Dept. for Speech, Music and Hearing

Quarterly Progress and Status Report

A study of speech intelligibility over a public address system

Lundin, F. J.

KTH Computer Science and Communication

journal: STL-QPSR
volume: 27
number: 1
year: 1986
pages: 001-030

I. ROOM ACOUSTICS

A. A STUDY OF SPEECH INTELLIGIBILITY OVER A PUBLIC ADDRESS SYSTEM

Fred J. Lundin*

Abstract

Speech intelligibility over the public address system at Arlanda Airport has been calculated by different methods. The articulation index method (AI) is based on frequency characteristics and provides merely a rough correction for room reverberation. On the other hand, a method suggested by Peutz (1971) and by Klein (1971) based on room acoustics, does not employ frequency characteristics. A compromise is the SRR-method presented in this paper, which utilizes the direct-to-reverberant sound intensity. It is based on the theory of Peutz and extended to handle the sound levels of the direct sound, of the reverberant sound, and of the noise. The analysis is performed in frequency bands and is applicable to rooms with multiple sources and ambient noise. Finally, the method of modulation transfer function (MTF) has been used. By this method the reduction in modulation depth of speech signals within separate octave bands caused by reverberation is calculated. It is more complex than the other methods. The outcome from these four prediction methods has been compared to measured values recorded by use of a dummy head in two rooms and evaluated by a listening group of ten people. The intelligibility is tested at two background noise levels (with a signal-to-noise ratio of 10 and 20 dB, respectively). The results show a fairly good agreement between measured and predicted data of lower speech levels but when both noise and reverberation interfere, the methods will underestimate the articulation loss. Under these conditions the MTF-method will give the most appropriate result. Our study also indicates that the more complex methods are not much superior to the simpler ones.

1. INTRODUCTION

When a new international terminal at Arlanda Airport, Stockholm, was projected, a high standard of speech intelligibility of the public address system was requested. Special attention was payed to room absorption properties and to the design of the sound reinforcement system. Speech intelligibility predictions were used as a tool for selecting acoustic materials as well as loudspeakers. The sound reinforcement system was equipped with an automatic gain control for compensation of the influence of background noise which resulted in a remarkable improvement in intelligibility.

An intelligibility test was made after the system had been adjusted and put into operation. This was done under laboratory conditions with recorded speech material. In the present study, the results from this test will be compared to different methods of prediction.

---

*Graduate Student at KTH. Working place: Televerket, Farsta, Sweden.

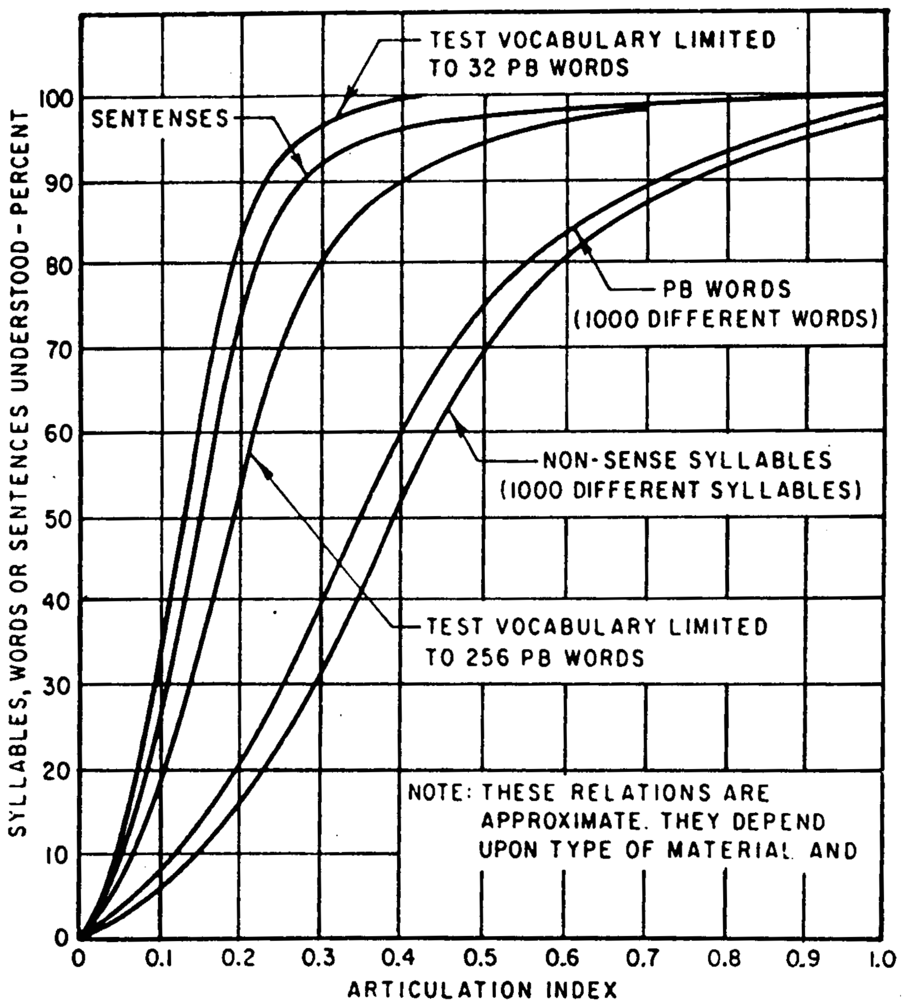

Fig. 1. Relation between AI and various measures of speech intelligibility.

Available theory generally relates to auditoria, normally with one source only. In our situation, the acoustic conditions of the public waiting halls differed considerably from auditoria and the sound distribution had to rely on multiple sources.

2. THEORY

2.1 Articulation Index

Previous speech researchers have worked along different lines. Some of them have studied the overall properties of speech by means of statistical tools. Accordingly, speech has been represented by its long-term spectrum and its amplitude distribution (Beranek, 1947; Fant, 1959; Fletcher, 1953). Other have studied speech from the phonetic point of view. Thus, the speech has been regarded as a dynamic process based on a string of phonemes (Fant, 1968).

The earliest attempt at a quantitative description of the influence of room reverberation on speech was reported by Knudsen & Harris (1950) who discussed the relationship between the reverberation time and the speech intelligibility. At that time, French & Steinberg (1947) developed the concept of articulation index (AI) for intelligibility predictions in the presence of noise and band-pass filtering, see also Beranek (1947). This index has been related to speech intelligibility for different speech materials (Fig. 1). Results from the Knudsen and Harris's studies were also included in the AI-method. The method has further been developed by Kryter (1962) and has become a standard (ANSI, 1969).

The AI-method is based on the long-term idealized speech spectrum for male voices, which is raised 12 dB to include peak amplitudes. The signal-to-noise ratio (SNR) is calculated for a number of frequency bands. The noise level is represented either by the ambient noise or the hearing threshold. Compensation for masking phenomena is also included in the method. The SNR-values are limited to a dynamic range of 30 dB and added together. The articulation index AI is the ratio between the calculated value of a specific case and the maximum value that could be attained. The method is based on 20 bands in the range 200-6100 Hz, which contribute equally to speech intelligibility. Alternatively, 15 third-octave bands or five-octave bands could be used in combination with weighting factors (see Table I).

For influence of room reverberation, the AI-value is corrected with an amount which depends only on the reverberation time of the room. Kryter specifies the reverberation time to the value at 512 Hz according to the result of Knudsen and Harris, but in the ANSI-standard no specific frequency is recommended.

Octave band mid frequency	Weighting factor
250 Hz	0.072
500 Hz	0.144
1000 Hz	0.222
2000 Hz	0.327
4000 Hz	0.234

Table I. Weighting factors for the octave bands.

2.2 Reflection Pattern Models

Research on sound reflection patterns and the balance between early and late reflections have been undertaken by Lochner & Burger (1961). Their studies have been focused on the determination of what combined effect the reverberation, the noise, and the reflection sequence would have on speech intelligibility. They found that the sound energy, received during the first 95 ms after the direct sound, is essential for speech perception, while reflections received after 95 ms are regarded as detrimental. Latham (1979) has modified this theory to take into account background noise. He has formed a signal-to-noise index

$$S/N_{eff} = 10 \cdot \log \frac{\int_{0}^{95ms} w(p,t)\ p^2(t)\ dt}{\int_{95ms}^{\infty} p^2\ (t)\ dt + p_{PNC}^2\ \tau} \tag{1}$$

where w(p,t) is the weighting function for integration properties of the hearing system, p(t) is the instantaneous value of sound pressure, t is time in ms, τ is the period of the speech intelligibility test passage, and p_{PNC} is the level of the background noise specified by preferred noise criterion (PNC) curves.

A similar reasoning is found in Kuttruff (1973) where he forms the log ratio between useful sound intensity and detrimental intensity including also noise. He states that this measure should be greater than or equal to zero as a criterion of good intelligibility.

2.3 Peutz's Method

By defining articulation loss of consonants (AL_{cons}) as a criterion of speech transmission in a room, Peutz (1971) has introduced a more sensitive parameter for intelligibility compared with syllable or word intelligibility, especially when room reverberation is the influential factor. He has performed a series of listening tests under various conditions by using word lists with CVC words, to find the relation between AL_{cons} and reverberation time.

In rooms with reverberation he has found that the AL_{cons} is a function of the reverberation time (T), the room volume (V), and the distance (d) between the speaker and the listener up to a specific distance, the critical distance (d_c)

$$d_c = 0.2 \sqrt{V/T} . \tag{2}$$

Further away from the speaker the AL_{cons}-measures depend on T only.

$$AL_{cons} = \frac{200\, d^2\, T^2}{V} + a \ (\%) \qquad (\text{for } d<d_c) \tag{3}$$

$$AL_{cons} = 9\,T + a \qquad (\%) \qquad (\text{for } d>d_c) \tag{4}$$

A correction a has been added to the AL_{cons}-value that depends on the listener's skill. In Peutz's studies this varied between 1.5 and 12.5%.

In the case of interfering noise, the articulation loss of consonants was a function of the SNR-value in the range between -10 dB and 25 dB. For values of SNR less than -10 dB, the articulation loss of consonants was 100% and above 25 dB the articulation loss of consonants did not vary with the SNR-value.

Thus, Peutz has stated that the intelligibility in terms of AL_{cons} can be predicted for different source-listener distances in rooms where noise and reverberation influence the speech. However, the significance of AL_{cons}-measures has not been well established. For a claimed high intelligibility, an AL_{cons}-value of less than 10%-15% seems acceptable but Peutz does not provide any clear guide-lines. However, the AL_{cons}-measure is now widely accepted.

Klein (1971) has extended the theory of Peutz to be applicable for design and judgement of sound reinforcement systems. When n sources in a room contribute to the sound intensity and their directivity factor is Q, the critical distance will be

$$d_c = 0.2 \sqrt{Q\, V / n\, T} . \tag{5}$$

This expression is valid only if the sound field of the room is diffuse, and all of the sources contribute equally to the reverberant sound. We know that both Q and T in the expression usually vary with the frequency. Sometimes, T might vary with a factor of 5, and Q might vary with a factor of 100 in the speech range. Therefore, the variation of the parameters in the frequency range should not be neglected.

2.4 Direct-to-Reverberant Intensity Method (SRR)

For a sound source in a room, the intensity of the direct sound will be

$$I_d = Q\,P / (4\pi d^2) \qquad (6)$$

and the intensity of the reverberant sound

$$I_r = 4\,P / A\,, \qquad (7)$$

where Q is the directivity of the source, d is the distance between the source and the listener, P is the acoustical power in W, and A is the absorption of the room in m^2 sabine.

The logarithm of the direct-to-reverberant sound intensity ratio we denote SRR (signal-to-reverberation ratio) in accordance with SNR.

$$SRR = 10 \cdot \log\,(I_d/I_r) \qquad (8)$$

The distance between the sound source and the point, where the direct sound is equal to the reverberant sound (SRR = 0 dB), is called the reverberation radius r_r. From Eqs. (6) and (7) and Sabines formula we get

$$r_r = \sqrt{QA / 16\pi} = 0.057\,\sqrt{Q\,V / T} \qquad (9)$$

From a comparison with Eq. (5) the critical distance of Peutz and of Klein will be equal to 3.51 r_r and should emphasize the importance of r_r for the speech intelligibility. The listening distance should be normalized to this reverberation radius, and from Peutz's result we observe that the intelligibility is a function of this measure up to 3-4 r_r.

The direct-to-reverberant sound ratio SRR we express by Eqs. (6), (7), (8), and (9) as a function of the distance to the source d, and the reverberation radius r_r by

$$SRR = -20 \cdot \log\, d/r_r. \qquad (10)$$

At Peutz's critical distance, the direct-to-reverberant sound ratio SRR is -10.9 dB. Using Eq. (3) generalized by Klein together with Eqs. (9) and (10) gives the articulation loss of consonants inside the critical distance (with zero-correction a=0) as

$$AL_{cons} = 0.65 \cdot T \cdot 10^{-SRR/10} \quad (\%). \tag{11}$$

Outside the critical distance, AL_{cons} is still 9 T (%).

For multiple sources in a room at different distances from the listener, it is possible to calculate the sum of the individual intensities from each of the sources and the intensity of the composited reverberant sound, and finally the SRR-value. Hereby, the AL_{cons}-value can be calculated by Eq. (11), which was not possible by Eq. (3).

In the presence of ambient noise with an intensity I_n, the signal-to-noise ratio SNR will be defined by

$$SNR = 10 \cdot \log((I_d + I_r)/I_n) \tag{12}$$

From the diagrams of Peutz, we deduce the following relation between the SNR-value and the articulation loss of consonants

$$AL_{cons} = (AL')^{(SNR+10)/35} \cdot 10^{(50-2 \cdot SNR)/35} \tag{13}$$

in the range -10 dB<SNR<25 dB. AL′ is the articulation loss of consonants, as in Eq. (11), when only reverberation but not noise (i.e., SNR>25 dB) will reduce the intelligibility.

The SRR-value and the SNR-value can be calculated from the frequency response of speech, the transmission characteristics of the sound reinforcement system, the frequency response of the loudspeakers, and from the noise spectrum. Preferably, the calculation should be done in several frequency bands and a final AL_{cons}-value can be obtained by weighting the results from each band. Thus, the SRR-method utilizes the AL_{cons}-prediction from Peutz, the frequency analysis from the AI-method, and intensity considerations as in the reflection pattern models.

2.5 Modulation Transfer Function

A more complex way of calculating the speech intelligibility is by the modulation transfer function (MTF). The method was first proposed by Houtgast & Steeneken (1973) and has later been revised (Houtgast, Steeneken, & Plomp, 1980). Here, speech is regarded as a modulated signal with a modulation frequency F in the range F = 0.4 Hz to F = 20 Hz. The influence of the room acoustics will smear out the modulation depth (see Lundin, 1982). This reduction in modulation can be interpreted as an apparent signal-to-noise ratio (SNR_{app}). It is in accordance with the theory behind the AI-method or the SRR-method. If the modulation is m(F), then

$$SNR_{app} = 10 \log \frac{m(F)}{1 - m(F)} \quad . \tag{14}$$

It can be shown (Houtgast & Steeneken, 1978) that the modulation transfer function in a room with reverberation time T is

$$m(F) = \left(1 + (2\pi F\, T /13.8)^2\right)^{-\frac{1}{2}} \tag{15}$$

The m(F)-values should be calculated for modulation frequencies F from 0.4 Hz up to 20 Hz in third-octave band steps and for octave bands from 125 Hz up to 8000 Hz. This forms 126 values, which may be transformed to apparent SNR-values by Eq. (14).

In accordance with the limitation to 30 dB of the dynamic range of speech used in the articulation index calculations, the apparent SNR-values are limited to the range $-15\ \text{dB} < SNR_{app} < 15\ \text{dB}$. Next, the apparent signal-to-noise ratios are averaged over the modulation frequencies, resulting in one value for each of the seven-octave bands from 125 Hz to 8000 Hz. Finally, a speech transmission index STI (in accordance with the articulation index AI) is calculated for each octave band by

$$STI_{oct} = (\overline{SNR_{app}} + 15) \,/\, 30 \tag{16}$$

and the bands are weighted together to form the final STI-value.

The relation between the speech transmission index STI and the articulation loss of consonants AL_{cons} has been studied under various conditions with reverberation and noise. From a chart presented by Houtgast, Steeneken, & Plomp (1980), we deduce the following empirical relation for nonsense CVC words

$$AL_{cons} = 10^{-2.3\cdot STI + 2.2} \quad . \tag{17}$$

A specific feature of the modulation transfer function method is the possibility to measure the STI-value (Steeneken & Houtgast, 1980). Today, there is even a portable instrument for measuring STI by a fast method (Steeneken & Houtgast, 1985). This method (RASTI) utilizes only nine of the 126 testing points.

3. METHODS

3.1 Recording of the Test Material

In the international terminal at Arlanda Airport there are three large halls and two piers with 20 gates in total. The largest hall is the departure hall with a volume of 37,000 m^3. The floor is 22 m x 198 m and the height is 8.7 m. From the ceiling 39 horn loudspeakers (JBL 2390/2440 with lenses) cover the listening area. The loudspeakers are positioned in two rows (Fig. 2a) in a zig-zag pattern 8.0 m above the floor. The loudspeakers are oriented to give a slice-like radiation pattern in the cross section and not along the hall. For good coverage the horns are tilted 30° towards the center line of the hall (Fig. 2b).

The intelligibility of speech from the loudspeaker system in this very absorbent hall was of specific interest. The reverberation time was as low as 1.1 s. For the test two positions were selected in the room (Fig. 2a). The first position (H), with good sound coverage, was chosen under a loudspeaker in a row, and the second one (J), with poor sound coverage, between two of loudspeakers in a row.

Another room of interest was one of the waiting areas at a gate. Many of these areas were arranged as open plan areas, but a few of them were delimited by walls. The selected room (Fig. 3) had a volume of 1080 m^3 (19 m x 14 m x 4 m), and the reverberation time was 0.6 s. The distributed loudspeaker system in the room was composed of three rows with six loudspeakers (JBL 2110); each located in a square network and mounted 3.0 m above the floor. The two selected points in this room were under one of the loudspeakers (P) and in the middle of four of the loudspeakers (Q).

In these four positions the intelligibility test was carried out (Hagerman & Lindblad, 1978). The speech material comprised 13 phonetically balanced lists of 50 nonsense CVC words each. They were recorded in an anechoic chamber from a female speaker. These lists were played back over the sound reinforcement system and a test material was recorded in stereo by using a dummy head (Kleiner, 1976). The ears of the dummy head were situated 1.25 m over the floor. For calibration and setting of the spectral balance, the same person used for recording of the speech material read the lists with the ordinary microphone of the announcement center. This was done as an alternative to the playback of the lists. The sound was monitored at the recording positions through the microphones of the dummy head and headphones, and was compared with the sound of the loudspeakers. This enabled the balance setting for the recording to be as natural as possible.

In addition to the recorded test material, a reference tape was made by copying the original tape through a filtering that was equal to the frequency response of the sound reinforcement system and the dummy head together. Hence, the reference tapes should be equivalent to the test conditions except for room reverberation.

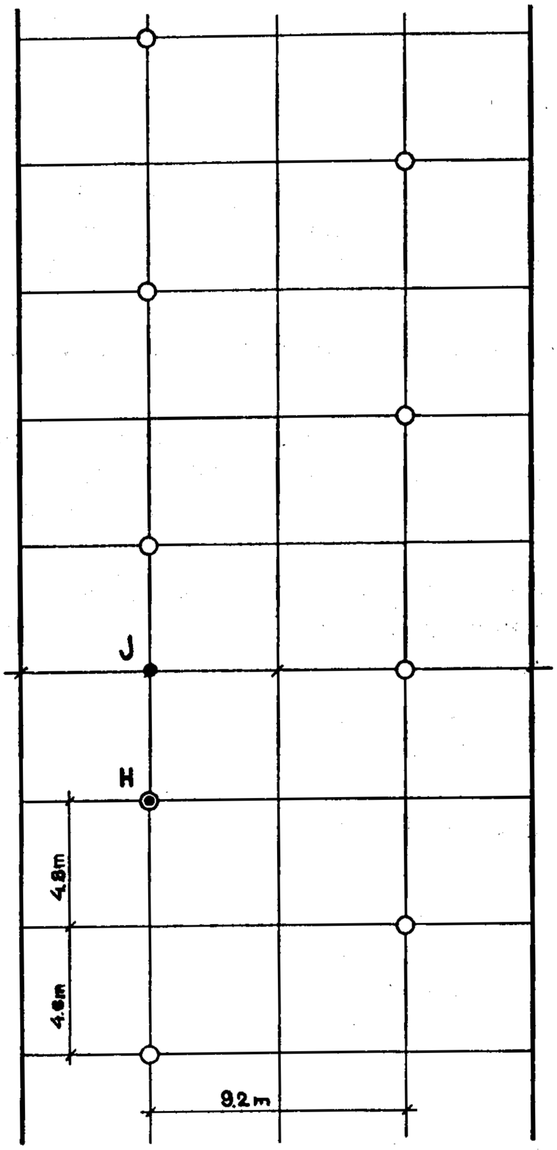

Fig. 2a. Part of a plan-drawing of the departure hall. The loudspeakers are indicated by circles, and the positions H and J by dots.

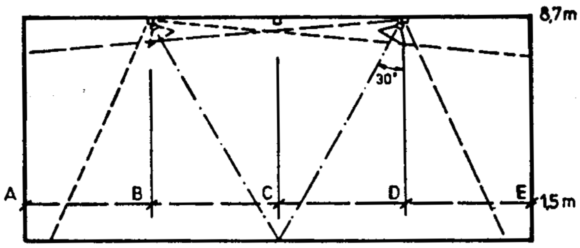

Fig. 2b. Cross-section of the departure hall.

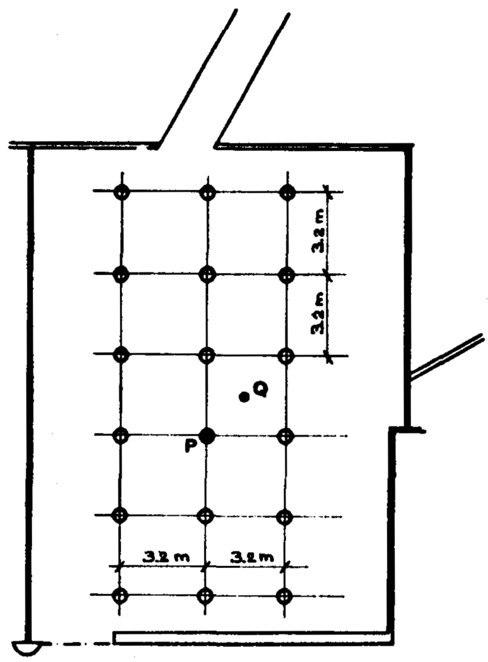

Fig. 3. Plan-drawing of the room at the gate. The loudspeakers are indicated by circles, and the positions P and Q by dots.

The calibration tone on the test tapes could not be used for level control due to standing waves in the rooms. Therefore, the speech level of the recorded material had to be measured. For every listening position the levels of all the words of one list were plotted. The speech level was determined as the mean values of the peaks of the 50 words in that position.

To compensate for the influence of ambient noise, the sound reinforcement system was equipped with an automatic gain control, which was controlled by the noise during the pauses between the announcements. This unit had a dynamic range of 20 dB, and the gain was set by the noise level according to the curve in Fig. 4. On this curve two points were chosen. The sound level of 70 dB(A) with a background noise level of 50 dB(A) represented normal conditions. The other point represented a noisy condition with a background noise level of 75 dB(A), which would adjust the speech level to 85 dB(A). Consequently, the intelligibility test was performed at the two signal-to-noise ratios of 10 dB and of 20 dB, respectively.

The recording was done in the night under quiet conditions. A stationary background noise was preferred to be used in the intelligibility test. Some recordings of the noise were done inside the terminal in the middle of the day, and the long-term spectra of these recordings were analyzed. The average spectrum was flat up to 400 Hz and for higher frequencies the fall was 6 dB/octave (Fig. 5). Two noise generators with this long-term spectrum were built and connected to each of the stereo channels to give uncorrelated noise between the channels. Thus, the noise was mixed into the speech material to be tested.

3.2 The Test Procedure

Ten students at the Royal Institute of Technology with normal hearing formed the listening group. The test material from the four listening positions were presented to the listeners at two signal-to-noise ratios. In addition, the reference material was presented without any noise and with noise that corresponded to an SNR-value of 10 dB. Everyone of the test persons listened to one list for each of the ten conditions. The test material was presented through ear-phones (Yamaha HP1).

4. RESULTS

4.1 Intelligibility Test

Confusion matrices from the test were drawn for the initial consonants, the vowels, and the final consonants. Some of these are presented in Appendix I. In broad outline the confusion matrices show specific problems in perceiving the consonants /b/, /v/, /p /, /m/, /n/, /f/, /j/, and /ʃ / compared to the other consonants.

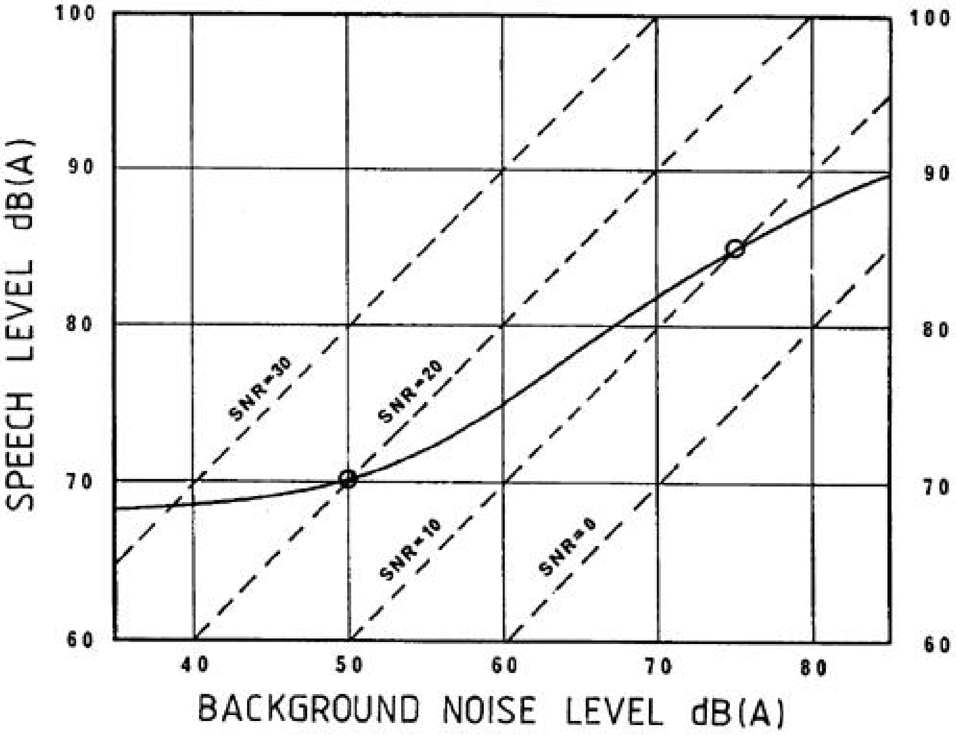

Fig. 4. The gain adjustment of the automatic gain control.

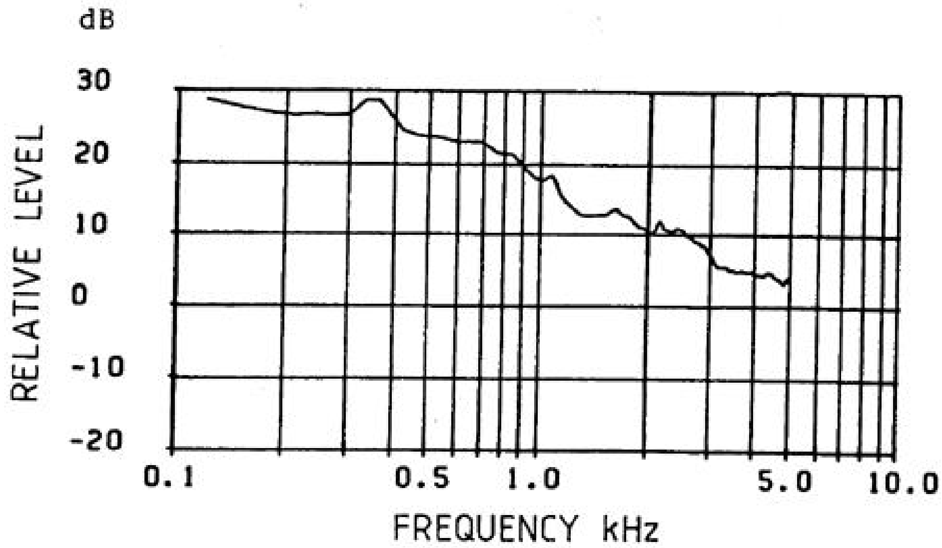

Fig. 5. Long-term-average-spectrum from noise measurements at Arlanda Airport.

Some typical confusions were made between voiced and unvoiced stop consonants, such as between /g/ and /k/, or /t/ and /d/ but also between the voiced consonants /v/ and /b/, especially in the initial position. In the reverberant and noisy situation (with SNR=10 dB) the discrimination between the nasals /n/, /m/, and /ŋ/ was very hard and the average articulation loss was as high as 30%. Many of the nasals were also perceived as /v/, /d/, or /l/.

The fricative /f/ was often perceived as /v/, /p/, or /k/, and the /ʃ/-sound tended sometimes to be perceived as a /ʒ/-sound. the confusions were observed at the background noise level of 50 dB(A), as expected.

For the vowels the most frequent confusions were between long and short vowels. We also observed confusions between the front vowels on one hand and between the back vowels on the other hand.

For all of the ten conditions the average intelligibility and standard deviation of the test subjects were calculated. This was done for whole words, initial and final consonants, and vowels. The complete results are shown in Appendix III. From these data the articulation loss of consonants was calculated as the mean value of the initial and the final consonants.

4.2 Comparison to Predicted Values

We calculated the sound intensity of the direct sound in the departure hall from loudspeaker data (such as the sound pressure levels in various directions for different frequency bands), distance from the listening points to each of the loudspeakers, and the equalization of the sound reinforcement system (Lundin, 1983).

In a similar way, the reverberant sound level was calculated with respect to the room absorption. The intensities of the direct and the reverberant sounds were added, weighted by the long-term spectrum of a male speaker, and adjusted in level to an A-weighted speech level of 70 dB(A) and 85 dB(A), respectively. According to the AI-calculation scheme, the speech level was raised 12 dB to include the peaks of the speech.

The background noise was frequency balanced, based on the measurements in Fig. 5, and the level was set to 50 dB(A) and 75 dB(A), respectively. The signal-to-noise ratios of the five-octave bands were weighted and added to get the final AI-value for a specific position and background noise. The corresponding intelligibility or articulation loss values of the ten conditions were found from the chart of the relation between AI and various measures of speech intelligibility (Fig. 1), where the curve for the phonetically balanced (PB) 1000 words was used. Fig. 6a shows a comparison between predicted and measured data for the different positions at the speech level of 70 dB(A), and in Fig. 6b the corresponding data for the speech level of 85 dB(A) are shown.

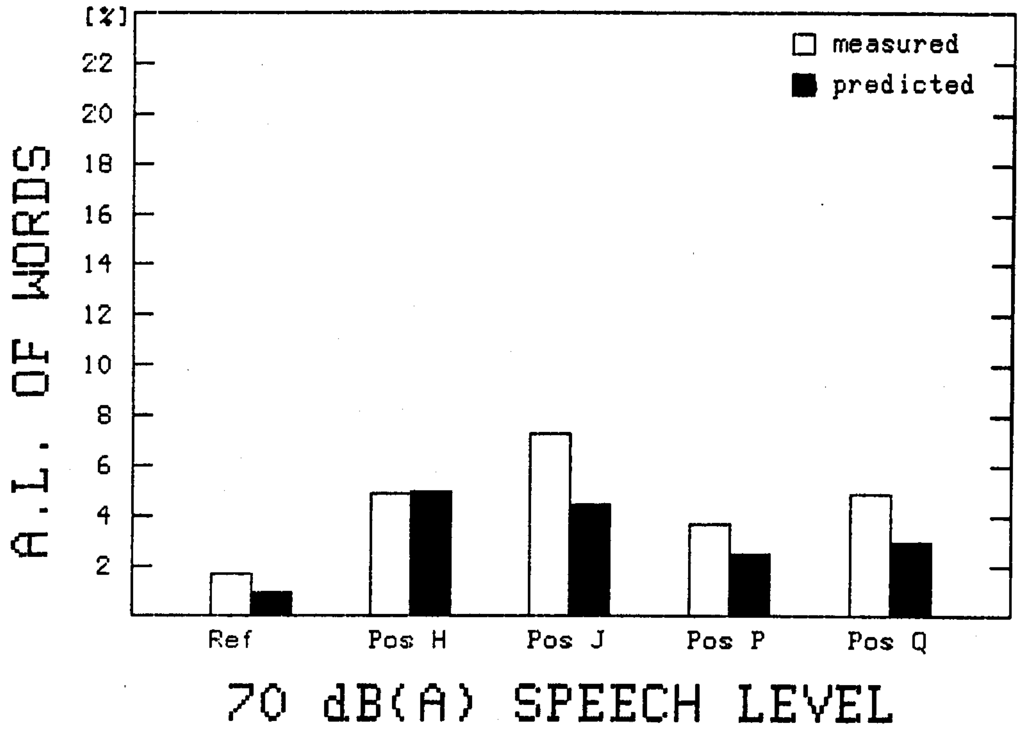

Fig. 6a. Predicted intelligibility by the AI-method in terms of articulation loss of words compared to measured values in the different positions at a speech level of 70 dB(A) and a noise level of 50 dB(A).

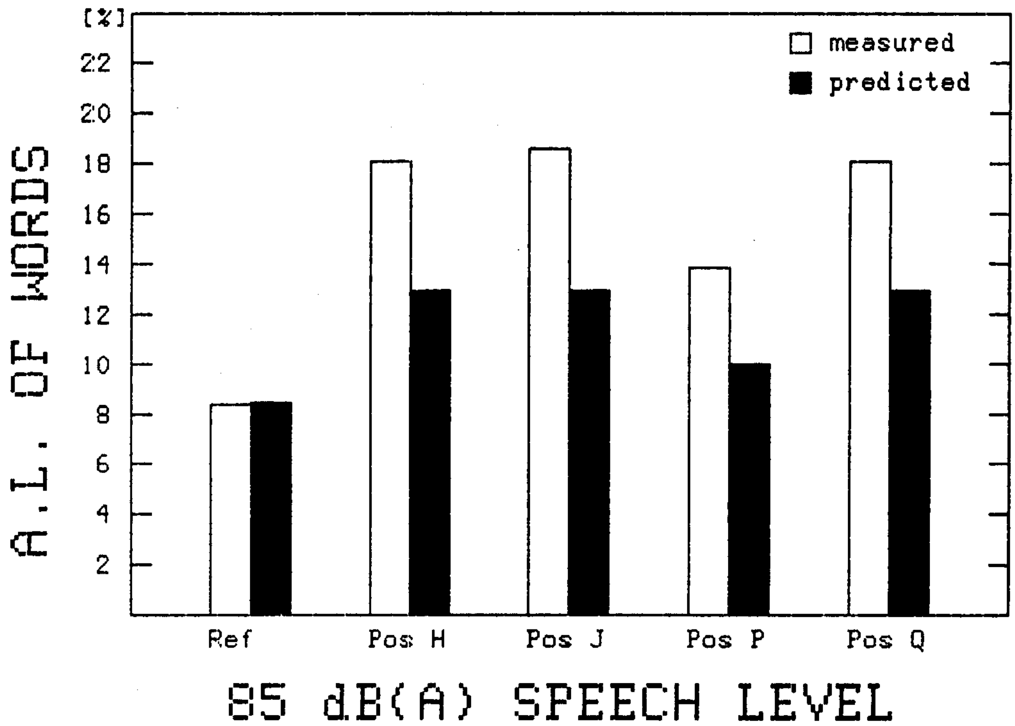

Fig. 6b. Predicted intelligibility by the AI-method in terms of articulation loss of words compared to measured values in the different positions at a speech level of 85 dB(A) and a noise level of 75 dB(A).

When calculating the articulation loss of consonants by the method suggested by Peutz and by Klein, we used for the reverberation time T and the directivity Q the average values of the 1000 Hz and the 2000 Hz octave bands. Peutz has used the reverberation time at 1400 Hz. In the departure hall there was 39 loudspeakers of the same type. Only the distance to the nearest one was used in the calculations. The SNR-value was set to 20 dB and 10 dB, respectively.

When using the proposed SRR-method the direct sound level, the reverberant sound level, and the level of the ambient noise were calculated as in the AI-method. The long-time-average-speech spectrum was used as an input. From the octave band values of the direct-to-reverberant sound ratio (SRR) and the direct+reverberant-to-noise ratio (SNR), weighted SRR- and SNR-values were derived. The weighting was done by the same factors as in the AI-method (Table I), depending on the importance of every octave band for the speech intelligibility. Then the AL_{cons}-value was determined from Eq. (11) and Eq. (13).

Before applying the MTF-method, in our case with the multiple sources (n), the ratio Q/d^2 had to be calculated in the general expression for calculation of the m(F) (Houtgast & Steeneken, 1980, Appendix 2) by

$$Q/d^2 = \sum_{i}^{n} Q_i \ / \ d_i^2 \quad . \tag{18}$$

The SNR-values, defined by Eq. (12), were 20 dB and 10 dB, respectively, in the MTF-calculations. The directivity factor of the listener (Q_1=1.5), reflecting the binaural enhancement of the direct sound field in relation to the reverberant sound field was used in the calculations by the MTF-method. For the MTF-calculations no indication about the input speech spectrum was given. In our calculations we have used both the wide-band noise spectrum (M1) and the speech long-time-average spectrum (M2). There will be a difference in the spectral balance for the same SNR-value, and the MTF-predictions will differ.

For the waiting room at the gate (positions P and Q), the same procedure was used for calculation of the articulation loss of consonants as described for the departure hall. In this case there were 18 loudspeakers.

The diagrams in Fig. 7 show the comparison between the measured data and the predicted data of AL_{cons} by Peutz's method (P), by the SRR-method (S), and by the alternative MTF-methods (M1 and M2) for the four listening positions. The speech level was 70 dB(A) and the noise level was 50 dB(A).

In Fig. 8 the corresponding diagrams for the positions are shown when the speech level was 85 dB(A) and the noise level was 75 dB(A).

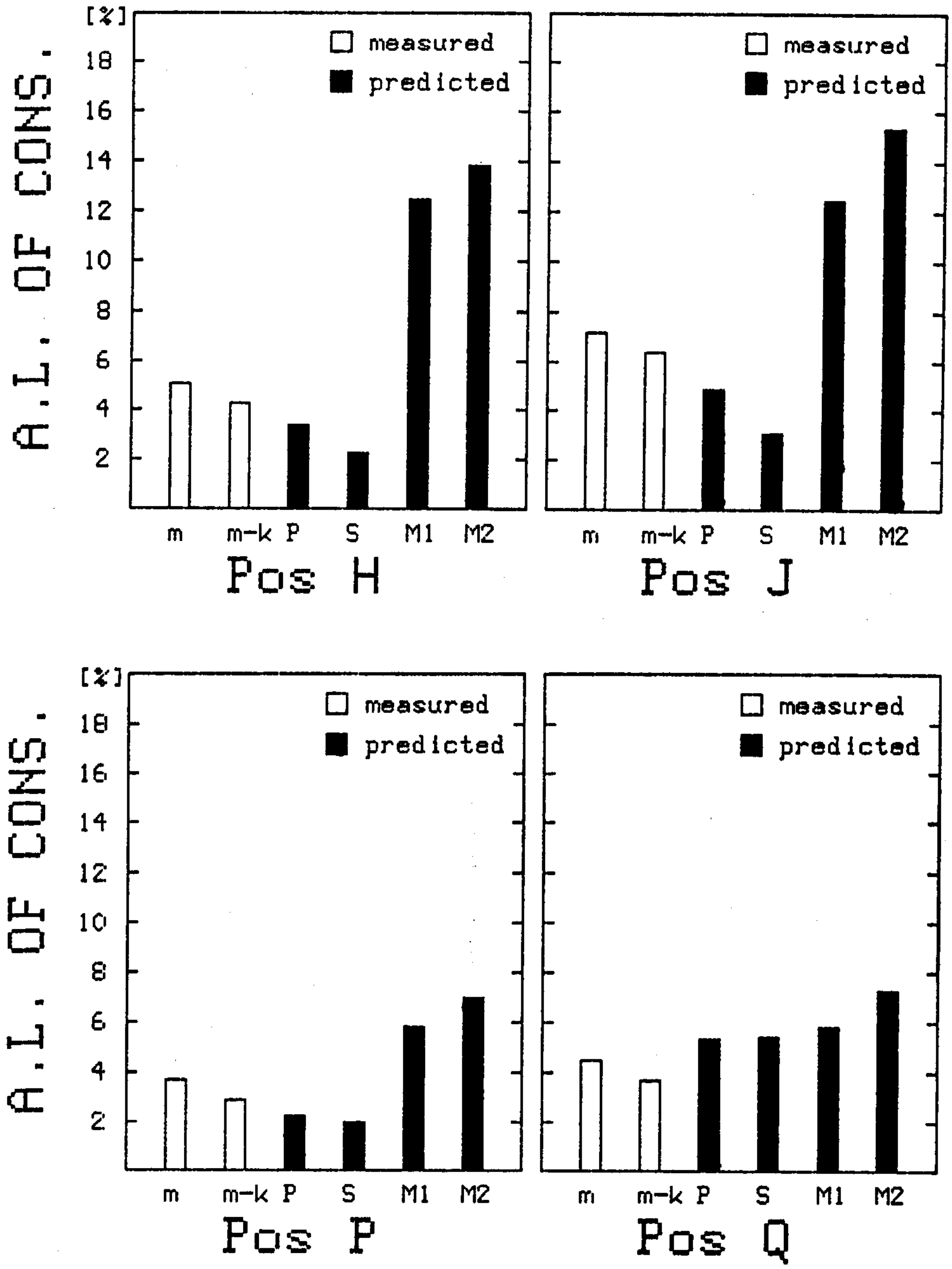

Fig. 7. Comparison between measured and predicted data of articulation loss of consonants in the different positions and at a speech level of 70 dB(A) and a noise level of 50 dB(A).
The columns represent:
m = measured articulation loss of consonants
m-k = as above but corrected by the reference (0.8%)
P = prediction by Peutz's method
S = prediction by the SRR-method
M1 = prediction by the MTF-method using wide-band noise spectrum as signal
M2 = prediction by the MFT-method using long-term-average speech spectrum as signal

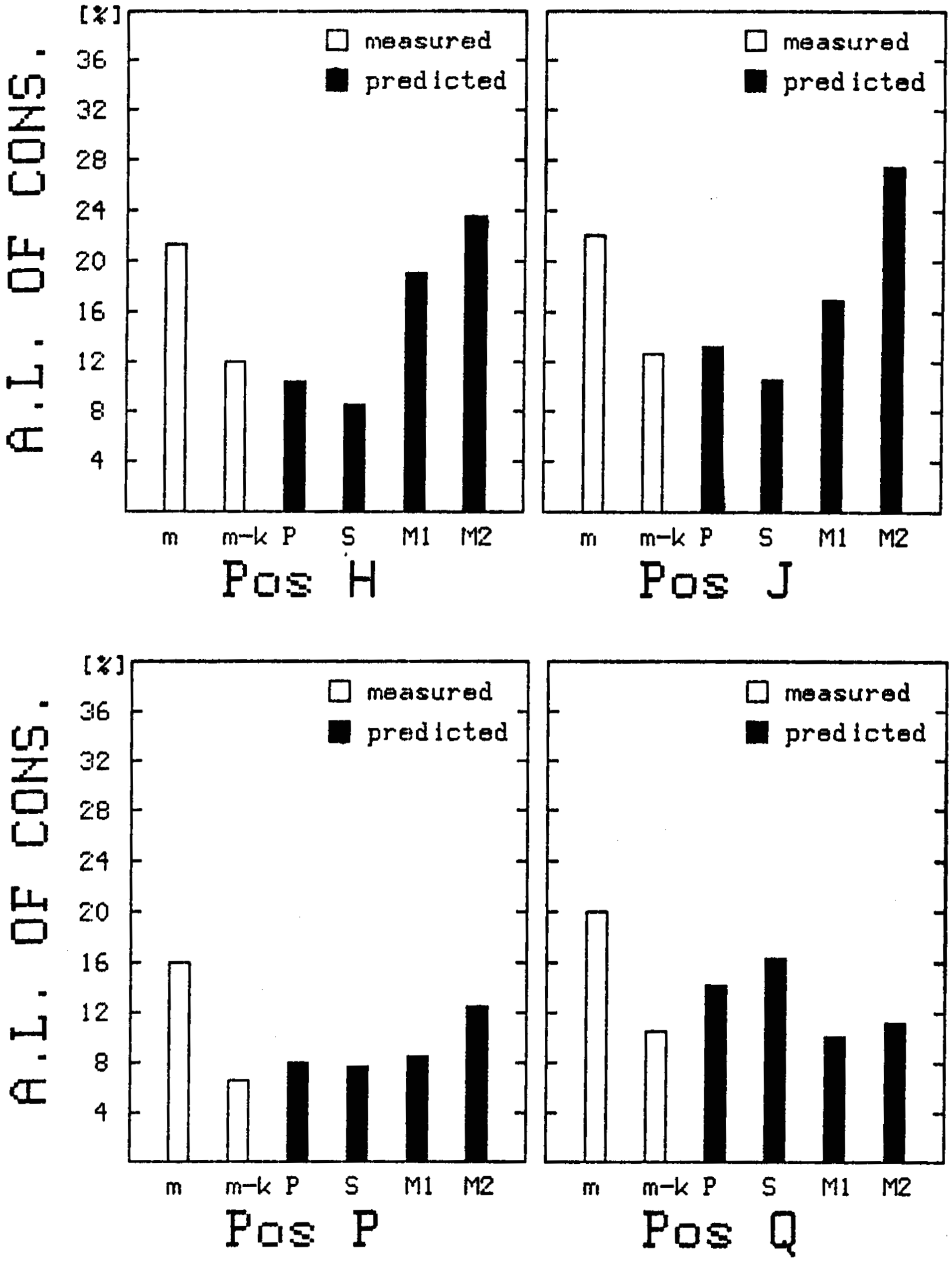

Fig. 8. Comparison between measured and predicted data of articulation loss of consonants in the different positions and at a speech level of 85 dB(A) and a noise level of 75 dB(A).
The columns represent:
m = measured articulation loss of consonants
m-k = as above but corrected by the reference (9.4%)
P = prediction by Peutz's method
S = prediction by the SRR-method
M1 = prediction by the MTF-method using wide-band noise spectrum as signal
M2 = prediction by the MTF method using long-term average speech spectrum as signal

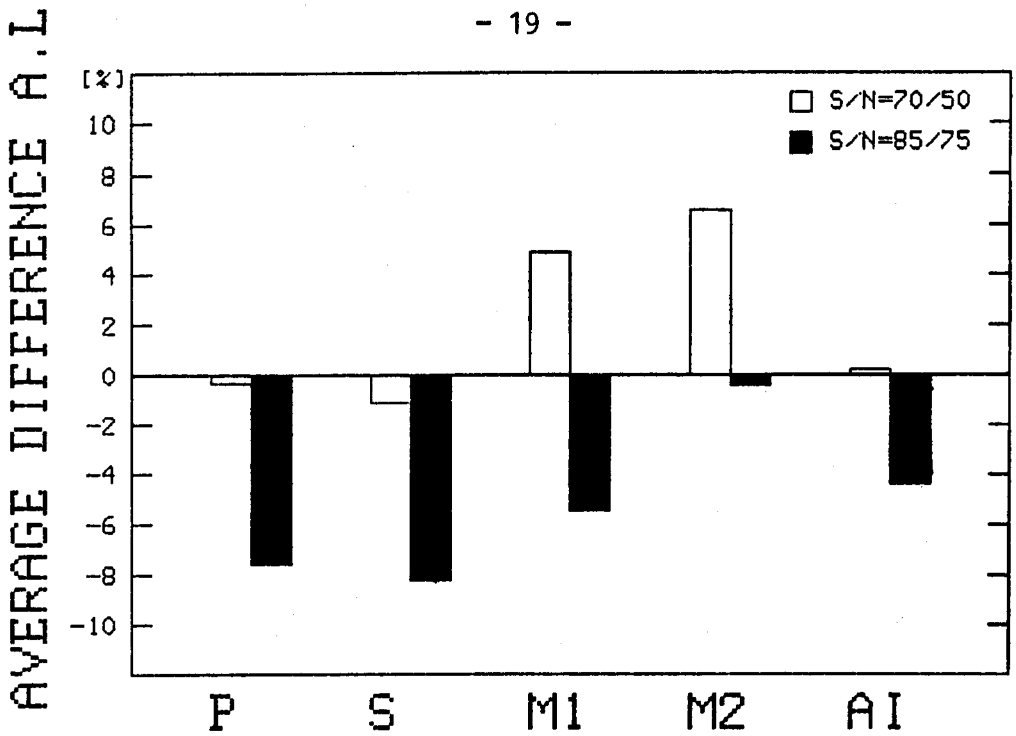

Fig. 9. Average differences for the prediction methods in the study. For symbol key, see Fig. 8.

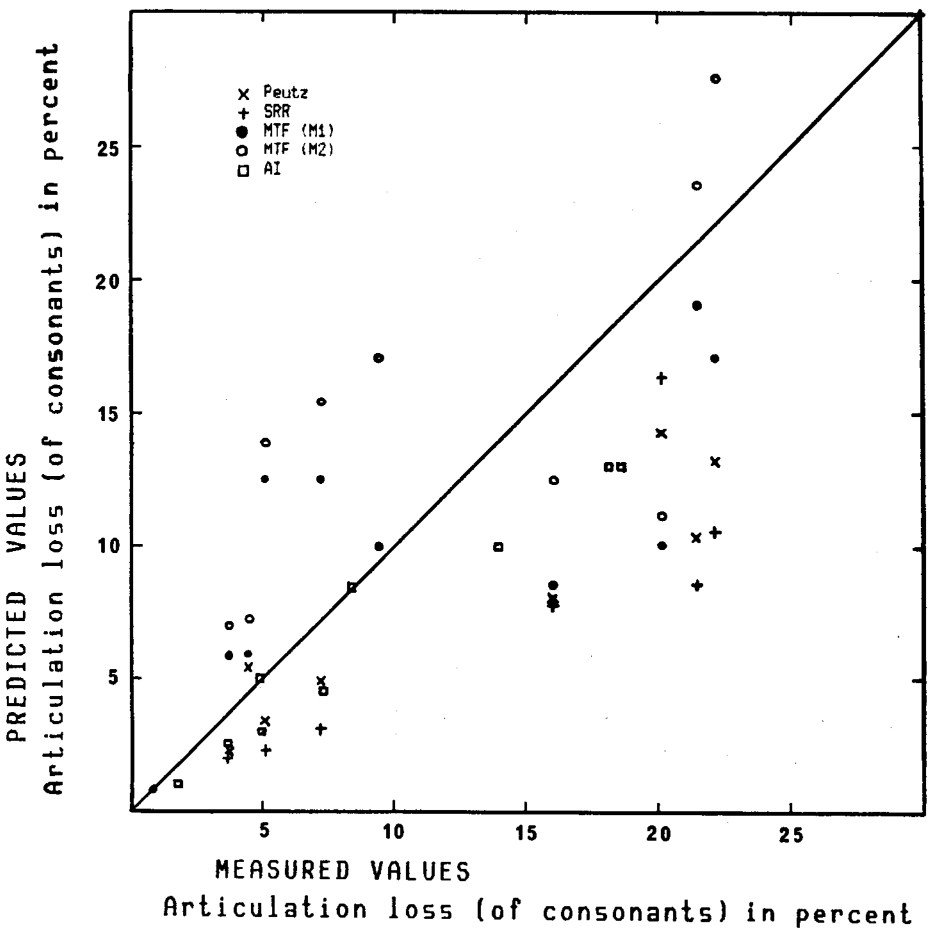

Fig. 10. Scattergram for a comparison of predicted and measured data.

The average differences between measured and predicted values for the methods are shown in Fig. 9 where the open columns represent the 70 dB(A) speech level situation and the filled columns represent the 85 dB(A) speech level situation.

Another way of representing the differences is in a scattergram. This is shown i Fig. 10 for all the comparisons.

In Appendix II the measured and predicted values are presented in tabular form.

5. DISCUSSION AND SUMMARY

At the situation when the speech level was 70 dB(A) and the noise level was 50 dB(A), the results between predicted and measured values show a fairly good agreement. However, the SRR-method predicts values that are closer to 2/3 of the measured ones, while the MTF-method predicts values that are nearly twice as high as the measured ones.

At the situation when the speech level was 85 dB(A) and the noise level was 75 dB(A), the measured values were 7% higher than the predicted ones as an average. In this case the MTF-prediction is the closest one in the departure hall. For the room with shorter reverberation time, we see a smaller spread between the methods.

The reference recording of the intelligibility test (R70) with neither reverberation nor noise gave an AL_{cons}-value of 0.8%. This value can be regarded as the correction a in Eqs. (3) and (4). It depends on the listeners' skills in the test group. However, Peutz has achieved higher correction values. For the reference case when SNR is 10 dB, we observe an AL_{cons}-value of 9.4% with noise but without reverberation. This is close to the predicted values from the AI-method or the MTF-method (with wide-band noise signal input). If we subtract these values from the measured articulation loss values, we see a closer agreement to the predicted values, especially for the Peutz's method and the SRR-method. In Figs. 7 and 8 there is a special column for this case (m-k). The question is, what combined effect do the noise and the reverberation have on the AL_{cons}-value?

The prediction by the AI-method gave results concerning the whole word, not only the consonants, and is not fully comparable to the predicted AL_{cons}-values. The correction term depending on the reverberation is for the departure hall 0.11 AI-units and for the gate room 0.06. For the lower speech level this prediction seems to be good, but in the case when the speech level is 85 dB(A), an additional correction of 0.08 for all the positions should give a more accurate result which is an increase of the correction term by 73%.

In spite of our doubt regarding the variation of the Q-value and the reverberation time, the results of Peutz's method seem to be fairly good. For prediction of situations with only interfering noise but without reverberation, Peutz's method is not applicable since it assumes a finite reverberation time.

The wide spectral representation of the suggested SRR-model does not show any great advantages in our test. However, the frequency response of the direct sound, the reverberant sound, and the ambient noise level are well defined. The weighting function for the different bands should be a subject for further studies. Another interesting subject is the determination of the reverberant sound. In the SRR-model we used the sum of the reverberant parts determined by the acoustical power and the room absorption. The separation into useful and detrimental reflections might be a way to extend the SRR-model. The appropriate factors of the formulae for a better agreement to measured data should be considered.

The difference between AI and MTF on one hand and SRR and Peutz's method on the other hand is that of introducing an index between the signal-to-noise calculations and the intelligibility. In the SRR- and Peutz's method the AL_{cons}-values are calculated directly. From the exponential relations in Eq. (11) and Eq. (13), we realize the sensivity to incorrect settings of SRR and SNR, which probably seems to be the reason for the deviation from the measured data.

The more complex MTF-method did not show significantly better accuracy in the prediction of the intelligibility in our test. However, the method is useful when both noise and reverberation interfere with speech. The prediction of the intelligibility for the reference (R85) was very good. In the calculation we have used two different input signals. If we utilize the long-term spectra of speech as input, the AL_{cons}-values will come out 27% greater as an average cmpared with a wide-band noise signal.

In a similar comparison of prediction methods, Smith (1981) reports the articulation index method to be most accurate up to a source-to-listener distance less than the critical distance, and for greater distances the signal-to-reverberation method to be the most accurate. He also recommends the articulation index method for distributed loudspeaker systems. We have also found a good agreement between the articulation index method and the measured values but with some underestimation of the articulation loss, especially when the noise influence is high.

In a recent study on the influence of loudspeaker directivity on the intelligibility (Jacob, 1985), some prediction methods were also compared. The result showed a scattering of data at the prediction by Peutz's method. In rooms with high reverberation, the deviation was high. Because of the short reverberation times in our study, we did not observe a high deviation.

The study by Jacob was performed in different auditoria. Using the signal-to-noise procedure based on the theory of Lochner & Burger (1961) he found an underestimation of the articulation loss of half of the measured value. In our calculation we do not have quite the same measure, but the signal-to-reverberation method will give a similar underestimation. However, we have regarded all the reverberant sound to be detrimental, while in the theory of Lochner and Burger, it is only the late reflections that influence the intelligibility. If only a fraction of the reverberant sound will be taken into account, a greater underestimation will be the result, in accordance with the result of Jacob.

Concerning the modulation transfer function, Jacob has found an overestimation compared to measured values. We have found similar results for our positions except for the room with short reverberation time at the high noise level.

One of our problems in the intelligibility test was to get a correct measure of the speech level, as described in Section 3. The setting of a correct value is of importance for the SNR and a few dB's difference will highly influence the test result.

When we carried out the recordings of the test tapes in the rooms with a very high performance, we were facing the question whether we really did measure the acoustics of the room or rather the limitations of the testing method. However, from the results it is obvious that the reverberation will increase the articulation loss. The recording method, with a dummy head and laboratory evaluation of test data by playback of the test material over headphones, with such a small test group as ten people seems to be a sensitive instrument for evaluation of high performance speech communication.

The acoustics of the departure hall are specific in the sense that there is a considerable amount of absorbing materials. A sound-wave propagating along the hall will be attenuated 2 dB more than in open space, when exceeding the distance of 4 r_r from the source. Thus, it must be considered if all of the 39 loudspeakers should be regarded as contributing to the reverberant sound. Some calculations based on the five or seven closest loudspeakers did not show values closer in agreement to the measurements in this case with the exception of the MTF-method.

The aim when constructing the rooms and optimizing the reinforcement system was to achieve an AL_{cons}-value not exceeding 10%. This should be valid for normal conditions. At the very high noise level of 75 dB(A) and the compensated speech level of 85 dB(A), the speech is still intelligible even if the AL_{cons}-value is inceased to 20%.

Without the automatic gain control (Fig. 4) the SNR-value at 75 dB(A) ambient noise should be -5 dB, and the AL_{cons}-value 55%-60%, a value too high for comprehension of speech. If we consider the AL_{cons}-value of 20% (SNR>10 dB) as the maximum for acceptable speech intelligibility, the limit for the background noise without any noise compensation should then be 60 dB(A). This is 15 dB lower than the noise level with the use of the compensation.

To summarize, in the study we have found some differences in the predicted values from the different methods compared to the measured values. The most complex predicting method is the MTF-method. The SRR-method suggested by us is a compromise between the AI- and Peutz's methods. However, we have found the simpler methods from Peutz and the articulation index calculations to be as good as the more complex methods in this study. All the methods underestimate the articulation loss for most of the cases.

We have observed a high quality of speech from the sound reinforcement system in these locations. The sounds which are frequently confused when both noise and reverberation influence the speech are mainly /v/, /b/, /p/, /m/, /n/, /f/, /j/, and /ʃ/.

6. ACKNOWLEDGMENT

The intelligibility test was financed by Ericsson Telemateriel AB, and performed by the staff of the Department of Technical Audiology at the Karolinska Institute in Stockholm. Among all the people that have been involved in this study I will mention Ann-Cathrine Lindblad and Håkan Sjögren, who made the test recordings, and Björn Hagerman who has analyzed the intelligibility data. I also want to thank Stellan Dahlstedt at formerly Akustik-Konsult AB, who made the reverberation time and the frequency response measurements. A special thank is given to my collegues at Ericsson, especially Tage Andersson and Claes von Rothstein.

References

A N S I (1969): "American national standard methods for the calculation of the articulation index", American National Standards Institute, Inc., 1430 Broadway, New York, N.Y. 10018, Jan., ANSI S3.5-1969.

Beranek L.L. (1947): "The design of speech communication system", Proc. of the I.R.E., Sept.

Fant, G. (1968): "Analysis and synthesis of speech processes", pp. 173-277 in (B. Malmberg, ed.), Manual of Phonetics. North-Holland Publ.Co., Amsterdam.

Fant, G. (1959): "Acoustic analysis and synthesis of speech with applications to Swedish", Ericsson Technics 1.

Fletcher, H. (1953): Speech and Hearing in Communication, R.E. Krieger Publ.Co., New York.

French, N.R. & Steinberg, J.C. (1947): "Factors governing the intelligibility of speech sounds", J.Acoust.Soc.Am. 19, p. 90.

Hagerman, B. & Lindblad, A-C. (1978): "Speech intelligibility of the loudspeaker system at Arlanda international terminal", Report TA 91, Dept. of Technical Audiology, the Karolinska Institute, Stockholm (in Swedish).

Houtgast, T. & Steeneken, H.J.M. (1973): "The modulation transfer function in room acoustics as a predictor of speech intelligibility", Acustica 28:1, p. 66.

Houtgast, T. & Steeneken, H.J.M. (1978): "Applications of the modulation transfer function in room acoustics", Report IZF 1978-20, Inst. for Perception TNO, Soesterberg, The Netherlands.

Houtgast, T., Steeneken, H.J.M., & Plomp, R. (1980): "Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics", Acustica 46, p. 60.

Jacob, K.D. (1985): "Subjective and predictive measures of speech intelligibility - The role of loudspeaker directivity", J.Audio Eng.Soc. 33, Dec., p. 950.

Klein, W. (1971): "Articulation loss of consonants as a basis for the design and judgment of sound reinforcement systems", J.Audio Eng.Soc. 19:11, Dec., p. 920.

Kleiner, M. (1976): "Problems in the design and use of 'dummy-heads'", Proc. of the Audio Eng.Soc. 75th Conv., March.

Knudsen, V.O. & Harris, C.M. (1950): Acoustical Designing in Architecture, John Wiley & Sons, Inc., New York.

Kryter, K.D. (1962): "Methods for the calculation and use of the articulation index", J.Acoust.Soc.Am. 34:11, Nov., p. 1689.

Kuttruff, H. (1973): Room Acoustics, Applied Science Publ. Ltd., London.

Latham, H.G. (1979): "The signal-to-noise ratio for speech intelligibility - an auditorium acoustics design index", Appl.Acoust. 12, July, p. 253.

Lochner, J.P.A. & Burger, J.F. (1961): "The intelligibility of speech under reverberant conditions", Acustica 11:4, p. 195.

Lundin, F.J. (1982): "The influence of room reverberation on speech - an acoustical study of speech in a room", STL-QPSR 2-3/1982, p. 24.

Lundin, F.J. (1983): "Modeling the intelligibility of a central loudspeaker cluster in a reverberant area", Proc. Audio Eng.Soc. 74th Convention, New York, Oct.

Peutz, V.M.A. (1971): "Articulation loss of consonants as a criterion for speech transmission in a room", J.Audio Eng.Soc. 19:11, Dec. p. 915.

Smith, H.G. (1981): "Acoustic design considerations for speech intelligibility", J. Audio Eng. Soc. 29:6, p. 408.

Steeneken, H.J.M. & Houtgast, T. (1980): "A physical method for measuring speech-transmission quality", J.Acoust.Soc.Am. 67:1, p.318.

Steeneken, H.J.M. & Houtgast, T. (1985): "RASTI: A tool foor evaluating auditoria", Brüel & Kjaer Tech. Rev. 3.

INITIAL CONSONANT CONFUSION MATRIX

GROUP DATA
KEY= 75

50 LISTS

COLUMNS ARE ORIGINAL, ROWS ARE ANSWERS
S/N = 85/75

	V	B	P	T	D	G	K	Π	F	S	SJ	TJ	J	M	N	NG	L	R		?
V	177	46	12	.	2	4	.	2	17	.	.	.	.	12	2	.	2	3	2	.
B	18	150	7	.	3	.	.	1	5	.	.	.	.	2	.	.	.	.	.	.
P	1	2	160	9	.	.	2	.	14	1	.	.	.	1	1	.	.	.	.	.
T	.	.	1	255	6	.	3	.	1	.	.	.	.	.	.	.	.	.	.	.
D	.	.	.	.	102	.	.	.	.	.	.	.	.	.	1	.	.	.	1	.
G	.	4	.	.	.	110	9	.	.	.	.	.	.	1	.	.	1	.	.	.
K	2	.	3	.	1	8	289	1	1	.	.	.	.	.	1	.	.	.	.	.
Π	1	.	.	.	.	.	1	133	1	.	.	.	.	.	.	.	.	.	1	.
F	.	.	3	1	.	.	.	.	142	3	.	.	.	.	.	.	.	.	.	.
S	.	.	.	.	.	.	.	.	.	466	.	.	.	.	.	.	.	.	.	.
SJ	.	.	.	.	.	.	.	.	.	1	50	.	.	.	.	.	.	.	.	.
TJ	.	.	.	.	.	.	.	.	.	.	6	49	.	.	.	.	.	.	.	.
J	.	.	.	.	.	.	.	2	.	.	.	.	104	.	.	.	3	.	.	.
M	7	4	2	.	.	.	.	.	.	.	.	.	.	101	8	.	2	.	1	.
N	.	.	.	1	.	.	.	.	.	.	.	.	.	11	135	.	3	.	2	.
NG	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
L	.	.	.	.	.	.	.	.	.	.	.	.	2	.	7	.	384	1	1	.
R	1	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	6	365	.	.
	2	5	1	2	.	.	.	.	.	1	.	.	1	.	1	.	6	.	.	.
?	2	3	3	.	1	3	3	2	5	3	.	.	.	1	3	.	4	1	.	.
	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

FINAL CONSONANT CONFUSION MATRIX

GROUP DATA
KEY= 50 40 LISTS

COLUMNS ARE ORIGINAL, ROWS ARE ANSWERS
S/N = 70/50

	V	B	P	T	D	G	K	H	F	S	SJ	TJ	J	M	N	NG	L	R		?
V	49	3	2	.	.	1	.	.	5	.	.	.	.	6	3	.	1	.	12	.
B	2	12	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
P	.	.	119	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.
T	.	.	.	363	4	.	.	.	.	1	.	.	.	.	.	.	.	.	2	.
D	.	.	.	5	160	.	1	.	.	.	.	.	.	.	.	.	.	2	2	.
G	.	.	.	.	.	104	7	.	.	.	.	.	.	.	.	2	.	1	.	.
K	.	.	.	.	.	.	323	.	.	.	.	.	.	.	.	.	.	.	.	.
H	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F	.	.	2	.	.	.	1	.	48	2	.	.	.	.	.	.	.	.	.	.
S	.	.	.	.	.	.	.	.	.	335	1	.	.	.	.	.	.	.	.	.
SJ	.	.	.	1	.	.	.	.	.	.	32	.	.	.	.	.	.	.	.	.
TJ	.	.	.	2	.	.	.	.	.	.	6	.	.	.	.	.	.	.	.	.
J	.	.	.	.	.	.	.	.	.	.	.	.	59	.	.	.	1	.	.	.
M	.	.	1	.	.	.	.	.	.	.	.	.	.	89	1	.	.	.	.	.
N	.	.	.	1	.	.	.	.	.	.	.	.	.	16	249	.	.	.	1	.
NG	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	74	.	.	.	.
L	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	271	3	1	.
R	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	291	1	.
	5	.	3	3	.	.	4	.	.	3	.	.	4	5	2	.	1	3	.	.
?	4	.	1	1	.	1	.	.	1	2	.	.	.	.	1	.	.	.	.	.

VOWEL CONFUSION MATRIX

GROUP DATA
KEY= 75   50 LISTS

COLUMNS ARE ORIGINAL, ROWS ARE ANSWERS
S/N = 85/75

	A	O	U	A	E	I	Y	Ä	Ü	A:	O:	U:	A:	E:	I:	Y:	Ä:	Ü:		?
A	377	.	.	.	.	.	.	.	.	6	.	.	.	.	.	.	.	.	.	.
O	.	39	3	14	.	.	.	.	.	.	11	.	2	.	.	.	.	.	.	.
U	.	9	153	17	.	.	.	.	7	.	5	3	1	.		.	.	.	.	.
A	.	5	3	168	.	.	.	.	1	.	1	.	.	.	.	.	.	.	.	.
E	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
I	.	.	.	.	.	212	3	13	.	.	.	.	.	4	1	.	.	.	.	.
Y	.	.	1	.	.	3	71	7	3	.	.	4	.	.	.	1	.	2	.	.
Ä	.	.	.	.	.	1	.	306	5	.	.	.	.	2	.	.	5	.	.	.
Ü	1	.	4	.	.	.	.	3	67	.	.	.	.	.	.	.	.	2	.	.
A:	2	.	.	4	.	.	.	.	.	144	.	.	.	.	.	.	.	.	.	.
O:	.	20	5	.	.	.	.	.	.	.	83	.	8	.	.	.	.	.	.	.
U:	.	.	2	.	.	3	3	3	2	.	2	103	1	.	.	.	.	2	.	.
Ä:	.	2	1	4	.	.	.	.	.	.	4	.	139	.	.	.	.	.	.	.
E:	.	.	.	.	.	2	.	1	.	.	.	.	.	65	.	.	21	.	.	.
I:	.	.	.	.	.	3	.	2	.	.	.	.	.	1	120	3	.	.	.	.
Y:	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	31	.	.	.	.
Ä:	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	63	.	.	.
Ü:	.	.	1	1	.	.	.	.	2	.	.	.	.	.	.	.	.	87	.	.
	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
?	.	.	.	1	.	1	.	3	.	1	.	.	.	.	.	.	.	.	.	.

FINAL CONSONANT CONFUSION MATRIX

GROUP DATA
KEY= 75 50 LISTS

COLUMNS ARE ORIGINAL, ROWS ARE ANSWERS
S/N = 85/75

	V	B	P	T	D	G	K	H	F	S	SJ	TJ	J	M	N	NG	L	R		?
V	40	6	2	.	.	.	.	.	7	.	.	.	.	13	13	1	3	1	19	.
B	2	9	11	.	.	.	.	.	.	.	.	.	.	2	.	.	1	1	.	.
P	1	.	117	2	2	.	3	.	6	.	.	.	.	3		1	.	4	2	.
T	.	.	3	440	8	.	.	.	.	6	.	.	.	.	1	.	.	1	5	.
D	.	4	3	15	194	1	1	.	.	.	.	.	1	1	11	.	.	3	3	.
G	1	.	1	.	1	111	25	.	.	.	.	.	.	1	2	6	.	4	.	.
K	.	.	.	.	.	13	374	.	5	.	.	.	.	.	.	.	.	2	.	.
H	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
F	.	.	3	.	.	.	5	.	41	3	.	.	.	1	1	.	1	1	3	.
S	.	.	.	2	.	.	.	.	.	393	.	.	.	.	.	.	.	.	.	.
SJ	.	.	.	1	.	.	.	.	.	2	47	.	.	.		.	.	3	.	.
TJ	.	.	.	1	.	.	.	.	.	.	4	.	.	.	.	.	.	.	.	.
J	.	.	.	.	1	.	.	.	.	.	.	.	61	1	8	1	3	.	7	.
M	2	.	2	.	.	.	.	.	.	.	.	.	.	68	10	2	2	.	5	.
N	3	.	2	.	1	.	.	.	1	.	.	.	1	25	240	8	3	1	3	.
NG	.	.	.	.	.	.	.	.	.	.	.	.	.	2	1	60	.	.	.	.
L	2	.	1	.	.	2	1	.	.	.	.	.	3	7	14	8	293	13	3	.
R	.	1	1	.	.	.	.	.	.	1	.	.	1	1	1	2	9	324	2	.
	15	.	7	7	4	.	12	.	5	5	.	.	13	13	17	4	20	12	.	.
?	7	.	5	4	.	1	5	.	2	4	.	.	2	7	7	2	6	8	.	.

	Articulation loss of words		Articulation loss of consonants					
POS S/N	TEST	AI	TEST	TEST -REF	Peutz	SRR	MTF M1	MTF M2
R70/00	1.7	1.0	0.8	0.0			0.8	0.8
P70/50	3.7	2.5	3.7	2.9	2.3	2.0	5.9	7.0
Q70/50	4.9	3.0	4.5	3.7	5.4	5.5	5.9	7.3
H70/50	4.9	5.0	5.1	4.3	3.4	2.3	12.5	13.9
J70/50	7.3	4.5	7.2	6.4	4.9	3.1	12.5	15.4
R85/75	8.4	8.5	9.4	0.0			10.0	17.1
P85/75	13.9	10.0	16.0	6.6	8.1	7.8	8.6	12.5
Q85/75	18.1	13.0	20.1	10.7	14.3	16.4	10.1	11.2
H85/75	18.1	13.0	21.4	12.0	10.4	8.6	19.1	23.6
J85/75	18.6	13.0	22.1	12.7	13.3	10.6	17.1	27.6

Comparison between predicted and measured data (TEST).

POSITION	S/N	HELORDSRÄTT	INITIAL	VOKAL	FINAL	MEDEL	KONSMED		
(R) REF.INSP.	70/00	95.00	99.40	96.40	99.00	98.27	99.20	MEAN	N=10
		4.74	1.35	3.10	2.54	1.76	1.32	ST.DEV	
(P) PIR BÄTTRE	70/50	89.40	97.80	96.20	94.80	96.27	96.30	MEAN	N=10
		3.78	2.20	2.90	2.70	1.45	2.11	ST.DEV	
(Q) PIR SÄMRE	70/50	86.00	98.40	94.40	92.60	95.13	95.50	MEAN	N=10
		6.99	1.84	6.45	3.89	2.52	2.17	ST.DEV	
(H) HALL BÄTTRE	70/50	87.00	96.60	95.40	93.20	95.07	94.90	MEAN	N=10
		5.19	2.32	2.50	3.43	1.81	2.33	ST.DEV	
(J) HALL SÄMRE	70/50	81.20	94.40	92.40	91.20	92.67	92.80	MEAN	N=10
		8.95	2.46	7.17	3.55	3.68	2.25	ST.DEV	
(R) REF.INSP.	85/75	77.80	91.60	93.60	89.60	91.60	90.60	MEAN	N=10
		6.76	2.46	4.30	3.75	2.20	2.41	ST.DEV	
(P) PIR BÄTTRE	85/75	66.00	85.60	90.20	82.40	86.07	84.00	MEAN	N=10
		12.51	8.78	4.94	6.65	5.37	7.50	ST.DEV	
(Q) PIR SÄMRE	85/75	55.00	85.80	86.00	74.00	81.93	79.90	MEAN	N=10
		10.03	6.70	5.42	6.80	4.27	5.93	ST.DEV	
(H) HALL BÄTTRE	85/75	57.20	86.40	88.40	70.80	81.87	78.60	MEAN	N=10
		14.40	6.98	8.10	10.25	7.03	7.07	ST.DEV	
(J) HALL SÄMRE	85/75	56.20	83.80	88.40	72.00	81.40	77.90	MEAN	N=10
		8.66	5.12	5.15	7.54	4.12	5.53	ST.DEV	

Result of the intelligibility test at different positions and signal levels. The result is an average of the ten test subjects.