (C1, C2 in Figure 16.9 below) and the finite opamp bandwidth are likely to cause further degradation at high audio frequencies. If you are designing both ends of a balanced interconnection and you are spending money on a few precision resistors, you should most definitely put them in the input amplifier, not the balanced output. The LF gain of the opamp, and opamp CMRR, have virtually no effect.

In fact, balanced input amplifiers like Figure 16.4 and 16.9, built with four ordinary 1% resistors, are used very extensively in the professional audio business, and almost always prove to have adequate CMRR for the job; I have spent a lot of time designing mixing consoles and I do not recall a single occasion when this was not the case. When more CMRR is wanted, for example in high-end mixing consoles, one of the resistances is made trimmable with a preset, as shown in Figure 16.7. This can mean a lot of tweaking in manufacture, as there might easily be three or four of these balanced inputs per channel, but looking on the bright side, it is a quick set-and-forget adjustment that will never need to be touched again unless one of the four fixed resistors needs replacing, and that is extremely unlikely. CMRRs at LF of more than 70 dB can easily be obtained by this method, but the CMRR at HF will degrade due to the opamp gain roll-off and stray capacitances.

Figure 16.8 shows the CMRR measurements for a trimmable balanced input amplifier. The flat line at −50 dB was obtained from standard balanced input using four 1% 10 kΩ resistors straight out of the box, while the much better (at LF, anyway) trace going down to −85 dB was obtained from Figure 16.7 by using a multi-turn preset for PR1. Note that R4 is an E96 value so a 1 K preset can swing the total resistance of that arm both above and below the nominal 10 kΩ.

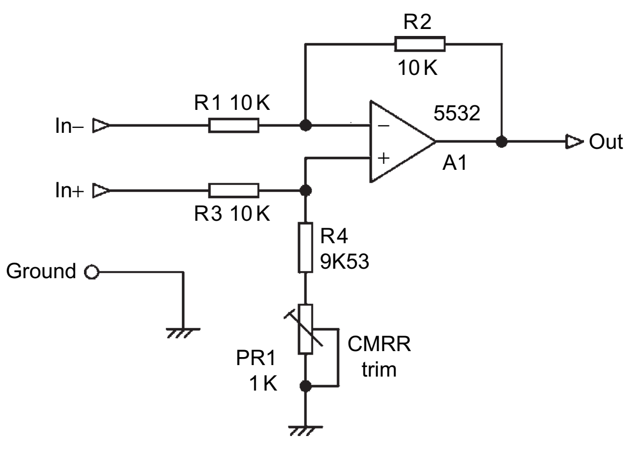


Figure 16.7: A balanced input amplifier, with preset pot to trim for best LF CMRR.

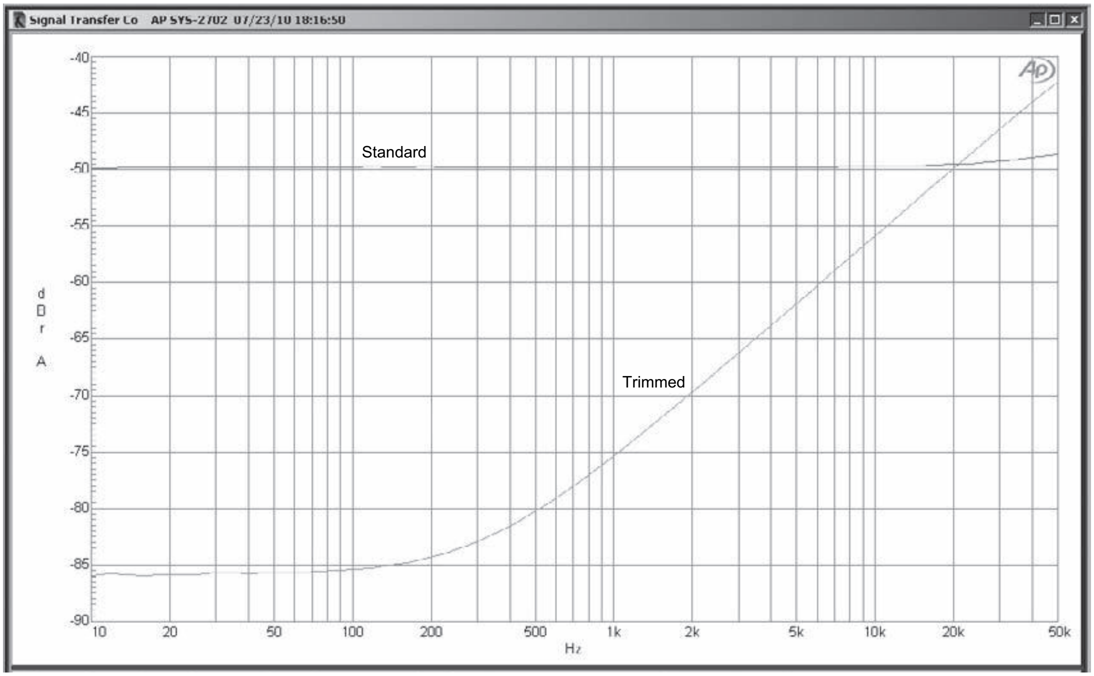


Figure 16.8: CMRR results from a standard balanced amplifier as in Figure 16.4, and from trimmed Figure 16.7. The opamp was a 5532 and all resistors were 1%. The trimmed version is better than 80 dB up to 500 Hz.

The CMRR is dramatically improved by more than 30 dB in the region 50–500 Hz where ground noise tends to intrude itself, and is significantly better across almost all the audio spectrum.

The sloping part of the trace in Figure 16.8 is partly due to the finite open-loop bandwidth of the opamp, and partly due to unbalanced circuit capacitances. The CMRR is actually worse than 50 dB above 20 kHz, due to the stray capacitances in the multiturn preset, and the fact that I threw the circuit together on a piece of prototype board. In professional manufacture the value of PR1 would probably be much smaller, and a small one-turn preset used with much less stray capacitance. Still, I think you get the point; for relatively small manufacturing quantities CMRR trimming is both economic and effective.

16.18 A Practical Balanced Input

The simple balanced input circuits shown in Figures 16.4 and 16.7 are not fit to face the outside world without additional components. Figure 16.9 shows a fully equipped version. Firstly, and most important, C1 has been added across the feedback resistor R2; this prevents stray capacitances from Pin 2 to ground causing extra phase-shifts that lead to HF instability.

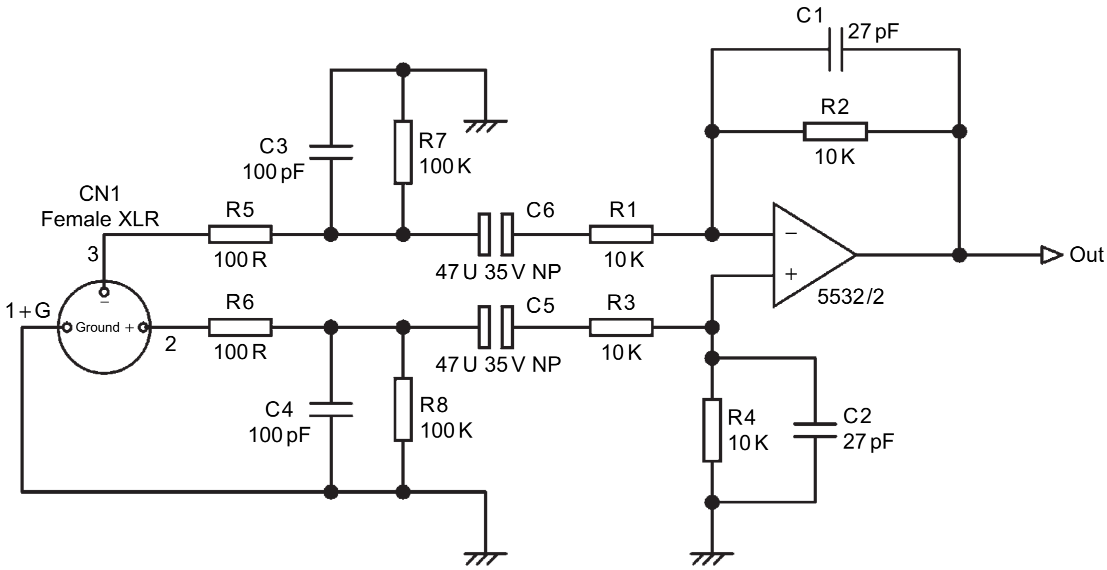


Figure 16.9: Balanced input amplifier with the extra components required for DC blocking and EMC immunity.

The value required for stability is small, much less than that which would cause an HF roll-off anywhere near the top of the audio band. The values here of 10 k and 27 pF give –3 dB at 589 kHz, and such a roll-off is only down by 0.005 dB at 20 kHz. C2, of equal value, must be added across R4 to maintain the balance of the amplifier, and hence its CMRR, at high frequencies.

C1 and C2 must not be relied upon for EMC immunity as C1 is not connected to ground, and there is every chance that RF will demodulate at the opamp inputs. A passive RF filter is therefore added to each input, in the shape of R5, C3 and R6, C4, so the capacitors will shunt incoming RF to ground before it reaches the opamp. Put these as close to the input socket as possible to minimise radiation inside the enclosure.

I explained earlier in this chapter when looking at unbalanced inputs that it is not easy to guess what the maximum source impedance will be, given the existence of "passive preamplifiers" and valve equipment. Neither are likely to have a balanced output, unless implemented by transformer, but either might be used to feed a balanced input, and so the matter needs thought.

In the unbalanced input circuit resistances had to be kept as low as practicable to minimise the generation of Johnson noise that would compromise the inherently low noise of the stage. The situation with a standard balanced input is however different from the unbalanced case as there have to be resistances around the opamp, and they must be kept up to a certain value to give acceptably high input impedances; this is why a balanced input like this one is much noisier. We could therefore make R5 and R6 much larger without a measurable noise penalty if we reduce R1 and R3 accordingly to keep unity gain. In Figure 16.9 R5 and R6 are kept at 100 Ω, so if we assume 50 Ω output resistances in both legs of the source equipment, then we have a total of 150 Ω, and 150 Ω and 100 pF give −3 dB at 10.6 MHz. Returning to a possible passive preamplifier with a 10 kΩ potentiometer, its maximum output impedance of 2.5 k plus 100 Ω with 100 pF gives −3 dB at 612 kHz, which remains well clear of the top of the audio band.

As with the unbalanced input, replacing R5 and R6 with small inductors will give much better RF filtering but at increased cost. Ideally a common-mode choke (two bifilar windings on a small toroidal core) should be used as this improves performance. Check the frequency response to make sure the LC circuits are well-damped and not peaking at the turnover frequency.

C5 and C6 are DC-blocking capacitors. They must be rated at no less than 35 V to protect the input circuitry, and are the non-polarised type as external voltages are of unpredictable polarity. The lowest input impedance that can occur with this circuit when using 10 kΩ resistors, is, as described above, 6.66 kΩ when it is being driven in the balanced mode. The low-frequency rolloff is therefore −3 dB at 0.51 Hz. This may appear to be undesirably low, but the important point is not the LF rolloff but the possible loss of CMRR at low frequencies due to imbalance in the values of C5 and C6; they are electrolytics with a significant tolerance. Therefore they should be made large so their impedance is a small part of the total input impedance. 47 uF is shown here but 100 uF or 220 uF can be used to advantage if there is the space to fit them in. The low-end frequency response must be defined somewhere in the input system, and the earlier the better, to prevent headroom or linearity being affected by subsonic disturbances, but this is not a good place to do it. A suitable time-constant immediately after the input amplifier is the way to go, but remember that capacitors used as time-constants may distort unless they are NP0 ceramic, polystyrene, or polypropylene. See the chapter on passive components for more on this.

R7, R8 are DC drain resistors to prevent charges lingering on C5 and C6. These can be made lower than for the unbalanced input as the input impedances are lower, so a value of say 100 kΩ rather than 220 kΩ makes relatively little difference to the total input impedance.

A useful property of this kind of balanced amplifier is that it does not go mad when the inputs are left open-circuit—in fact it is actually *less* noisy than with its inputs shorted to ground. This is the opposite of the "normal" behaviour of a high-impedance unterminated input. This is because two things happen; open-circuiting the hot input doubles the resistance seen by the non-inverting input of the opamp, raising its noise contribution by 3 dB. However, opening the cold input makes the noise gain drop by 6 dB, giving a net drop in noise output of

Table 16.7: Noise Output Measured from Simple Balanced Amps Using a 5532 Section

R Value Ω	50 Ω Terminated Inputs	Open-Circuit Inputs	Terminated/Open Difference
100 k	−95.3 dBu	−97.8 dBu	2.5 dBu
10 k	−104.8 dBu	−107.6 dBu	2.8 dBu
2 k0	−109.2 dBu	−112.0 dBu	2.8 dBu
820	−111.7 dBu	−114.5 dBu	2.8 dBu

approximately 3 dB. This of course refers only to the internal noise of the amplifier stage, and pickup of external interference is always possible on an unterminated input. The input impedances here are modest, however, and the problem is less serious than you might think. Having said that, deliberately leaving inputs unterminated is always bad practice.

If this circuit is built with four 10 kΩ resistors and a 5532 opamp section, the noise output is −104.8 dBu with the inputs terminated to ground via 50 Ω resistors. As noted above, the input impedance of the cold input is actually lower than the resistor connected to it when working balanced, and if it is desirable to raise this input impedance to 10 kΩ, it could be done by raising the four resistors to 16 kΩ; this slightly degrades the noise output to −103.5 dBu. Table 16.7 gives some examples of how the noise output depends on the resistor value; the third column gives the noise with the input unterminated, and shows that in each case the amplifier is about 3 dB quieter when open-circuited. It also shows that a useful improvement in noise performance is obtained by dropping the resistor values to the lowest that a 5532 can easily drive (the opamp has to drive the feedback resistor), though this usually gives unacceptably low input impedances. More on that at the end of the chapter.

16.19 Variations on the Balanced Input Stage

I now give a collection of balanced input circuits that offer advantages or extra features over the standard balanced input configuration. The circuit diagrams often omit stabilising capacitors, input filters, and DC blocking capacitors to improve the clarity of the basic principle. They can easily be added; in particular bear in mind that a stabilising capacitor like C1 in Figure 16.9 is often needed between the opamp output and the negative input to guarantee freedom from high-frequency oscillation.

16.20 Combined Unbalanced and Balanced Inputs

If both unbalanced and balanced inputs are required, it is extremely convenient if it can be arranged so that no switching between them is required. Switches cost money, mean more holes in the metalwork, and add to assembly time. Figure 16.10 shows an effective way

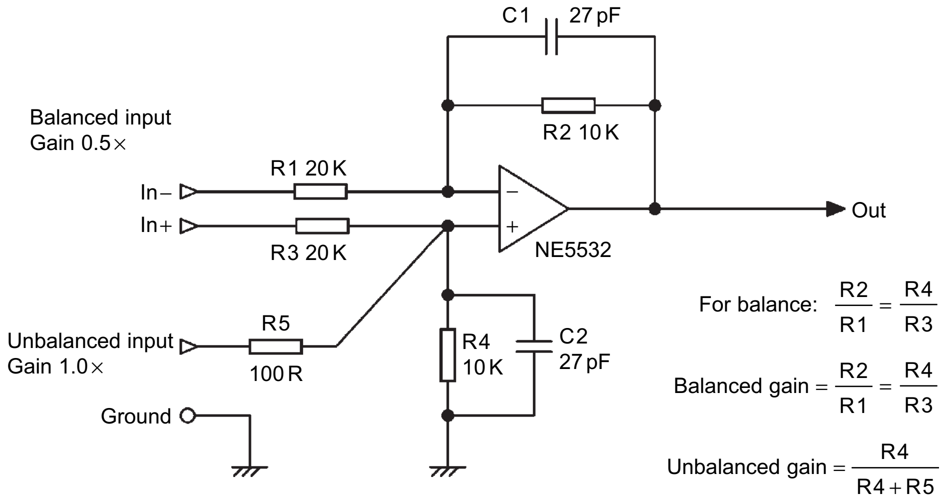


Figure 16.10: Combined balanced and unbalanced input amplifier with no switching required, but some performance compromises.

to implement this. In balanced mode, the source is connected to the balanced input and the unbalanced input left unterminated. In unbalanced mode, the source is connected to the unbalanced input and the balanced input left unterminated, and no switching is required. It might appear that these unterminated inputs would pick up extra noise, but in practice this is not the case. It works very well and I have used it successfully in high-end equipment for two prestigious manufacturers.

As described above, in the world of hi-fi, balanced signals are at twice the level of the equivalent unbalanced signals, and so the balanced input must have a gain of ½ or −6 dB relative to the unbalanced input to get the same gain by either path. This is done here by increasing R1 and R3 to 20 kΩ. The balanced gain can be greater or less than unity, but the gain via the unbalanced input is always one. The differential gain of the amplifier and the constraints on the component values for balanced operation are shown in Figure 16.10, and are not repeated in the text to save space. This applies to the rest of the balanced inputs in this chapter.

There are two minor compromises in this circuit which need to be noted. Firstly, the noise performance in unbalanced mode is worse than for the dedicated unbalanced input described earlier in this chapter, because R2 is effectively in the signal path and adds Johnson noise. Secondly, the input impedance of the unbalanced input cannot be very high because it is set by R4, and if this is increased in value all the resistances must be increased proportionally and the noise performance in balanced mode will be markedly worse. It is important that only one input cable should be connected at a time, because if an unterminated cable is left connected to an unused input, the cable capacitance to ground can cause frequency response anomalies and might in adverse circumstances cause HF oscillation. A prominent warning on the back panel and in the manual is a very good idea.

16.21 The Superbal Input

This version of the balanced input amplifier, shown in Figure 16.11, has been referred to as the "Superbal" circuit because it gives equal impedances into the two inputs for differential signals. It was originated by David Birt of the BBC; see [7]. With the circuit values shown the differential input impedance is exactly 10 kΩ via both hot and cold inputs. The common-mode input impedance is 20 kΩ as before.

In the standard balanced input R4 is connected to ground, but here its lower end is actively driven with an inverted version of the output signal, giving symmetry. The increased amount of negative feedback reduces the gain with four equal resistors to –6 dB instead of unity. The gain can be reduced below −6 dB by giving the inverter a gain of more than one; if R1, R2, R3, and R4 are all equal, the gain is 1/(A+1), where A is the gain of the inverter stage. This is of limited use as the inverter U1:B will now clip before the forward

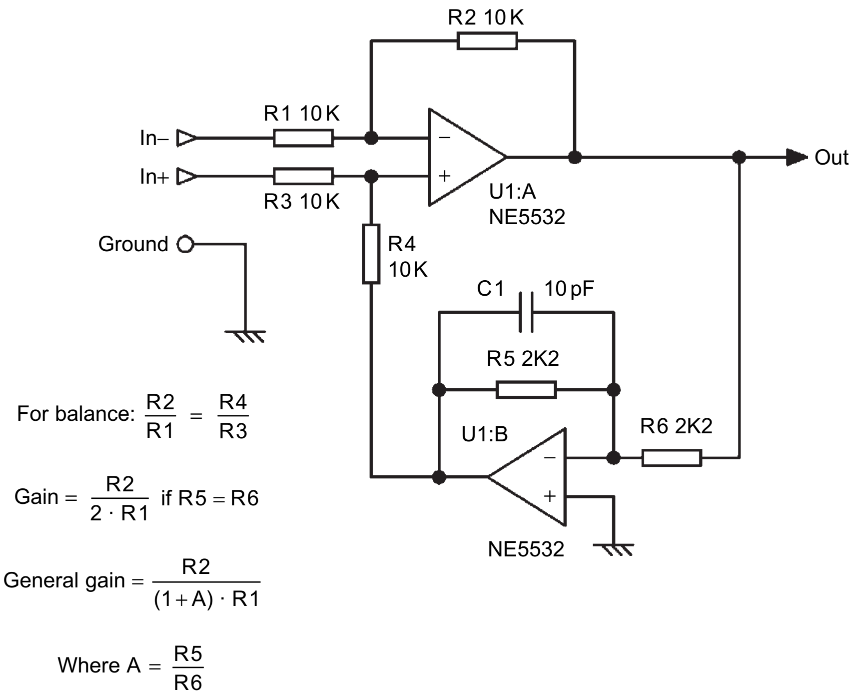


Figure 16.11: The Superbal balanced input requires another amplifier but has equal input impedances.

amplifier U1:A, reducing headroom. If the gain of the inverter stage is gradually reduced from unity to zero, the stage slowly turns back into a standard balanced amplifier with the gain increasing from −6 dB to unity and the input impedances becoming less and less equal. If a gain of less than unity is required it should be obtained by increasing R1 and R3.

R5 and R6 should be kept as low in value as possible to minimise Johnson noise; there is no reason why they have to be equal in value to R1, etc. The only restriction is the ability of U1:A to drive R6 and U1:B to drive R5, both resistors being effectively grounded at one end. The capacitor C1 will almost certainly be needed to ensure HF stability; the value in the figure is only a suggestion. It should be kept as small as possible because reducing the bandwidth of the inverter stage impairs CMRR at high frequencies.

16.22 Switched-Gain Balanced Inputs

A balanced input stage that can be switched to two different gains while maintaining CMRR is very useful. Equipment often has to give optimal performance with both semi-pro (–7.8 dBu) and professional (+4 dBu) input levels. If the nominal internal level of the system is in the normal range of −2 to −6 dBu, the input stage must be able to switch between amplifying and attenuating, while maintaining good CMRR in both modes.

The brute-force way to change gain in a balanced input stage is to switch the values of either R1 and R3, or R2 and R4, in Figure 16.4, keeping the pairs equal in value to maintain the CMRR; this needs a double–pole switch for each input channel. A much more elegant technique is shown in Figure 16.12. Perhaps surprisingly, the gain of a differential amplifier can be manipulating by changing the drive to the feedback arm (R2 etc.) only, and leaving the other arm R4 unchanged, without affecting the CMRR. The essential point is to keep the source resistance of the feedback arm the same, but drive it from a scaled version of the opamp output. Figure 16.12 does this with the network R5, R6, which has a source resistance

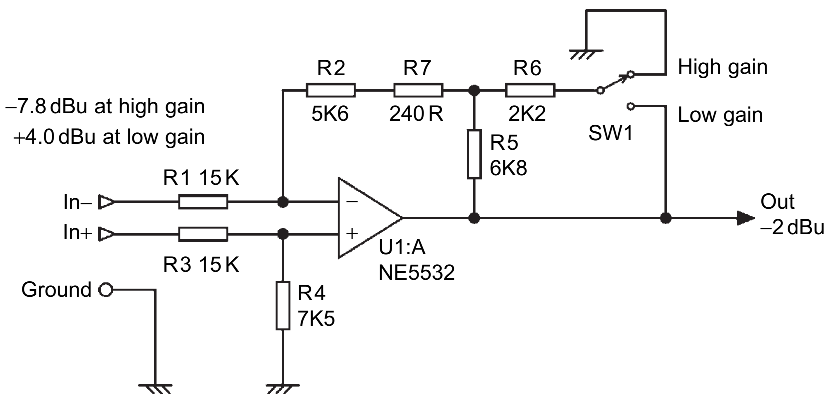


Figure 16.12: A balanced input amplifier with gain switching that maintains good CMRR.

made up of 6k8 in parallel with 2k2, which is 1.662 kΩ. This is true whether R6 is switched to the opamp output (low gain setting) or to ground (high gain setting), for both have effectively zero impedance. For low gain the negative feedback is not attenuated, but fed through to R2 and R7 via R5, R6 in parallel. For high gain R5 and R6 become a potential divider, so the amount of feedback is decreased and the gain increased. The value of R2 + R7 is reduced from 7k5 by 1.662 kΩ to allow for the source impedance of the R5, R6 network; this requires the distinctly non-standard value of 5.838 kΩ, which is here approximated by R2 and R7 which give 5.6 kΩ + 240 Ω = 5.840 kΩ. This value is the best that can be done with E24 resistors; it is obviously out by 2 Ω, but that is much less than a 1% tolerance on R2, and so will have only a vanishingly small effect on the CMRR.

Note that this stage can attenuate as well as amplify if R1, R3 are set to be greater than R2, R4, as shown here. The nominal output level of the stage is assumed to be −2 dBu; with the values shown the two gains are −6.0 and +6.2 dB, so +4 dBu and −7.8 dBu respectively will give −2 dBu at the output. Other pairs of gains can of course be obtained by changing the resistor values; the important thing is to stick to the principle that the value of R2 + R7 is reduced from the value of R4 by the source impedance of the R5, R6 network. With the values shown the differential input impedance is 11.25 kΩ via the cold and 22.5 kΩ via the hot input. The common-mode input impedance is 22.5 kΩ.

Switched-gain inputs like this one have the merit that there are no issues with balance between channels because the gain is defined by relatively precise fixed resistors, rather than ganged pots, as used in the next section. This neat little circuit has the added advantage that nothing bad happens when the switch is moved with the circuit operating. When the wiper is between contacts you simply get a gain intermediate between the high and low settings, which is pretty much the ideal situation. Make sure the switch is a break-before-make type to avoid shorting the opamp output to ground when the switch is moved.

16.23 Variable-Gain Balanced Inputs

The beauty of a variable-gain balanced input is that it allows you to get the incoming signal up or down to the nominal internal level as soon as possible, minimising both the risk of clipping and contamination with circuit noise. The obvious method of making a variable-gain differential stage is to use dual-gang pots to vary either R1, R3 or R2, R4 together, to maintain CMRR. This is clumsy, and gives a CMRR that is both bad and highly variable due to the inevitable mismatches between pot sections. For a stereo input the required 4-gang pot is an unappealing proposition.

There is however a way to get a variable gain with good CMRR, using a single pot section. The principle is essentially the same as for the switched-gain amplifier above; keep constant the source impedance driving the feedback arm, but vary the voltage applied. The principle is shown in Figure 16.13. To the best of my knowledge I invented this circuit in 1982; any comments on this point are welcome. The feedback arm R2 is driven by voltage-follower U1:B. This eliminates the variations in source impedance at the pot wiper, which would badly degrade the CMRR. R6 limits the gain range and R5 modifies the gain law to give it a more usable shape. When the pot is fully up (minimum gain) R5 is directly across the output of U1:A, so do not make it too low in value. If a centre-detent pot is used to give a default gain setting, this may not be very accurate as it partly depends on the ratio of pot track (no better than +/−10% tolerance, and sometimes worse) to 1% fixed resistors.

This configuration is very useful as a general line input with an input sensitivity range of −20 to +10 dBu. For a nominal output of 0 dBu, the gain of Figure 16.13 is +20 to −10 dB, with R5 chosen for 0 dB gain at the central wiper position. An opamp in a feedback path may appear a dubious proposition for HF stability, because of the extra phase-shift it introduces, but here it is working as a voltage-follower, so its bandwidth is maximised and in practice the circuit is dependably stable.

Circuitry like this is ideal for single-channel applications, but can create difficulties when used in stereo or other multi-channel formats, because of matching problems in ganged potentiometers. This version of the circuit has a fairly wide gain range, and it is necessary to use an RD (anti-log 10%) law pot to get a reasonable linear-in-dB control law. This introduces gain errors between nominal identical channels because firstly, the value of the pot track is not controlled anything like as closely as that of a 1% resistor and so the effect of R5 on the pot law will vary, which leads to gain differences between the channels. Secondly, any log or anti-log pot law is made up of dual resistance sections and this

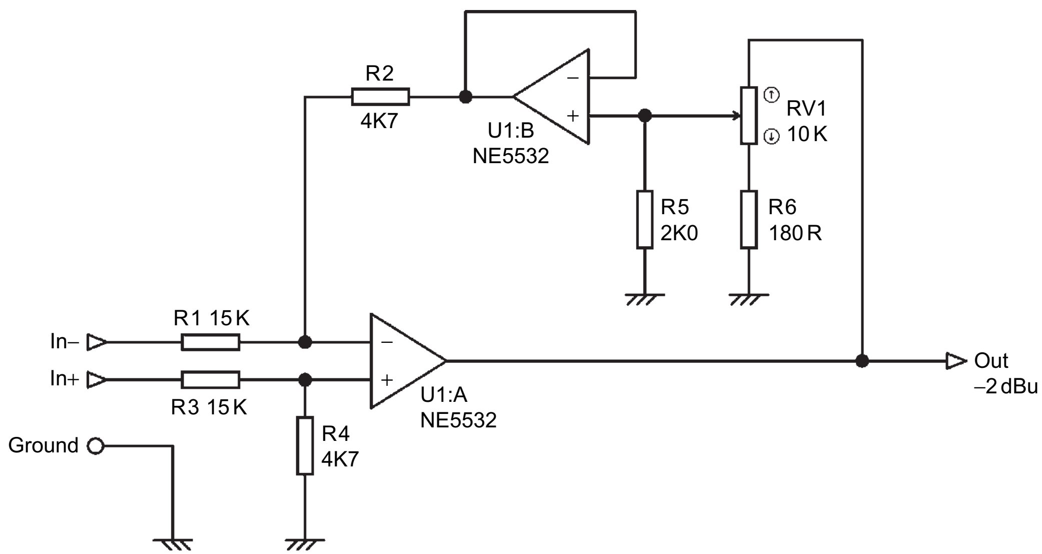


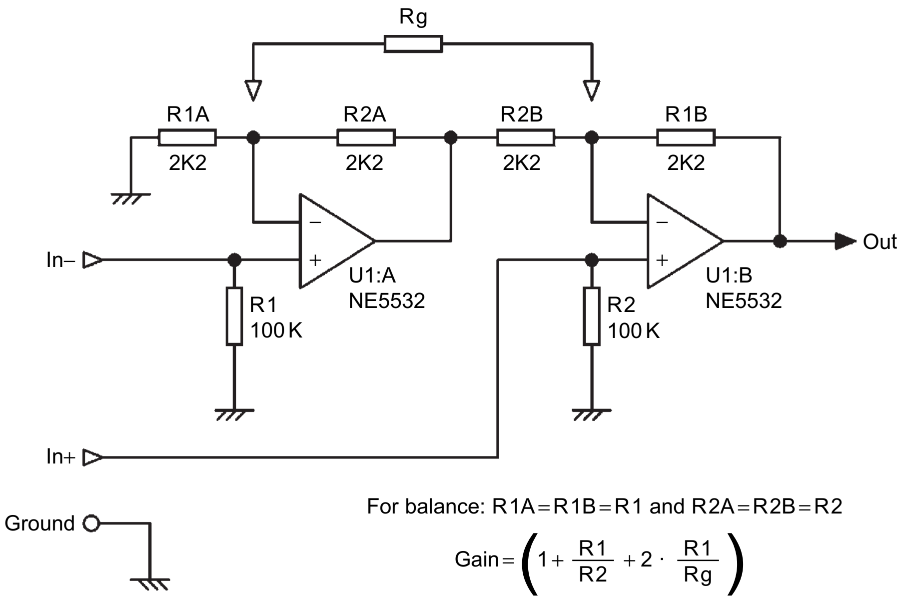
Figure 16.13: Variable gain balanced input amplifier.

introduces more errors. If a more restricted gain range is being used then it may be possible to use a linear pot and rely on the loading effect of R5 to give an acceptable control law. The whole problem of multi-channel gain control with potentiometers of limited accuracy is examined in detail in [8].

16.24 High Input Impedance Balanced Inputs

We saw earlier that high input impedances are required to maximise CMRR of a balanced interconnection, but the input impedances offered by the standard balanced circuit are limited by the need to keep the resistor values down to control Johnson noise. High-impedance balanced inputs are also useful for interfacing to valve equipment in the strange world of retro hi-fi. Adding output cathode-followers to valve circuitry is expensive and consumes a lot of extra power, and so the output is often taken directly from the anode of a gain-stage, and even a so-called bridging load of 10 kΩ may seriously compromise the distortion performance and output capability of the source equipment.

Figure 16.14 shows a configuration where the input impedances are determined only by the bias resistances R1 and R2. They are shown here as 100 kΩ, but may be considerably higher if opamp bias currents permit. A useful property of this circuit is that adding a single resistor Rg increases the gain, but preserves the circuit balance and CMRR. This configuration cannot be set to attenuate because the gain of an opamp with series feedback cannot be reduced below unity.



$$\text{Gain} = \left(1 + \frac{R1}{R2} + 2 \cdot \frac{R1}{Rg}\right)$$

Figure 16.14: High input impedance balanced input.

It is of course always possible to give a basic balanced input a high input impedance by putting unity-gain buffers in front of each input, but that uses three opamp sections rather than two. Sometimes, however, it is appropriate. Much more on that later.

We saw earlier that the simple balanced input is surprisingly quiet and well-behaved when its input are unterminated. This is not the case with this configuration, which because of its high input impedances will be both noisy and susceptible to picking-up external interference if either input is left open-circuit.

16.25 The Instrumentation Amplifier

Almost every book on balanced or differential inputs includes the three-opamp circuit of Figure 16.15 and praises it as the highest expression of the differential amplifier. It is called the instrumentation amplifier configuration because of its undoubted superiority for data-acquisition. (Specialised ICs exist that are sometimes also called instrumentation amplifiers or in-amps; these are designed for very high CMRR data-acquisition. They are expensive and in general not optimised for audio work.)

Like the low-noise balanced amplifiers described later, the instrumentation amplifier is split into a first and second stage. The differential input stage buffers the balanced line from the input impedances of the final differential stage; the four resistances around the latter can therefore be made much lower in value, reducing Johnson noise and the effects of current noise significantly, while keeping the CMRR benefits of presenting high input impedances to the balanced line. The other feature, which is usually much more emphasised because of its

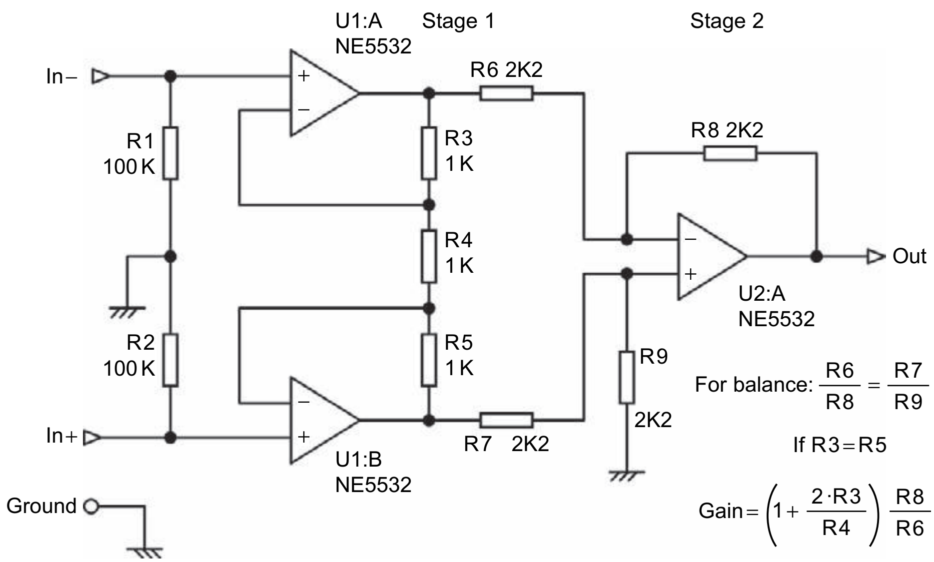


Figure 16.15: The instrumentation amplifier configuration. Gain here is 3 times.

unquestionable elegance, is that the dual input stage, with its shared feedback network R3, R4, R5, can be set to have a high differential gain by giving R4 a low value, but its common-mode gain is always unity; this property is not affected by mismatches in R3 and R5. The final amplifier then does its usual job of common-mode rejection, and the combined CMRR can be very good indeed if the first-stage gain is high.

Unfortunately a high first-stage gain is not very often useful for audio balanced line inputs. A data-acquisition application like ECG monitoring may need a gain of thousands of times, which will allow a stunning CMRR without using precision resistors, but the cruel fact is that in audio use gain in the input amplifier is often simply not wanted. In a typical opamp signal path, the nominal internal level is usually between −6 and 0 dBu, and if the level of the incoming balanced signal is at the professional level of +4 dBu, then what you need is 6 dB of attenuation rather than any gain. Gain now and attenuation later must introduce what can only be called a headroom bottleneck. If the incoming level was the semi-pro −7.8 dBu then a small amount of gain could be introduced, but then the CMRR advantage would be equally small, and certainly not worth the cost of the extra circuitry.

However, active crossovers are a special case. Chapter 14 shows how running active crossover circuitry at a much higher internal level than is usual in audio equipment is entirely practical. Chapter 19 demonstrates that the real-life noise benefits of a balanced input amplifier with a gain of four times (12 dB) and an internal level of 3 Vrms are impressive. If the balanced input is an instrumentation amplifier with all the four times gain in its first stage, there is a measured CMRR improvement of a very real 12 dB, and there is also a most useful 4.5 dB reduction in the noise output of the balanced amplifier compared with conventional methods. This is because the first stage works under better noise conditions than the second stage. This can be taken further by giving the first stage a gain of 8 times and the second stage 0.5 times without headroom penalty; CMRR is now improved by 18 dB and noise out reduced by 5.3 dB.

16.26 Transformer Balanced Inputs

When it is essential that there is no galvanic connection (i.e., no electrical conductor) between two pieces of equipment, transformer inputs are indispensable. They are also useful if EMC conditions are severe. Figure 16.16 shows a typical transformer input. The transformer usually has a 1:1 ratio, and should have an inter-winding screen, which must be earthed to optimise the high-frequency CMRR, and minimise noise pickup and EMC troubles. If the transformer is in a metal shielding can this needs to be grounded to reduce noise pickup; round cans can be held in a metal capacitor clip connected to ground.

The transformer secondary must see a high impedance as this is reflected to the primary and represents the input impedance; here it is set by R2, and a buffer drives the circuitry downstream. In addition, if the secondary loading is too heavy there will be increased transformer distortion at low frequencies. Line input transformers are built with small cores

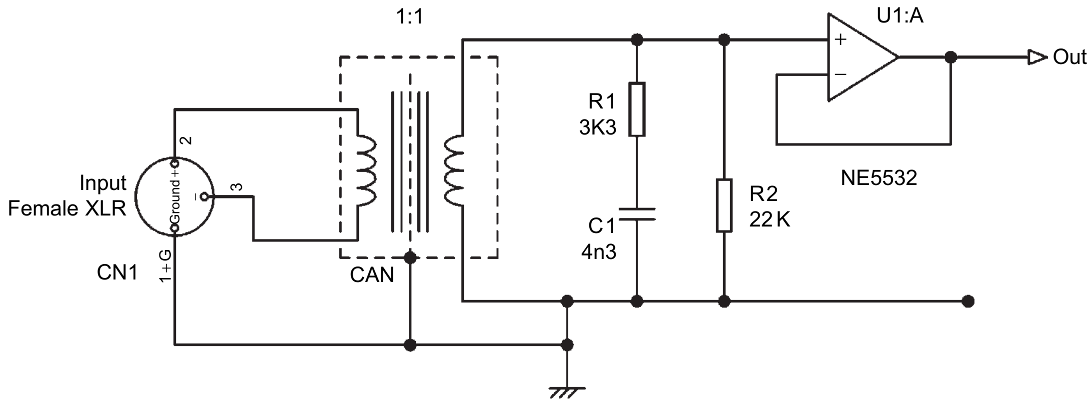


Figure 16.16: A transformer balanced input. R1 and C1 are the Zobel network that damps the transformer secondary resonance.

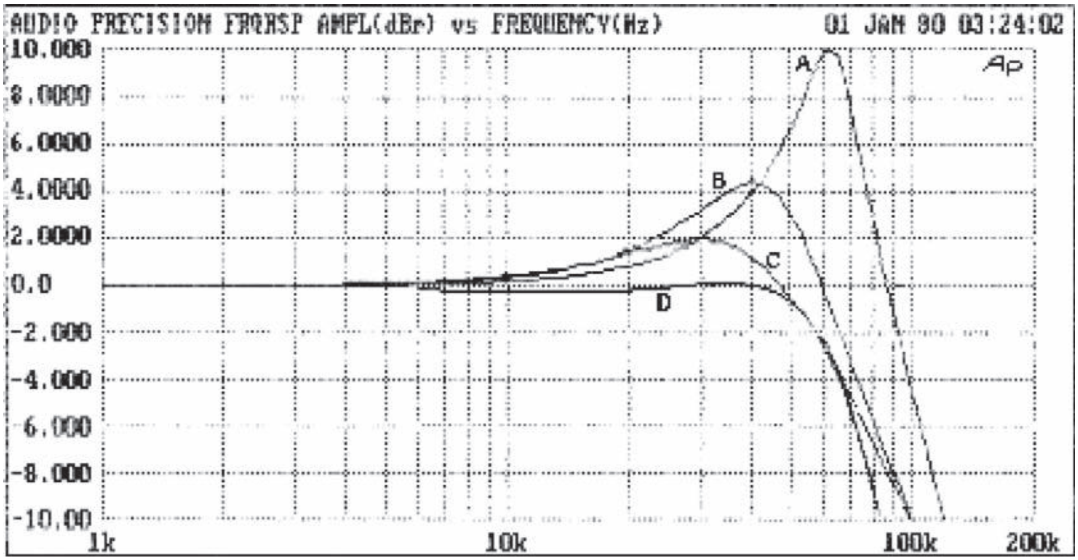


Figure 16.17: Optimising the frequency response of a transformer balanced input by placing a Zobel network across the secondary winding.

and are only intended to deliver very small amounts of power; they are *not* interchangeable with line output transformers. A most ingenious approach to dealing with this distortion problem by operating the input transformer core at near-zero flux was published by Paul Zwicky in 1986 [9]; unfortunately two transformers are required.

There is a bit more to correctly loading the transformer secondary. If it is simply loaded with a high-value resistor there will be peaking of the frequency response due to resonance between the transformer leakage inductance and the winding capacitance [10]. This is shown in Figure 16.17, where a Sowter 3276 line input transformer (a high-quality component) was given a basic resistive loading of 100 kΩ. The result was Trace A, which has a 10 dB peak around 60 kHz. This is bad not only because it accentuates the effect of out-of-band noise, but because it intrudes on the audio frequency response, giving a lift of 1 dB at 20 kHz. Reducing the resistive load R2 would damp the resonance, but it would also reduce the input impedance. The answer is to add a Zobel network, that is, a resistor and capacitor in series, across the secondary; this has no effect except at high frequencies. The first attempt used R1 = 2k7 and C1 = 1 nF, giving Trace B, where the peaking has been reduced to 4 dB around 40 kHz, but the 20 kHz lift is actually slightly greater. R1 = 2k7 and C1 = 2 nF gave Trace C, which is a bit better in that it only has a 2 dB peak. A bit more experimentation ended up with R1 = 3k3 and C1 = 4.3 nF (3n3 + 1 nF) and yielded Trace D, which is pretty flat, though there is a small droop around 10 kHz. The Zobel values are fairly critical for the flattest possible response, and must certainly be adjusted if the transformer type is changed.

No discussion on transformer coupling is complete without pointing out that transformers have poor linearity at low frequencies- orders of magnitude worse than any electronic circuitry. A line input transformer is lightly loaded (by 22 kΩ in Figure 16.16) but this does not reduce the LF distortion, which comes as the inharmonious third harmonic. What does affect it is the source impedance feeding it—50 Ω can easily double the distortion compared with a negligible source impedance, and you're not likely to see lower than 50 Ω unless the source equipment has been deliberately designed with so-called "zero-impedance" output stages; these are described in Chapter 17 on line outputs. Fortunately, the signal levels in the bottom octave of the audio band are usually something like 10–12 dB lower than the maximum amplitudes, which occur in the middle frequencies, and this helps to ease the situation a bit; see Chapter 14 on system design for more on the amplitude/frequency distribution of musical signals.

Transformers also have the traditional disadvantages of size, weight, and cost. With these issues to grapple with, you can see why people don't design in transformers unless they really have to.

16.27 Input Overvoltage Protection

Input overvoltage protection is not common in hi-fi applications, but is regarded as essential in most professional equipment. The normal method is to use clamping diodes, as shown in Figure 16.18, that prevent certain points in the input circuitry from moving outside the supply rails.

This is straightforward, but there are two points to watch. Firstly, the ability of this circuit to withstand excessive input levels is not without limit. Sustained overvoltages may burn out R5 and R6, or pump unwanted amounts of current into the supply rails; this sort of protection is mainly aimed at transients. Secondly, diodes have a non-linear junction capacitance when they are reverse biased, so if the source impedance is significant the diodes will cause distortion at high frequencies. To quantify this problem here are the

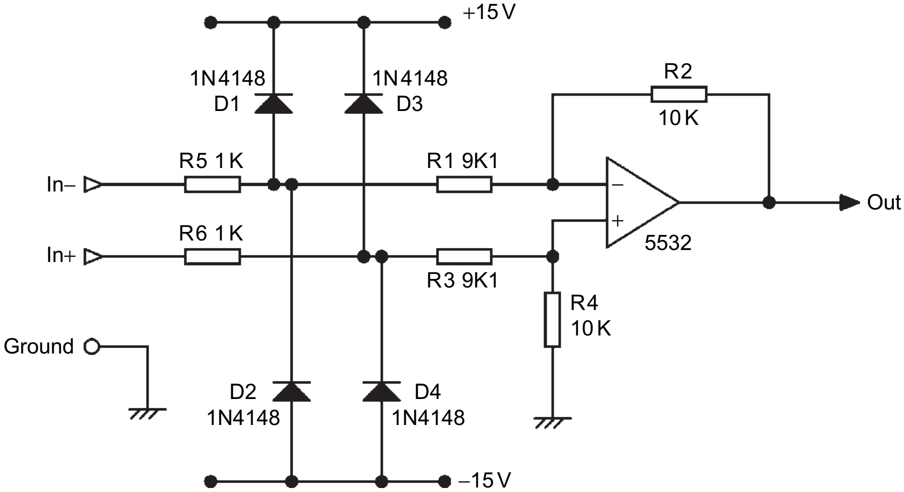


Figure 16.18: Input overvoltage protection for a balanced input amplifier.

results of a few tests. If Figure 16.18 is fed from the low impedance of the usual kind of line output stage, the impedance at the diodes will be about 1 kΩ and the distortion induced into an 11 Vrms 20 kHz input will be below the noise floor. However, if the source impedance is high so the impedance at the diodes is increased to 10 kΩ, with the same input level, the THD at 20 kHz was degraded from 0.0030% to 0.0044% by adding the diodes. I have thought up a rather elegant way to eliminate this effect completely, but being a mercenary fellow I hope to sell it to someone.

16.28 Noise and Balanced Inputs

So far we have not said much about the noise performance of balanced inputs, though on our way through the chapter we have noted that standard balanced input amplifier constructed with four 10 kΩ resistors and a 5532 section has the relatively high noise output of −104.8 dBu with both its inputs terminated by 50 Ω to ground. That value of 50 Ω is in no way critical because its value is so much lower than that of the 10 kΩ resistors.

When evaluating some sorts of input amplifier which have a well-defined input load, such as microphone preamplifiers or moving-magnet RIAA preamplifiers, it is useful to state the Noise Figure (NF). This is the difference in dB between what the Johnson noise from the input load would be if it was amplified by a theoretical noiseless amplifier, and the real noise output with the real amplifier. It is a powerful tool as it indicates at once how far short of perfection a design falls. Evaluating the NF of a microphone preamplifier is straightforward as the input load is usually treated as a pure resistance of 200 Ω; NFs of 1 or 2 dB can be obtained easily using a combination of discrete BJTs as input devices, though usually only at high gains because otherwise the Johnson noise from the gain pot resistance degrades things [12]. The situation with a moving-magnet RIAA preamplifier is more complex as the input load has a high value of inductance as well as series resistance, and this has powerful effects on the noise generated both by the amplifier and the input loading resistance [13]. Nonetheless, NFs of 3 dB or so are possible using 5532 or 5534 opamps for the input stage.

The situation is rather different if we try to apply this to a balanced input amplifier. Take the standard balanced input with its four 10 kΩ resistors, 5532 section, and both inputs terminated by 50 Ω to ground. The Johnson noise from each 50 Ω resistor is −135.2 dBu over a 22 kHz bandwidth at 25°C; the noise from both together is 3 dB more because the two noise sources are uncorrelated. This means that the noise voltage at the input is −132.2 dBu, a very low level in anyone's terms. Unfortunately, the noise output from the stage is −104.8 dBu, giving us a Noise Figure of 27.4 dB. In most fields of electronic endeavour this would be regarded as truly appalling, and fit only for the dustbin. However, simple balanced input amplifiers of this type are widely used in the professional audio industry, so clearly Noise Figures are in this case not that useful a figure of merit.

What NFs do is to show us what room there is for improvement in our design. 27.4 dB is a lot of room, and in the next section we will attempt to cut the NF down to size.

16.29 Low-Noise Balanced Inputs

I have remarked several times that the standard balanced input amplifier with four 10 kΩ resistors shown in Figure 16.19a is markedly noisier than an unbalanced input like that in Figure 16.1. The unbalanced input stage, with its input terminated by 50 Ω to ground, has a noise output of −119.0 dBu over the usual 22–22 kHz bandwidth. If the balanced circuit is built with 10 kΩ resistors and a 5532 section, the noise output is −104.8 dBu with the inputs similarly terminated. This is a big difference of 14.2 dB.

In the hi-fi world in particular, where an amplifier may have both unbalanced and balanced inputs, most people feel that this is the wrong way round. Surely the balanced input, with its professional XLR connector and its much-vaunted rejection of ground noise, should show a better performance in all departments? Well, it does—except as regards internal noise, and a 14 dB discrepancy is both clearly audible and hard to explain away. This section explains how to design it away instead.

We know that the source of the extra noise is the relatively high resistor values around the opamp (see Table 16.7 earlier in the chapter), but these cannot be reduced in the simple balanced input amplifier without reducing the input impedances below what is acceptable. The answer is to lower the resistor values but buffer them from the input with a pair of voltage-followers; this arrangement is shown in Figure 16.19b. 5532s are a good choice for this as they combine low voltage noise with low distortion and good load-driving capability. Since the input buffers are voltage-followers with 100% feedback, their gain is very

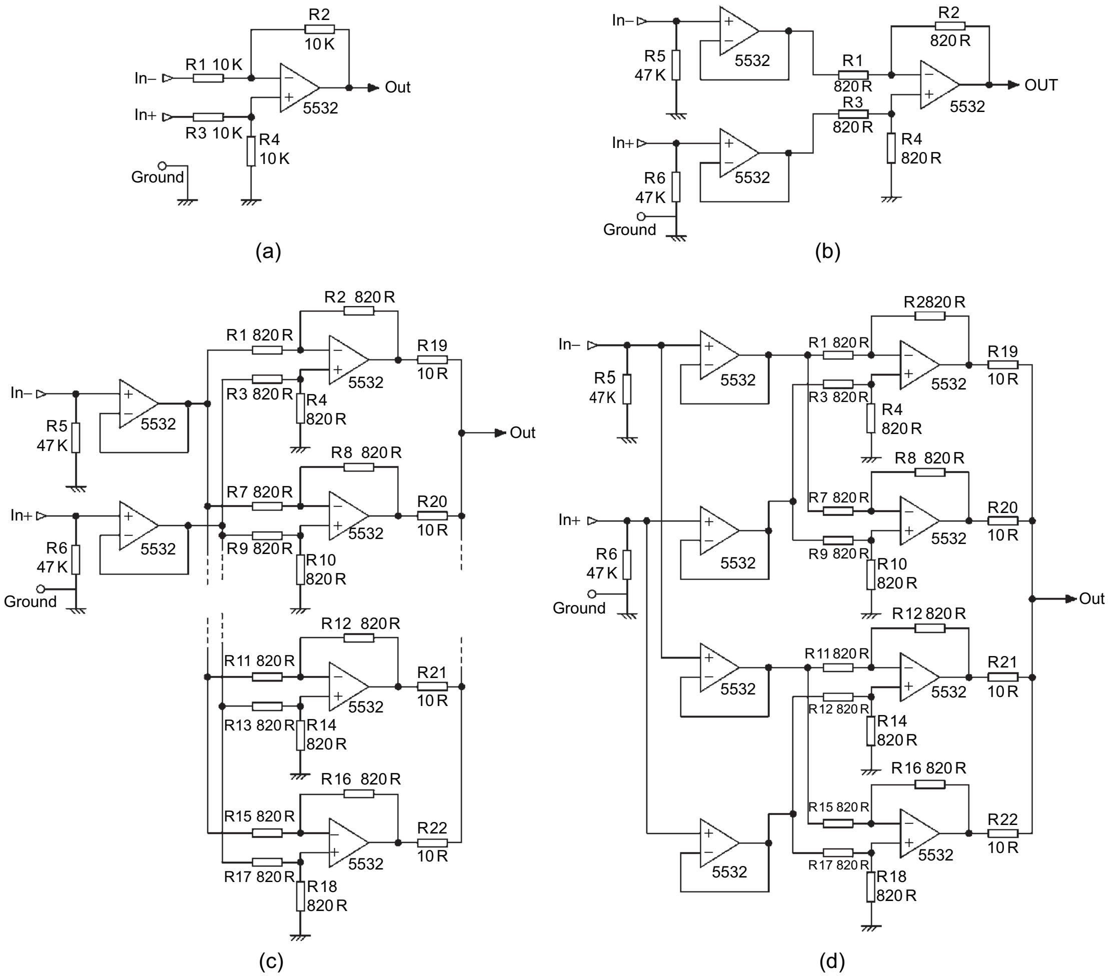


Figure 16.19: Low noise unity-gain balanced inputs using multiple 5532 buffers and differential amplifiers.

accurately defined at unity and the CMRR is therefore not degraded; CMRR is still defined by the resistor tolerances, and by the bandwidth of the differential opamp. In fact, the overall CMRR for the balanced link is likely to be improved, as we now have equal hot and cold input impedances in all circumstances, and we can make them much higher than the usual balanced input impedances. Figures 16.19b to 16.19d show 47 kΩ input resistors, but these could easily be raised to 100 kΩ. The offset created by the input bias current flowing through this resistance needs to be watched, but it should cause no external troubles if the usual blocking capacitors are used.

There is a limit to how far the four resistors can be reduced, as the differential stage has to be driven by the input buffers, and it also has to drive its own feedback arm. If 5532s are used a safe value that gives no measurable deterioration of the distortion performance is about 820 Ω, and an 5532 differential stage alone (without the buffers) and 4 × 820 Ω resistors gives a noise output of −111.7 dBu, which is 6.6 dB lower than the standard 4 × 10 kΩ version. Adding the two input buffers degrades this only slightly to −110.2 dB, because we are adding only the voltage noise component of the two new opamps, and we are still 5.4 dB quieter than the original 4 × 10 kΩ version. It is interesting point that we now have three opamps in the signal path instead of one, but we still get a significantly lower noise level.

This might appear to be all we can do; it is not possible to reduce the value of the four resistors around the differential amplifier any further without compromising linearity. However, there is almost always some way to go further in the great game that is electronics, and here are three possibilities. A step-up transformer could be used to exploit the low source impedance (remember we are still assuming the source impedances are 50 Ω) and it might well work superbly in terms of noise alone, but transformers are always heavy, expensive, susceptible to magnetic fields, and of doubtful low-frequency linearity. We would also very quickly run into headroom problems; balanced line input amplifiers are normally required to attenuate rather than amplify.

We could design a discrete-opamp hybrid stage with discrete input transistors, which are quieter than those integrated into IC opamps, coupled to an opamp to provide raw loop gain; this can be quite effective but you need to be very careful about high-frequency stability, and it is difficult to get an improvement of more than 6 dB. Thirdly, we could design our own opamp using all discrete parts; this approach tends to have less stability problems as all circuit parameters are accessible, but it definitely requires rather specialised skills, and the result takes up a lot of PCB area.

Since none of those three approaches are appealing, now what? One of the most useful techniques in low-noise electronics is to use two identical amplifiers so that the gains add arithmetically, but the noise from the two separate amplifiers, being uncorrelated, partially cancels. Thus we get a 3 dB noise advantage each time the number of amplifiers used is doubled. This technique works very well with multiple opamps; let us apply it and see how far it may be usefully taken.

Since the noise of a single 5532-section unity-gain buffer is only −119.0 dBu, and the noise from the 4 × 820 Ω differential stage (without buffers) is a much higher −111.7 dBu, the differential stage is clearly the place to start work. We will begin by using two identical 4 × 820 Ω differential amplifiers as shown in the top section of Figure 16.19c, both driven from the existing pair of input buffers. This will give partial cancellation of both resistor and opamp noise from the two differential stages if their outputs are summed. The main question is how to sum the two amplifier outputs; any active solution would introduce another opamp, and hence more noise, and we would almost certainly wind up worse off than when we started. The answer is however, beautifully simple. We just connect the two amplifier outputs together with 10 Ω resistors; the gain does not change but the noise output drops. The signal output of both amplifiers is nominally the same, so no current should flow from one opamp output to the other. In practice there will be slight gain differences due to resistor tolerances, but with 1% resistors I have never experienced any hint of a problem. The combining resistor values are so low at 10 Ω that their Johnson noise contribution is negligible.

The use of multiple differential amplifiers has another advantage—the CMRR errors are also reduced in the same way that this the noise is reduced. This is also similar to the use of multiple resistors or capacitors to improve the accuracy of the total value, as explained in Chapter 12.

We therefore have the arrangement of Figure 16.19c, with single input buffers, (i.e., one per input) and two differential amplifiers, and this reduces the noise output by 2.3 dB to −112.5 dBu, which is quieter than the original 4 × 10 kΩ version by an encouraging 7.4 dB. We do not get a full 3 dB noise improvement because both differential amplifiers are handling the noise from the input buffers, which is correlated and so is not reduced by partial cancellation. The contribution of the input buffer noise is further brought out if we take the next step of using four differential amplifiers. There is of course nothing special about using amplifiers in powers of two. It is perfectly possible to use three or five differential amplifiers in the array, which will give intermediate amounts of noise reduction. If you have a spare opamp section, then put it to work!

So, leaving the input buffers unchanged, we use them to drive an array of four differential amplifiers. These are added on at the dotted lines in the lower half of Figure 16.19c. We get a further improvement, but only by 1.5 dB this time. The output noise is down to −114.0 dBu, quieter than the original 4 × 10 kΩ version by 8.9 dB. You can see that at this point we are proceeding by decreasing steps, as the input buffer noise is starting to dominate, and there seems little point in doubling up the differential amplifiers again; the amount of hardware could be regarded as a bit excessive, and so would the PCB area occupied. The increased loading on the input buffers is also a bit of a worry.

A more fruitful approach is to tackle the noise from the input buffers, by doubling them up as in Figure 16.19d, so that each buffer drives only two of the four differential amplifiers. This means that the buffer noise will also undergo partial cancellation, and will be reduced by 3 dB. There is however still the contribution from the differential amplifier noise, and so the actual improvement on the previous version is 2.2 dB, bringing the output noise down to −116.2 dBu, which is quieter than the original 4 × 10 kΩ version by a thumping 11.1 dB. Remember that there are two inputs, and "double buffers" means two buffers per hot and cold input, giving a total of four in the complete circuit.

Since doubling up the input buffers gave us a useful improvement, it's worth trying again, so we have a structure with quad buffers and four differential amplifiers, as shown in Figure 16.20, where each differential amplifier now has its very own buffer. This improves

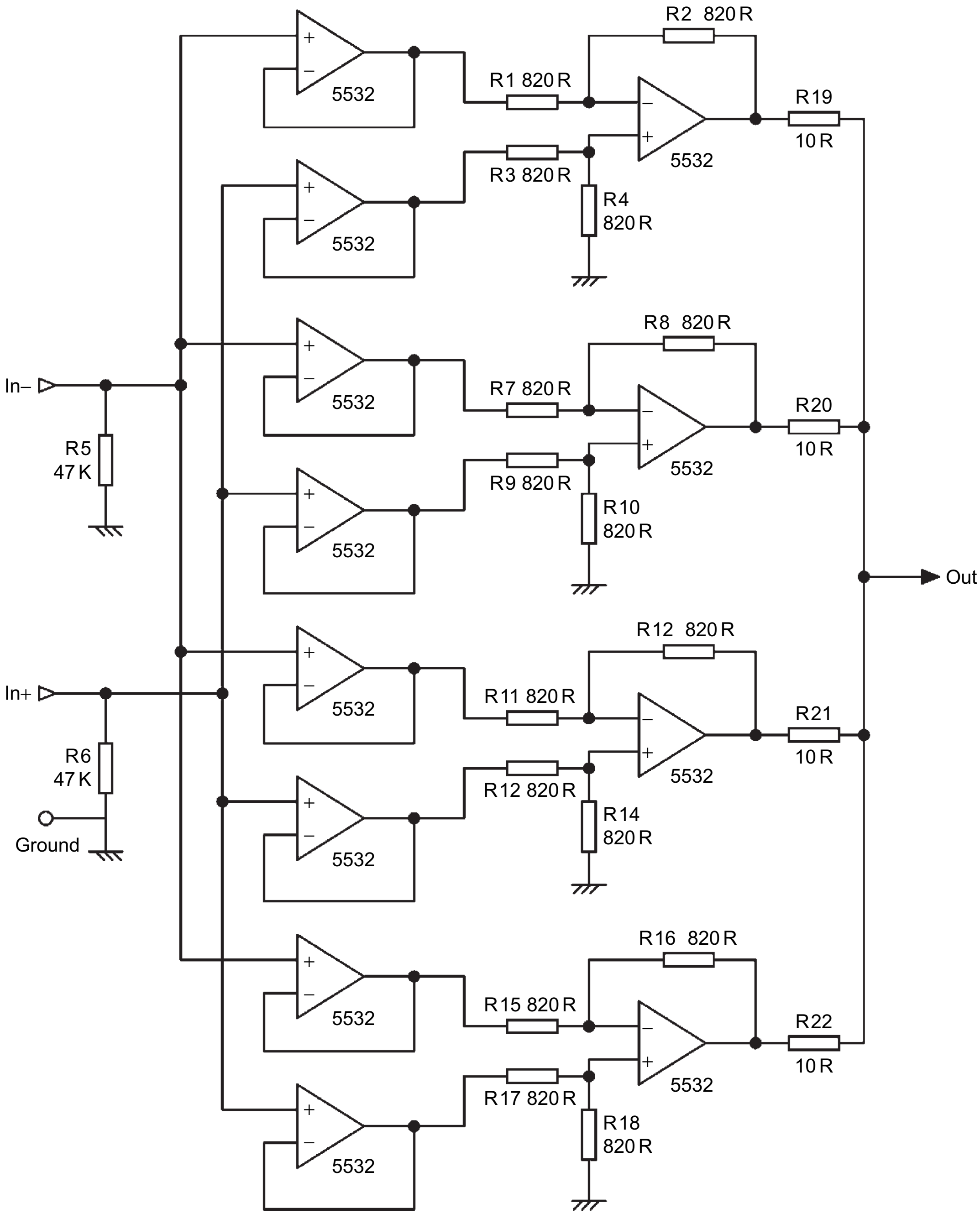


Figure 16.20: The final 5532 low noise unity-gain balanced input stage, with quad input buffers and four differential amplifiers. The noise output is only −117.0 dBu.

on the previous version by a rather less satisfying 0.8 dB, giving an output noise level of −117.0 dBu, quieter than the original 4 × 10 kΩ version by 11.9 dB. The small improvement we have gained indicates that the focus of noise reduction needs to be returned to the differential amplifier array, but the next step there would seem to be using eight amplifiers, which is not very appealing. Thoughts about ears of corn on chessboards tend to intrude at this point.

This is a good moment to pause and see what we have achieved. We have built a balanced input stage that is quieter than the standard balanced input by 11.9 dB, using standard components of low cost. We have used increasing numbers of them, but the total cost is still small compared with enclosures, power supplies, etc. On the other hand, the power consumption is naturally several times greater. The technology is highly predictable and the noise reduction reliable; in fact it is bullet-proof. The linearity is as good as that of a single opamp of the same type, and in the same way there are no HF stability problems.

What we have not done is build a balanced input that is quieter than the unbalanced one—we are still 2.0 dB short of that target, but at least we have reached a point where the balanced input is not obviously noisier. Earlier we evaluated the Noise Figure of the standard 4 × 10 kΩ balanced input and found it was a startling 27.4 dB; our new circuit here has a Noise Figure of 15.2 dB, which looks a good deal more respectable.

The noise results, including Noise Figures, are all summarised in Table 16.9 at the end of this chapter.

16.30 Low-Noise Balanced Inputs in Real Life

Please don't think that this examination of low-noise input options is merely a voyage off into pure theory. It has real uses and has been applied in practice. The Cambridge Audio 840 W power amplifier, a design of mine which, I might modestly mention in passing, won a CES Innovation Award in January 2008. This unit has both unbalanced and balanced inputs, and conventional technology would have meant that the balanced inputs would have been significantly the noisier of the two. Since the balanced input is the "premium" input, many people would think there was something amiss with this state of affairs. We therefore decided the balanced input had to be quieter than the unbalanced input. Using 5532s in an architecture similar to those outlined above, this requirement proved straightforwardly attainable, and the final balanced input design was both economical and quieter than its unbalanced neighbour by a dependable 0.9 dB. Two other versions were evaluated that made the balanced input quieter than the unbalanced one by 2.8 dB, and by 4.7 dB, at somewhat greater cost and complexity. These were put away for possible future upgrades.

The Signal Transfer Company [11] manufactures a low-noise balanced input card based on these principles which has 47 kΩ input impedances, unity gain, and a noise output of only −115 dBu.

16.31 Ultra-Low-Noise Balanced Inputs

In the section on low-noise balanced inputs above. we reclined briefly on our laurels, having achieved an economical balanced input stage with output noise at the extremely low level of −117.0 dBu. Regrettably this is still 2 dB noisier than a simple unbalanced input. It would be wrong to conclude from this that the resources of electronic design are exhausted. At the end of the noise-reduction sequence we were aware that the dominant noise source was currently the differential amplifier array, and we shrank from doubling up again to use eight amplifiers because of issues of cost and the PCB area occupied. We will take things another step by taking a much more relaxed view of cost (as one does at the "high-end"), and see how that changes the game. We will, however, retain some concern about PCB area.

An alternative way to make the differential amplifier array quieter is simply to use opamps that are quieter. These will inevitably be more expensive—much more expensive—than the ubiquitous 5532. Because of the low resistor values around the opamps we need to focus on low voltage-noise rather than low current noise, and there are several that are significantly better than the 5532, as shown by the typical noise density figures in Table 16.8.

Clearly moving to the 5534A will give a significant noise reduction, but since there is only a single opamp per package, and external compensation is needed, the board area used will be much greater. The new chip on the block is the LM4562, a bipolar opamp which has finally surpassed the 5532 in performance. The input voltage noise density is typically 2.7 nV/√Hz, substantially lower than the 5 nV/√Hz of the 5532. For applications with low source impedances, this implies a handy noise advantage of 5 dB or more. The LM4562 is a dual opamp will not take up more space. At the time of writing it is something like 10 times more expensive than the 5532.

Step One—we replace all four opamps in the differential amplifiers with LM4562s. They are a drop-in replacement with no circuit adjustments required at all. We leave the quad 5532

Table 16.8: Voltage and Noise Densities for Low-Noise Balanced-Input Opamp Candidates

Opamp	Voltage Noise Density nV√Hz	Current Noise Density pA√Hz
5532	5	0.7
5534A	3.5	0.4
LM4562	2.7	1.6
AD797	0.9	2
LT1028	0.85	1

input buffers in place. The noise output drops by an impressive 1.9 dB, giving an output noise level of −118.9 dBu, quieter than the original 4 × 10 kΩ version by 14.1 dB, and only 0.1 dB noisier than the unbalanced stage.

Step Two—replace the quad 5532 buffers with quad LM4562 buffers. Noise falls by only 0.6 dB, the output being −119.5 dBu, but at last we have a balanced stage that is quieter than the unbalanced stage, by a small but solid 0.5 dB.

One of the pre-eminent low-noise-with-low-source-resistance opamps is the AD797 from Analog Devices, which has a remarkably low voltage noise at 0.9 nV/√Hz (typical at 1 kHz) but it is a very expensive part, costing between 20 and 25 times more than a 5532 at the time of writing. The AD797 has is a single opamp, while the 5532 is a dual, so the cost per opamp is actually 40 to 50 times greater, and more PCB area is required, but the potential improvement is so great we will overlook that.

Step Three—we replace all four opamps in the differential amplifiers with AD797s, putting the 5532s back into the input buffers in the hope that we might be able to save money somewhere. The noise output drops by a rather disappointing 0.4 dB, giving an output noise level of −119.9 dBu, quieter than the original 4 × 10 kΩ version by 15.1 dB.

Perhaps putting those 5532s back in the buffers was a mistake? Our fourth and final move in this game of electronic chess is to replace all the quad 5532 input buffers with dual (not quad) AD797 buffers. This requires another four AD797s (two per input) and is once more not a cheap strategy. We retain the four AD797s in the differential amplifiers. The noise drops by another 0.7 dB yielding an output noise level of −120.6 dBu, quieter than the original 4 × 10 kΩ version by 15.8 dB, and quieter than the unbalanced stage by a satisfying 1.6 dB. You can do pretty much anything in electronics with a bit of thought and a bit of money.

If however, we look at the Noise Figure for this final design, we feel a bit less happy. Despite deploying some ingenious circuitry and a lot of premium opamps, we still have an NF of 11.6 dB. We have not even managed to get the NF down to single figures, and so there is plenty of scope yet for some creative design. The relatively high NF is essentially because the reference input loads we are using are 50 Ω resistors, which naturally generate a very low level of Johnson noise (−136.2 dBu each). If we want to make more progress in this direction we might start thinking about moving-coil preamp circuitry, which can achieve NFs of less than 7 dB with an input load as low as 3.3 Ω [14]. This can be achieved by using special low-Rb discrete transistors as input devices, with an opamp to provide open-loop gain. This sort of hybrid circuitry must be carefully designed to avoid HF stability problems, whereas simply plugging in more opamps always works.

You are probably wondering what happened to the LT1028 lurking at the bottom of Table 16.8. It is true that its voltage noise density is slightly better than that of the AD797, but there is a subtle snag. As described in Chapter 13, the LT1028 has bias-current cancellation circuitry which injects correlated noise currents into the two inputs. These will cancel if the impedances seen by the two inputs are the same, but in moving-magnet amplifier use the impedances differ radically and the LT1028 is not useful in this application. The input conditions here are more benign, but the extra complication is unwelcome and I have never used the LT1028 in audio work. In addition, it is a single opamp with no dual version.

This is not of course the end of the road. The small noise improvement in the last step we made tells us that the differential amplifier array is still the dominant noise source, and further development would have to focus on this. A first step would be to see if the relatively high current-noise of the AD797s is significant with respect to the surrounding resistor values. If so, we need to see if the resistor values can be reduced without degrading linearity at full output. We should also check the Johnson noise contribution of all those 820 Ω resistors; they are generating −123.5 dBu each at room temperature, but of course the partial cancellation effect applies to them as well.

All these noise results are also summarised in Table 16.9 below.

Table 16.9: A Summary of the Noise Improvements Made to the Balanced Input Stage

Buffer Type	Amplifier	Noise Output dBu	Improvement on Previous Version dB	Improvement Over 4 × 10 kΩ Diff Amp dB	Noisier Than Unbal Input by: dB	Noise Figure Ref 2 × 50 Ω dB
	5532 voltage-follower	−119.0			0 dB ref	
None	Standard diff amp 10 K 5532	−104.8	0	0.0 dB ref	14.2	27.4
None	Single diff amp 820 R 5532	−111.7	6.9	6.9	7.3	20.5
Single 5532	Single diff amp 820 R 5532	−110.2	5.4	5.4	8.8	22.0
Single 5532	Dual diff amp 820 R 5532	−112.5	2.3	7.4	6.5	19.7
Single 5532	Quad diff amp 820 R 5532	−114.0	1.5	9.2	5.0	18.2
Dual 5532	Quad diff amp 820 R 5532	−116.2	2.2	11.4	2.8	16.0
Quad 5532	Quad diff amp 820 R 5532	−117.0	0.8	12.2	2.0	15.2
Quad 5532	Quad diff amp 820 R LM4562	−118.9	1.9	14.1	0.1	13.3
Quad LM4562	Quad diff amp 820 R LM4562	−119.5	0.6	14.7	−0.5	12.7
Quad 5532	Quad diff amp 820 R AD797	−119.9	0.4	15.1	−0.9	12.3
Dual AD797	Quad diff amp 820 R AD797	−120.6	0.7	15.8	−1.6	11.6

References

[1] N. Muncy, Noise susceptibility in analog and digital signal processing systems, JAES 3 (6) (1995) 447.
[2] B. Whitlock, Balanced lines in audio systems: fact, fiction, & transformers, JAES 3 (6) (1995) 454–465.
[3] T. Williams, EMC for Product Designers, Newnes (Butterworth-Heinemann), Boston, MA, (1992), ISBN 0 7506 1264 9, p. 176.
[4] T. Williams, As [2] above, p. 173.
[5] S. Winder, What's The Difference? Electronics World, London, 1995, p. 989.
[6] J. Graeme, Amplifier Applications of Opamps, McGraw-Hill, New York, 1999, p. 16, ISBN 0-07-134643-0.
[7] D. Birt, Electronically Balanced Analogue-Line Interfaces, Proceedings of the Institute of Acoustics Conference, Windermere, UK, 1990.
[8] D. Self, Small Signal Audio Design, Chap. 9, Focal Press, Boston, MA, 2009, ISBN 978-0-240-52177-0.
[9] P. Zwicky, Low-Distortion Audio Amplifier Circuit Arrangement, US Patent No. 4,567,443, 1986.
[10] G.A.V. Sowter, Soft magnetic materials for audio transformers: history, production, and applications, JAES 35 (10) (1987) 774.
[11] http://www.signaltransfer.freeuk.com/ (low-noise balanced input PCB), 2010 (accessed 12.10).
[12] D. Self, Small Signal Audio Design, Chap. 13, Focal Press, Boston, MA, 2009, p. 325, ISBN 978-0-240-52177-0.
[13] D. Self, Small Signal Audio Design, Chap. 7, Focal Press, Boston, MA, 2009, p. 191, ISBN 978-0-240-52177-0.
[14] D. Self, Small Signal Audio Design, Chap. 8, Focal Press, Boston, MA, 2009, p. 217, ISBN 978-0-240-52177-0.

CHAPTER 17

Line Outputs

17.1 Unbalanced Outputs

There are only two electrical output terminals for an unbalanced output—signal and ground. However, the unbalanced output stage in Figure 17.1a is fitted with a three-pin XLR connector, to emphasise that it is always possible to connect the cold wire in a balanced cable to the ground at the output end and still get all the benefits of common-mode rejection if you have a balanced input. If a two-terminal connector is fitted, the link between the cold wire and ground has to be made inside the connector, as shown in Figure 16.2 in the chapter on line inputs.

The output amplifier in Figure 17.1a is configured as a unity-gain buffer, though in some cases it will be connected as a series feedback amplifier to give gain. A non-polarised DC blocking capacitor C1 is included; 100 uF gives a −3 dB point of 2.6 Hz with one of those notional 600 Ω loads. The opamp is isolated from the line shunt-capacitance by a resistor R2, in the range 47–100 Ω, to ensure HF stability, and this unbalances the hot and cold line impedances. A drain resistor R1 ensures that no charge can be left on the output side of C1; it is placed *before* R2, so it causes no attenuation. In this case the loss would only be 0.03 dB, but such minor errors can build up to an irritating level in a large system and it costs nothing to avoid them.

If the cold line is simply grounded as in Figure 17.1a, then the presence of R2 degrades the CMRR of the interconnection to an uninspiring −43 dB even if the balanced input at the other end of the cable has infinite CMRR in itself and perfectly matched 10 kΩ input impedances.

To fix this problem, Figure 17.1b shows what is called an impedance-balanced output. The cold terminal is neither an input nor an output, but a resistive termination R3 with the same resistance as the hot terminal output impedance R2. If an unbalanced input is being driven, this cold terminal is ignored. The use of the word "balanced" is perhaps unfortunate as when taken together with an XLR output connector it implies a true balanced output with anti-phase outputs, which is *not* what you are getting. The impedance-balanced approach is not particularly cost-effective, as it requires significant extra money to be spent on an XLR connector. Adding an opamp inverter to make it a proper balanced output costs little more, especially if there happens to be a spare opamp half available, and it sounds much better in the specification.

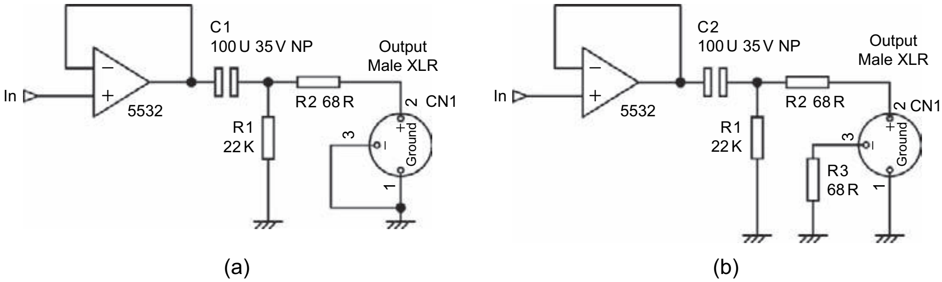


Figure 17.1: Unbalanced outputs, (a) simple output, and (b) impedance-balanced output for improved CMRR when driving balanced inputs.

There is an example of the use of impedance-balancing in the active crossover design example in Chapter 19; here the outputs come directly from low-impedance level-trim controls, with output impedances that vary somewhat with the level settings, so compromise values for the impedance-balancing resistances must be used.

Active crossover output stages may also incorporate level trim controls, mute switches, and phase-invert switches. These features are covered in Chapter 14 on crossover system design.

17.2 Zero-Impedance Outputs

Both the unbalanced outputs shown in Figure 17.1 have series output resistors to ensure stability when driving cable capacitance. This increases the output impedance and can lead to increased crosstalk in some situations, notably when different signals are being passed down the same signal cable. This is a particular problem when two or more layers of ribbon cable are laid together in a "lasagne" format for neatness or to save space. In some cases layers of grounded screening foil are interleaved with the cables, but this is rather expensive and awkward to do, and does not greatly reduce crosstalk between conductors in the same piece of ribbon. The only way to do this is to reduce the output impedance.

A simple but very effective way to do this is the so-called "zero-impedance" output configuration. Figure 17.2a shows how the technique is applied to an unbalanced output stage with 10 dB of gain. Feedback at audio frequencies is taken from outside isolating resistor R3 via R2, while the HF feedback is taken from inside R3 via C2 so it is not affected by load capacitance and stability is unimpaired. Using a 5532 opamp, the output impedance is reduced from 68 Ohms to 0.24 Ohms at 1 kHz—a dramatic reduction that will reduce purely capacitive crosstalk by an impressive 49 dB. The output impedance increases to 2.4 Ohms at 10 kHz and 4.8 Ohms at 20 kHz as opamp open-loop gain falls with frequency. The impedance-balancing resistor on the cold pin has been replaced by a link to match the near-zero output impedance at the hot pin.

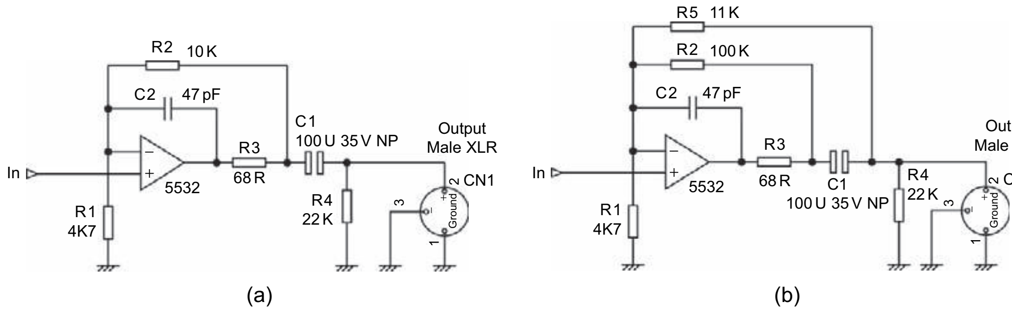


Figure 17.2: (a) Zero-impedance output; (b) zero-impedance output with NFB around output capacitor.

Figure 17.2b shows a refinement of this scheme with three feedback paths. Electrolytic coupling capacitors can introduce distortion if they have more than a few tens of milliVolts of signal across them, even if the time-constant is long enough to give a virtually flat LF response. (This is looked at in detail in the chapter on passive components.) In Figure 17.2b most of the feedback is now taken from outside C1, via R5, so it can correct capacitor distortion. The DC feedback goes via R2, now much higher in value, and the HF feedback goes through C2 as before to maintain stability with capacitive loads. R2 and R5 in parallel come to 10 kΩ so the gain is the same. Any circuit with separate DC and AC feedback paths must be checked carefully for frequency response irregularities, which may happen well below 10 Hz.

17.3 Ground-Cancelling Outputs

This technique, also called a ground-compensated output, appeared in the early 1980s in mixing consoles. It allows ground voltages to be cancelled out even if the receiving equipment has an unbalanced input; it prevents any possibility of creating a phase error by mis-wiring; and it costs virtually nothing in itself though it does require a three-pin output connector.

Ground-cancelling (GC) separates the wanted signal from the unwanted ground voltage by addition at the output end of the link, rather than by subtraction at the input end. If the receiving equipment ground differs in voltage from the sending ground, then this difference is added to the output signal so that the signal reaching the receiving equipment has the same ground voltage superimposed upon it. Input and ground therefore move together and the ground voltage has no effect, subject to the usual effects of component tolerances. The connecting lead is wired differently from the more common unbalanced-out, balanced-in situation, as now the cold line is be joined to ground at the *input* or receiving end.

An inverting unity-gain ground-cancel output stage is shown in Figure 17.3a. The cold pin of the output socket is now an input, and has a unity-gain path summing into the main signal going to the hot output pin to add the ground voltage. This path R3, R4 has a

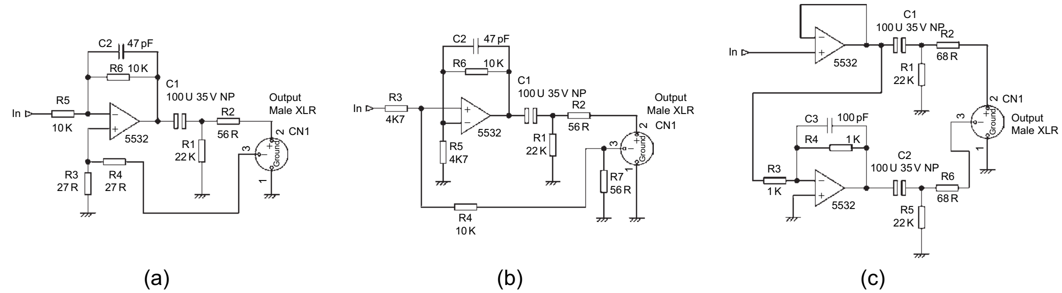


Figure 17.3: (a) Inverting ground-canceling output; (b) non-inverting ground-canceling output; (c) a true balanced output.

very low input impedance equal to the hot terminal output impedance so if it *is* used with a balanced input, then the line impedances will be balanced, and the combination will still work effectively. The 6 dB of attenuation in the R3-R4 divider is undone by the gain of two set by R5, R6. It is unfamiliar to most people to have the cold pin of an output socket as a low impedance input, and its very low input impedance minimises the problems caused by mis-wiring. Shorting it locally to ground merely converts the output to a standard unbalanced type. On the other hand, if the cold input is left unconnected then there will be a negligible increase in noise due to the very low input resistance of R3.

This is the most economical GC output, but obviously a phase-inversion is not always convenient. Figure 17.3b shows a non-inverting GC output stage with a gain of 6.6 dB. R5 and R6 set up a gain of 9.9 dB for the amplifier, but the overall gain is reduced by 3.3 dB by attenuator R3, R4. The cold line is now terminated by R7, and any signal coming in via the cold pin is attenuated by R3, R4 and summed at unity gain with the input signal. A non-inverting GC stage must be fed from a very low impedance such as an opamp output to work properly. There is a slight compromise on noise performance here because attenuation is followed by amplification.

Ground-cancelling outputs are an economical way of making ground-loops innocuous when there is no balanced input, and it is rather surprising they are not more popular; perhaps people find the notion of an input pin on an output connector unsettling. In particular, GC outputs would appear to offer the possibility of a quieter interconnection than the standard balanced interconnection because a relatively noisy balanced input is not required; see Chapter 16 on line inputs. Ground-cancelling outputs can also be made zero-impedance using the techniques described earlier.

17.4 Balanced Outputs

Figure 17.3c shows a balanced output, where the cold terminal carries the same signal as the hot terminal but phase-inverted. This can be arranged simply by using an opamp stage to invert the normal in-phase output. The resistors R3, R4 around the inverter should be as low in value as possible to minimise Johnson noise, because this stage is working at a noise-gain of two, but bear in mind that R3 is effectively grounded at one end and its loading, as well as the external load, must be driven by the first opamp. A unity-gain follower is shown for the first amplifier, but this can be any other shunt or series feedback stage as convenient. The inverting output if not required can be ignored; it must *not* be grounded, because the inverting opamp will then spend most of its time clipping in current-limiting, almost certainly injecting unpleasantly crunching distortion into the crossover grounding system. Both hot and cold outputs must have the same output impedances (R2, R6) to keep the line impedances balanced and the interconnection CMRR maximised.

It is vital to realise that this sort of balanced output, unlike transformer balanced outputs, by itself gives no common mode rejection at all. It must be connected to a balanced input that can subtract one output from another if ground noise is to be cancelled.

The advantage that this kind of balanced output has over an unbalanced output is that the total signal level on the interconnection is increased by 6 dB, which if correctly handled can improve the signal-to-noise ratio, especially if the balanced input amplifier is relatively noisy, which the standard version certainly is. The extra noise from the inverting output amplifier is negligible compared with this. It is less likely to crosstalk to other lines even if they are unbalanced, as the currents injected via the stray capacitance from each line will tend to cancel; how well this works depends on the physical layout of the conductors. All balanced outputs give the facility of correcting phase errors by swapping hot and cold outputs. This is, however, a two-edged sword, because it is probably how the phase got wrong in the first place.

There is no need to worry about the exact symmetry of level for the two output signals; ordinary 1% tolerance resistors are fine. Slight gain differences between the two outputs only affect the signal-handling capacity of the interconnection by a very small amount. This simple form of balanced output is the norm in hi-fi balanced interconnection, but is less common in professional audio, where the quasi-floating output, which emulates a transformer winding, gives both common mode rejection and more flexibility in situations where temporary connections are frequently being made.

17.5 Transformer Balanced Outputs

If true galvanic isolation between equipment grounds is required, this can only be achieved with a line transformer, sometimes called a line-isolating transformer; don't confuse them with mains-isolating transformers. You don't, as a rule, use line transformers unless you really have to because the much-discussed cost, weight, and performance problems are very real, as you will see shortly. However, they are sometimes found in big sound reinforcement systems and in any environment where high RF field strengths are encountered. They are unlikely to be used in active crossovers for domestic hi-fi. A basic transformer balanced

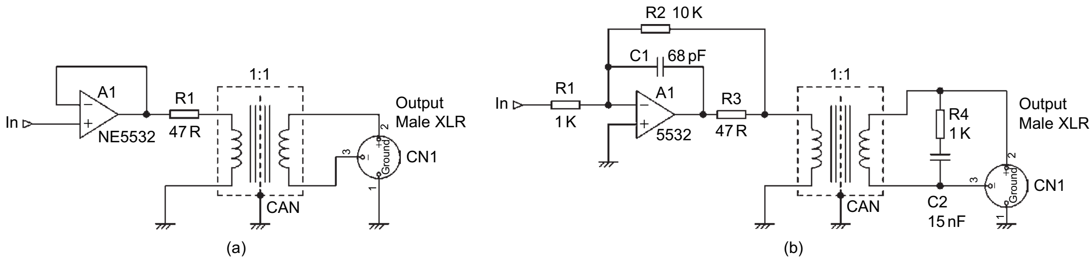


Figure 17.4: Transformer balanced outputs; (a) standard circuit; (b) zero-impedance drive to reduce LF distortion, with Zobel network across secondary.

output is shown in Figure 17.4a; in practice A1 would probably also have some other function, such as providing gain or filtering. In good-quality line transformers there will be an inter-winding screen, which must be earthed to minimise noise pickup and general EMC problems. In most cases, this does *not* ground the external can and you have to arrange this yourself, possibly by mounting the can in a metal capacitor clip. Make sure the can is earthed as this definitely does reduce noise pickup.

Be aware that the output impedance will be higher than usual because of the ohmic resistance of the transformer windings. With a 1:1 transformer, as normally used, both the primary and secondary winding resistances are effectively in series with the output. A small line transformer can easily have 60 Ω per winding, so the output impedance is 120 Ω plus the value of the series resistance R1 added to the primary circuit to prevent HF instability due to transformer winding capacitances and line capacitances. The total can easily be 160 Ω or more, compared with, say, 47 Ω for non-transformer output stages. This will mean a higher output impedance and greater voltage losses when driving heavy loads.

DC flowing through the primary winding of a transformer is bad for linearity, and if your opamp output has anything more than the usual small offset voltages on it, DC current flow should be stopped by a blocking capacitor.

17.6 Output Transformer Frequency Response

If you have looked at the section in Chapter 16 on the frequency response of line input transformers, you will recall that they give a nastily peaking frequency response if the secondary is not loaded properly, due to resonance between the leakage inductance and the stray winding capacitances. Exactly the same problem afflicts output transformers, as shown in Figure 17.5; with no output loading there is a frightening 14 dB peak at 127 kHz. This is high enough in frequency to have very little effect on the response at 20 kHz, but this high-Q resonance isn't the sort of lurking horror you want in your circuitry. It could easily cause some nasty EMC problems.

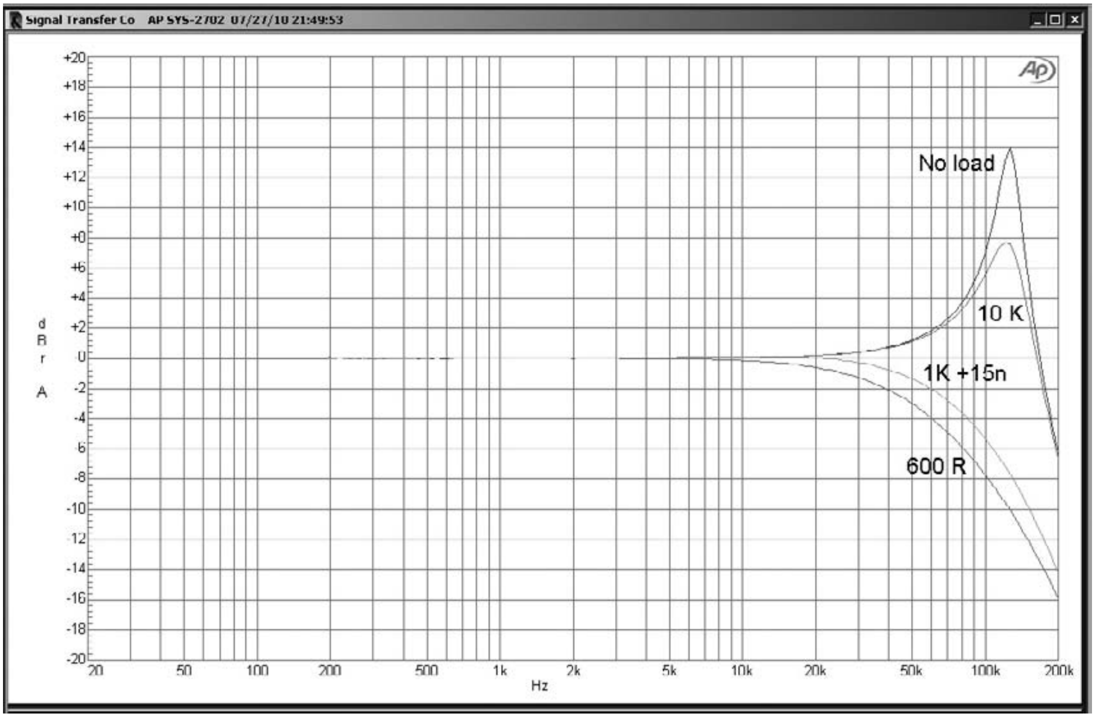


Figure 17.5: Frequency response of a Sowter 3292 output transformer with various loads on the secondary. Zero-impedance drive as in Figure 17.4b.

The transformer measured was a Sowter 3292 1:1 line isolating transformer. Sowter are a highly respected company, and this is a quality part with a mumetal core and housed in a mumetal can for magnetic shielding. When used as the manufacturer intended, with a 600 Ω load on the secondary, the results are predictably quite different, with a well-controlled roll-off that I measured as −0.5 dB at 20 kHz.

The difficulty is that there are very few if any genuine 600 Ω loads left in the world, and most output transformers are going to be driving much higher impedances. If we are driving a 10 kΩ load, the secondary resonance is not much damped and we still get a thoroughly unwelcome 7 dB peak above 100 kHz, as shown in Figure 17.5. We could of course put a permanent 600 Ω load across the secondary, but that will heavily load the output opamp, impairing its linearity, and will give us unwelcome signal loss due in the winding resistances. It is also profoundly inelegant.

A better answer, as in the case of the line-input transformer, is to put a Zobel network, that is, a series combination of resistor and capacitor, across the secondary, as in Figure 17.4b. The capacitor required is quite small and will cause very little loading except at high frequencies where signal amplitudes are low. A little experimentation yielded the values of 1 kΩ in series with 15 nF, which gives the much improved response shown in Figure 17.5. The response is almost exactly 0.0 dB at 20 kHz, at the cost of a very gentle 0.1 dB rise around 10 kHz; this could probably be improved by a little more tweaking of the Zobel values. Be aware that a different transformer type will require different Zobel values.

17.7 Transformer Distortion

Transformers have well-known problems with linearity at low frequencies. This is because the voltage induced into the secondary winding depends on the rate of change of the magnetic field in the core, and so the lower the frequency, the greater the change in magnitude must be for transformer action [1]. The current drawn by the primary winding to establish this field is non-linear, because of the well-known non-linearity of iron cores. If the primary had zero resistance, and was fed from a zero source impedance, as much distorted current as was needed would be drawn and no one would ever know there was a problem. But… there is always some primary resistance, and this alters the primary current drawn so that third-harmonic distortion is introduced into the magnetic field established, and so into the secondary output voltage. Very often there is a series resistance R1 deliberately inserted into the primary circuit, with the intention of avoiding HF instability; this makes the LF distortion problem worse. An important point is that this distortion does not appear only with heavy loading—it is there all the time, even with no load at all on the secondary; it is not analogous to loading the output of a solid-state power amplifier, which invariably increases the distortion. In fact, in my experience transformer LF distortion is slightly better when the secondary is connected to its rated load resistance. With no secondary load, the transformer appears as a big inductance, so as frequency falls the current drawn increases, until with circuits like Figure 17.4a, there is a sudden steep increase in distortion around 10–20 Hz as the opamp hits its output current limits. Before this happens the distortion from the transformer itself will be gross.

To demonstrate this I did some distortion tests on the same Sowter 3292 transformer. The winding resistance for both primary and secondary is about 59 Ω. It is, however, quite a small component, 34 mm in diameter and 24 mm high and weighing 45 gm, and is obviously not intended for transferring large amounts of power at low frequencies. Figure 17.6 shows the LF distortion with no series resistance, driven directly from a 5532 output, (there were no HF stability problems in this case, but it might be different with cables connected to the secondary) and with 47 and 100 Ω added in series with the primary. The flat part to the right is the noise floor.

Taking 200 Hz as an example, adding 47 Ω in series increases the THD from 0.0045% to 0.0080%, figures which are in exactly the same ratio as the total resistances in the primary circuit in the two cases. It’s very satisfying when a piece of theory slots right home like that. Predictably, a 100 Ω series resistor gives even more distortion, namely 0.013% at 200 Hz, and once more proportional to the total primary resistance.

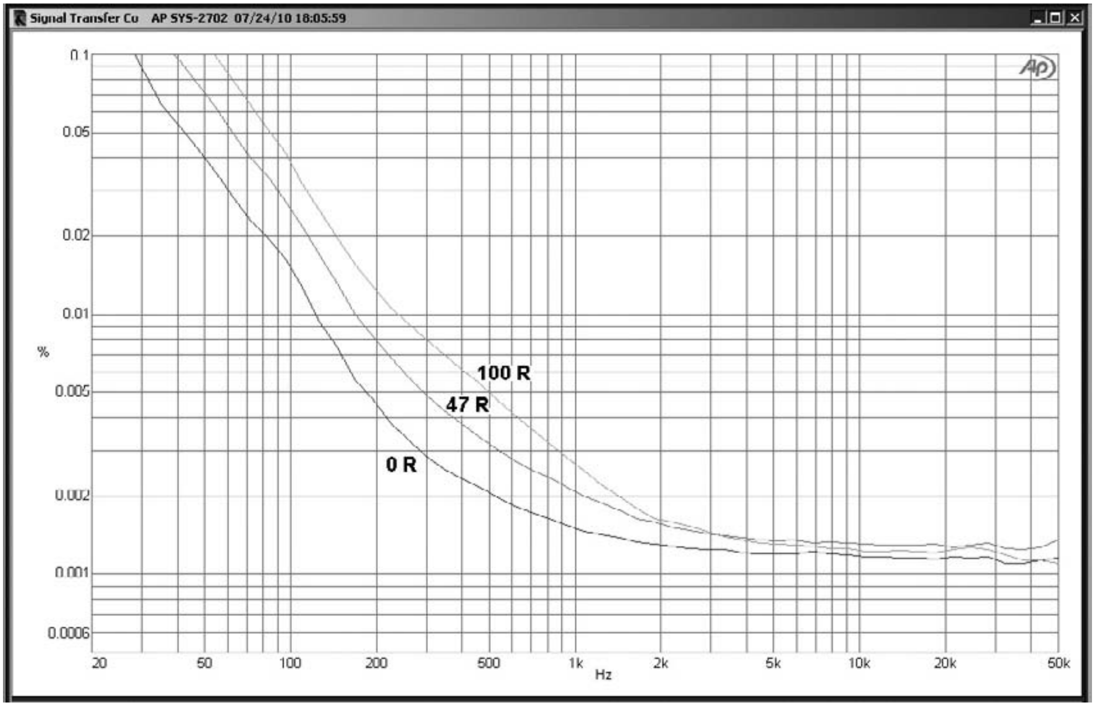


Figure 17.6: The LF distortion rise for a 3292 Sowter transformer, without (0R) and with (47 Ω and 100 Ω) extra series resistance. Signal level 1 Vrms.

If you're used to the near-zero LF distortion of opamps, you may not be too impressed with Figure 17.6, but this is the reality of output transformers. The results are well within the manufacturer's specifications for a high-quality part. Note that the distortion rises rapidly towards the LF end, roughly tripling as frequency halves. It also increases fast with level, roughly quadrupling as level doubles. Having gone to some pains to make electronics with very low distortion, this non-linearity at the very end of the signal chain is distinctly irritating.

The situation is somewhat eased in actual use as signal levels in the bottom octave of audio are normally about 10–12 dB lower than the maximum amplitudes at somewhat higher frequencies; see Chapter 14 for more on this.

17.8 Reducing Transformer Distortion

In electronics, as in so many other areas of life, there is often a choice between using brains or brawn to tackle a problem. In this case "brawn" means a bigger transformer, such as the Sowter 3991, which is still 34 mm in diameter but 37 mm high, weighing in at 80 gm. The extra mumetal core material improves the LF performance, but you still get a distortion plot very much like Figure 17.6 (with the same increase of THD with series resistance), except now it occurs at 2 Vrms instead of 1 Vrms. Twice the metal, twice the level—I suppose it makes sense. You can take this approach a good deal further with the Sowter 4231, a much bigger open-frame design tipping the scales at a hefty 350 gm. The winding resistance for the primary is 12 Ω and for the secondary 13.3 Ω, both a good deal lower than the previous figures.

Figure 17.7 shows the LF distortion for the 4231 with no series resistance, and with 47 and 100 Ω added in series with the primary. The flat part to the right is the noise floor. Comparing it with Figure 17.5 the basic distortion at 30 Hz is now 0.015% compared with about 0.10% for the 3292 transformer. While this is a useful improvement it is gained at considerable expense. Now adding 47 Ω of series resistance has dreadful results-distortion increases by about 5 times. This is because the lower winding resistances of the 4231 mean that the added 47 Ω has increased the total resistance in the primary circuit to five times what it was. Predictably, adding a 100 Ω series resistance approximately doubles the distortion again. In general, bigger transformers have thicker wire in the windings, and this in itself reduces the effect of the basic core non-linearity, quite apart

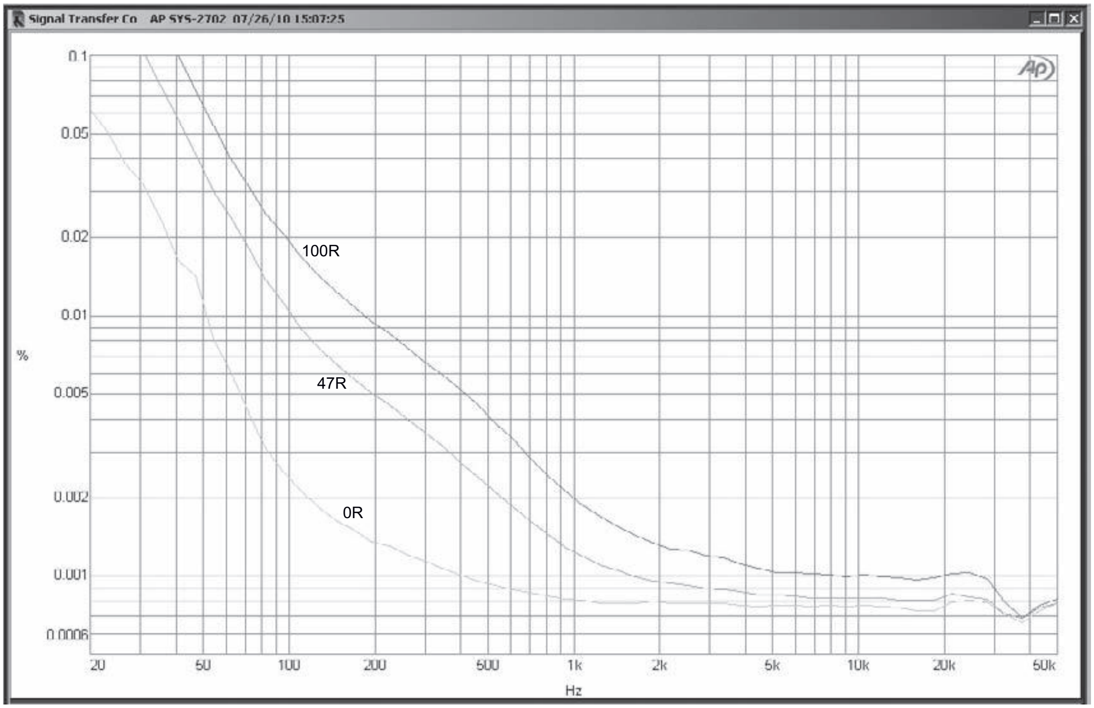


Figure 17.7: The LF distortion rise for a much larger 4231 Sowter transformer, without and with extra series resistance. Signal level 2 Vrms.

from the improvement due to more core material. A lower winding resistance also means a lower output impedance.

The LF non-linearity in Figure 17.7 is still most unsatisfactory compared with that of the electronics. Since the "My policy is copper and iron!" [2] approach does not solve the problem, we'd better put brawn to one side and try what brains we can muster.

We have seen that adding series resistance to ensure HF stability makes things definitely worse, and a better means of isolation is a low value inductor of say 4 uH paralleled with a low-value damping resistor of around 47 Ω. However inductors cost money, and a more economic solution is to use a zero-impedance output as shown in Figure 17.4b above. This gives the same results as no series resistance at all, but with dependable HF stability. However, the basic transformer distortion remains because the primary winding resistance is still there, and its level is still too high. What can be done?

The LF distortion can be reduced by applying negative feedback via a tertiary transformer winding, but this usually means an expensive custom transformer, and there may be some interesting HF stability problems because of the extra phase-shift introduced into the feedback by the tertiary winding; this approach is discussed in [3]. However, what we really want is a technique that will work with off-the-shelf transformers.

A better way is to cancel out the transformer primary resistance by putting in series an electronically generated negative resistance; the principle is shown in Figure 17.8,

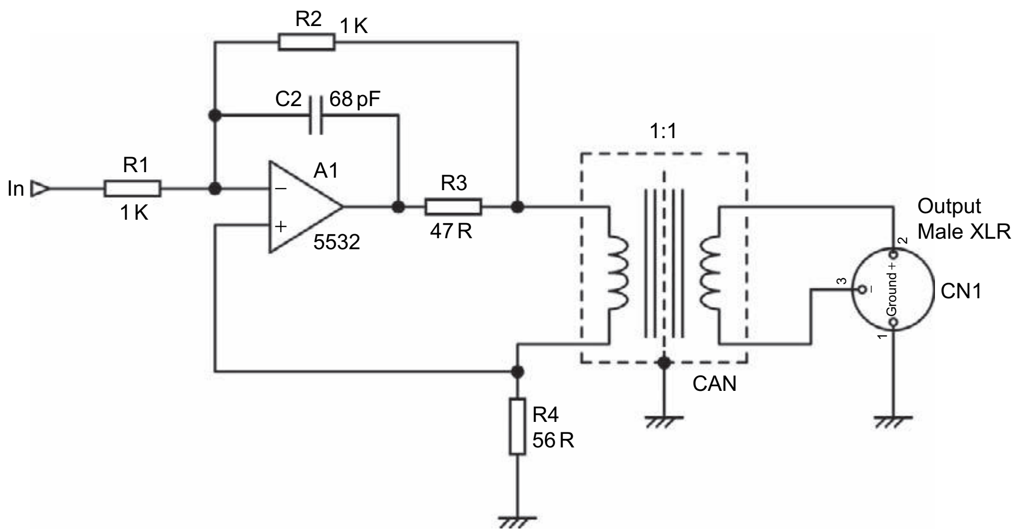


Figure 17.8: Reducing LF distortion by canceling out the primary winding resistance with a negative resistance generated by current-sensing resistance R4. Values for Sowter 3292 transformer.

where a zero-impedance output is used to eliminate the effect of the series stability resistor. The 56 Ω resistor R4 senses the current through the primary, and provides positive feedback to A1, proportioned so that a negative output resistance of twice the value of R4 is produced, which will cancel out both R4 itself and the primary winding resistance. As we saw earlier, the primary winding resistance of the 3292 transformer is approx 59 Ω, so if R4 was 59 Ω we should get complete cancellation. But…

…it is always necessary to use positive feedback with caution. Typically it works, as here, in conjunction with good old-fashioned negative feedback, but if the positive exceeds the negative (this is one time you do *not* want to accentuate the positive) then the circuit will typically latch up solid, with the output jammed up against one of the supply rails. R4 = 56 Ω in Figure 17.8 worked reliably in all my tests, but increasing R4 to 68 Ω caused immediate problems, which is precisely what you would expect. No input blocking capacitor is shown in Figure 17.8 but it can be added ahead of R1 without increasing the potential latch-up problems.

This circuit is only a basic demonstration of the principle of cancelling primary resistance, but as Figure 17.9 shows it is still highly effective. The distortion at 100 Hz is reduced by a

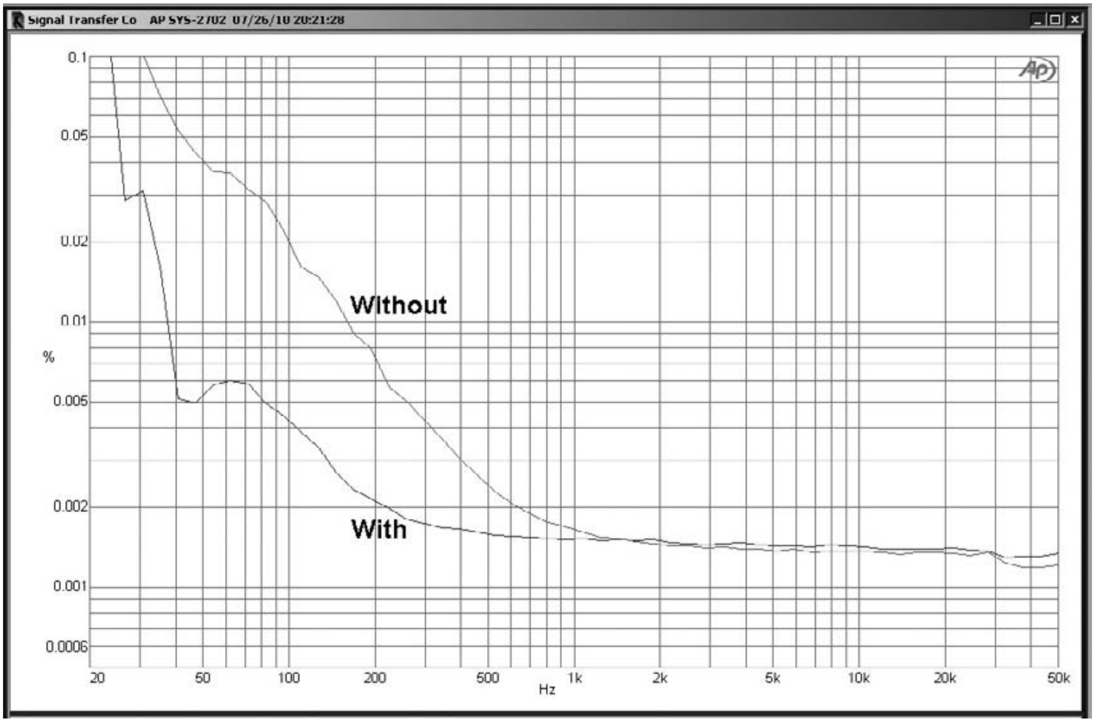


Figure 17.9: The LF distortion rise for a 3292 Sowter transformer, without and with winding resistance cancellation as in Figure 17.7. Signal level 1 Vrms.

factor of five, and at 200 Hz by a factor of four. Since this is achieved by adding one resistor, I think this counts as a triumph of brains over brawn, and indeed confirmation of the old adage that size is less important than technique.

The method is sometimes called “mixed feedback” as it can be looked at as a mixture of voltage and current feedback. The principle can also be applied when a balanced drive to the output transformer is used. Since the primary resistance is cancelled, there is a second advantage as the output impedance of the stage is reduced. The secondary winding resistance is however still in circuit, and so the output impedance is usually only halved.

If you want better performance than this—and it is possible to make transformer non-linearity effectively invisible down to 15 Vrms at 10 Hz—there are several deeper issues to consider. The definitive reference is Bruce Hofer’s patent, which covers the transformer output of the Audio Precision measurement systems [4]. There is also more information in the *Analog Devices Opamp Applications Handbook* [5].

References

[1] G.A.V. Sowter, Soft magnetic materials for audio transformers: history, production, and applications, JAES 35 (10) (1987) 769.
[2] Otto von Bismarck Speech, Speech before the Prussian Landtag’s Budget Committee, 1862. (Actually, he said blood and iron.)
[3] T. Finnern, Interfacing Electronics and Transformers, AES preprint #2194 77th AES Convention. Hamburg 1985.
[4] B. Hofer, Low-distortion transformer-coupled circuit. US Patent 4,614,914, 1986.
[5] W. Jung (Ed.), Opamp Applications Handbook, Chap. 6, Newnes, Boston, MA, 2004, pp. 484–491.

This page intentionally left blank

CHAPTER 18

Power Supply Design

"We thought, because we had power, we had wisdom." Stephen Vincent Benet: Litany for Dictatorships, 1935

18.1 Opamp Supply Rail Voltages

Running opamps at ±17 V rather than ±15 V gives an increase in headroom and dynamic range of 1.1 dB for virtually no cost and with no reliability penalty. This assumes that the opamps concerned have a maximum supply voltage rating of ±18 V, which is the case for the old Texas TL072, the new LM4562, and many other types.

The 5532 is (as usual) in a class of its own. Both the Texas and Fairchild versions of the NE5532 have an Absolute Maximum power supply voltage rating of ±22 V (though Texas also gives a "recommended supply voltage" of ±15 V), but I have never met any attempt to make use of this capability. The 5532 runs pretty warm on ±17 V when it is simply quiescent, and my view is that running it at any higher voltage is pushing the envelope. Moving from ±17 V rails to ±18 V rails only gives 0.5 dB more headroom, while stretching things to ±20 V would give a further 1.4 dB. Running at the full ±22 V would yield a more significant 2.2 dB improvement over the ±17 V case, but that is just asking for trouble. Anything above ±17 V is also going to cause difficulties if you want to run opamps with maximum supply ratings of ±18 V on the same supply rails.

We will therefore concentrate here on ±17 V supplies. They can be conveniently built with TO-220 regulators; this usually means an output current capability that does not exceed 1.5 Amps, but that is plenty for even complicated active crossovers.

An important question is: how low does the noise and ripple on the supply output rails need to be? Opamps in general have very good Power Supply Rejection Ratios (PSRR) and some manufacturer's specs are given in Table 18.1.

The PSRR performance is actually rather more complex than the bare figures given in the table imply; PSRR is typically frequency-dependent (deteriorating as frequency rises) and different for the +V and −V supply pins. It is however rarely necessary to get involved in this degree of detail. Fortunately even the cheapest IC regulators like the classic 78xx/79xx series have low enough noise and ripple outputs that opamp PSRR performance is rarely an issue.

Table 18.1: PSRR Specs for Common Opamps

Opamp Type	PSRR Minimum dB	PSRR Typical dB
5532	80	100
LM4562	110	120
TL072	70	100

Table 18.2: Typical Additional Supply Rails for Opamp Based Systems

Supply Voltage	Function
+5 V	Housekeeping microcontroller
+9 V	Relays
+24 V	LED bar-graph metering systems, discrete audio circuitry, relays

There is however another point to ponder; if you have a number of electrolytic-sized decoupling capacitors between rail and ground, enough noise and ripple can be coupled into the non-zero ground resistance to degrade the noise floor. Intelligent placing of the decouplers can help—putting them near where the ground and supply rails come onto the PCB means that ripple will go straight back to the power supply without flowing through the ground tracks on the rest of the PCB.

Apart from the opamp supply rails, audio electronics may require additional supplies, as shown in Table 18.2.

It is often convenient to power relays from a +9 V unregulated supply that also feeds the +5 V microcontroller regulator.

18.2 Designing a ±15 V Supply

Making a straightforward ±15 V 1 Amp supply for an opamp-based system is very simple, and has been ever since the LM7815/7915 IC regulators were introduced (which was a long time ago). They are robust and inexpensive parts with both overcurrent and over-temperature protection, and give low enough output noise for most purposes. We will look quickly at the basic circuit because it brings out a few design points which apply equally to more complex variations on the theme. Figure 18.1 shows the schematic, with typical component values; a centre-tapped transformer, a bridge rectifier, and two reservoir capacitors C1, C2 provide the unregulated rails that feed the IC regulators. The secondary fuses must be of the slow-blow type. The small capacitors C7-C9 across the input to the bridge reduce RF emissions from the rectifier diodes; they are shown as X-cap types not because they have to withstand 230 Vrms, but to underline the need for them to be rated to withstand continuous AC stress. The capacitors C3, C4 are to ensure HF

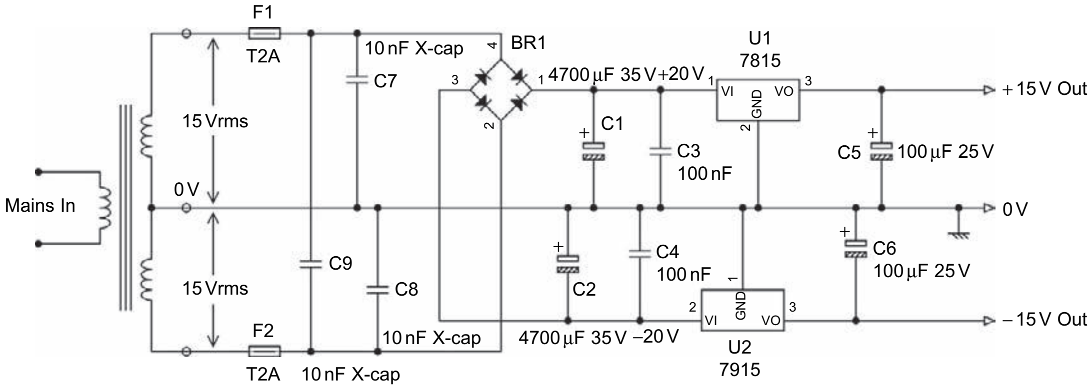


Figure 18.1: A straightforward ±15 V power supply using IC regulators.

stability of the regulators, which like a low AC impedance at their input pins, but these are only required if the reservoir capacitors are not adjacent to the regulators, ie more than 10 cm away. C5, C6 are not required for regulator stability with the 78/79 series—they are there simply to reduce the supply output impedance at high audio frequencies.

There are really only two electrical design decisions to be made; the AC voltage of the transformer secondary and the size of the reservoir capacitors. As to the first, you must make sure that the unregulated supply is high enough to prevent the rails dropping out (i.e., letting hum through) when a low mains voltage is encountered, but not so high that either the maximum input voltage of the regulator is exceeded, or it suffers excessive heat dissipation. How low a mains voltage it is prudent to cater for depends somewhat on where you think your equipment is going to be used, as some parts of the world are more subject to brown-outs than others. You must consider both the minimum voltage-drop across the regulators (typically 2 V) and the ripple amplitude on the reservoirs, as it is in the ripple troughs that the regulator will first "drop out" and let through unpleasantness at 100 Hz.

In general, the RMS value of the transformer secondary will be roughly equal to the DC output voltage.

The size of reservoir capacitor required depends on the amount of current that will be drawn from the supply. The peak-to-peak ripple amplitude is normally to be in the region of 1 to 2 Volts; more ripple than this reduces efficiency as the unregulated voltage has to be increased to allow for unduly low ripple troughs, and less ripple is usually unnecessary and gives excessive reservoir capacitor size and cost. The amount of ripple can estimated with adequate accuracy by using Equation 18.1

$$\mathrm{Vpk-pk} = \frac{I \cdot \Delta t \cdot 1000}{C} \tag{18.1}$$

where:

Vpk-pk is the peak-to-peak ripple voltage on the reservoir capacitor
I is the maximum current drawn from that supply rail in Amps
Δt is the length of the capacitor discharge time, taken as 7 milliseconds
C is the size of the reservoir capacitor in microFarads
The "1000" factor simply gets the decimal point in the right place

Note that the discharge time is strictly a rough estimate, and assumes that the reservoir is being charged via the bridge for 3 msec, and then discharged by the load for 7 msec. Rough estimate it may be, but I have always found it works very well.

The regulators must be given adequate heatsinking. The maximum voltage drop across each regulator (assuming 10% high mains) is multiplied by the maximum output current to get the regulator dissipation in Watts, and a heat sink selected with a suitable thermal resistance to ambient (in °C per Watt) to ensure that the regulator package temperature does not exceed, say, 90 °C. Remember to include the temperature drop across the thermal washer between regulator and heatsink.

Under some circumstances it is wise to add protective diodes to the regulator circuitry, as shown in Figure 18.2. The diodes D1, D3 across the regulators are reverse-biased in normal operation, but if the power supply is driving a load with a large amount of rail decoupling capacitance, it is possible for the output to remain higher in voltage than the regulator input as the reservoir voltage decays. D1, D3 prevent this effect from putting a reverse voltage across the regulators.

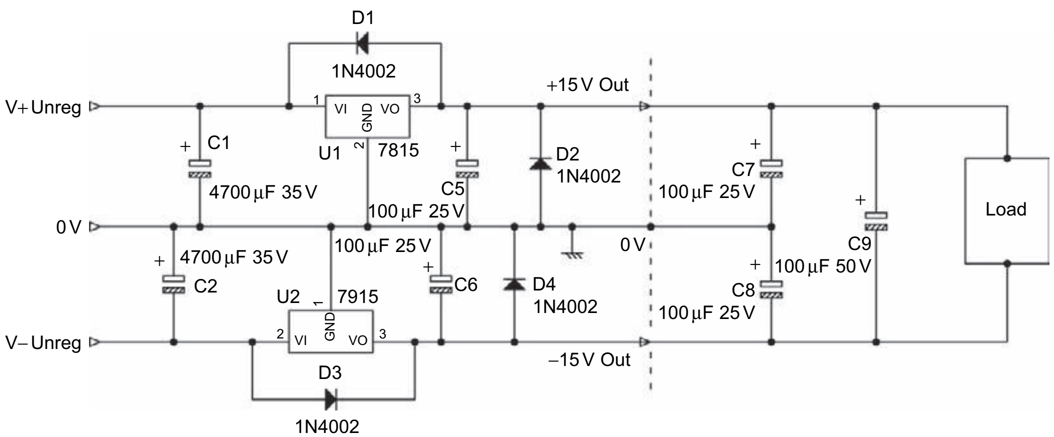


Figure 18.2: Adding protection diodes to a ±15 V power supply. The load has decoupling capacitors to both ground (C7, C8) and between the rails (C9); the latter can cause start-up problems.

The shunt protection diodes D2, D4 are once again reverse-biased in normal operation. D2 prevents the +15 V supply rail from being dragged below 0 V if the −15 V rail starts up slightly faster, and likewise D4 protects the −15 V regulator from having its output pulled above 0 V. This can be an important issue if rail-to-rail decoupling such as C9 is in use; such decoupling can be useful because it establishes a low AC impedance across the supply rails without coupling supply rail noise into the ground, as C7, C8 are prone to do. However, it also makes a low-impedance connection between the two regulators. D2, D4 will prevent damage in this case, but leave the power supply vulnerable to start-up problems; if its output is being pulled down by the −15 V regulator, the +15 V regulator may refuse to start. This is actually a very dangerous situation, because it is quite easy to come up with a circuit where start-up will only fail one time in twenty or more, the incidence being apparently completely random, but presumably controlled by the exact point in the AC mains cycle where the supply is switched on, and other variables such as temperature, the residual charge left on the reservoir capacitors, and the phase of the moon. If even one start-up failure event is overlooked or dismissed as unimportant, then there is likely to be serious grief further down the line. *Every power supply start-up failure must be taken seriously*.

18.3 Designing a ±17 V Supply

There are 15 V IC regulators, (7815, 7915) and there are 18 V IC regulators, (7818, 7918) but there are no 17 V IC regulators. This problem can be effectively solved by using 15 V regulators and adding 2 Volts to their output by manipulating the voltage at the REF pin. The simplest way to do this is with a pair of resistors that divide down the regulated output voltage and apply it to the REF pin as shown in Figure 18.3a. (The transformer and AC input components have been omitted in this and the following diagrams, except where they differ from those shown above.) Since the regulator maintains 15 V between the OUT and REF pin, with suitable resistor values the actual output with respect to 0 V is 17 V.

The snag with this arrangement is that the quiescent current that flows out of the REF pin to ground is not well controlled; it can vary between 5 and 8 mA, depending on both the input voltage and the device temperature. This means that R1 and R2 have to be fairly low in value so that this variable current does not cause excessive variation of the output voltage, and therefore power is wasted.

If a transistor is added to the circuit as in Figure 18.3b, then the impedance seen by the REF pin is much lower. This means that the values of R1 and R2 can be increased by an order of magnitude, reducing the waste of regulator output current and reducing the heat liberated. This sort of manoeuvre is also very useful if you find that you have a hundred thousand 15 V regulators in store, but what you actually need for the next project is an 18 V regulator, of which you have none.

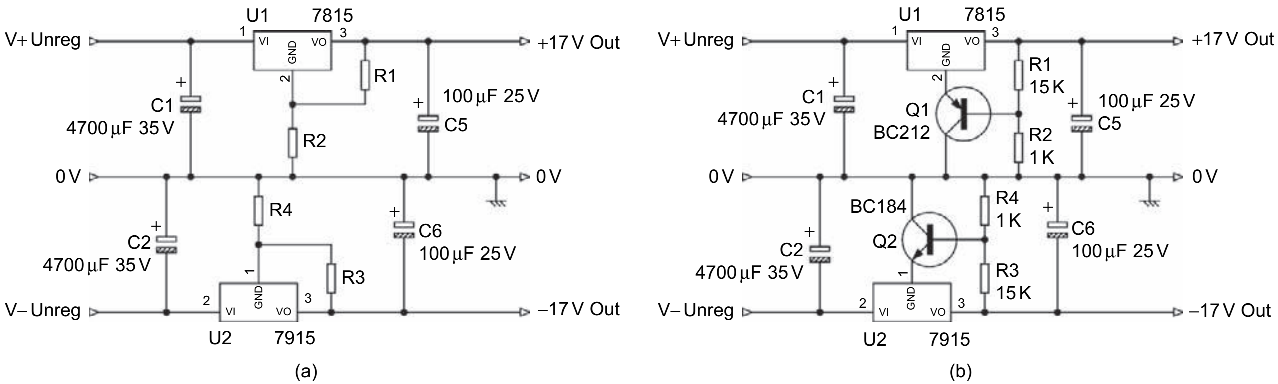


Figure 18.3: Making a ±17 V power supply with 15 V IC regulators. (a) Using resistors is inefficient and inaccurate; (b) adding transistors to the voltage-determining resistor network makes the output voltage more predictable and reduces the power consumed in the resistors.

What about the output ripple with this approach? I have just measured a power supply using the exact circuit of Figure 18.3b, with 2200 uF reservoirs, and I found −79 dBu (87 uV rms) on the +17 V output rail, and −74 dBu (155 uV rms) on the 17 V rail, which is satisfyingly low for inexpensive regulators, and should be adequate for almost all purposes; note that these figures include regulator noise as well as ripple. The load current was 110 mA. If you *are* plagued by ripple troubles, the usual reason is a rail decoupling capacitor that is belying its name by coupling rail ripple into a sensitive part of the ground system, and the cure is to correct the grounding rather than design an expensive ultra-low ripple PSU. Note that doubling the reservoir capacitance to 4400 uF only improved the figures to −80 dBu and −76 dBu respectively; just increasing reservoir size is not a cost-effective way to reduce the output ripple.

18.4 Using Variable-Voltage Regulators

It is of course also possible to make a ±17 V supply by using variable output voltage IC regulators such as the LM317/337. These maintain a small voltage (usually 1.2 V) between the OUTPUT and ADJ (shown in figures as GND) pins, and are used with a resistor divider to set the output voltage. The quiescent current flowing out of the ADJ pin is a couple of orders of magnitude lower than for the 78/79 series, at around 55 uA, and so a simple resistor divider gives adequate accuracy of the output voltage, and transistors are no longer needed to absorb the quiescent current. A disadvantage is that this more sophisticated kind of regulator is somewhat more expensive than the 78/79 series; at the time of writing they cost something like 50% more. The 78/79 series with transistor voltage-setting remains the most cost-effective way to make a non-standard-voltage power supply at the time of writing.

It is clear from Figure 18.4 that the 1.2 V reference voltage between ADJ and out is amplified by many times in the process of making a 17 V or 18 V supply; this not only increases output ripple, but also output noise as the noise from the internal reference is being amplified. The noise and ripple can be considerably reduced by putting a capacitor C7 between the ADJ pin and ground. This makes a dramatic difference; in a test PSU with a 650 mA load the output noise and ripple was reduced from −63 dBu (worse than the 78xx series) to −86 dBu (better than the 78xx series), and so such a capacitor is usually fitted as standard. If it is fitted, it is then essential to add a protective diode D1 to discharge C7, C8 safely if the output is short-circuited, as shown in Figure 18.5.

The ripple performance of the aforementioned test PSU, with a 6800 uF reservoir capacitor and a 650 mA load, is summarised for both types of regulator in Table 18.3. Note that the exact ripple figures are subject to some variation between regulator specimens.

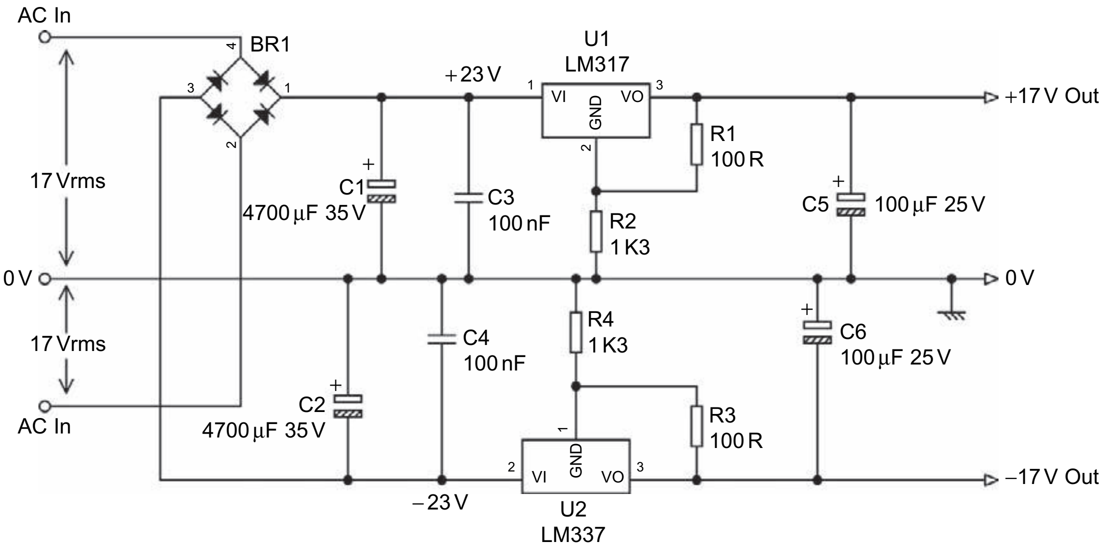


Figure 18.4: Making a ±17V power supply with variable-voltage IC regulators.

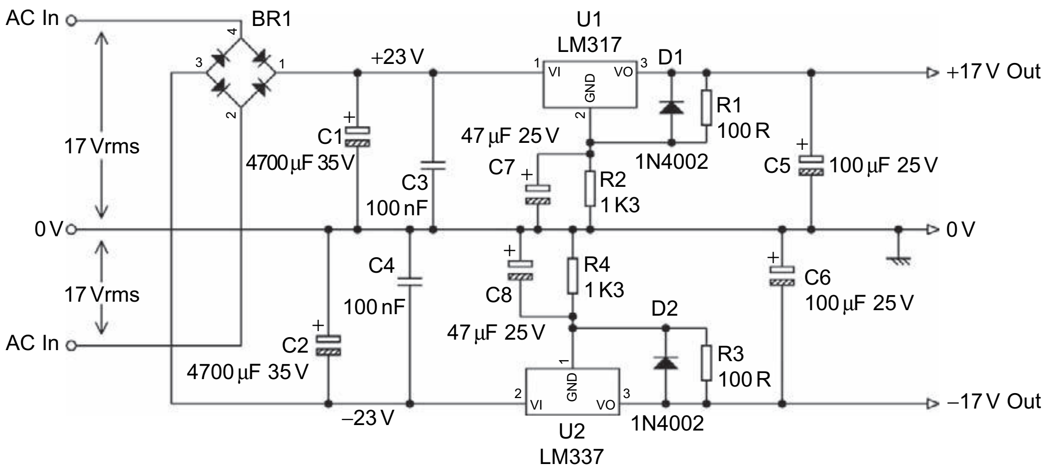


Figure 18.5: Adding ripple-improvement capacitors and protective diodes to a variable-voltage IC regulator.

Table 18.3: Comparing the Noise and Ripple Output of Various Regulator Options

	7815 + Transistor	LM317
No C on LM317 ADJ pin	−73 dBu (all ripple)	−63 dBu (ripple & noise)
47 uF on LM317 ADJ pin	−73 dBu (all ripple)	−86 dBu (ripple & noise)
Input filter 2.2 Ω & 2200 uF	−78 dBu (ripple & noise)	−89 dBu (mostly noise)
Input filter 2.2 Ω & 4400 uF	−79 dBu (mostly noise)	−90 dBu (all noise)

18.5 Improving Ripple Performance

Table 18.3 shows that the best noise and ripple performance that can be expected from a simple LM317 regulator circuit is about −86 dBu (39 uV rms) and this still contains a substantial ripple component. The reservoir capacitors are already quite large at 4700 uF, so what is to be done if lower ripple levels are needed? The options are:

1. Look for a higher-performance IC regulator. They will cost more and there are likely to be issues with single sourcing.
2. Design your own high-performance regulator using discrete transistors or opamps. This is not a straightforward business, especially if all the protection that IC regulators have is to be included. There can be distressing issues with HF stability.
3. Add an RC input filter between the reservoir capacitor and the regulator. This is simple and pretty much bullet-proof, and preserves all the protection features of the IC regulator, though the extra components are a bit bulky and not that cheap. There is some loss of efficiency due to the voltage drop across the series resistor; this has to be kept low in value so the capacitance is correspondingly large.

The lower two rows of Table 18.3 show what happens. In the first case the filter values were 2.2 Ω and 2200 uF. This has a −3 dB frequency of 33 Hz and attenuates the 100 Hz ripple component by 10 dB. This has a fairly dramatic effect on the visible output ripple, but the dB figures do not change that much as the input filter does not affect the noise generated inside the regulator. Increasing the capacitance to 4400 uF sinks the ripple below the noise level for both types of regulator.

18.6 Dual Supplies from a Single Winding

It is very convenient to use third-party “wall-wart” power supplies for small pieces of equipment, as they come with all the safety and EMC approvals already done for you, though admittedly they do not look appropriate with high-end equipment.

The vast majority of these supplies give a single AC voltage on a two-pole connector, so a little thought is required to derive two supply rails. Figure 18.6 shows how it is done in a ±18 V power supply; note that these voltages are suitable only for a system that uses 5532s throughout. Two voltage-doublers of opposite polarity are used to generate the two unregulated voltages. When the incoming voltage goes negative, D3 conducts and the positive end of C1 takes up approximately 0 V. When the incoming voltage swings positive, D1 conducts instead and the charge on C1 is transferred to C3. Thus the whole peak-to-peak voltage of the AC supply appears across reservoir capacitor C3. In the same way, the peak-to-peak voltage, but with the opposite polarity, appears across reservoir C4.

Since voltage-doublers use half-wave rectification, they are not suitable for high current supplies. When choosing the value of the reservoir capacitor values, bear in mind that the

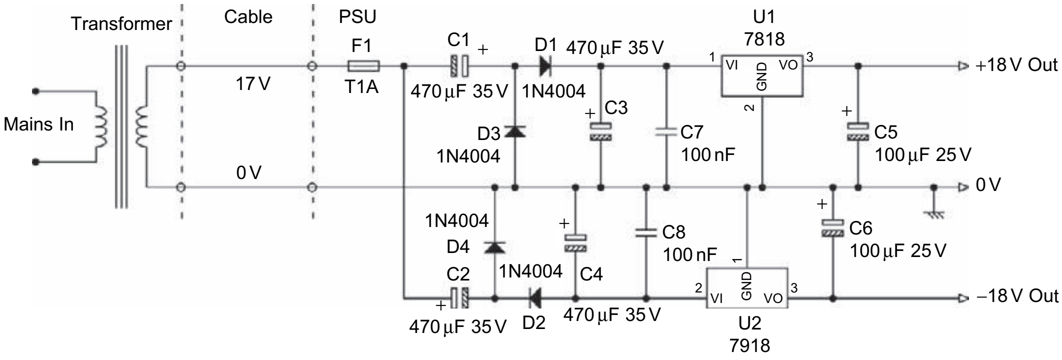


Figure 18.6: A ±18 V power supply powered by a single transformer winding.

discharge time in Equation 18.1 above must be changed from 7 msec to 17 msec. The input capacitors C1, C2 should be the same size as the reservoirs.

18.7 Mutual Shutdown Circuitry

The 5532 opamp is in general a tough item but it has an awkward quirk. If one supply rail is lost and collapses to 0 V, while the other rail remains at the normal voltage, a 5532 can under some circumstances get into an anomalous mode of operation that draws large supply currents and it ultimately destroys itself by overheating. Other opamps may suffer the same problem but I am not currently aware of any. To prevent damage from this cause opamp supplies should be fitted with a mutual shutdown system. This ensures that if one supply rail collapses, because of overcurrent, over-temperature or any other cause, the other rail will be promptly switched off. A simple way to implement this is shown in Figure 18.7, which also demonstrates how to make a ± supply using only positive regulators. The use of a second positive regulator to produce the negative output rail looks a little strange at first sight but I can promise you it works. The regulators in this case are the high-current Linear Tech LT1083, which can handle 7.5 Amps, but is not available in a negative version. That should be more than enough current for even the most complex active crossover…

The extra circuitry to implement mutual-shutdown is very simple; R5, D3, R6, and Q1 and Q2. Because R5 is equal to R6, D3 normally sits at around 0 V in normal operation. If the +17 V rail collapses, Q2 is turned on by R6, and the REF pin of U2 is pulled down to the bottom rail, reducing the output to the reference voltage (1.25 V). This is not completely off, but it is low enough to prevent any damage to opamps.

If the −17 V rail collapses, Q1 is turned on by R5, pulling down the REF pin of U1 in the same way. Q1 and Q2 do not operate exactly symmetrically, but it is close enough for our purposes.

Note that this circuit can only be used with variable output voltage regulators, because it relies on their low reference voltages to achieve what is effectively switch-off.

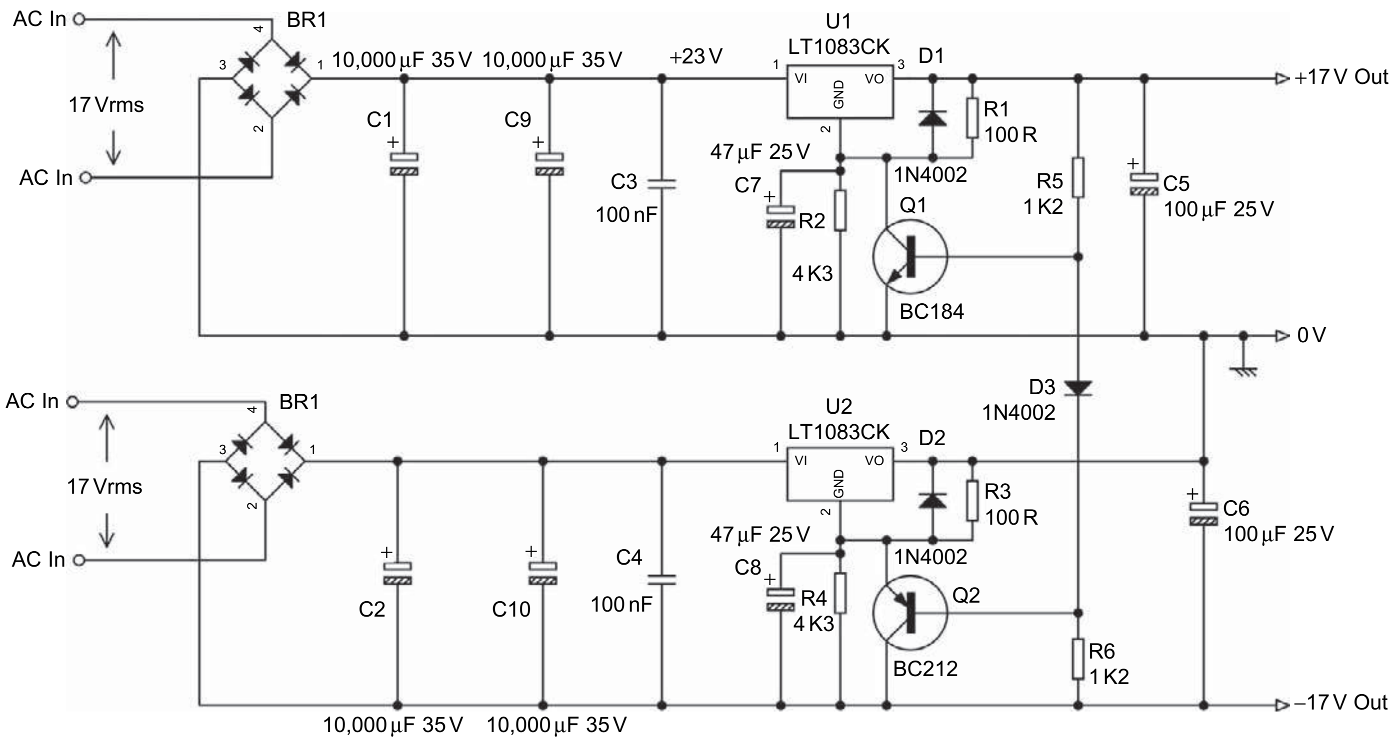


Figure 18.7: A high-current ±17 V power supply with mutual rail shutdown, and using only positive regulators.

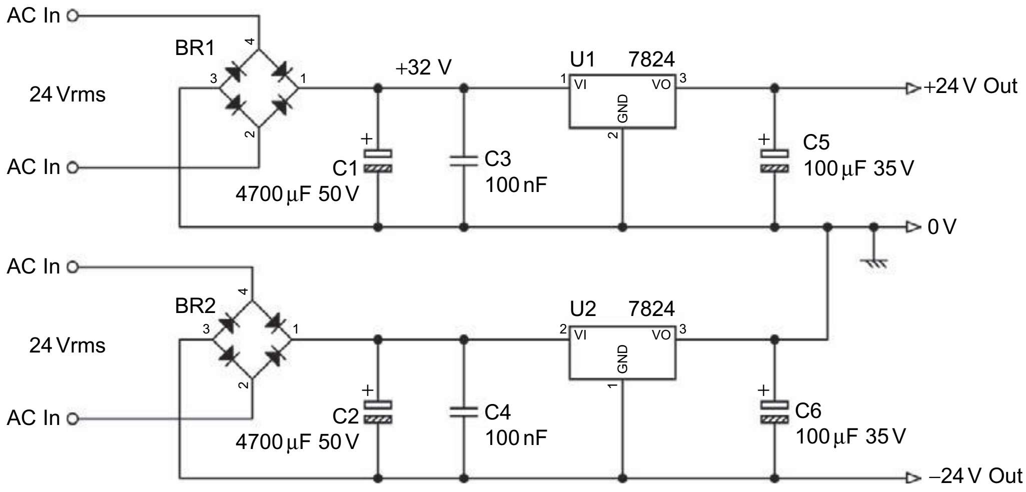


Figure 18.8: A ±24 V power supply using only positive regulators.

18.8 Power Supplies for Discrete Circuitry

One of the main reasons for using discrete audio electronics is the possibility of handling larger signals than can be coped with by opamps running off ±17 V rails. The use of ±24 V rails allows a 3 dB increase in headroom, which is probably about the minimum that justifies the extra complications of discrete circuitry. A ±24 V supply can be easily implemented with 7824/7924 IC regulators.

A slightly different approach was used in my first published preamplifier design [1]. This preamp in fact used two LM7824 +24 V positive regulators as shown in Figure 18.8 because at the time the LM7924 −24 V regulator had not yet reached the market. This configuration can be very useful in the sort of situation where you have a hundred thousand positive regulators in store, but no negative regulators.

Note that this configuration does however require two separate transformer secondary windings; it cannot be used with a single centre-tapped secondary.

Reference

[1] D.R.G. Self, An Advanced Preamplifier Design, Wireless World, London, 1976.

CHAPTER 19

An Active Crossover Design

Here I present an active crossover design that will illustrate a good number of the principles and techniques put forward in this book. A design that demonstrated all of them would be a cumbersome beast, so I have aimed instead to show the most important of them while also providing a practical and adaptable design which will be of use in the real world.

The design is a generic crossover that can be adapted to a wide range of applications by changing its parameters and its configuration. The crossover frequencies can be changed simply by altering component values, and circuit blocks for equalisation or time delay can be added or removed, with due care to preserve absolute phase, of course. Since widely varying drive units may be used I have made no attempt to add equalisation or integrate the driver response into the filter operation.

The crossover design is primarily aimed at hi-fi applications rather than sound reinforcement. It does not, for example, in its basic form have variable crossover frequencies or balanced outputs, though instructions are given for adding the latter.

19.1 Design Principles

The aim of this chapter is not just to provide a complete design, but to demonstrate the use of various design principles expounded in the body of this book.

- Low impedance design for low noise, Chapter 14
- Elevated internal levels, Chapter 12
- Further elevated internal level for the HF path, Chapter 14
- Low-noise balanced inputs, Chapter 16
- Optimised filter order for best noise, Chapter 14
- Mixed capacitor types in filters, Chapter 8
- Use of multiple resistors to improve accuracy, Chapter 12
- Delay compensation using third-order allpass filters, Chapter 10

19.2 Example Crossover Specification

This is the basic specification for an active crossover that we will use as an example. (As noted in Chapter 10, the path lengths to be compensated for were carefully chosen to give nice round figures for the delay times required.)

No of bands:	Three
Type	Linkwitz-Riley fourth order
Mid/HF crossover frequency:	3 kHz
LF/Mid crossover frequency:	400 Hz
HF path time delay compensation	80 usec, tolerance +/−5% (path length 27 mm)
Mid path time delay compensation	400 usec, tolerance +/−5% (path length 137 mm)
	Gain: 0dBu in for 1 Vrms out

The crossover has balanced inputs, but in its basic form the outputs are unbalanced, the assumption being that if a balanced interconnection is required, this will be provided at the power amplifier inputs. The possibility of adding balanced output stages is examined at the end of the chapter. If the overall audio system can be satisfactorily designed with balanced crossover inputs but unbalanced crossover outputs, as opposed to unbalanced crossover inputs but balanced crossover outputs, then the former is clearly more economical as for a stereo crossover it requires two balanced input stages rather than six balanced output stages.

Low-impedance design is used throughout; the circuit impedances are designed to be as low as possible without causing extra distortion by overloading opamp outputs. This reduces the effect of opamp current noise and resistor Johnson noise but has no effect on opamp voltage noise.

The design assumes that there is no level control between the crossover and power amplifier, or if there is, it is set to maximum and left there; this allows us to use elevated internal levels in the crossover to improve the noise performance, as described in Chapter 14.

No equalisation stages are included in the signal paths.

19.3 The Gain Structure

In Chapter 14, I explained how with suitable system design the internal levels of an active crossover could be significantly raised to reduce the effect of circuit noise. Here I have decided to go for an internal level of 3 Vrms, 12 dB higher than the assumed input voltage of 775 mV rms (0 dBu.) This retains 10 dB of headroom to accommodate maladjusted input levels. We also saw in Chapter 14 that the distribution of amplitude with frequency in music is such that the levels in the top few octaves are much lower than in the rest of the audio spectrum. We decided conservatively that with an upper crossover point around 3 kHz, it was safe to raise the HF channel level by a further 6 dB.

It is therefore necessary to have an input amplifier with a gain of close to four times (+11.9 dB) to raise the internal level to a nominal 3 Vrms as soon as possible, and another +6 dB of amplification will be required as early as possible in the HF path.

The bandwidth definition filter is shown in Figure 19.1 as having unity gain for simplicity at this point. In fact, as we shall shortly see, it has a small loss of 0.3 dB, which has to be recovered elsewhere in the circuitry to keep the levels spot-on.

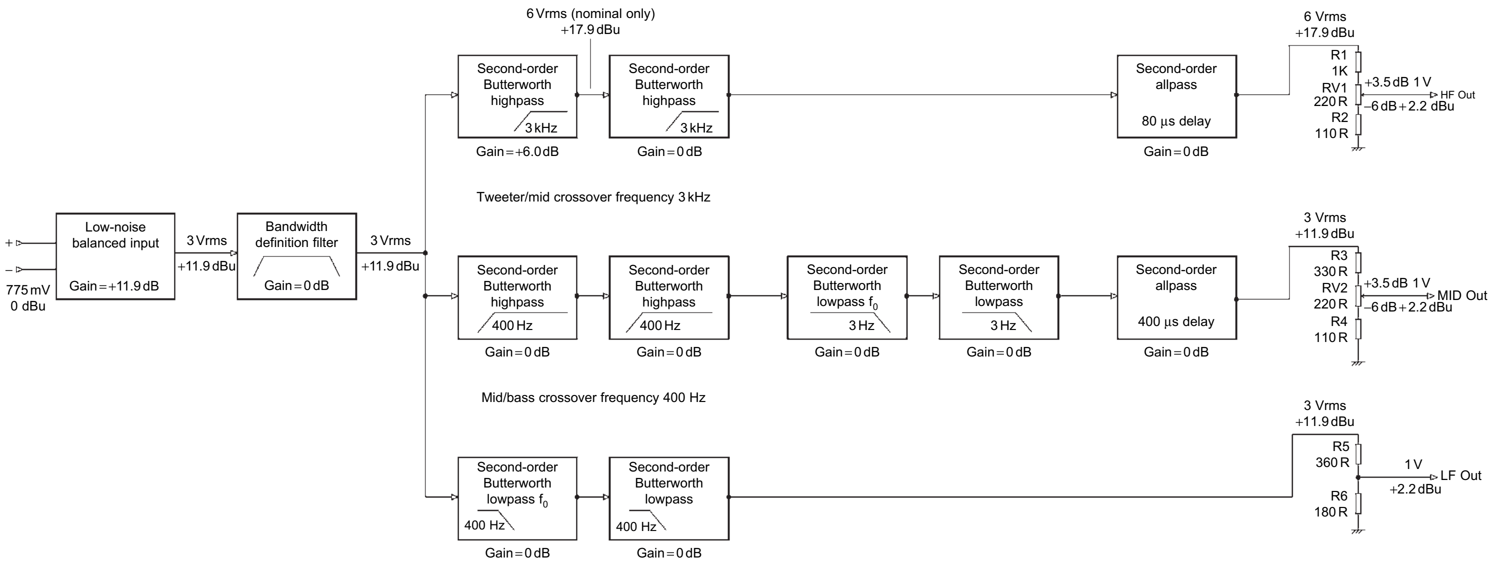


Figure 19.1: Block diagram of three-way Linkwitz–Riley fourth-order subtractive crossover.

19.4 Resistor Selection

The basic policy for component selection is that followed throughout this book—to use as few capacitor values as possible, which means sticking to the sparse E3 or E6 series, and let the resistor values come out as they may. I decided that no more than two resistors from the E24 series would be used in series or parallel to get as close as possible to the desired value. Where it could be done the resistors would be of near-equal value to get the best possible improvement in accuracy by using multiple resistors, as explained in Chapter 12. An error window for the nominal value (excluding tolerances) is set at ±0.5%. 1% tolerance resistors are assumed.

Resistors in parallel rather than in series are generally to be preferred as it makes the PCB layout easier; assuming they are lying side by side you just have to join the adjacent pads with very short tracks. It could also be argued that the parallel connection makes for better reliability, as a dry solder-joint that fails completely, or a resistor that goes open-circuit, is less likely to cause oscillation or stop the signal altogether. Film resistors do not normally fail short-circuit; open-circuit is at least 10 times more likely. In some sound-reinforcement circumstances any sound is better than none at all.

19.5 Capacitor Selection

I decided to use polypropylene capacitors wherever they would have a measurable effect on the distortion performance. They get bulky and expensive quickly as their value increases, and so I further concluded that 220 nF should be the largest value employed. It is assumed that only the E3 series (10, 22, 47) values are available. The number of different capacitor values (not counting small ceramics) has been kept down to four—2n2, 10 nF, 47 nF, and 220 nF. It makes sense to put some effort into this as they are probably the most expensive electronic components in the crossover, and so you want to gain as much cost advantage from purchasing in quantity as possible. No particular assumptions are made about capacitor tolerance.

19.6 The Balanced Line Input Stage

It was explained in Chapter 16 that the conventional unity-gain balanced input stage made with four 10 kΩ resistors is relatively noisy, especially when compared with an unbalanced input using a simple voltage-follower; the balanced stage noise output is about –104.8 dBu. This problem can be addressed by reducing the value of the 10 kΩ resistors around the balanced stage, and driving the balanced stage with 5532 unity-gain buffers to keep the input impedance up. In the unity-gain case this reduces the noise output to –110.2 dBu, a very useful 5.4 dB quieter.

For our crossover design we do not want a unity-gain stage, but one that gives a gain of four times, (+11.9 dB) so that a nominal input of 775 mV rms is raised to 3 Vrms. We use the same strategy of input buffers and a balanced amplifier working at low impedances, but the required gain is obtained by altering the ratio of the resistors R7, R8 to R9, R10 in Figure 19.2. The output noise from this stage with the inputs terminated by 50 Ω resistors

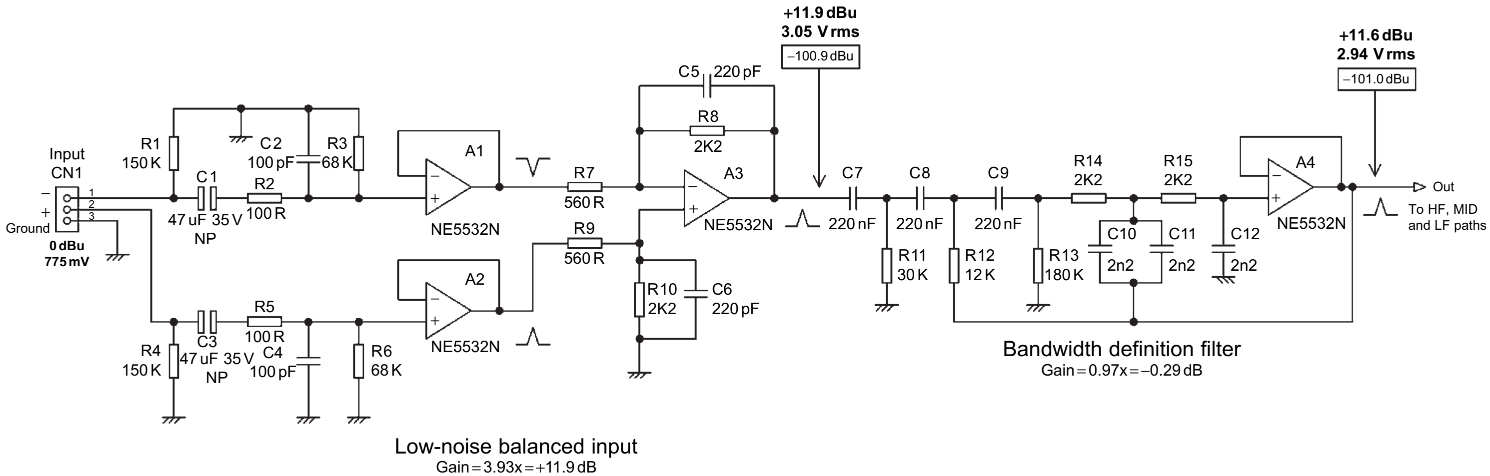


Figure 19.2: Schematic diagram of the input circuitry: the balanced input stage and bandwidth-definition filter.

is –100.9 dBu, so its Equivalent Input Noise (EIN) is –100.9 – 11.9 = –112.8 dBu. For what it's worth, this is the same as the Johnson noise from a 9.6 kΩ resistor. The usual EMC filter and DC-blocking networks are placed at the input.

This is a rare case where a significant amount of gain is required in a balanced input amplifier, so the instrumentation amplifier configuration is an interesting alternative, giving a CMRR improved by 11.9 dB, the gain of the stage. I made some tests just before going to press, and this really does work.

19.7 The Bandwidth Definition Filter

The bandwidth definition stage consists of a third-order Butterworth high-pass filter with a cutoff frequency of 20 Hz to remove subsonic signals and protect loudspeakers, and a second-order Butterworth low-pass filter with a cutoff frequency of 50 kHz. Unlike the example in Chapter 8, E24 resistor values have been used in the subsonic section to get the cutoff frequency exactly at 20 Hz. The physically large 220 nF capacitors in the filter are susceptible to electrostatic hum pickup and this must be considered in the physical layout of the crossover.

The bandwidth definition filters are placed as early as possible in the signal path, immediately after the balanced input amplifier, to prevent headroom being eaten up by large subsonic signals. This should also minimise the generation of intermodulation distortion by large ultrasonic signals.

The combined filter has a small midband loss of –0.29 dB when designed with 220 nF capacitors. Redesigning it to use 470 nF capacitors (keeping the capacitors in the lowpass section the same) would reduce this to an even more negligible loss of –0.15 dB, but it's hard to argue that the result is worth the significant extra cost of the capacitors. The only real problem with the 0.29 dB loss is that when tracking a test signal through the HF path, you will encounter 2.90 Vrms at the combined filter output, instead of a nice round 3.0 Vrms. Rather than propagate this inelegance through the rest of the HF path, the signal is restored to 3.0 Vrms in the first HF filter. This attenuate-then-amplify process inevitably incurs a noise penalty, but in this case it is very small indeed. If an extra stage was used to recover the loss this would be inelegant and uneconomical, but since the next stage in the HF path is already configured to give gain, there is no cost penalty at all.

19.8 The HF Path: 3 kHz Linkwitz–Riley Highpass Filter

The HF signal path in Figure 19.3 includes two second-order Butterworth filters that make a fourth-order Linkwitz–Riley filter. Both are of the Sallen & Key type. The first filter is configured for a gain of +6.3 dB, to raise the nominal level and make the signal less vulnerable to circuit noise, and also to recover that 0.3 dB loss in the combined filter. The nominal level in the block diagram of Figure 19.1 is shown as "6 Vrms" which if was a

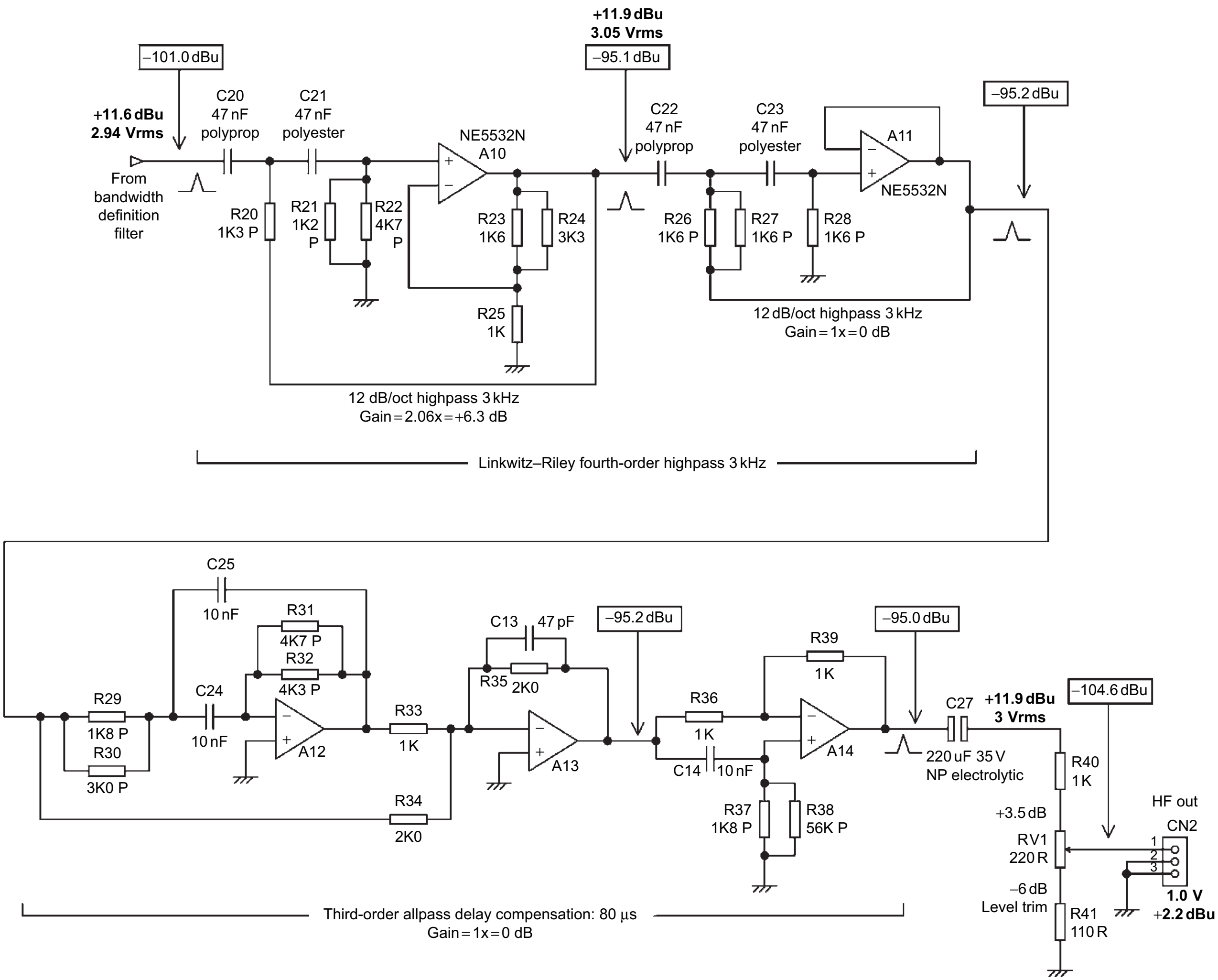


Figure 19.3: Schematic diagram of the HF path, with Linkwitz–Riley highpass filter and third-order time delay compensator. Noise levels shown at each stage.

real level would be too high, but the relatively low amount of HF energy in musical signals gives a level of more like 3 Vrms in practical use. A level of 6 Vrms will however be obtained if you apply a test signal at 775 mV rms to the input.

The design of this stage is straightforward—see Chapter 8 for more information on designing Sallen & Key filters with whatever gain you want. The value of R20 comes very conveniently as the E24 preferred value of 1K3, but the value of the second resistance in the filter comes out as 970.7 Ω, well away from either 910 Ω or 1 kΩ. R2 is therefore made up of a parallel pair R21 and R22. The optimal method for selecting these was described in Chapter 12. If the two resistors are approximately equal in value, the accuracy of their combined resistance is improved by a factor of $\sqrt{2}$ as the errors tend to cancel. Thus a 1% tolerance becomes a 0.707% tolerance. As the values become more unequal the combined tolerance increases until it is effectively 1% as the tolerance is decided by only one resistor. The procedure is therefore to start with near-equal values and pick a resistor pair to give a result just above the target value, and then pick a resistor pair to give a result just below. If neither result is within the desired error window, one of the resistors is incremented and we try again. As this algorithm proceeds the resistor values become more and more unequal, and the improvement in tolerance diminishes, so the sooner we find a satisfactory answer the better. In some respects you have to balance the improvement in tolerance against the accuracy of the combined resistance value obtained. The process is mechanical and tedious, but once set up on a spreadsheet it is fairly quick to do. I have written a Javascript program that does it instantly.

In this case our target value is 970.7 Ω and the error window is ±0.5%. Our progress is set out in Table 19.1.

Step 1: We start with 1800 Ω (R21) in parallel with 2000 Ω, (R22) the combined resistance of which is 2.4% low. The next value for R22 above 2000 Ω is 2200 Ω, which

Table 19.1: Selecting the Best Parallel Combination of R21 and R22 to Get the Target Value of 970.7 Ω

Step	R21 Ohms	R22 Ohms	Combined Ohms	Error %
1	1800	2000	947.368	−2.40%
	1800	2200	990.000	1.99%
2	1600	2400	960.000	−1.10%
	1600	2700	1004.651	3.50%
3	1500	2700	964.286	−0.66%
	1500	3000	1000.000	3.02%
4	1300	3600	955.102	−1.61%
	1300	3900	975.000	0.44%
5	1200	4700	955.932	−1.52%
	1200	5100	971.429	0.08%

with 1800 Ω gives a combined resistance 1.99% high. Our target value is therefore irritatingly about halfway between these two resistor combinations. In each step we find the two resistor values that 'bracket' the target, in other words one value gives a combined resistance that is too low, and the other a combined resistance that is too high,
Step 2: We reduce the 1800 Ω resistor to 1600 Ω and try again. Putting 2400 Ω in parallel gives a combined resistance that is only 1.1% low, our best shot so far, but nowhere near good enough. Putting 2700 Ω in parallel gives a result 3.5% high. Are we downhearted? Yes, but a faint heart never built a fair crossover so we will persist.
Step 3: We reduce the 1600 Ω resistor to 1500 Ω, and find that 2700 Ω in parallel is only 0.66% low, which is temptingly close to 0.5%, but there is no point in setting error windows if you're not going to stick to them. We proceed, but with the nagging awareness that each step is reducing the accuracy improvement we will get from using multiple resistors.
Step 4: We reduce the 1500 Ω resistor to 1300 Ω, and with 3600 Ω in parallel the result is 1.61% low. However, with 3900 Ω in parallel, bingo! The combined value is only 0.44% high, inside our ±0.5% error window, but not, it must be said, a very long way inside it. Our work here is done. Or is it? There is always the nagging doubt that you might get a much better result if you go just a step or two further. In this case that doubt is very much justified.
Step 5: We reduce the 1300 Ω resistor to 1200 Ω, and with 4700 Ω in parallel the result is 1.52% low. But with 5100 Ω in parallel—cracked it! The result is only 0.08% high, and more than good enough, considering the tolerances of the components. The accuracy improvement is much impaired, as one resistor is more than four times the other, (see Table 12.7) but the 1300 Ω and 3900 Ω resistor combination in Step 4 is not much better in that respect.

We therefore select R21 as 1200 Ω and R22 as 5100 Ω. I realise that this process sounds a bit long-winded when every step is described, but a spreadsheet version is reasonably quick. (I should warn you at this point that the Goalseek function in Excel apparently can't cope with the mathematical expression for the value of parallel resistors, and tends to unhelpfully offer "solutions" up in the PetaOhm regions.) The custom Javascript solution is of course the best.

The next step is to check by calculation or simulation (not measurement because the tolerances of the real components used will confuse things) that the filter cutoff frequency falls within the required error window.

In the Case of the Second Filter (Quick, Watson, the game's afoot!), we are much luckier. Using the process just described, we find that taking the value of the first resistor as 800 Ω gives an error of –0.23% in the filter cutoff frequency which is less than the resistor tolerance (even after the full √2 accuracy improvement, which we get in this case) and probably *much* less than the capacitor tolerances. Thus, R26 and R27 are 1K6, and R28 is also 1K6. It happens to work out very nicely, though in terms of component count we only save one resistor.

In each filter, only the first capacitor (C20, C22) is a polypropylene type, while the second (C21, C23) is polyester. Only one capacitor needs to be polypropylene to avoid capacitor distortion, but it must be the first one in the filter. This intriguing state of things is described more fully in Chapter 8.

19.9 The HF Path: Time Delay Compensation

When we looked at the question of time delay compensation in Chapter 10, we noted that mercifully it is not necessary to maintain an absolutely accurate delay over the whole audio spectrum. Instead the delay only has to be constant over each crossover region. The HF delay only needs to be maintained around the 3 kHz crossover point for so long as both drive units are radiating significantly, and likewise the MID delay needs only to be constant around the 400 Hz crossover. One of the many advantages of the Linkwitz–Riley crossover configuration is that the slopes are steep at 24 dB/octave, and so these regions of overlap are relatively narrow, simplifying the problem.

When allpass filters are used as delay elements they give a constant group delay at low frequencies but it begins to fall off at high frequencies. This makes the design of the HF delay more complicated than that of the MID delay. A first-order allpass filter designed for a delay of 80 usec unfortunately starts to roll-off quite early; as frequency rises the delay is down by 10% at 2.4 kHz, before we even reach the 3 kHz crossover point, and sinks to 50% at 9.3 kHz, slowly approaching zero above 100 kHz. A first-order allpass filter clearly won't do the job.

One solution is to cascade three first-order filters in series. The delay is now spread out over three sections, with each one set to a 80 usec/3 = 26.7 usec delay. The total 80 usec delay is now sustained up to three times the frequency, being 10% down at 7.5 kHz, and not down 50% until 28 kHz, well outside the audio spectrum. So, is this good enough? The MID and HF drivers will be contributing equally at 3 kHz, both of them being 6 dB down. 7.5 kHz is only 1.3 octaves away from the 3 kHz MID-HF crossover frequency, but the high slope of the Linkwitz-Riley crossover means that the signal to the MID drive unit will be 32 dB down, though its acoustical contribution is less certain as it depends on the drive unit frequency response outside its intended band.

Another solution would be a second-order allpass filter; the delay of an 80 usec version falls by 10% at 4.79 kHz, better than the first-order filter, (10% down at 2.4 kHz) but worse than the triple first-order filter (10% down at 7.5 kHz). The second-order filter as described in Chapter 10 has the disadvantage of phase-inverting in the low-frequency range where its delay is constant. It is also 3.2 dB noisier than the third-order solution we are about to look at. All in all, a second-order allpass filter does not look promising.

The last delay filter examined in detail in Chapter 10 is the third-order allpass filter, which is made up of a second-order allpass cascaded with a first order allpass. When designed for an 80 usec delay it is 10% down at 12.7 kHz, giving almost twice the flat delay frequency

Table 19.2: The Best Parallel Combinations for Non-Preferred Resistance Values in the HF Allpass Filter

Resistor	Target Ohms	Ra Ohms	Rb Ohms	Combined Ohms	Error %
R29 & R30	1126	1800	3000	1125.0	−0.09%
R31 & R32	2251	4700	4300	2245.6	−0.24%
R37 & R38	1746	1800	56000	1743.9	−0.12%

range of the three cascaded first-order filters, at a fractionally lower cost, as it actually uses one less resistor. It uses the same number of potentially expensive capacitors. The –10% point for the delay is now 2.1 octaves above our 3 kHz crossover frequency, and the signal sent to the MID driver will be down by 50 dB, so whatever the response of the driver itself we can be pretty sure that the fall-off in time delay will have no audible consequences. When designed for 80 usec it is 3.2 dB quieter than the equivalent second-order filter, and it does not inconveniently phase-invert in the low-frequency range. The third-order filter is clearly the better solution, and so it is chosen for the HF path delay.

The 80 usec third-order allpass filter is studied in detail in Chapter 10, with full consideration of its quite subtle noise and distortion characteristics, so I will not repeat that here. Suffice it to say that the circuit of Figure 10.22 is cut and pasted into our schematic of Figure 19.3. In the MFB filter, both capacitors are polypropylene, though the mixed capacitor phenomenon applies to MFB filters as well as Sallen & Key filters, as described in Chapter 8.

The parallel resistor combinations for the awkward resistance values in the HF allpass filter are given in Table 19.2. We will get a good improvement in precision with R32 and R32, as the values are near-equal, not much for R29 and R30, and virtually none for R37 and R38, where our luck was *right* out.

The schematic of the complete HF path is shown in Figure 19.3. Note that the component numbering starts at 20 for resistors and capacitors, and 10 for opamps; this is to allow additions to other parts of the schematic without renumbering everything. The signal and noise levels are given for the output of each stage.

The allpass stage can of course be omitted if time delay compensation is done by other means, such as the physical cabinet construction.

19.10 The MID Path: Topology

The MID path contains a 400 Hz fourth-order Linkwitz–Riley highpass filter and a 3 kHz fourth-order Linkwitz–Riley lowpass filter. In this situation the order of the filters in the signal path needs to be considered. In Chapter 14 I demonstrated that a noise advantage can be gained by putting the lowpass filter second in the path, as it removes some of the noise from the previous highpass filter. With the filter frequencies we are using here the noise advantage is 1.1 dB; not an enormous amount, but achieved at no cost whatsoever. Let's do it.

You will recall that in the HF path the first filter was configured to give +6.3 dB of gain, recovering the loss in the bandwidth definition filter. While it would be possible to configure the first MID filter to give +0.3 dB of gain (we do not want the +6.0 dB in this path) it does not seem worth the extra complication as +0.3 dB represents only a very small change in the position of the output level control.

19.11 The MID Path: 400 Hz Linkwitz–Riley Highpass Filter

The MID path begins with two 400 Hz second-order Butterworth highpass filters that make a fourth-order Linkwitz–Riley highpass filter. Both are of the unity-gain Sallen & Key type, since there is no need for gain, and in fact they are identical.

We might have been lucky with the resistor values of the second filter in the HF path, but things pan out differently here. The target value for the exact 400 Hz cutoff frequency is 1278.9 Ω. That is temptingly close to 1300 Ω, but shoving that value in alone gives a cutoff frequency error of –1.6%; clearly we could simply add quite a high value in parallel to tweak the resistance just a little, but there will be effectively no improvement in accuracy. Perhaps there are two more nearly equal values that will give the combined value we want? Regrettably not. Playing the combinations, we find that the first pair of values that puts the cutoff frequency within the ±0.5%. error window is R50 = 1300 Ω in parallel with R51 = 82 kΩ, so that's what we have to use. The error in the cutoff frequency is only –0.07%.

The second resistance in the filter has a target value of 1278.9 Ω times two, which is 2557.7 Ω. Our first attempt at a parallel combination is R52 = 5100 Ω in parallel with R53 = 5100 Ω, which gives 2550 Ω, which is only 0.30% low. Looks like there's no need to need to look any further. And yet there is the nagging doubt that there might be a better solution … and there is. The very next step is 5600 Ω in parallel with 4700 Ω, giving 2555.3 Ω, which is only 0.09% low, and gives us almost all the accuracy improvement possible. In this case there is not much point in looking further still. The same resistor values are used in both filters.

As for the filters in the HF path, only the first capacitors (C50, C52) needs to be polypropylene for low distortion; the second capacitors (C51, C53) can be polyester.

19.12 The MID Path: 3 kHz Linkwitz–Riley Lowpass Filter

After the 400 Hz highpass filter come the two second-order Butterworth lowpass filters that make up a fourth-order Linkwitz–Riley lowpass filter. Both are of the unity-gain Sallen & Key type. Using 47 nF capacitors, the target value for the two resistors is 798.15 Ω. We can see at once that two 1600 Ω resistors in parallel is going to be very close, and the combined resistance of 800 Ω is in fact only 0.23% low. We also get the maximum possible improvement in accuracy, so we look no further.

19.13 The MID Path: Time Delay Compensation

The time delay required in the MID path is five times longer at 400 usec, but on the other hand it covers the 400 Hz crossover point so the delay does not need to be constant up to such a high frequency as in the HF path. Let's see if that makes thing any easier.

Since a first-order allpass filter designed for a delay of 80 usec has the delay down by 10% at 2.4 kHz, we would expect the same filter designed for a 400 usec delay to be down by 10% at 480 Hz, and simulation confirms that this is so; the relationship is just simple proportion. 480 Hz is much too close to the 400 Hz crossover point so once more we are going to have to look at a more sophisticated solution. We will aim for 10% down at two octaves above crossover, in other words at 1.6 kHz, because this gave pretty convincing results in the HF path.

Looking at the options, we can say that:

Three first-order filters in series will be 10% down at 7.5 kHz/5 = 1.5 kHz. This is 1.9 octaves away from crossover and is very close to our target, but this configuration uses a relatively large number of components. Four cascaded first-order filters would certainly do the job, but the component count is looking excessive compared with the other options, because four capacitors and four opamp sections are required, while the third-order allpass filter only uses three of each.

A second-order allpass filter designed for 400 usec will be down 10% at 4.79 kHz/5 = 958 Hz. This is not even close to our two-octave target, and as we have seen the second-order allpass filter has performance issues compared with the third-order, and an unwanted phase-inversion. We can rule this approach out.

A third-order allpass filter designed for 400 usec, will be 10% down at 12.7 kHz/5 = 2.54 kHz, and simulation conforms this figure is correct. It is 2.7 octaves away from crossover and gives us a healthy safety margin—in fact it looks almost a bit too healthy, as all the evidence is that a two octave spacing is ample. However, it is the only alternative that meets our requirements, and there is no obvious way to save a few parts by cutting the spacing down a bit. It is therefore once more selected for our design.

The 400 usec third-order allpass filter is very similar to that of the 80 usec version in the HF path, the main difference being that the capacitors have increased from 10 nF to 47 nF, and the resistor values adjusted accordingly to get the desired delay time. The schematic of the complete MID path is shown in Figure 19.4. Note that the component numbering starts at

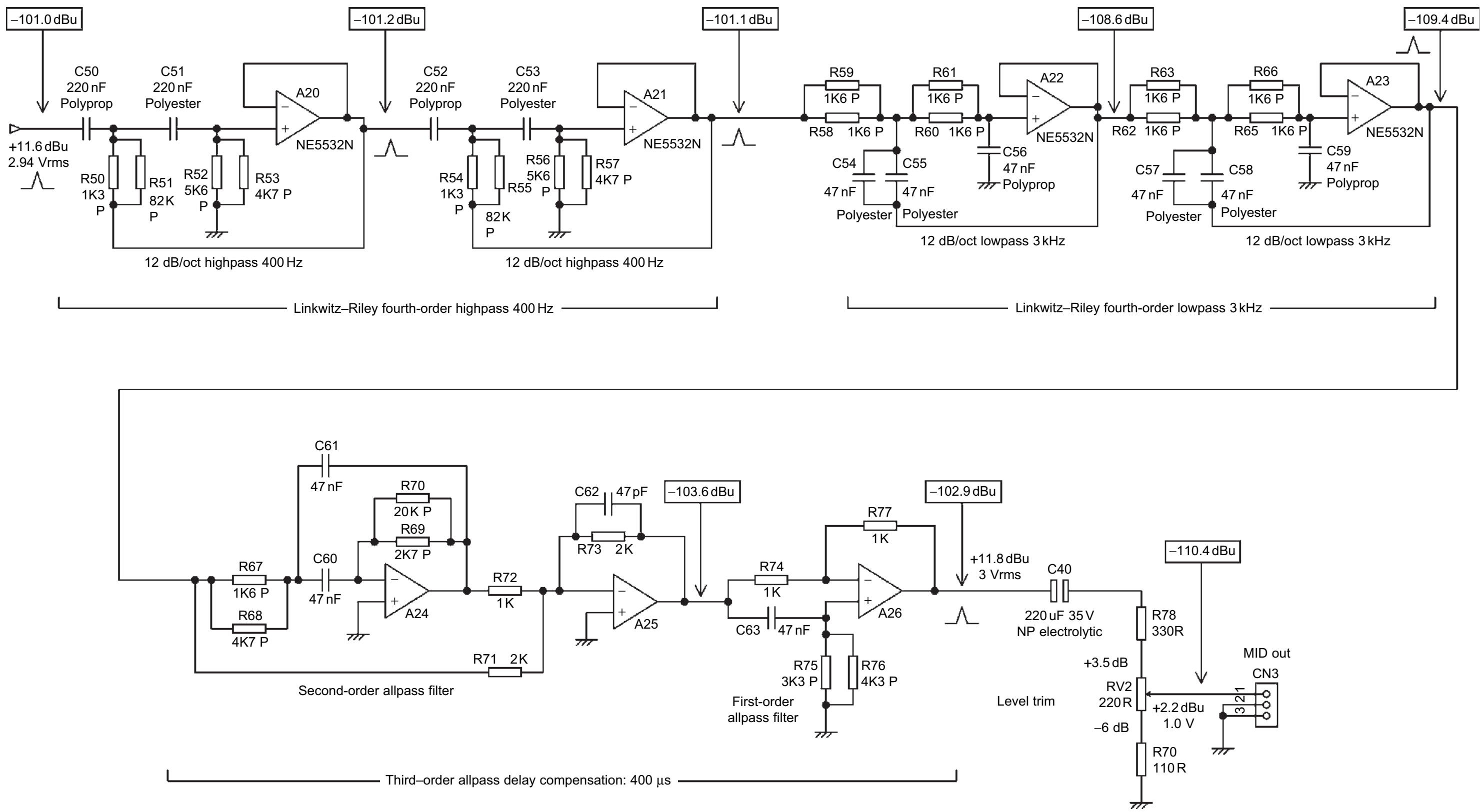


Figure 19.4: Schematic diagram of the MID path, with Linkwitz–Riley highpass and lowpass filters, followed by the third-order time delay compensator. Noise levels shown at each stage.

Table 19.3: The Best Parallel Combinations for Non-Preferred Resistance Values in the HF Allpass Filter

Resistor	Target Ohms	Ra Ohms	Rb Ohms	Combined Ohms	Error %
R67, R68	1189	1600	4700	1193.7	0.39%
R69,R70	2378	2700	20000	2378.9	0.04%
R75,R76	1868	3300	4300	1867.1	−0.05%

50 for resistors and capacitors, and 20 for opamps; this permits additions to other parts of the schematic without global renumbering. The signal and noise levels are given for the output of each stage.

The parallel resistor combinations for the three non-preferred resistance values in the MID allpass filter are given in Table 19.3. We will get a reasonable improvement in precision with R67 and R68, as the values are near-equal, not much for R75 and R76, and virtually none for R69 and R70.

The allpass stage can be omitted if time delay compensation is implemented by alternative means such as the physical construction of the enclosure.

19.14 The LF Path: 400 Hz Linkwitz–Riley Lowpass Filter

The LF path is the simplest of the three paths in the crossover. It consists of two 400 Hz second-order Butterworth lowpass filters that make a fourth-order Linkwitz–Riley lowpass filter; both are of the unity-gain Sallen & Key type. There is no delay compensation. The target resistance value for the two resistors is 1278.9 Ω, which not surprisingly is the same value as the first resistor in the MID highpass filter, as both use a capacitance value of 220 nF. As before, the first pair of values that puts the cutoff frequency within the ±0.5% error window is 1300 Ω in parallel with 82 kΩ. The error in the cutoff frequency is only –0.07%.

19.15 The LF Path: No Time Delay Compensation

No time delay compensation is required in the LF path because the physical distance between the acoustic centres of the LF and MID drive units is compensated for by the delay in the MID path. In fact, if the delay compensation was in the LF path, it would have to have a negative delay; in other words the output would emerge before the input arrived. Such circuits, though they would be extremely useful for predicting lottery numbers if you wired enough of them in series, are notoriously hard to design.

The schematic of the complete LF path is shown in Figure 19.5. The component numbering starts at 80 for resistors and capacitors, and 40 for opamps to allow additions to other parts

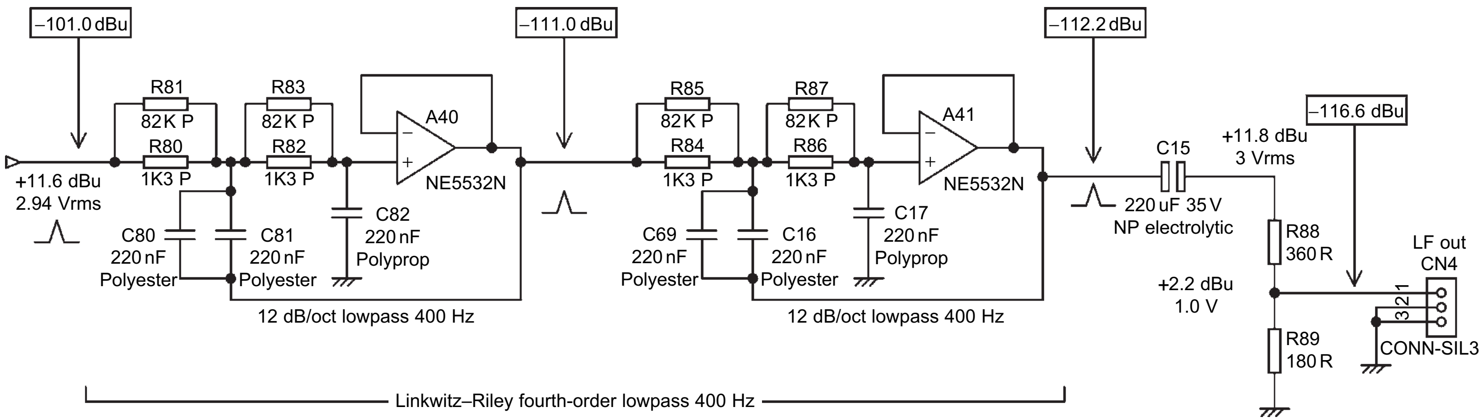


Figure 19.5: Schematic diagram of the LF path, which consists simply of a 400 Hz Linkwitz–Riley lowpass filter and a fixed output attenuator.

of the schematic without renumbering. The signal and noise levels are given for the output of each stage.

19.16 Output Attenuators and Level Trim Controls

The output level controls used here have a deliberately limited range because they are not intended to be used as volume controls. The range about the nominal 1 Vrms output is +3.5 dB to –6.0 dB, and this should be more than enough to allow for power amplifier gain tolerances (assuming a nominally identical set of power amplifiers) or drive unit sensitivity variations. If, however, if the crossover is intended to work with a wide range of power amplifiers of varying sensitivities the range may need to be extended by reducing the end-stop resistors at the top and bottom of the preset control.

The values chosen give the nominal output with the preset wiper central in its track, so the system can be lined up pretty well just by eye; this won't of course work with multi-turn preset pots. The exactness of this unfortunately depends on the track resistance of the preset in relation to the fixed end-stop resistors above and below it, and its tolerance is unlikely to be better than 10%; it may be 20%. These tolerances are horribly wide compared with the 1% of fixed resistors. If really precise levels are required the output will have to be measured during the trimming operation.

The output networks are configured with the lowest possible resistances that will not load the opamp upstream excessively, to keep the final output impedance low enough to drive a reasonable amount of cable without HF losses. The assumption in this design is that the power amplifiers will not be too far away and their inputs will be driven directly. Putting a unity-gain buffer after the level trim network would allow a lower output impedance, especially if the "zero-impedance" type of buffer is used, but this is likely to compromise the noise performance.

An alternative approach would be to reduce the resistor values in the output networks so the maximal output impedance is as low as desired, and then drive this with a suitable number of 5532 unity-gain buffers connected in parallel via 10 Ω current-sharing resistors. The effect on the noise performance will in this case be negligible, especially since the noise contributions of the added buffers will partially cancel as they are uncorrelated. You might need to keep an eye on the power dissipation in the output network resistors.

An unbalanced output will give better common-mode rejection if it is configured as impedance-balanced, by using a three-pin output connector with the cold (–) pin connected to ground through an impedance that approximates as closely as possible to the output

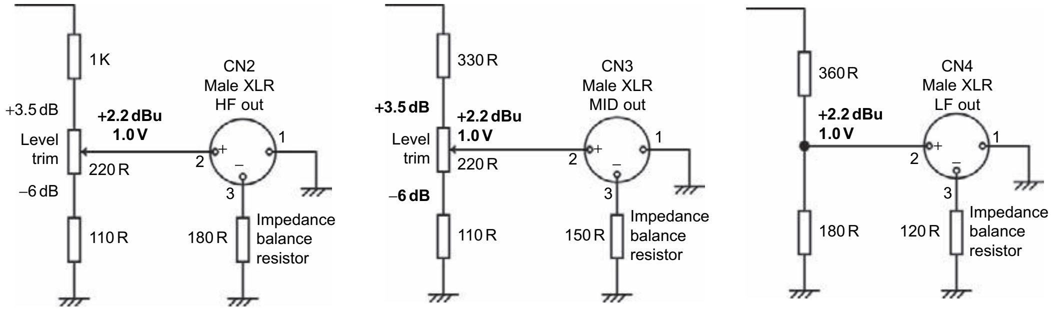


Figure 19.6: Reconfiguring the output networks for impedance-balanced operation with power amplifiers fitted with balanced inputs.

impedance driving the hot (+) output pin. This improves the balance of the interconnection, as fully described in Chapter 16. In our circumstances here the output impedance of the level-trim network varies as the wiper of the preset is moved from one end of its travel to the other, and so the impedance from the cold pin to ground is inevitably a compromise.

In the case of the HF path, the output impedance varies from 248 Ω with the trimmer at maximum and 101 Ω with the trimmer at minimum. The middle setting, which gives the nominal output level, gives an output impedance of 184 Ω. An impedance-balance resistor of 180 Ω will give negligible common-mode error.

The MID path output impedance varies from 165 Ω at maximum to 92 Ω at minimum, with a middle value of 147 Ω, so an impedance-balance resistor of 150 Ω will give a very small error.

The LF path has no output level trim and its output impedance is fixed at 120 Ω, so here the impedance-balancing can be made exact.

The impedance-balanced outputs for the three paths are shown in Figure 19.6.

19.17 Balanced Outputs

Adding balanced output stages is not wholly straightforward. If placed after the output attenuator and trim networks, as in Figure 19.7a, they are likely to compromise the noise performance, while if placed *before* the output networks, as in Figure 19.7b, a dual-gang trim preset is required; good luck with sourcing that. A solution to this dilemma is to separate the gain trim and output attenuation functions as shown in Figure 19.7c.

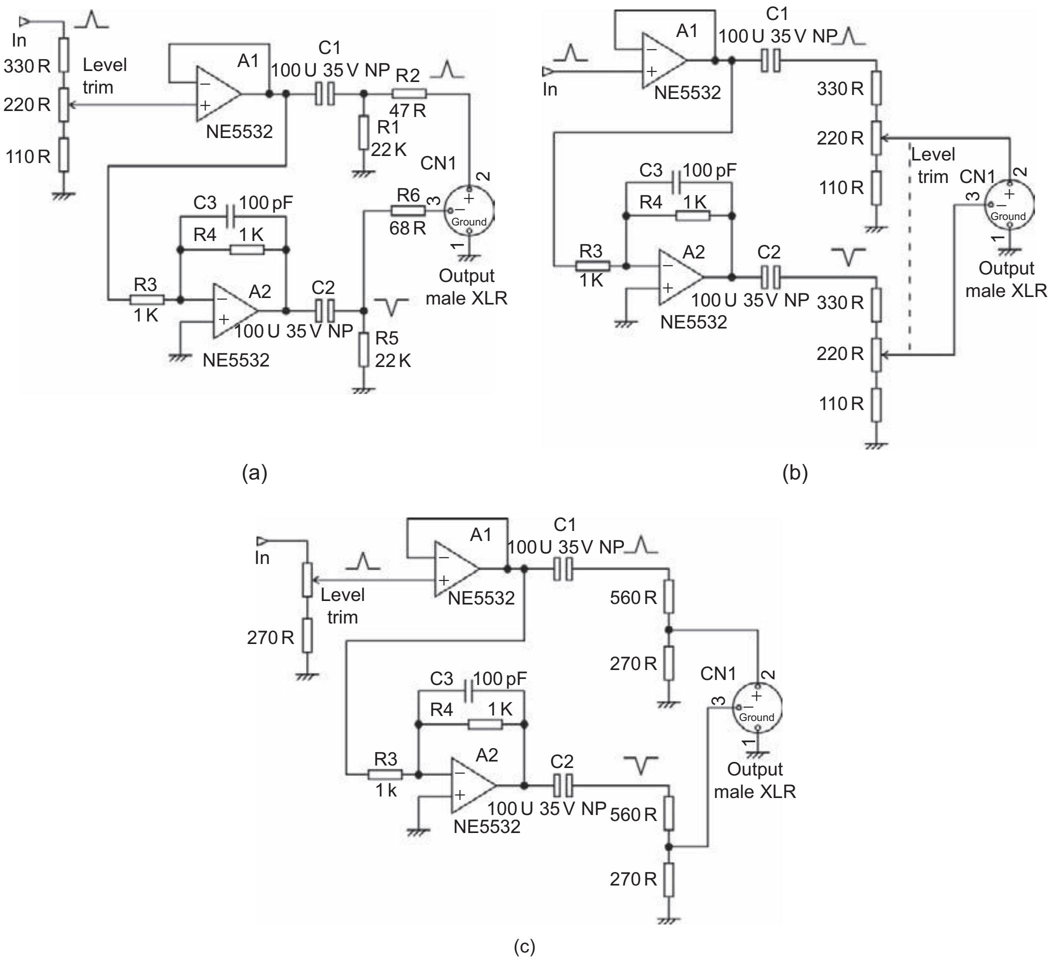


Figure 19.7: Options for balanced outputs for crossovers with elevated internal levels and output attenuators. a) balanced output stage after gain trim; b) balanced output stage before dual gain trim; c) gain trim and attenuator functions separated.

If the prospective power amplifiers have unbalanced inputs, consideration should be given to the possibility of using ground-cancelling outputs, as described in Chapter 17.

19.18 Crossover Programming

Many of the resistors in the schematics if this chapter have the letter "P" next to their resistance value. These are the components that need to be altered to change the characteristics of the crossover. The idea is that these resistor positions on a prototype PCB can be fitted with single-way turned-pin sockets like those that ICs are plugged into. It is then possible to very easily plug resistors in and out during development. This is much quicker than pulling the PCB out of the case, desoldering one resistor and then re-soldering another one in, and PCBs will only take so much of this before the pads are damaged. When the design is finalised the PCBs will have resistors soldered into the same positions in the usual way.

Single-pin sockets can be obtained by cutting up a strip of sockets. Make sure you use the kind of socket strip intended for this sort of thing—deconstructing IC sockets does not usually work well as the plastic is more brittle and tends to shatter.

19.19 Noise Analysis: Input Circuitry

The schematics of the input circuitry, the HF path, the MID path, and the LF path all have the measured noise levels at the output of each stage indicated by rectangular boxes with arrows. This information is essential for performing a noise analysis, in which the noise contribution of each stage is assessed to see if it is what we expect, and how it relates to the noise generated by the other stages in the path we are examining. Because of the way that uncorrelated noise sources add in a RMS fashion, the largest noise source tends to dominate the final result—the noise at the end of the path—and so we want to identify that source and see if it is worthwhile to make it quieter and so improve the overall noise performance.

As throughout this book, all the noise measurements given are unweighted and in a 22 kHz bandwidth.

The measured noise output from the balanced input stage with its +11.9 dB of gain is –100.9 dBu. We terminate both input pins with 50 Ω to ground, and the Johnson noise from each termination resistor is –135.2 dBu (22 kHz bandwidth at 25°C). The noise from both together is 3 dB more, not 6 dB, because the two noise sources are uncorrelated. The noise voltage at the input is thus a very very low –132.2 dBu. When amplified by +11.9 dBu this becomes –120.3 dBu, which is almost 20 dB below the stage output noise and is therefore making a negligible contribution to the total. Despite its special low-noise design, virtually all of its output noise is generated by the balanced input amplifier itself.

The Equivalent Input Noise (EIN) of the balanced input stage is −100.9 dBu – 11.9 dB = −112.8 dBu, giving us a Noise Figure of 22.4 dB. This would usually be considered pretty poor, but as is discussed in Chapter 16 on line inputs, Noise Figures are not a very useful figure of merit for balanced inputs. What is unquestionably true is that –100.9 dBu is a very low value of noise to find in audio circuitry, so we seem to have made a good start.

The noise level at the output of the bandwidth definition filter might be expected to be a bit higher, because of its noise contribution, but it is actually a tad lower at –101.0 dBu. This is because of the lowpass action of the filter. If the 22 kHz measurement filter in the testgear was a brick-wall job this would not happen, but it has in fact only a 18 dB/octave roll-off so the bandwidth definition filtering above 22 kHz does have enough effect to outweigh the noise contribution of the stage.

At this point we cannot say that either stage is in urgent need of improvement, though we thoughtfully note that effectively all the noise from the input circuitry is generated internally by the balanced input amplifier.

19.20 Noise Analysis: HF Path

The measured noise output from the first 3 kHz highpass filter is –95.1 dBu. This filter has a gain of +6.3 dB, and the noise level at its input is –101.0 dBu, so if the stage was noiseless we might expect a noise output of –101.0 dBu + 6.3 dB = –94.6 dBu; in fact the measured noise output is *less* than this, at –95.1 dBu. This is of course because we are dealing with a 3 kHz highpass filter which is rejecting a substantial chunk of the audio spectrum. Calculating the noise contribution from this stage is therefore quite complicated. For the moment we will simply note that this stage does not appear to be excessively noisy.

The measured noise output from the second 3 kHz highpass filter is –95.2 dBu. This filter has a gain of 0 dB, but again the output noise level is fractionally less than the input noise, due to the filtering action.

The final stage in the HF path is the 80 usec third-order allpass filter time delay compensator, which we might expect to be relatively noisy compared with the filters because of its greater complexity. However, the noise at its output is only 0.2 dB higher than at its input, measuring –95.0 dBu. If we subtract the input noise from the output noise we get an estimate for the stage contribution of –108.5 dBu; unfortunately we are subtracting two figures with only a small difference between them. That difference is fairly near the limit of measurement, and so the result will be very inaccurate. If you want to know the noise from a stage, then the correct method is to measure the noise output of the stage by itself. We did this for the third-order allpass filter back in Chapter 10, and the actual output noise was –102.8 dBu, which shows you how dangerous it is to use the subtraction method inappropriately. If we add the internal noise of –102.8 dBu to the input noise of –95.2 dBu, in theory we get an output noise of –94.5 dBu, an increase of 0.7 dB. (The small discrepancy is almost probably due to minor frequency response irregularities in the allpass filter, caused by capacitor tolerances.) We make a note that the noise contribution of the third-order allpass filter is small, but not negligible. More importantly, we can only conclude that the noise performance of the HF path is dominated by the noise being fed to it from the input circuitry.

The final part of the HF path is the output attenuator and level trim network, which at its nominal setting has a 15.6 dB loss to undo the doubly-elevated level in the HF path.

Since it is just a resistive attenuator, I expected no problems here. But, the noise level measured at the network output was –105.1 dBu, when it should have been –95.0 dBu – 15.6 dBu = –110.6 dBu. A 5 dB discrepancy cannot be ignored; investigation showed that the problem was the output impedance of the network, which at 184 Ω is somewhat higher than usual. The Audio Precision SYS-2702 is designed to be fed from low-impedance sources, and its input amplifiers appear to have high current noise as a consequence of attaining very low voltage noise; this is never a problem in normal use, but here the source impedance is higher than usual. Driving the AP input via a 5532 unity-gain buffer (which has its own noise) and the usual 47 Ω cable-isolating resistor gave a more realistic output noise reading of –109.4 dBu. That is quiet.

19.21 Noise Analysis: MID Path

The measured noise output from the first unity-gain 400 Hz highpass filter is –101.2 dBu. This is again less than the input noise because of the filtering action. The measured noise output from the second unity-gain 400 Hz highpass filter is –101.1 dBu, a tiny increase at the limit of measurement. We conclude that neither filter is making a significant contribution to the noise in the MID path.

After the two highpass filters come the two 3 kHz lowpass filters, fed with –101.1 dBu of noise from the second 400 Hz highpass filter. The noise output from the first lowpass filter is –108.6 dBu, and from the second lowpass filter –109.4 dBu. Both these figures are much lower than the input noise of –101.1 dBu as a consequence of the lowpass filtering which cuts the bandwidth from 22 kHz to 3 kHz.

The final stage in the MID path, as in the HF path, is a third-order allpass filter time delay compensator, this time designed for a 400 usec delay. We saw when we looked at the HF allpass filter in isolation that it was relatively noisy, with a measured output noise of –102.8 dBu; in this case the different circuit values for 400 usec give us a slightly lower allpass-filter-only noise output of –104.1 dBu. (The second-order allpass filter that makes up the first section of the complete third-order filter has a noise output of –104.9 dBu, so that is clearly where most of the noise comes from.) When the third-order allpass filter is placed in the MID path, the measured noise output is –102.9 dBu. Since this is more than 6 dB greater than the –109.4 dBu noise level going in, it is clear that by far the greater part of the noise is generated internally by the third-order allpass filter. This is because the nominal level in the MID path is 6 dB lower than that in the HF path. This time we make a note that the noise contribution of the third-order allpass filter is dominant, and might repay some attention.

The final part of the MID path is the output attenuator and level trim network, which at its nominal setting has a 9.5 dB loss to undo the elevated level in the MID path. The noise level measured directly at its output was initially –107.9 dBu, but as we saw at the end of the HF path, that reading is exaggerated by the AP input current noise. Driving the AP input via a 5532 unity-gain buffer and 47 Ω cable-isolating resistor reduced the output noise reading to –110.4 dBu. That too, is rather quiet; the noise level is lower than that at the HF path output because of the presence of the 3 kHz lowpass filters.

19.22 Noise Analysis: LF Path

This path consists only of the two 400 Hz lowpass filters and the fixed output attenuator. Given the 400 Hz cutoff frequency of the lowpass filters we expect a good noise performance, and we get it. The noise at the input of the LF path is the –101.0 dBu coming from the input circuitry. This is reduced to –111.0 dBu at the output of the first lowpass filter, and further to –112.2 dBu at the output of the second lowpass filter.

The final part of the LF path is the fixed output attenuator network, which has a 9.5 dB loss to undo the elevated level in the MID path. This in theory gives an output noise of –121.7 dBu. This is going to be hard to measure as it is below the noise floor of the AP measuring equipment, which is, on my example, –119.6 dBu with the input short-circuited. Once more we have to drive the AP input via a 5532 unity-gain buffer and 47 Ω cable-isolating resistor to avoid misleadingly high readings due to current noise, and that 5532 will add its own voltage noise. The reading we get is –113.8 dBu. This is reduced to –115.1 dBu, after we subtract the known AP noise floor. We then calculate the noise added by the 5532 buffer, as it is too low to measure accurately, using its typical input noise density of 5 nV/√Hz and a 22 kHz bandwidth (the effect of the 5532 current noise in the attenuator output impedance is negligible). The answer is that the 5532 buffer contributes –120.38 dBu. If we subtract that from our figure of –115.1 dBu, we get a stunningly low –116.6 dBu. This may not be the most accurate reading in the history of audio but it is good evidence that everything is working properly, and that we have a *very* low noise output from the LF path.

19.23 Improving the Noise Performance: The MID Path

On our journey down the MID path, we noted that the third-order allpass filter time delay compensator was generating internally most of the noise that was measured at its output. It is the dominant noise generator in the MID path, so let's see if we can do something about that.

There is a very simple fix, which you may have already seen coming. The third-order allpass filter is at the end of the MID path, so all its noise heads for the output unmolested.

If, however, it is moved so it is after the 400 Hz highpass filters but before the 3 kHz lowpass filters, the latter will have a dramatic effect on the noise level. Making the change, the noise output of the third-order allpass filter is now –99.6 dBu as it adds its own noise to the –101.1 dBu reaching it from the highpass filters upstream. The noise output of the first lowpass filter is pleasingly lower at –107.3 dBu, and the noise output of the second lowpass filter is even lower at –108.2 dBu.

After the output attenuator network, the noise at the final output is reduced from –110.4 dBu to –113.3 dBu, an improvement of 2.9 dB that costs us nothing more than a moment's thought.

The true noise output must be –108.2 – 9.5 dBu = –117.7 dBu, because of the effect of the 9.6 dB output attenuator, but we have already described the difficulties of measuring noise at such low levels. The revised schematic is shown in Figure 19.8.

In case you're wondering why the MID path wasn't configured in this way from the start, the answer is that until I built the prototype circuit I didn't know how the noise contributions of the stages were going to work out. It could have been predicted by a lot of calculation, but that gets complicated when you're dealing with stages like filters that have a non-flat-frequency response. Sometimes it's quicker to just get out the prototype board and start plugging bits in. You're going to be doing it sooner or later.

19.24 Improving the Noise Performance: The Input Circuitry

The other major point we noted in our noise analysis of the crossover was that the noise performance of the HF path was dominated by the noise being fed to it from the input circuitry, the incoming level being –101.0 dBu. Almost all of this is coming from the balanced input amplifier, the contribution of the bandwidth definition filter being very small. Therefore the only way to improve things is to take a hard look at the balanced input amplifier; clearly our original design decision to make it a special low-noise type was sound, but what are the options for making it even quieter? Improvements to this stage are going to be value for money as they will reduce the noise level being fed to all three signal paths.

Now, you might be questioning whether it is worth spending any more money at all on noise reduction, because the circuitry is already rather quiet, and you might expect the equipment upstream, the preamplifier, mixing console or whatever, to generate a higher noise level than the crossover. This to my mind is a matter of design philosophy rather than dogged pragmatism. We may suspect that the source equipment will be noisier than our circuitry, but that is the responsibility of whoever designed it. If we make our device as good as we can without making obviously uneconomic decisions, then we not only get a

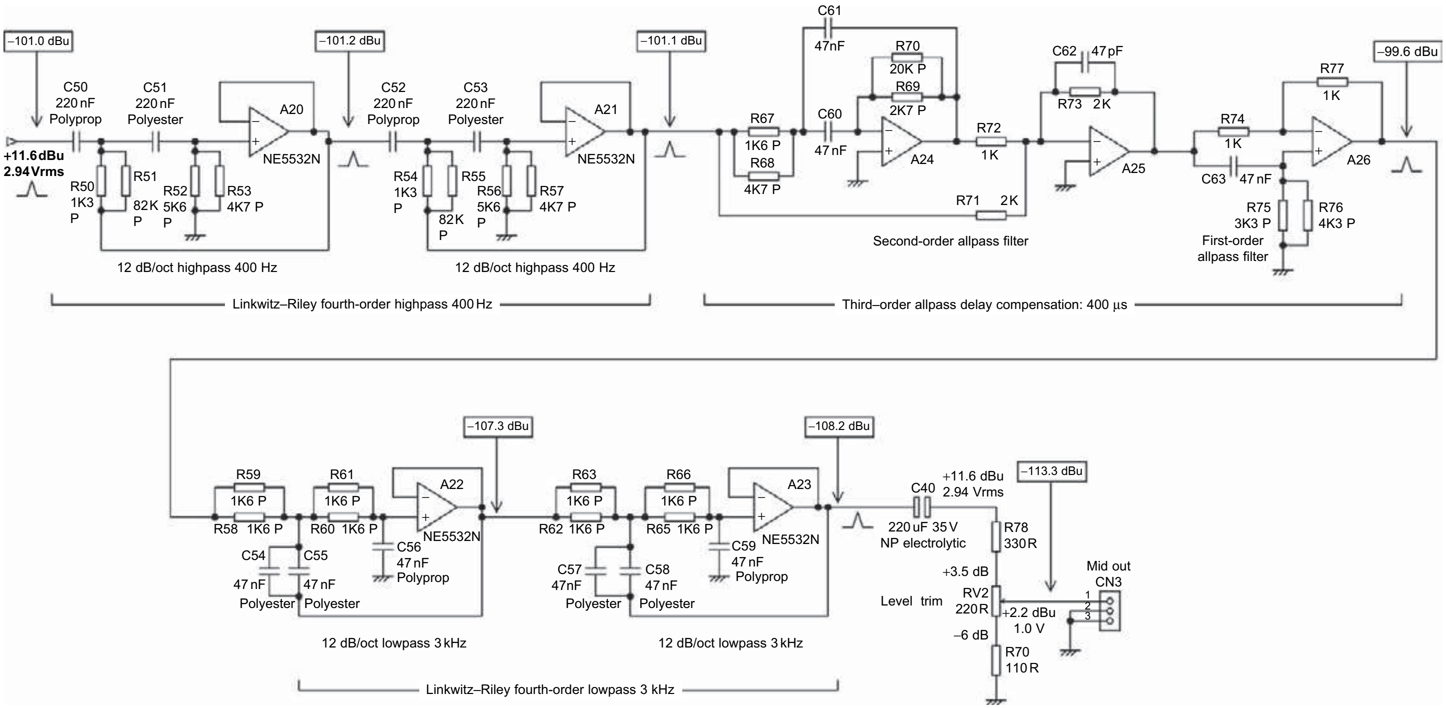


Figure 19.8: Schematic diagram of the improved MID path, with the third-order time delay compensator moved to before the Linkwitz–Riley lowpass filter to reduce noise.

virtuous warm glow, but we can be sure we are relatively future-proof in terms of improvements that may be brought about in the source equipment. There is also the point that we will get some excellent numbers for our specifications, and they may impress prospective customers.

In Chapter 16 we looked at a series of improvements that could be made to unity-gain balanced input amplifiers to improve their noise performance, all of which, regrettably but perhaps inevitably, required more numerous or more costly opamps. Here we are dealing with a gain of +11.9 dB so the optimal path of improvement may be somewhat different. The original balanced input stage consists of two unity-gain buffers and a low-impedance balanced (differential) amplifier. From here on I shall just refer to them as “the buffers” and “the balanced amplifier.” Given the difficulty of measuring the very low noise levels at the outputs of the attenuator network, the noise was measured just before the attenuator and the output noise calculated. In each case the input noise of the AP measuring system has been subtracted in an attempt to obtain accurate numbers.

We proceed as follows:

Step 1: Replace the 5532 opamp A3 in the balanced amplifier with the expensive but quiet LM4562. This gives a rather disappointing reduction in the noise from the balanced stage of 1.1 dB, which feeds through to even smaller improvements in the path outputs

Step 2: We deduce that the input buffers A1, A2 must be generating more of the noise than we thought, so we put the 5532 back in the A3 position, and replace both input buffers with a LM4562; since the LM4562 is a dual opamp this makes layout simple. The noise from the balanced stage is now 1.9 dB better than the original design, which is more encouraging, and two of the paths have a 2.2 dB improvement. This sounds impossible, but actually results from the three crossover paths dealing with different parts of the audio spectrum. At any rate there is no doubt that the modification of Step 2 is more effective than Step 1.

Step 3: We enhance the balanced amplifier by putting two identical 5532 stages in parallel and averaging their outputs by connecting them together with 10 Ω resistors. The LM4562 input buffers are retained as their superior load-driving capability is useful for feeding two balanced stages in parallel. This drops the balanced input stage noise to 3.9 dB below the original design. The improvement at the output of the bandwidth definition filter is less at 3.1 dB. The filter makes its own noise contribution and this is now more significant. Once more there are useful reductions in the noise out from the three paths.

Step 4: LM4562s are not cheap, but we decide to splurge on two of them. We return to a single balanced amplifier using an LM4562, and the LM4562 input buffers are retained. This is very little better than Step 3, and significantly more expensive.

Step 5: It appears that abandoning the double balanced amplifier was not a good move, so we bring it back, this time with both of its sections employing an LM4562. We retain the LM4562 buffers. This gives a very definite improvement, with the balanced input stage noise now 5.8 dB below the original design. The path output noise measurements also improve, though by a lesser amount, because, as with bandwidth definition filter, the noise generated by the filter circuitry has become more significant as the noise from the balanced input stage has been reduced. We have only replaced two opamp packages with the LM4562 (which is about ten times more costly than the 5532) so the cost increase is not great.

The reduction in balanced input amplifier noise could be pursued much further, by more extensive paralleling of opamps, as described in Chapter 16. However, is it worthwhile? This is questionable given the noise that is generated downstream of it.

These results are summarised in Table 19.4., together with their effect on the noise output (before the output attenuator) of each of the three crossover paths. You will note a few minor inconsistencies in the last decimal place of some of the figures; this is due partly to the fact that the three paths are handling different parts of the audio spectrum, and to some extent due to the difficulties of measuring the very low noise levels accurately. Nonetheless, the overview it gives of the worth of the various modifications is correct.

Looking at the bottom two rows of the table which gives the final noise outputs, and the signal/noise ratios for a 1 Vrms output (after the output attenuators) we can see that the

Table 19.4: Improvements in Noise Performance

		Bal Stage	BW Defn	HF	MID	LF
Step		**Noise Out dB**	**Noise Out dB**	**Noise Out dB**	**Noise Out dB**	**Noise Out dB**
0	Original design	0 dB ref	0 dB ref	0 dB ref	0 dB ref	0 dB ref
1	Bal amp A3 = LM4562 Buffers A1, A2 = 5532	1.1	1.2	1.0	1.4	1.8
2	Bal amp A3 = 5532 Buffers A1, A2 = LM4562	1.9	1.8	2.2	1.5	2.2
3	2 × Bal amp A3 = 5532 Buffers A1, A2 = LM4562	3.9	3.1	3.5	1.9	2.3
4	Bal amp A3 = LM4562 Buffers A1, A2 = LM4562	4.0	3.3	2.3	2.0	2.5
5	2 × Bal amp A3 = LM4562 Buffers A1, A2 = LM4562	5.8	4.7	4.7	2.0	2.6
5	Path output noise dBu			−115.3	−120.0	−125.2
5	Path output signal/noise ratio dB			117.5	122.2	127.4

HF path is the noisiest by a long way, despite its 6 dB higher internal level. This is because it contains a relatively noisy third-order allpass filter, but no lowpass filters to discriminate against its noise. Clearly that 6 dB extra level is a very good idea. The MID path is quieter because it does have a lowpass filter which can be placed after its third-order allpass filter. The LF path consists only of a lowpass filter and is consequently the quietest of the lot.

The final version (Step 5) of the input circuitry is shown in Figure 19.9. Note that the bandwidth definition filter has not been modified at any point.

19.25 The Noise Performance: Comparisons with Power Amplifier Noise

I hope you will agree that the noise performance of this active crossover is rather good, especially after the modifications. But how does it compare with the noise from a well-designed power amplifier? The EIN of a Blameless power amplifier is –120 dBu [1] if the input signal is applied directly to the power amplifier, rather than through a balanced input stage. As we have seen, balanced input stages are relatively noisy and are certainly much noisier than a power amplifier alone.

The figure of –120 dBu is therefore the most demanding case to compare the crossover noise against. What are we trying to achieve? If we accept that the noise from the loudspeakers can go up by 3 dB when we connect the crossover, its output noise must be no more than –120 dBu, at the same level as the power amplifier EIN. If, however, we are more demanding and will only stand for a noise increase of 1 dB, which is at the limit of audibility, then the crossover output noise must be reduced to –126 dB. We will take that figure as a target and see what can be done.

We will look at the noise output of the MID path because the ear is most sensitive in this part of the audio spectrum (400 Hz–3 kHz). The noise output after the performance optimisation is –120.0 dBu. This is very quiet indeed for any piece of audio equipment, but regrettably 6.0 dB higher than the ambitious target we have just adopted. You may be doubting if the crossover noise performance could be improved that much, and while I am sure it could be done, I am less sure that it could be done economically.

The resistance values in the crossover have already been reduced as much as possible without introducing extra distortion or demanding large and expensive capacitors, reducing the effects of opamp current noise and resistor Johnson noise, but this has no effect on opamp voltage noise. The effect of this can only be reduced by using more expensive opamps with a lower input voltage noise density, or by paralleling opamp stages. Neither technique promises a radical reduction in noise when practically applied.

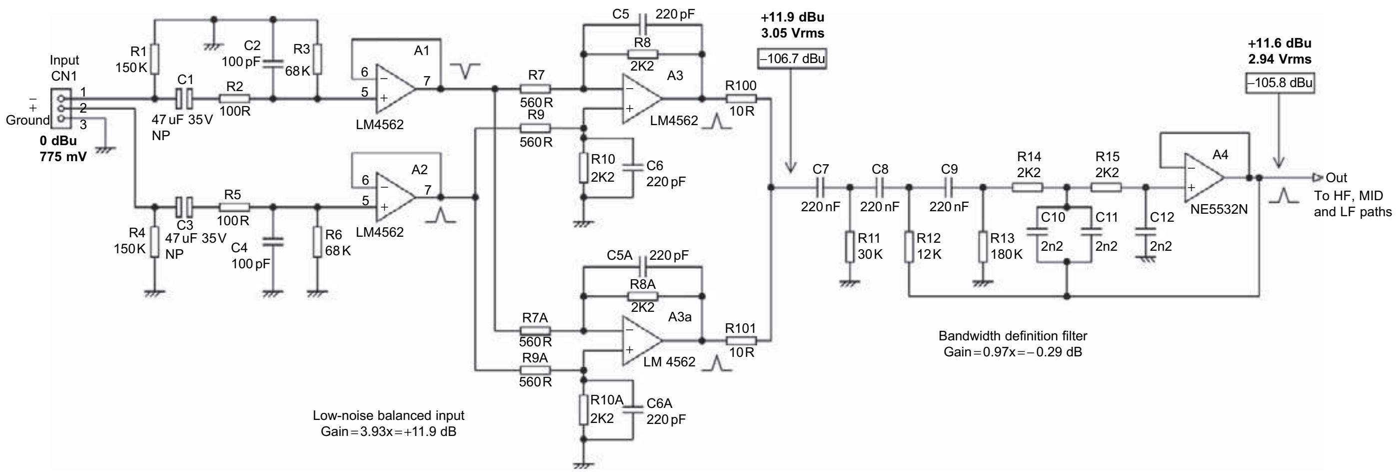


Figure 19.9: Schematic diagram of the improved balanced input amplifier in its final state, using two paralleled balanced input amplifiers and two LM4562 opamps. The noise output from the balanced input stage has been reduced by 5.8 dB.

Nonetheless, let us conduct a thought experiment. We know that whatever amplifier technology you use, putting two amplifiers in parallel and averaging their outputs (usually by connecting them together with 10 Ω resistors) reduces the noise output by 3 dB; putting four in parallel reduces the noise by 6 dB, and so on. The same applies to putting two identical active crossovers in parallel, so… if we stacked up four of them we could unquestionably get a 6 dB noise reduction and meet our target. This is perhaps not very sensible, but it does prove one thing—it is physically possible to accomplish the formidable task of meeting our very demanding noise target, even if it is hardly economical to do so.

19.26 Conclusion

The description of this active crossover has hopefully demonstrated the techniques and design principles described in this book. The use of elevated internal levels, plus other low-noise techniques, has allowed us to achieve a remarkable noise performance.

Reference

[1] D. Self, Audio Power Amplifier Handbook, fifth ed., Newnes, Boston, MA, p. 104 (power amp input noise) W ISBN: 978-0-240-52162-6.

APPENDIX 1

Crossover Design References

Not all these papers and articles are referenced in the main text of this book. They are of varying value and significance. These references are to crossover design itself; references relating to peripheral matters like filter design are not included. Papers relating solely to digital crossovers have also not been included. The majority of these papers come from the JAES, the *Journal of the Audio Engineering Society*. These can be purchased online at http://www.aes.org/. The cost is currently $5 per paper for members and $20 per paper for non-members.

Early Stuff	
Hilliard & Kimball	"Dividing Networks for Loud Speaker Systems" Academy Research Council Technical Bulletin, Mar 1936 Reprinted *JAES* Vol 26 #11, Nov 1978, pp. 850–855
1960–69	
Robert Ashley	"On the Transient Response of Ideal Crossover Networks" *JAES* Vol 10 #3, July 1962, pp. 241–244
1970–79	
Richard H Small	"Constant Voltage Crossover Network Design" *JAES* Vol 19 #1, Jan 1971, pp. 12–19
Richard H Small	"Phase and Delay Distortion in Multiple-Driver Loudspeaker Systems" *JAES* Vol 19 #1, Jan 1971, p. 56
Richard H Small	"Crossover Networks and Modulation Distortion" *JAES* Vol 19 #1, Jan 1971, pp. 55–56
Ashley & Kaminsky	"Active and Passive Filters as Loudspeaker Crossover Networks" *JAES* Vol 19 #6, June 1971, pp. 494–502
Don Keele	"What's So Sacred About Exponential Horns?" AES Preprint #1038, May 1975
Siegfried Linkwitz	"Active Crossover Networks for Non-coincident Drivers" *JAES* Vol 24 #1, Jan/Feb 1976, pp. 2–8
Erik Baekgaard	"A Novel Approach to Linear Phase Loudspeakers Using Passive Crossover Networks" *JAES* Vol 25 #5, 284 1977 (filler-driver)
Siegfried Linkwitz	"Passive Crossover Networks for Non-coincident Drivers" *JAES* Vol 26 #3, March 1978, pp. 149–150
1980–89	
W. Marshall Leach	"Loudspeaker Driver Phase Response: The Neglected Factor in Crossover Network Design" *JAES* Vol 28 #6, June 1980, pp. 410–421
Garde, P	"All-pass Crossover Systems" *JAES* Vol 28 #9, Sept 1980

(*Continued*)

(*Continued*)

Dennis G Fink	"Time Offset and Crossover Design" *JAES* Vol 28, September 1980, pp. 601–611
Robert Bullock	"Loudspeaker-Crossover Systems: An Optimal Crossover Choice" *JAES* Vol 30 #7/8, Aug 1982, pp. 486–495
Adams & Roe	"Computer-Aided Design of Loudspeaker Crossover Networks" *JAES* Vol 30 #7/8, Aug 1982, pp. 496–503
Richard Greiner	"Tone Burst Testing on Selected Electronic Crossover Networks" *JAES* Vol 30 #7/8, Aug 1982, pp. 522–527
Lipshitz, Pocock & Vanderkooy	"Audibility of Midrange Phase Distortion in Audio Systems" *JAES* #9, Sept 1982, pp. 580–595
Wieslaw R Woszczyk	"Bessel Filters as Loudspeaker Crossovers" 72nd AES Convention, Oct 1982, Preprint 1949
Lipshitz SP & Vanderkooy J	"A Family of Linear Phase Crossover Networks of High Slope Derived by Time Delay" *JAES* Vol 31 #1/2, Jan/Feb 1983, pp. 2–20
Greiner & Schoessow	"Electronic Equalisation of Closed-Box Loudspeakers" *JAES* Vol 31 #3, Mar 1983, pp. 125–134
Shanefield	Comment on "Audibility of Midrange Phase Distortion in Audio Systems" plus reply by Lipshitz, Pocock & Vanderkooy *JAES* Vol 31 #6, Jun 1983, p. 447
Robert Bullock	"Satisfying Loudspeaker Crossover Constraints with Conventional Networks" *JAES* Vol 31 #7, Jul/Aug 1983, pp. 489–499
James Moir	Comment on "Audibility of Midrange Phase Distortion in Audio Systems" *JAES* Vol 31 #12, Dec 1983, p. 939
B J Sokol	"Practical Subwoofer Design" (crossover filter) *Electronics World*, Dec 1983, p. 41
Robert Bullock	"Passive 3-Way Allpass Crossover Networks" *JAES* Vol 32 #9, Sept 1984, p. 626
Joseph D'Appolito	"Active Realisation of Multi-Way All-Pass Crossover Systems" 76th AES Convention, Oct 1984, Preprint 2125
Lipshitz & Vanderkooy	"Phase Linearisation of Crossover Networks Possible by Time Offset & Equalisation" *JAES* Vol 32 #12, Dec 1984, p. 946
Lipshitz SP & Vanderkooy J	"Use of Frequency Overlap and Equalization to Produce High-Slope Linear Phase Loudspeaker Crossover Networks" *JAES* Vol 33 #3, Mar 1985, pp. 114–126
Robert Bullock	Corrections to "Passive Three-Way All-Pass Crossover Networks" *JAES* Vol 33 #5, May 1985, p. 355
Catrysse	"On the Design of Some Feedback Circuits for Loudspeakers" *JAES* Vol 33 #6, Jun 1985, p. 430
L R Fincham	"Subjective Importance of Uniform group Delay at Low Frequencies" *JAES* Vol 33 #6, Jun 1985, p. 436
Recklinghausen	"Low-Frequency Range Extension of Loudspeakers" *JAES* Vol 33 #6, Jun 1985, p. 440
Mitra & Damonte	"Tunable Active Crossover Networks" *JAES* Vol 33 #10, Oct 1985, p. 762
Deer, Bloom & Preis	"Perception of Phase Distortion in All-Pass Filters" *JAES* Vol 33 #10, Oct 1985, pp. 782–786
de Wit, Kaizer & Op de Beek	"Numerical Optimization of the Crossover Filters in a Multiway Loudspeaker System" *JAES* Vol 34 #3, March 1986, pp. 115–123

Vanderkooy & Lipshitz	"Power Response of Loudspeakers with Non-coincident Drivers—The Influence of Crossover Design" *JAES* Vol 34 #4, April 1986, pp. 236–244
Robert Bullock	"A New 3-Way Allpass Crossover Network Design" *JAES* Vol 34 #5, May 1986, p. 315
Rudolf Chalupa	"A Subtractive Implementation of Linkwitz–Riley Crossover Design" *JAES* Vol 34 #7/8, July/Aug 1986, p. 556
Lipshitz & Vanderkooy	"In-Phase Crossover Network Design" *JAES* Vol 34 #11, Nov 1986, p. 889
Regalia, Fujii, Mitra, Sanjit, Neuvo	"Active RC Crossover Networks with Adjustable Characteristics" *JAES* Vol 35 #1/2, Jan/Feb 1987, pp. 24–30
M Hawksford	"A Family of Circuit Topologies for the Linkwitz–Riley (LR-4) Crossover Alignment C14" 82nd AES Convention, Mar 1987, Preprint 2468
J D'Appolito	"Active Realization of Multiway All-Pass Crossover Systems" *JAES* Vol 35 #4, April 1987, pp. 239–245
Harry Baggen	"Active Phase-Linear Crossover Network" *Elektor*, Sept 1987, p. 61
Maffioli & Nicolao	"Another Approach to the Ideal Crossover: The Energy Filler" AES Convention 84 (March 1988) Paper #2642
Ronald Aarts	"A New Method for the Design of Crossover Filters" *JAES* Vol 37 #6, June 1989, pp. 445–454
1990–99	
M Hawksford	"Asymmetric All-Pass Crossover Alignments" *JAES* Vol 41 #3, March 1993, pp. 123–134
Neville Thiele	"Precise Passive Crossover Networks Incorporating Loudspeaker Driver Parameters" *JAES* Vol 45 #7/8, July 1997, pp. 585–594
Bill Hardman	"Precise Active Crossover" Electronics World, Aug 1999, p. 652
Alex Megann	http://www.soton.ac.uk/~apm3/diyaudio/Filler_drivers.html, December 1999 (Second and fourth-order crossovers with filler-drivers)
2000–2009	
Christhof Heinzerling	"Adaptable Active Speaker System" (subtractive) *Electronics World*, Feb 2000, p. 105
John Watkinson	"Speaker's Corner: Crossover Summation" *Electronics World* Sept 2000, p. 720
Neville Thiele	"Loudspeaker Crossovers with Notched Responses" *JAES* Vol 48 #9, Sept 2000, p. 784
Neville Thiele	"Passive All-Pass Crossover System of Order 3 (Low Pass) + 5 (High Pass), Incorporating Driver Parameters" *JAES* Vol 50 #12, Dec 2002, pp. 1030–1038
Cochenour, Chai, Rich, & David	"Sensitivity of High-Order Loudspeaker Crossover Networks with All-Pass Response" *JAES* Vol 51 #10, Oct 2003, pp. 898–911
Neville Thiele	"An Active Biquadratic Filter for Equalising Overdamped Loudspeakers" AES Convention Paper 6153, 116th convention, May 2004
Neville Thiele	"Crossover Filter System and Method" US Patent 6,854,005 Feb 2005 (Assigned to Techstream Pty Ltd, Victoria, AU)
R Christensen	"Active All-Pass Crossover Networks with Equal Resistors and Equal Capacitors" *JAES* Vol 54 #1/2, Jan/Feb 2006
Neville Thiele	"Implementing Asymmetrical Crossovers" *JAES* Vol 55 #10, Oct 2007, pp. 819–832

This page intentionally left blank

APPENDIX 2

Loudspeaker Design References

These references are only a small sample of the many that are applicable to loudspeaker design. These have been selected because they either relate closely to crossover design, for example in the equalising of diffraction effects, or are of special significance in themselves, such as the classic papers by Richard Small.

Early stuff	
Rice & Kellogg	"Notes on the Development of a New Type of Hornless Loudspeaker" AIEE Transactions, Sept 1925 (The first description of the modern moving-coil direct radiator cone loudspeaker.)
Muller, Black & Davis	"The Diffraction Produced by Cylindrical and Cubical Obstacles and by Circular and Square Plates" *JAES* Vol 10 #1, July 1938, pp. 6–13
1960–69	
H F Olson	"Direct Radiator Loudspeaker Enclosures" *JAES* Vol 17 #1, Jan 1969, pp. 22–29 (Classic paper on frequency response anomalies due to diffraction.)
1970–79	
Richard H Small	"Direct Radiator Loudspeaker System Analysis" *JAES* Vol 20 #5 June 1972, pp. 383–395
Richard H Small	"Closed-Box Loudspeaker Systems Part 1: Analysis" *JAES* Vol 20 #10, Dec 1972, pp. 798–808
Richard H Small	"Closed-Box Loudspeaker Systems Part 2: Synthesis" *JAES* Vol 21 #1, Feb 1973, pp. 11–18
Richard H Small	"Vented-Box Loudspeaker Systems Part 1: Small-Signal Analysis" *JAES* Vol 21 #5, June 1973, pp. 363–372
Richard H Small	"Vented-Box Loudspeaker Systems Part 2: Large-Signal Analysis" *JAES* Vol 21 #6, Aug 1973, pp. 438–444
Richard H Small	"Vented-Box Loudspeaker Systems Part 3: Synthesis" *JAES* Vol 21 #7, Sept 1973, pp. 549–554
Richard H Small	"Vented-Box Loudspeaker Systems Part 4: Appendices" *JAES* Vol 21 #8, Oct 1973, pp. 635–639
Richard H Small	"Passive-Radiator Loudspeaker Systems Part 1: Analysis" *JAES* Vol 22 #8, Oct 1974, pp. 592–601
Richard H Small	"Passive-Radiator Loudspeaker Systems Part 2: Synthesis" *JAES* Vol 22 #9, Nov 1974, pp. 683–689
1980–89	
Bews & Hawksford	"Application of Geometric Theory of Diffraction (GTD) to Diffraction at the Edges of Loudspeaker Baffles" *JAES* Vol 34 #10, Oct 1986, pp. 771–779

(*Continued*)

(*Continued*)

Juha Backman	"Computation of Diffraction for Loudspeaker Enclosures" *JAES* Vol 37 #5, May 1989, pp. 353–362
James Porter and Earl Geddes	"Loudspeaker Cabinet Edge Diffraction" *JAES* Vol 37 #11, Nov 1989, pp. 908–918
1990–99	
Ralph E Gonzalez	"A Dual-Baffle Loudspeaker Enclosure for Balanced Reverberant Response" 91st AES Convention, Oct 1991, Preprint No. 3203
John Vanderkooy	"A Simple Theory of Cabinet Edge Diffraction" *JAES* Vol 39 #12, Dec 1991, pp. 923–933
Ian Hegglun	"Speaker Feedback" (motional feedback) *Electronics World*, May 1996, p. 378
Russel Breden	"Roaring Subwoofer" (motional feedback) *Electronics World*, Feb 1997, p. 104
J R Wright	"Fundamentals of Diffraction" *JAES* Vol 45 #5, May 1997, pp. 347–356
2000–2009	
John Watkinson	"Speaker's Corner: Voltage or Current Drive?" *Electronics World*, May 2001, p. 354
John Watkinson	"Speaker's Corner: Motional Feedback" *Electronics World*, Sept 2001, p. 698

Index

Page numbers in *italics* indicate figures, tables and footnotes

M

N

T

This page intentionally left blank