TECHNICAL AUDIOLOGY

Report TA No. 101
September 1981

SENTENCES FOR TESTING SPEECH INTELLIGIBILITY IN NOISE

Björn Hagerman

Errata

In Report TA No. 101 "Sentences for testing speech intelligibility in noise".

Abstract, line 13, 22 should read 25.

Page 6, lines 26 and 29, 22.1 should read 25.3.

Page 5, lines 22 and 23, 6.2 should read 5.7.

Page 5, line 22, 60 should read 40.

Karolinska Institutet
Technical Audiology
KTH, S-100 44 Stockholm
Sweden

Report TA101
September 1981

SENTENCES FOR TESTING SPEECH INTELLIGIBILITY IN NOISE

Björn Hagerman

ABSTRACT

A list of ten spoken Swedish sentences were computer edited to achieve new lists with exactly the same content of sound, but with new sentences. A noise was synthesized from the speech material by the computer to get exactly the same spectrum of the speech and the noise. The noise was furthermore amplitude modulated by a low frequency noise to make it sound more natural. This material was tested monaurally on 20 normal hearing subjects. The equality in intelligibility of some of the lists was investigated. Repeated threshold measurements in noise showed a standard deviation of 0.44 dB when the learning effect was outbalanced. Only a small part of the learning effect was due to learning of the word material. Intelligibility curves fitted to the data points in noise and without noise showed maximum steepnesses of 22 and 10 %/dB respectively. At constant signal to noise ratio (S/N) the best performance was achieved at a speech level of 53 dB.

This work was supported by the National Swedish Board for Technical Development.

Table of Contents

1 INTRODUCTION

Hearing impaired people (with or without hearing aid) often complain about the difficulty of understanding speech in a noisy environment (Plomp 1978). This is shown also by speech intelligibility measurements of hearing impaired subjects for different kind of noises (Aniansson, 1974) and for reverberation (Duquesnoy & Plomp, 1980). Routinely made clinical measurements of this ability, however, is still not very common. In Sweden the main reason probably is the lack of standardized methods and speech materials to make fast and reliable clinical measurements of the speech intelligibility in noise. Some other countries, however, are better off in this respect (Plomp & Mimpen, 1979, and Niemeyer, 1976).

2 DEVELOPMENT OF THE SPEECH MATERIAL

2.1 General requirements

The main purpose of this work was to develop a speech material suitable for clinical measurements of speech intelligibility in noise. It was intended to be used especially for hearing aid evaluation ("free-field" measurement), but also to measure the patient´s speech intelligibility in noise with earphones before the hearing aid is fitted. This should be a way to predict the possible benefit of a hearing aid (Plomp 1978).

For clinical applications it is necessary to have many different lists of equal difficulty. It is also very important that the test is fast and still gives a reliable result. This requires many test items per time unit. Thus it was decided to use sentences with each word as an item to score, which would also be more close to reality than a single word test. The ability to understand speech in noise can be measured as the intelligibility score for a certain signal to noise ratio (S/N) or as a threshold value of the S/N for a certain intelligibility, for example 50 %. The threshold method was chosen here, which requires a steep slope of the intelligibility curve to get a reliable measure of the threshold value.

The noise should have the same spectral content as the speech signal. With different spectral contents in the speech and in the noise, some frequencies would get more weight than others which is undesirable (Plomp & Mimpen, 1979). A steeper slope of the intelligibility curve is probably also achieved with equal spectral content in the speech and in the noise.

2.2 Speech material

Since our laboratory have access to a computer with very good capability to edit and mix digitally recorded sound, a new way was tried to get several different lists out of only one recorded list. In this way all lists would get exactly the same content of sounds, which would increase the probability that the different lists should be of equal difficulty for any patient. Ordinary test lists with different recordings and/or words can only be shown to be equal for groups of patients.

Ten sentences of five words each were chosen to constitute the original list shown in Table I. A fairly good phonetic balance was achieved. The semantic content of some of the sentences may be puzzling however. This was supposed to cause no problems after having heard the training list. The original list was recorded in a non-reverberant room with a Nagra tape recorder. It was read by a female speaker who tried to avoid transitions between the words. This material was recorded digitally on the computer and each word was cut out. Since all the sentences have identical structure new lists could be made by choosing the words randomly in each column from the original list. This was done by the computer, which also automatically mixed them together to form new lists. When listening to the new lists it was found that some words were not ideally cut. Some of them needed more or less silence in the beginning or ending to make the sentences sound more natural. After recutting, some lists were generated again and checked. This procedure was repeated seven times. After this the generated lists were considered good enough.

2.3 Noise

The noise signal was made by the computer using the sound of the original word list. A program for periodic filtering was used to make seven periodic noises with periodicities between 10 and 30 Hz which were mixed together to a noise without noticable periodicity but with exactly the same spectral content as the word lists (Fig. 1).

To get away from the steady state noise and make it sound somewhat more like coctail-party noise, it was also slightly amplitude modulated with a low-frequency signal consisting of a noise filtered around 2.1 Hz. The frequency curve of the shaping filter used is shown in Fig. 2.

2.4 Recording of master tape

The calibration tone was set to be the same as for the Swedish spondee-lists recorded at our department in 1966. This means that the calibration tone was set to be 3.4 dB lower than the level not exceeded 90 % of the time measured with 125 ms integration time and with the pauses between the sentences omitted. The 90 % level of the noise was 9 dB lower than for the speech but the peak sample value was about the same for the speech and for the noise. The same calibration tone was thus used for both. The speech and the noise levels mentioned later on refer to the level of this calibration tone measured on an IEC 303 coupler.

13 lists of sentences (shown in the Appendix) and 13 replicas of the noise were recorded on each channel of a stereo tape recorder (Revox B77). Exactly the same noise signal was thus used for each list. The pause between the sentences were 7 seconds and the total time of each list was 110 seconds. The calibration tone was recorded in the beginning of each channel with a recording level of -2 dB.

2.5 Pilot study

A pilot study was performed on 6 normal hearing subjects to form the details in the experimental design. As the slope of the intelligibility curve was considered to be too flat (about 11 %/dB) the level of the words were changed to equalize their difficulty. The desirable changes were found to have a range of ±4 dB. In order not to get the speech sound unnatural, however, no word was changed more than ±1.3 dB. After this a new master tape was made in the same way as before.

3 LISTENING TESTS ON NORMAL HEARING SUBJECTS

To achieve normative data of the speech material a listening test was performed on normal hearing subjects. The following issues were considered most interesting to study:

1. Homogeneity of the lists.

2. Reliability.

3. Learning effects.

4. Intelligibility curves

5. Influence of the speech level with constant S/N ratio.

3.1 Methods

The age of the 20 normal hearing subjects ranged between 18 and 32 years with a mean value of 26 years. The tone thresholds did not exceed 20 dB at any frequency between 125 and 8000 Hz, except of two subjects at 6 and 8 kHz respectively. The subjects were divided into two groups for a counterbalanced design of the experiment.

The tape used in the experiment was a copy of the master tape run at 19 cm/s. The signals from the Revox B77 tape recorder were fed through two Hewlett & Packard attenuators which were used to set the S/N ratio. The test equipment, also including our permanent psycho-acoustical measuring device with mixer, amplifier and attenuator, was calibrated before the experiment and checked in the middle and after the whole experiment. The type of earphone used was TDH39 with a MX41AR cushion. The subjects were tested monaurally one at a time, using the right ear, repeating verbally the percieved words of the sentence to the experimentor, who marked the result. Before the test the subject read the following instruction:

> We want to investigate how difficult it is to percieve speech in background noise. You will now listen to sentences consisting of 5 words.
>
> Example: "Bertil got eight white socks"
>
> Depending on the background noise it is sometimes easy and some-

times difficult to percieve the words. Please repeat clearly the words you have percieved. You may well guess, but please do not hesitate as long with your answer, as to miss the beginning of the next sentence. Don´t say anything about the words you cannot percieve.´

The experiment was divided into two sessions, run on separate days with an interval ranging from one day to a couple of weeks. The first session were divided into two parts, denoted A and B. It took about 40 minutes including a complete tone-audiogram. The second session were divided into four parts, denoted C,D,E and F. It took about 55 minutes including a 5 minutes intermission. Table II shows a summary of the experimental design. Part A,C and D started with a training list. The level was varied after each sentence during this list to get a rough idea of the threshold. The purpose of the thresholds measured in part A and F was to estimate the reliability and the learning effect. These thresholds were all measured in a descending way, the 50 % level interpolated from two points at each side of this level. The purpose of part B was to check the homogeneity of the lists. Since the lists all had exactly the same content of sound, it was strongly suggested that the homogeneity of the lists was guaranteed. Only five lists were thus checked in this respect. The intelligibility curves in part C were all measured in an ascending way. The corresponding curves without noise measured in part D, however, were achieved with the levels more randomly ordered, since the positions of the curves were not known good enough from the training list, to know which levels to choose from the beginning.

There were less problems than expected to perform the test. Some of the subjects mentioned that for some of the sentences they could not understand the meaning. Two of the subjects showed significantly longer response time and needed almost the whole pause between the sentences to answer.

3.2 Results

3.2.1 Homogeneity of the lists

Table III shows the mean intelligibility score of both groups together for each of the lists 1-5. Since five lists are already presented in part A, errors due to learning effects can be linearized over the lists in part B. Provided also that the learning effect is the same for both groups this effect is outbalanced and will thus not affect the result. Unexpectedly list No. 1 was shown to be significantly more difficult than the other lists. The long time average levels of the speech and the noise were therefore rechecked on the tape used. The level difference between list No. 1 and list No. 2 was less than 0.1 dB, both regarding the speech and the noise, which is too small to explain the difference between the mean scores. Studying the result of each word separately, however, it was found that the three most difficult sentences in list No. 1 were all placed in the beginning. This might have reduced the subject´s motivation and thus decreased the total result of this list. No one of the other eleven lists showed a bad pattern in this respect and they are considered to be of equal difficulty. It was also noted that the result of a certain word depended on the actual list. The reason for this is probably that the words are disturbed more or less depending on the actual value of the noise. Of course the result of a certain word may also depend on the words in the neighbourhood to a certain degree.

3.2.2 Discrimination reliability

From part A and part F it is possible to get information about the reliability in the discrimination measure. Part F was made in exactly the same way as the last threshold in part A, regarding the lists chosen and the S/N ratio. The differences between the corresponding discrimination scores were calculated for each subject. The mean and the standard deviation of these differences were 16.4 and 8.2 % respectively (N=40), showing the learning effect to be 16.4 %. Since the standard deviation of the difference is 8.2 % the standard deviation of the discrimination itself will be $8.2/\sqrt{2}$ %= 5.8 %. Identical lists were used here for the discrimination scores compared. Note however, that this standard deviation also includes the variability in the learning effect among subjects. From part B it is possible to get a corresponding value but with the learning effect outbalanced. In this case, however, different lists are used, and the corresponding standard deviation is then 6.2 % when the results of list No. 1 are omitted (N=60). The standard deviations 5.8 and 6.2 % estimated are in good agreement with the corresponding value of a 50-word list of single words (Hagerman, 1976), which should be between 5.3 and 6.2 % for discriminations between 25 and 75 %, perhaps somewhat higher for the sentences however, since the difficulties of the different words should be more equal in this case.

This means that the result of a certain word is not very much affected by the other words in the neighbourhood, since such a condition would give a higher value of the standard deviation. The influence by linguistic factors then must be small, which is considered as desirable for the intended field of application for this speech material. This was expected since all sentences have identical structure, which will very soon be learned by the subject tested.

3.2.3 Threshold reliability

Results of the threshold mesurements in part A and F are shown in Table IV. Threshold means and standard deviations from the intelligibility curves in part C are added here. To estimate the reliability of the threshold measurements the standard deviations of the differences are shown to the right in this table. From these three values the standard deviation of repeated measurements of a threshold in noise was calculated to be 0.44 dB, including the variability in the learning effect, but with the mean of the learning effect outbalanced. These thresholds were based on two lists, which take 3 minutes and 45 seconds together.

3.2.4 Learning effects

The mean thresholds in Table IV are plotted in Fig. 3, which shows the learning effect. This result will be discussed in detail later on.

3.2.5 Intelligibility curves

Part C and D were intended to give intelligibility curves in noise and without noise. Fig. 4 and 5 show the intelligibility scores plotted against the S/N ratio and the speech level respectively. Note that these levels are related to the actual threshold of each subject, which makes the picture more clear. The zero point of the X-axis thus corresponds to the mean threshold which is -8.15 dB S/N in noise (shown in Table IV) and 21.6 dB SPL without noise. The standard deviation of the threshold without noise was 2.49 dB. For speech audiometry the calibration tone is standardized to be 22 dB, which agrees very well with the achieved threshold value of 21.6 dB.

The curves were fitted on a least square basis and they consist of two exponential functions:

Y=-R+Q*(100+R+S)/(P+Q)*exp(P*(X-T)) when X<T

and (1)

Y=100+S-(50+S)*exp(-Q*X) when X>T where

T=-1/Q*ln(P*(100+R+S)/(50+S)/(P+Q))

The constants R and S show the deviation of the horizontal asymptotes from 0 and 100 % respectively. The maximum steepness of the two curves in Fig. 4 and 5 is 22.1 and 9.6 %/dB respectively, and the reason that the slope is much steeper with noise is probably that the spectra of the speech and the noise are identical. The slope values shown in Table IV differ from 22.1 due to the fact that they are estimated from two points on the curve on each side of the 50 % intercept and some of them are also achieved from a descending threshold method, while the intelligibility curve in noise was achieved from an ascending method.

3.2.6 Degree of difficulty of the words

Table V and VI show the mean intelligibility score of each word in noise and without noise. The tables differ considerably. According to Plomp (1978) the threshold without external noise can be interpreted as if it was determined by an internal noise in the ear. The reason that the tables differ may be that the spectral shape of the internal noise is quite different from the spectral shape of the noise used in the test. Note also that the last word has got a higher score than the first two words without noise, in contrast to the condition with noise. The reason for this is not clearly understood.

3.2.7 Influence of the speech level

In part E the influence of the absolute level of the speech on the intelligibility at constant S/N ratio was investigated. Fig. 6 shows the result. It is obvious that the discrimination value depends on the speech level chosen. Since the slope of the intelligibility curve is very steep, however, the S/N threshold will not depend very much on the speech level. For hearing impaired people the speech level often should have a greater influence though. Note that the maximum performance was found at a considerably lower level than the normal conversational speech level.

3.3 Discussion

The result of list No. 1 was significantly lower than the corresponding results of lists No. 2-5. In terms of threshold value, however, this difference is less than 0.5 dB. After all it was decided not to use this list in the future. It was also considered desirable to check the other lists in this respect.

The reliability of repeated measurements is so good that a threshold method based on only one list may be considered in the clinical application, providing that the spread of thresholds for different hearing impairments is high enough for this speech and noise material. For clinical use it is also important to consider the precision of the amplifiers and the attenuators in the audiometers if the speech and the noise are presented through different channels. When comparing results from two different hearing aids on a patient it is advisable to subtract the learning effect in order not to get false improvements.

No information was given to the subject during the test that all lists contained the same words. This fact, however, is probably rather soon realized by the subject. When all 50 words are known, any word could be guessed from the 10 words in the corresponding column, which will give a 10 % chance of guessing the correct word when it is not percieved. As the word content of the lists are learned the discrimination score at constant conditions then may rise 10 % at very low discrimination levels. Guessing is however not utilized at very low levels in this experiment since no forced choice procedure was used. This is evident from the intelligibility curves in Fig. 4 and Fig. 5. As the words are learned they will get more predictable and this fact should give a steeper slope according to Kalikow & Stevens (1977). From Table IV it is clear that the slope rises from 13.5 to 15.2 %/dB during the whole experiment, which should give a 0.4 dB improvement of the threshold. Learning of the word material will then explain only a small part of the total learning effect. This part is probably based on central processing of the speech. The rest of the learning effect is supposed to be due to more peripheral processes based on information of the spectral nature of the speech and the noise. In fact the result here agrees very well with the result of Plomp & Mimpen (1979) who used different words in the different lists, which supports the suggestion that only a small part of the learning effect is due to learning of the word material. Note that the learning effect then may be different when the test is run without noise, since it probably depends on the information achieved about the relation between the spectral properties of the speech and the noise.

The interesting and perhaps unexpected result that the maximum intelligibility score at constant S/N ratio was achieved for speech levels lower than normal conversational level, is supported by Dirks, Morgan and Velde (1980) whose results show the same tendency for spondees and monosyllabic words.

4 CONCLUSIONS

The sentences and the noise reported on here constitute a test material, which makes it possible to perform rapid and reliable measurements of the speech intelligibility for different purposes. It may be regarded as a compromise between a single word test and a normal sentence test. It remains to be seen however, if the validity of the test material is good enough. That is to say, that it really measures what we want to measure. This means for example, that the real speech discrimination of an arbitrary patient for different hearing aids in his home should be highly correlated to the measured intelligibilities with the same hearing aids at the clinic.

5 ACKNOWLEDGEMENTS

The author wants to express his gratitude to Ann-Cathrine Lindblad, Björn Lindström and Åke Olofsson for valuable discussions and for check-reading the manuscript.

6 REFERENCES

Aniansson, G. (1974). Methods for assessing high frequency hearing loss in every-day listening situations. Acta otolaryngol, Suppl 320.

Dirks, D.D., Morgan, E.M. and Velde, T.M. (1980). Effects of interfering noises on speech recognition for the hearing impaired. Paper presented at the annual meeting of the American Speech-Language-Hearing Association in Detroit 1980.

Duquesnoy, A.J. & Plomp, R. (1980). Effect of reverberation and noise on the intelligibility of sentences in case of presbyacusis. J Acoust Soc Amer 68(2), 537.

Hagerman, B. (1976). Reliability in the determination of speech discrimination. Scand Audiol 5, 219.

Kalikow, D.N & Stevens, K.N. (1977). Development of a test of speech intelligibility in noise using sentence material with controlled word predictability. J Acoust Soc Amer 61(5), 1337.

Niemeyer, W. (1976). Speech audiometry and fitting of hearing aids in noise. Audiology 15, 421.

Plomp, R. (1978). Auditory handicap of hearing impairment and the limited benefit of hearing aids. J Acoust Soc Amer 63(2), 533.

Plomp, R. & Mimpen, A.M. (1979). Improving the reliability of testing the speech reception threshold for sentences. Audiology 18, 43.

7 APPENDIX

The appendix shows the different lists generated and tables of the random number used in this process.

TRÄNINGSLISTA

1. Britta flyttar åtta svarta ringar
2. Elsa gav sex nya vantar
3. Peter köpte sju ljusa skålar
4. Karin ägde fyra vackra knappar
5. Bosse visar tre lätta dukar
6. Anna höll två mörka korgar
7. Jonas lånar elva hela lådor
8. Svante tog arton gamla bollar
9. Gustav ser nio fina pennor
10. Märta har tolv stora mössor

LISTA 1

1. Britta tog åtta hela vantar
2. Karin har fyra nya ringar
3. Märta visar nio mörka dukar
4. Bosse köpte sju vackra bollar
5. Anna flyttar sex gamla korgar
6. Elsa ägde arton svarta knappar
7. Peter ser elva stora lådor
8. Jonas höll tre lätta pennor
9. Svante gav två fina skålar
10. Gustav lånar tolv ljusa mössor

LISTA 2

1. Elsa ägde nio fina pennor
2. Gustav lånar två gamla lådor
3. Bosse gav sju mörka bollar
4. Britta visar arton stora knappar
5. Peter flyttar åtta ljusa skålar
6. Svante tog tolv hela dukar
7. Märta har elva lätta korgar
8. Karin höll fyra vackra mössor
9. Jonas köpte sex nya vantar
10. Anna ser tre svarta ringar

LISTA 3

1. Gustav ser tolv svarta skålar
2. Jonas flyttar sju ljusa lådor
3. Märta köpte fyra hela mössor
4. Elsa lånar tre mörka vantar
5. Karin visar åtta gamla dukar
6. Bosse ägde sex vackra ringar
7. Peter gav två stora bollar
8. Britta tog arton nya pennor
9. Svante har nio fina knappar
10. Anna höll elva lätta korgar

LISTA 4

1. Peter köpte två fina knappar
2. Bosse gav sex nya dukar
3. Britta flyttar fyra lätta vantar
4. Anna höll tre vackra ringar
5. Svante visar åtta stora pennor
6. Karin tog tolv mörka lådor
7. Elsa lånar arton hela bollar
8. Jonas ser elva svarta mössor
9. Märta ägde nio gamla korgar
10. Gustav har sju ljusa skålar

LISTA 5

1. Jonas visar arton lätta korgar
2. Anna gav sju fina ringar
3. Gustav köpte sex hela knappar
4. Svante tog nio ljusa dukar
5. Bosse höll tre mörka vantar
6. Britta har fyra vackra skålar
7. Karin ägde åtta nya lådor
8. Märta lånar tolv stora mössor
9. Elsa ser två svarta pennor
10. Peter flyttar elva gamla bollar

LISTA 6

1. Elsa flyttar tre stora mössor
2. Märta tog nio vackra skålar
3. Britta lånar tolv lätta ringar
4. Gustav har sju hela vantar
5. Svante höll arton nya bollar
6. Jonas köpte två gamla lådor
7. Bosse visar åtta fina dukar
8. Karin ägde fyra mörka korgar
9. Anna gav sex svarta pennor
10. Peter ser elva ljusa knappar

LISTA 6

1. Elsa flyttar tre stora mössor
2. Märta tog nio vackra skålar
3. Britta lånar tolv lätta ringar
4. Gustav har sju hela vantar
5. Svante höll arton nya bollar
6. Jonas köpte två gamla lådor
7. Bosse visar åtta fina dukar
8. Karin ägde fyra mörka korgar
9. Anna gav sex svarta pennor
10. Peter ser elva ljusa knappar

LISTA 7

1. Karin flyttar åtta fina korgar
2. Britta ser fyra lätta pennor
3. Gustav höll tre svarta mössor
4. Peter ägde elva stora vantar
5. Märta tog sex mörka dukar
6. Svante lånar sju hela ringar
7. Jonas visar två ljusa lådor
8. Bosse har arton nya bollar
9. Anna gav nio vackra knappar
10. Elsa köpte tolv gamla skålar

LISTA 8

1. Svante tog åtta fina bollar
2. Anna lånar nio lätta korgar
3. Märta har elva stora vantar
4. Peter ägde tre hela mössor
5. Jonas gav tolv gamla skålar
6. Karin visar sex nya dukar
7. Gustav flyttar arton vackra pennor
8. Bosse köpte sju ljusa knappar
9. Elsa höll två svarta lådor
10. Britta ser fyra mörka ringar

LISTA 9

1. Bosse köpte sju mörka bollar
2. Jonas flyttar tre stora korgar
3. Svante har tolv hela pennor
4. Anna tog åtta ljusa knappar
5. Britta ägde arton svarta ringar
6. Peter höll nio nya lådor
7. Elsa ser elva gamla skålar
8. Karin visar fyra fina vantar
9. Märta gav sex lätta dukar
10. Gustav lånar två vackra mössor

LISTA 10

1. Bosse lånar fyra vackra korgar
2. Svante höll tolv svarta dukar
3. Märta gav åtta ljusa knappar
4. Anna flyttar sju gamla pennor
5. Karin ser tre fina bollar
6. Britta visar sex hela mössor
7. Peter tog elva nya vantar
8. Jonas har två lätta skålar
9. Gustav köpte nio mörka lådor
10. Elsa ägde arton stora ringar

LISTA 11

1. Svante lånar sex ljusa lådor
2. Peter höll sju fina skålar
3. Bosse köpte nio hela bollar
4. Jonas visar arton nya dukar
5. Karin ser två mörka ringar
6. Gustav flyttar fyra gamla korgar
7. Märta tog tolv svarta vantar
8. Anna har elva stora pennor
9. Britta gav åtta lätta knappar
10. Elsa ägde tre vackra mössor

LISTA 12

1. Gustav gav elva ljusa lådor
2. Svante lånar fyra nya ringar
3. Märta ser arton gamla dukar
4. Bosse köpte sex hela vantar
5. Jonas har två lätta skålar
6. Anna flyttar åtta vackra pennor
7. Elsa visar tre fina bollar
8. Karin ägde nio stora knappar
9. Peter tog tolv mörka mössor
10. Britta höll sju svarta korgar

LISTA 0

1	1	1	1	1
2	2	2	2	2
3	3	3	3	3
4	4	4	4	4
5	5	5	5	5
6	6	6	6	6
7	7	7	7	7
8	8	8	8	8
9	9	9	9	9
10	10	10	10	10

LISTA 1

2	10	6	2	3
1	9	3	4	9
3	8	7	6	8
9	4	5	5	2
8	7	4	1	5
7	6	10	10	1
4	3	8	3	10
6	2	2	9	4
5	1	1	8	6
10	5	9	7	7

LISTA 2

7	6	7	8	4
10	5	1	1	10
9	1	5	6	2
2	8	10	3	1
4	7	6	7	6
5	10	9	2	8
3	9	8	9	5
1	2	3	5	7
6	4	4	4	3
8	3	2	10	9

LISTA 3

10	3	9	10	6
6	7	5	7	10
3	4	3	2	7
7	5	2	6	3
1	8	6	1	8
9	6	4	5	9
4	1	1	3	2
2	10	10	4	4
5	9	7	8	1
8	2	8	9	5

LISTA 4

4	4	1	8	1
9	1	4	4	8
2	7	3	9	3
8	2	2	5	9
5	8	6	3	4
1	10	9	6	10
7	5	10	2	2
6	3	8	10	7
3	6	7	1	5
10	9	5	7	6

LISTA 5

6	8	10	9	5
8	1	5	8	9
10	4	4	2	1
5	10	7	7	8
9	2	2	6	3
2	9	3	5	6
1	6	6	4	10
3	5	9	3	7
7	3	1	10	4
4	7	8	1	2

LISTA 6

7	7	2	3	7
3	10	7	5	6
2	5	9	9	9
10	9	5	2	3
5	2	10	4	2
6	4	1	1	10
9	8	6	8	8
1	6	3	6	5
8	1	4	10	4
4	3	8	7	1

LISTA 7

1	7	6	8	5
2	3	3	9	4
10	2	2	10	7
4	6	8	3	3
3	10	4	6	8
5	5	5	2	9
6	8	1	7	10
9	9	10	4	2
8	1	7	5	1
7	4	9	1	6

LISTA 8

5	10	6	8	2
8	5	7	9	5
3	9	8	3	3
4	6	2	2	7
6	1	9	1	6
1	8	4	4	8
10	7	10	5	4
9	4	5	7	1
7	2	1	10	10
2	3	3	6	9

LISTA 9

9	4	5	6	2
6	7	2	3	5
5	9	9	2	4
8	10	6	7	1
2	6	10	10	9
4	2	7	4	10
7	3	8	1	6
1	8	3	8	3
3	1	4	9	8
10	5	1	5	7

LISTA 10

9	5	3	5	5
5	2	9	10	8
3	1	6	7	1
8	7	5	1	4
1	3	2	8	2
2	8	4	2	7
4	10	8	4	3
6	9	1	9	6
10	4	7	6	10
7	6	10	3	9

LISTA 11

5	5	4	7	10
4	2	5	8	6
9	4	7	2	2
6	8	10	4	8
1	3	1	6	9
10	7	3	1	5
3	10	9	10	3
8	9	8	3	4
2	1	6	9	1
7	6	2	5	7

LISTA 12

10	1	8	7	10
5	5	3	4	9
3	3	10	1	8
9	4	4	2	3
6	9	1	9	6
8	7	6	5	4
7	8	2	8	2
1	6	7	3	1
4	10	9	6	7
2	2	5	10	5

LISTA 20
(TRÄNING)

2	7	6	10	9
7	1	4	4	3
4	4	5	7	6
1	6	3	5	1
9	8	2	9	8
8	2	1	6	5
6	5	8	2	10
5	10	10	1	2
10	3	7	8	4
3	9	9	3	7

List no. 0

1.	Karin	gav	två	gamla	knappar
2.	Britta	höll	tre	hela	bollar
3.	Märta	ser	fyra	stora	vantar
4.	Peter	köpte	sex	nya	pennor
5.	Svante	lånar	sju	vackra	korgar
6.	Jonas	ägde	åtta	mörka	skålar
7.	Elsa	flyttar	nio	ljusa	mössor
8.	Anna	visar	elva	fina	dukar
9.	Bosse	har	tolv	lätta	ringar
10.	Gustav	tog	arton	svarta	lådor

1.	Karin	gave	two	old	buttons
2.	Britta	kept	three	whole	balls
3.	Märta	sees	four	big	gloves
4.	Peter	bought	six	new	pencils
5.	Svante	borrows	seven	nice	baskets
6.	Jonas	owned	eight	dark	bowls
7.	Elsa	moves	nine	light	caps
8.	Anna	shows	eleven	fine	cloths
9.	Bosse	has	twelve	light	rings
10.	Gustav	took	eighteen	black	boxes

Table I. Original list of sentences and its translation to English.

		First session		Second session			
Part		A	B	C	D	E	F
Purpose		Two thresholds	List homo-geneity	Intellig. curve in noise	Intellig. curve with-out noise	Speech level dependence	Threshold
Speech level (dB)		65	65	65	4 dB steps	40 53 65 77 90	65
S/N (dB)		3 dB steps	konstant	2 dB steps	-----	konstant	3 dB step
List no	Grp I	6 7 8 9	5 4 3 2 1	10 9 8 7 6	1 2 3 4 5	6 7 10 11 12	8 9
	Grp II	9 8 7 6	1 2 3 4 5	6 7 8 9 10	5 4 3 2 1	8 9 10 11 12	7 6

↑ Training list (before A) ↑ Training list (before C) ↑ Training list (before D) ↑ 5 min intermission (before E)

Table II. Overview of the experimental design.

List no.	1	2	3	4	5
Mean score %	44.5	51.4	50.5	49.2	50.5

Table III. Mean intelligibility score of lists No 1-5. N=20. Learning effects outbalanced. From part B.

		A1	A2	C	F	A2-A1	A2-C	A2-F
Threshold (dB)	MEAN ST.D	-7.13 0.79	-7.66 0.73	-8.15 0.70	-8.83 0.62	-0.53 0.67	0.49 0.60	1.17 0.60
Slope (%/dB)	MEAN ST.D	13.5 2.84	14.2 3.13	19.1 5.42	15.2 2.48			

Table IV. Means and standard deviations of thresholds and slopes from part A, C and F. The corresponding values of the differences between the thresholds to the right.

1.	75	34	78	66	28
2.	37	39	69	62	55
3.	52	49	44	42	18
4.	53	86	90	53	34
5.	89	31	49	63	32
6.	66	63	63	35	51
7.	69	41	41	65	59
8.	52	43	58	28	55
9.	62	59	45	50	24
10.	70	56	54	64	64
(means)	62.5	50.0	59.0	52.7	41.9

Table V. Mean intelligibility score (%) of each word averaged over all subjects and all conditions with noise (N=420). The positions of the numbers refer to the word positions in the original list shown in Table I.

1.	87	57	62	54	41
2.	45	32	48	61	77
3.	38	21	58	49	48
4.	45	52	73	72	36
5.	64	46	52	65	66
6.	38	45	66	47	80
7.	45	38	47	27	29
8.	43	28	46	31	34
9.	35	63	72	65	32
10.	46	48	71	58	74
(means)	48.6	43.0	59.5	52.9	51.7

Table VI. Mean intelligibility score (%) of each word averaged over all the subjects and the five lists without noise (N=100). The positions of the numbers refer to the word positions in the original list shown in Table I.

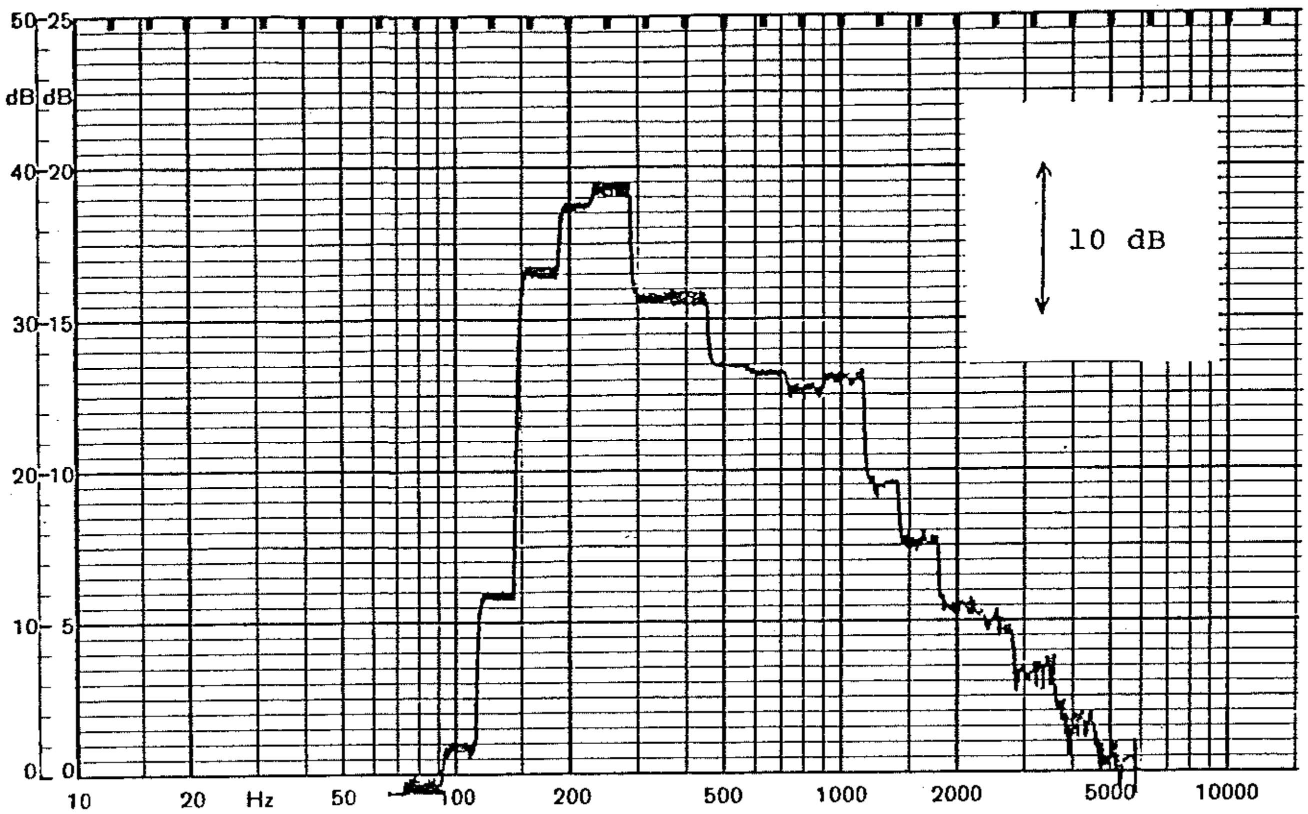

Fig. 1 1/3 octave spectrum of the masking noise before low-frequency modulation. This spectrum is valid also for the speech signal.

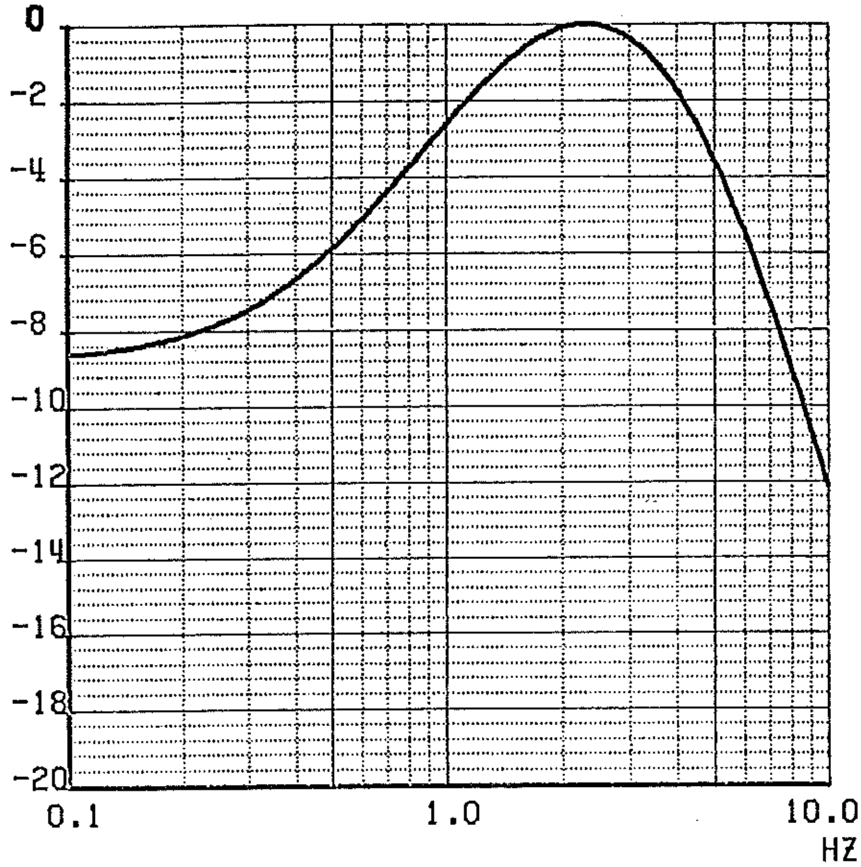

Fig. 2 Frequency curve of the filter used for shaping the modulating noise.

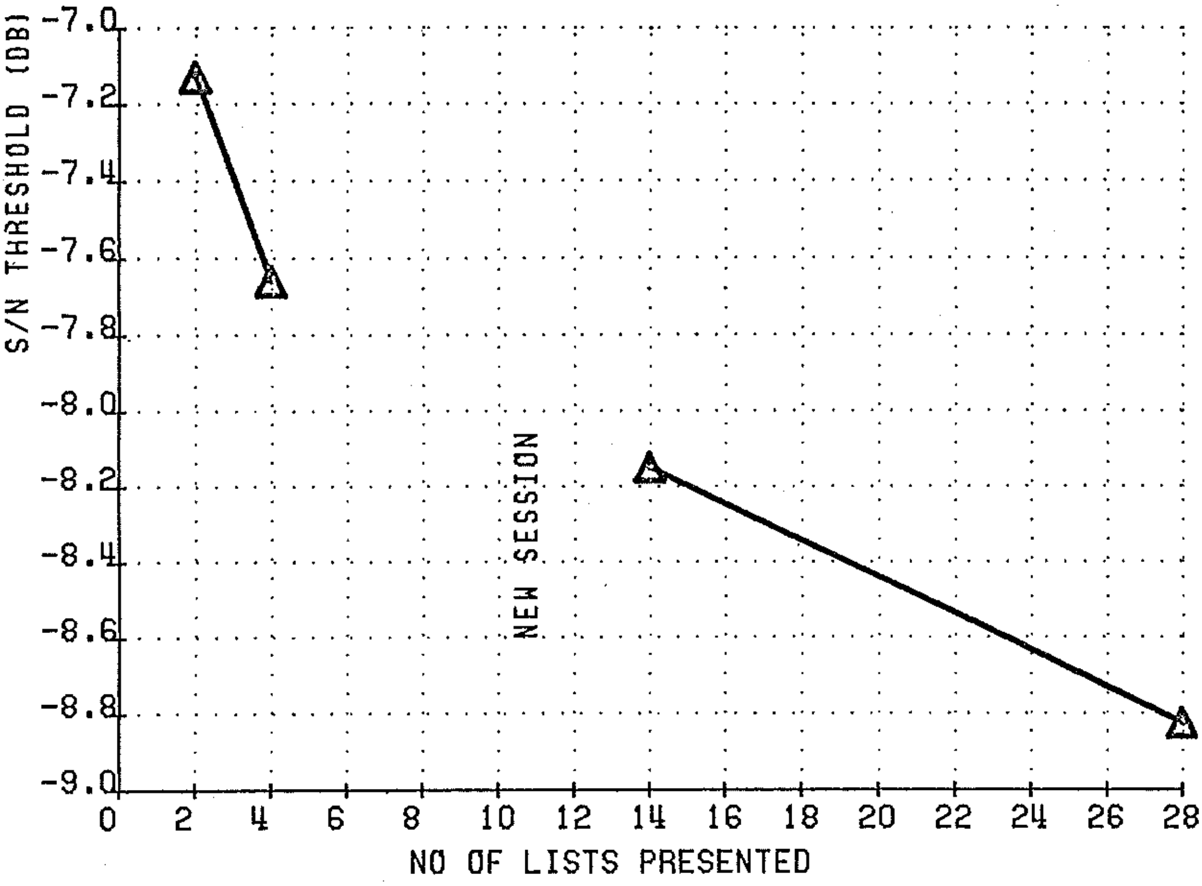

Fig. 3 Mean thresholds as a function of the number of lists already presented to the subjects, showing the learning effect. N = 20. The second and third point are not connected, since they refer to different sessions.

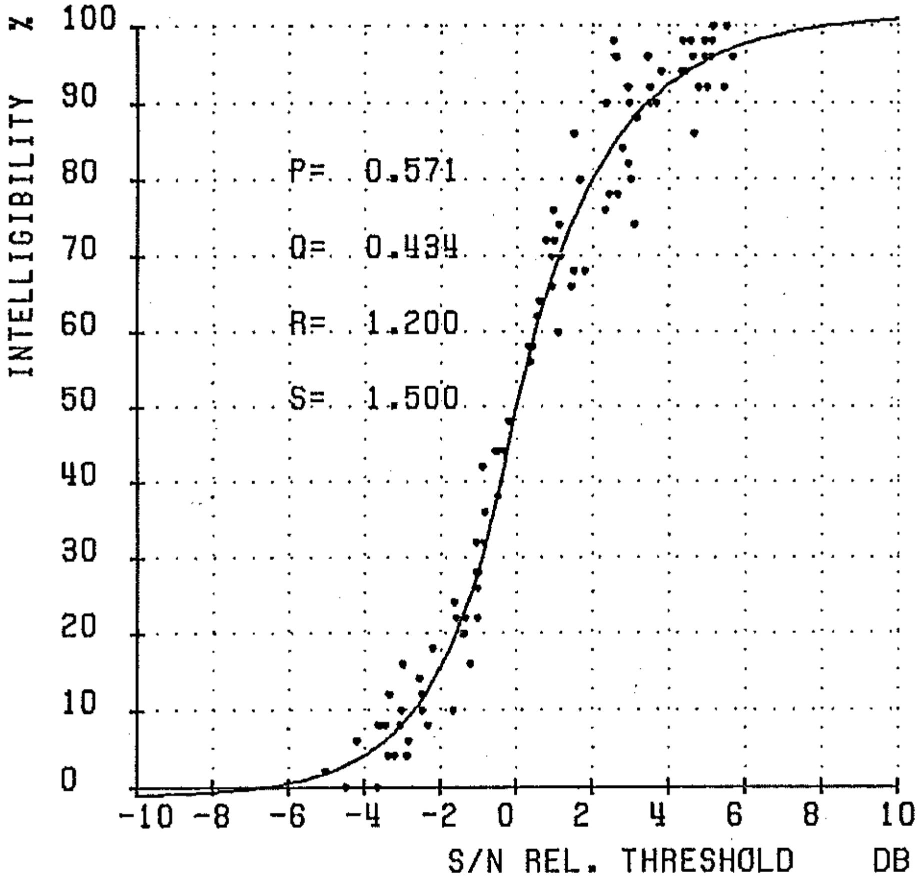

Fig. 4 Intelligibility curve in noise. From part C. Constants P-S according to equation (1).

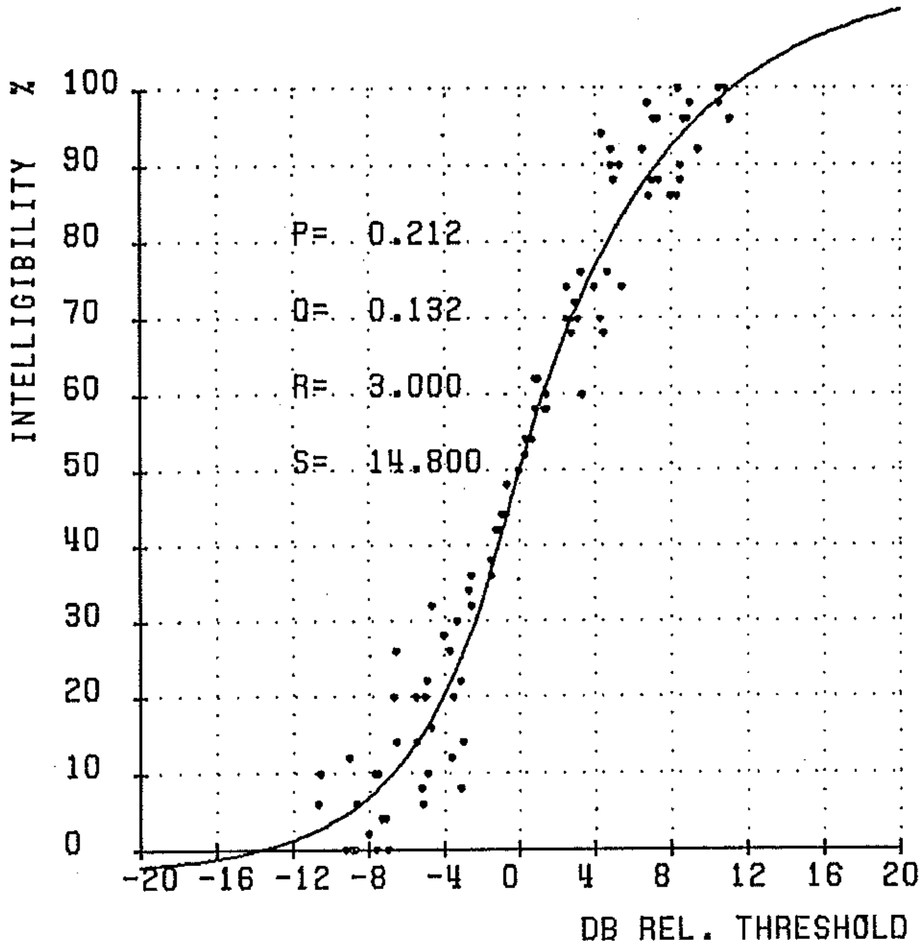

Fig. 5 Intelligibility curve without noise. From part D. Constants P-S according to equation (1).

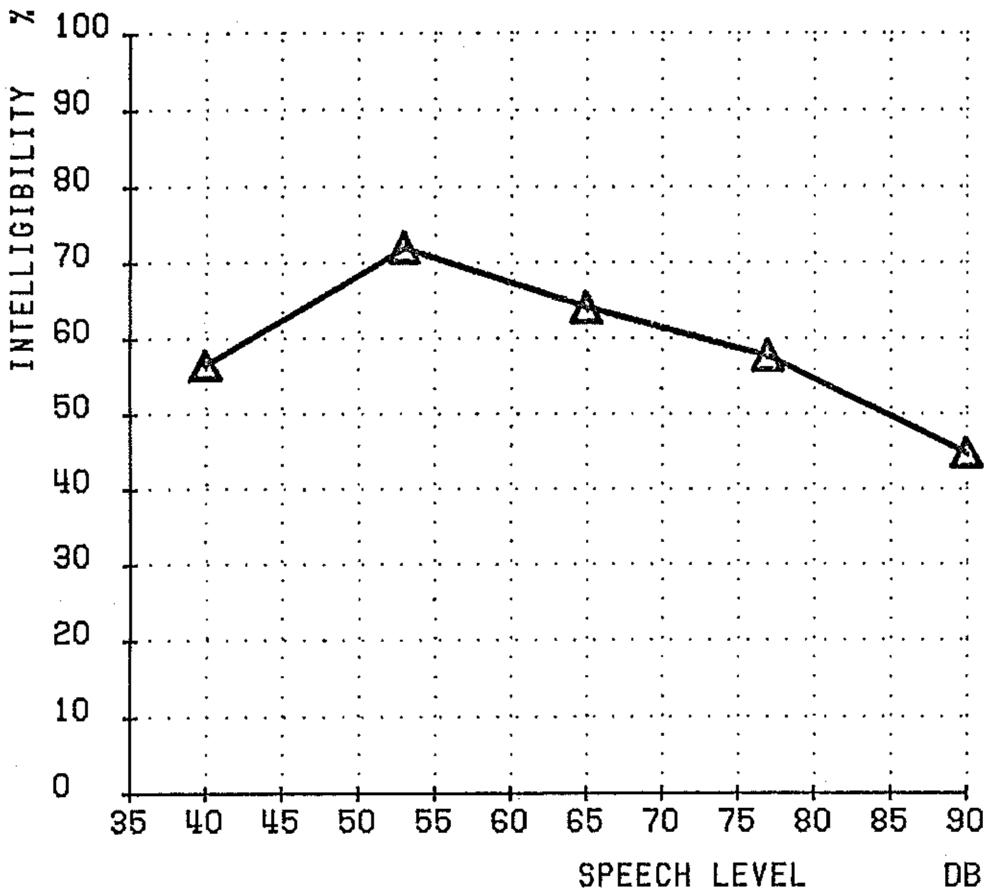

Fig. 6 Mean intelligibility score as a function of speech level. Constant S/N for each subject. N = 20. From part E.

Karolinska Institutet
Technical Audiology
KTH, S-100 44 Stockholm
Sweden

Appendix 2 of Report TA 101:

FURTHER INVESTIGATION ON LIST HOMOGENEITY IN SENTENCES FOR TESTING SPEECH INTELLIGIBILITY IN NOISE

Björn Hagerman

ABSTRACT

In the main experiment only lists No. 1-5 were checked for homogeneity (Report TA101). The deviating result of list No. 1 led to a second investigation which is accounted for in this appendix. Here lists No. 6-12 were tested, but lists No. 1 and 5 were also included to get a possibility to compare the two investigations.
The mean result of list No. 8 was shown to be 6 % higher than the mean result of all lists. However, the deviation of list No. 8 does not change the conclusions of the main experiment. Lists No. 2-12 except list No. 8 are accepted for usage.
Regarding the learning effect the results here correspond well to the results of the main experiment.

This work was supported by the National Swedish Board for Technical Development.

1 INTRODUCTION

In the main experiment only lists No. 1-5 were checked for homogeneity (Report TA101). The deviating result of list No. 1 led to a second investigation which is accounted for in this appendix. Here lists No. 6-12 were tested, but lists No. 1 and 5 were also included to get a possibility to compare the two investigations.

2 METHODS

9 normal hearing subjects participated in this test. All of them had listened to at least 10 lists about 8 months earlier in the pilot study preceeding the main experiment. They were tested separately in a session not exceeding half an hour. The session started with a training list. Table I shows the order of the test lists presented for each subject. This order was chosen to counterbalance the learning effect when comparing the mean results of different test lists. The level of the calibration tone was set to 65 dB SPL. The S/N was chosen individually between -7 and -9 dB according to the results of the pilot study to get approximately 50 % discrimination. No change of the S/N was done during the session. The tape and the technical equipment was the same as in the main experiment.

3 RESULTS

Table II shows the mean scores of the different lists averaged over the 9 subjects. The total mean is also shown to the right in the table. Since the standard deviation of repeated discrimination measurements in the main experiment was shown to be around 6 % for discrimination values around 50 %, the 95 % confidence limits for the total mean will be $1.96*6/\sqrt{9}$ % = 4 % from the total mean. The 95 % confidence interval will thus be 49.3 - 57.3 %. The result of list No. 8 falls outside this interval.

Fig. 1 shows the learning effect. It follows from Tab. I that inequalities of the lists will not affect the learning effect shown as the means taken over the results of the columns corresponding to Tab. I. Over the 9 lists heard the average score is raised 11 %. In the main experiment the slope of the intelligibility curve was shown to be 22 %/dB. A learning effect of 11 % then corresponds to 0.5 dB change of the S/N threshold, which agrees well with the result of the main experiment (Fig. 3 in the main report), considering that the subjects had participated in the pilot study earlier.

4 DISCUSSION

List No. 1 was the most difficult one in both investigations and even though its result fell within the confidence interval this time, there is no reason for changing the decision to reject this list. The result of list No. 8 fell outside the confidence interval and this list is not accepted for further usage. List No. 8 did not show any bad points regarding the difficulties of the sentences as did list No. 1 in the main experiment. The reason that list No. 8 deviate from the others is supposed to be the following:
All lists have exactly the same content of sound both regarding the words and the noise. However, a particular word is masked by different parts of the noise in different lists which may lead to the fact that different lists are masked more or less effectively.

The deviating result of list No. 8 does not change the conclusions of the main experiment, due to the balanced design. This was checked also for the intelligibility curve in noise (Fig 4 in the main report), which changes insignificantly if the results of list No. 8 in the main experiment are reduced by 6 %.

Subj. No.	List No.								
1	1	5	6	7	8	9	10	11	12
2	12	1	5	6	7	8	9	10	11
3	11	12	1	5	6	7	8	9	10
4	10	11	12	1	5	6	7	8	9
5	9	10	11	12	1	5	6	7	8
6	8	9	10	11	12	1	5	6	7
7	7	8	9	10	11	12	1	5	6
8	6	7	8	9	10	11	12	1	5
9	5	6	7	8	9	10	11	12	1

Table I. Order of the test lists presented for each subject.

List No.	1	5	6	7	8	9	10	11	12	Mean
Mean score %	50.7	55.1	55.1	53.8	59.8	51.1	51.6	54.7	53.1	53.28

Table II. Mean intelligibility scores and total mean of test lists. N=9.

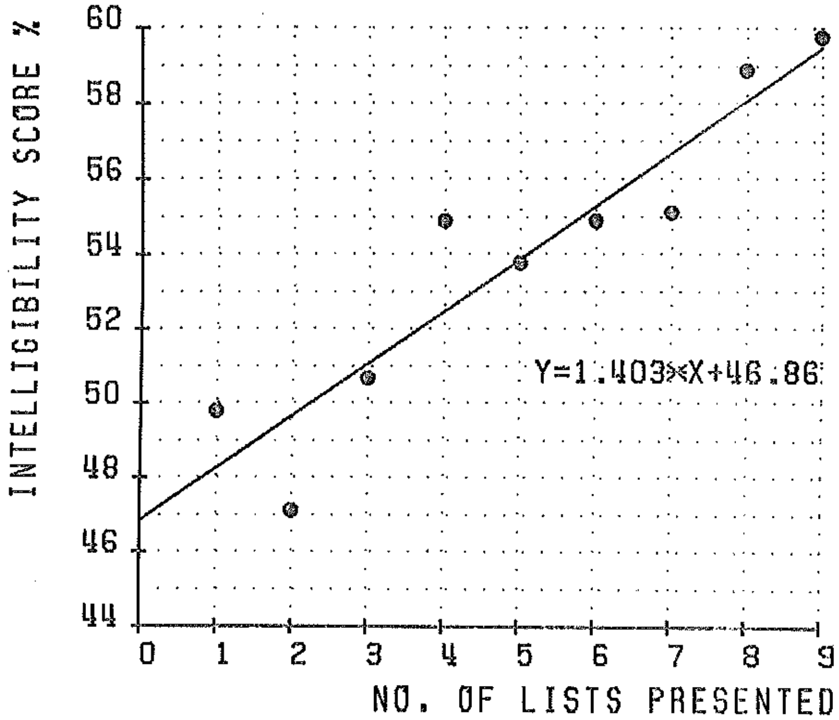

Fig. 1 Learning effects in a new session after eight months. Mean intelligibility score as a function of the number of lists presented to the subject. N=9. Regression line.