ISSN 0280-6819

TECHNICAL AUDIOLOGY

Report TA No. 107
June 1983

CLINICAL MEASUREMENTS OF SPEECH RECEPTION THRESHOLD IN NOISE

Björn Hagerman

ISSN 0280-6819
Report TA107
June 1983

Karolinska Institutet
Dept. of Technical Audiology
KTH
S-100 44 Stockholm
Sweden
Tel: +46-8-11 66 60

CLINICAL MEASUREMENTS OF SPEECH RECEPTION THRESHOLD IN NOISE

Björn Hagerman

ABSTRACT

New lists of spoken sentences, edited in a computer word by word, were tested clinically together with a noise, spectrally shaped as the speech. The purpose was to investigate the reliability and the learning effect of the speech reception threshold in noise, and also to get the relations of this threshold to ordinary audiometric measures. The threshold values ranged from -7 to +7 dB signal to noise ratio for the 97 ears investigated. Reliability, expressed as standard deviation for repeated measurements, deteriorated from 0.7 to 1.1 dB as the threshold deteriorated. Learning effect between the first and the second threshold increased from 0 to 1 dB as the threshold deteriorated. No ordinary audiometric test showed a high correlation to this threshold measure.

The discrimination of the sentences without noise, subjective rating of speech recognition in noise and optimum level for speech reception threshold in noise, were also investigated.

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

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 faculty, however, are still not very common. In Sweden the main reason probably has been the lack of standardized methods and speech materials to make fast and reliable clinical measurements of the speech intelligibility in noise.

A new Swedish speech material for this purpose, was earlier developed and tested on normal hearing subjects (Hagerman, 1982). The speech material was constructed in the following way. A list of ten spoken Swedish sentences were computer edited word by word to achieve new lists with exactly the same content of sound, but with new sentences. Each sentence consists of 5 words (e.g. "Peter bought seven dark balls") and each word is scored. A full list thus consists of 50 words. A noise was synthesized from the speech material by the computer, to give the noise exactly the same spectrum as the speech. The noise was furthermore amplitude modulated by a low frequency noise to make it sound more natural.

There are several possibilities for clinical use of this speech and noise material. It might be used as:

1. A prognostic test to predict the benefit of a hearing aid (Plomp, 1978).

2. A test to find the best ear for the hearing aid.

3. A test to compare different hearing aids

4. A test for final evaluation of the fitted hearing aid.

5. A complement to the ordinary SRT measurement in assessment of the social handicap index for insurance cases.

6. A diagnostic test.

The purpose of the present investigation was not primarily to evaluate the suitability of the material in these different respects. Further studies are needed to make this clear. The purpose of the clinical study reported on here was more basic, namely:

1. To investigate the relation between discrimination scores achieved with the sentences without noise and discrimination scores achieved with PB-lists.

2. To investigate the relation between the speech discrimination score measured in the conventional way with PB-words and the speech reception threshold in noise measured with the new speech material. (If the correlation is very high, there is no need of measuring both routinely.)

3. To test if threshold measurements in noise on hearing impaired subjects, fitted into the clinical routine, give the same reliability and learning effect as the earlier measurements on normal hearing subjects.

4. To test the hypothesis that the optimum level for hearing impaired, when listening to speech in noise at a fixed signal to noise ratio, is lower than the optimum level when listening to speech in quiet. (For the normal hearing subjects the optimum level in noise was lower than expected.)

5. To test if big differences exist between the two ears for some patients, regarding the speech reception threshold in noise.

6. To test roughly if measured thresholds in noise correlate with subjectively rated ability to recognize speech in noisy environments.

METHODS

The investigation was divided into two experiments, one regarding discrimination of sentences without noise and one regarding discrimination of sentences in the presence of noise. In both experiments the speech was presented monaurally, and the ear was chosen so as to avoid masking of the contralateral ear, when testing in noise. The speech level was always the same for the PB-discrimination test and the sentence test. The level was chosen just in the same way as the technician normally does in the clinical work, which usually means a level preferred by the patient. The speech and noise material was delivered from a Revox A77 tape recorder through an audiometer (specified below) and TDH39 earphones with MX41AR cushions.

Experiment 1 (without noise)

The audiometer used was a Madsen OB 70. Data from 89 unselected patients were collected by one technician. About one third of these were insurance cases and about one third were coming for a hearing aid. First a conventional tone audiogram was made and the speech reception threshold (SRT) for spondees was measured. Then the speech discrimination was measured with 50 PB-words in the conventional way and with 5 lists of the sentences without noise. Half of the patients, however, were tested with the sentences before the PB-words. No training of the sentences were allowed, but some PB-words were presented before the start to find a proper speech level, which was always the same for both speech materials.

Experiment 2 (with noise)

Data were collected by four technicians at four different clinics. The audiometers used were Madsen OB 822, which made it possible to mix speech and noise to one output channel without any undesired change of the levels. At one clinic, however, a Madsen OB 70 was used. In this case the change of the levels, achieved when the channels were mixed together at the output, were compensated for when calculating the results.

In experiment 2 the PB-word discrimination test always preceeded the sentence test, which started with a training list where the noise level was successively increased. Then the 50 % threshold in noise could be found with two or sometimes three sentence lists. The speech level was held constant and the same as for the PB-words. The noise level was changed in 3 dB steps between the lists, until one result on each side of the 50 % borderline was achieved. Two consecutive threshold measurements were made for each patient. At two clinics the second threshold was used to evaluate the learning effect and the reliability (N=70). At another clinic the second threshold was measured with the speech level lowered 7 dB (N=10). Here the PB-discrimination score was also measured at both levels. At the last clinic the second threshold was measured at the opposite ear (9 patients, 17 ears), to find out if big differences between the ears might exist.

Before the sentence test started, the patient was verbally asked to answer the following question:

> How well do You manage to recognize speech in noisy environments? Please compare Yourself with a normal hearing person in the corresponding situation. Do you hear

1	2	3	4	5
Quite as good	Slightly worse	Definitely worse	Considerably worse	Very much worse

The answers were then denoted by figures 1 to 5 in the results.

A total number of 89 patients were tested. Most of the patients were coming to get an hearing aid. Due to complex behaviour of the OB 70 attenuators at high setting when routed to the same output, we restricted the maximum speech level to 66 dB above normal threshold for the patients with this audiometer. Patients requiring higher speech levels were thus rejected. However, this concerns only the patients tested at two different speech levels. Otherwise no systematic selection of the patients were made. For 6 additional patients the sentence discrimination score in noise was measured at three or four noise levels to check the configuration of the intelligibility curve.

RESULTS

Experiment 1 (without noise)

The total material was divided into five classes according to the PB-word discrimination score. For each class the means of the PB-word score and for the five sentence list scores were calculated. See Table I which is also illustrated in Fig. 1. The learning effect is obviously not neglectable, but the second sentence list heard shows good agreement with the PB-word result. This holds for group mean results. The individual deviations are however considerable, which is shown in Fig. 2. Here the mean result of the five sentence lists is plotted against the PB-word result for each patient. The standard deviation of the difference between these two measures ranges between 11.6 and 20.4, except for the highest discrimination class with a standard deviation of only 3.4.

These standard deviations are too big to be explained by random fluctuations only (Hagerman, 1976), which means that the results of the sentences seem to be affected by some unknown factor. This factor might be related to the short term memory for example (Risberg & Agelfors, 1979). Comparing the results from patients falling above and below the 45 degree line in Fig. 2. respectively, no difference between the groups were found, regarding age, SRT, audiograms or speech levels used. The relative difficulty between the first and the last word in the sentence was also about the same for the two groups, indicating that the short time memory is not the crucial factor.

Experiment 2 (with noise)

The mean audiogram for the 80 ears with sensorineural impairment is shown in Fig. 3 with $\pm$ standard deviations indicated. The mean audiogram for the patients with conductive impairment (N=10) was 35 dB $\pm$2 dB SD at 125-4000 Hz.

Intelligibility curves

Fig. 4 shows the intelligibility curves for six extra patients. The most obvious difference between the curves are the different horizontal shifts compared to the leftmost curve concerning normal subjects. Curves more to the right, however, also seem to have somewhat less steep slopes, but the clear horizontal shifts justify the method of measuring the ability to recognize speech in noise, as a signal to noise threshold (S/N-threshold).

Relation to PB-discrimination

Fig. 5 shows the relation between the discrimination score of PB-words without noise and the S/N-threshold for sentences. (If not otherwise mentioned the first of the two measured S/N-thresholds is concerned further on in text and figures). The correlation is statistically significant ($r=-0.62$, $p<0.001$) but the S/N-threshold is not very well predicted by the discrimination score without noise, especially not for patients with 100 % discrimination score.

Relation to SRT without noise

For the relation between the SRT without noise for spondees and the S/N-threshold, the correlation is much lower ($r=0.36$) but still significant ($p<0.001$). See Fig. 6. Only sensorineural impairments are included here. The regression line would certainly be more steep if data for normal hearing subjects were included. Unfortunately the SRT was not measured with spondees in the earlier investigation of normal subjects (Hagerman, 1982). In consideration of this fact, the result corresponds rather well to the result of Plomp & Mimpen (1979a). The variability around the line is also very big in both investigations.

Relation to tone thresholds

In order to further investigate whether ordinary audiometric data could predict the ability to recognize speech in noise, the correlation coefficient between the S/N-threshold and the different tone thresholds respectively, were calculated. The result is shown in Fig. 7 (continuous line and filled symbols), together with the corresponding result from Smoorenburg et al., 1982 (dashed line and filled symbols). The agreement between these curves is very good, although Smoorenburgs data originates only from noise induced hearing losses. Open symbols also show the correlation between the SRT without noise and the tone thresholds. The discrepancy between these curves is probably due to different types of speech materials. In the present investigation spondees were used, recorded with a male speaker, while Smoorenburg used the same sentences both with and without noise, and a female speaker.

Reliability and learning effect

To examine the reliability and the learning effect the differences between the first and the second threshold were plotted against the value of the first threshold. See Fig. 8. The trend of the plot can be interpreted as the mean of the learning effect, which is in fact slightly negative for patients with good S/N-thresholds and seem to increase with increasing thresholds. Since the variability around the trend also increases with increasing thresholds, it was not considered appropriate to draw a regression line and use the residual variance as a measure of the reliability. Therefore the material was divided into two groups, with S/N-thresholds <0 dB and >0 dB respectively.

The mean learning effects of the two groups, calculated as the mean of the threshold differences, amounted to -0.1 dB and 1.1 dB respectively. The standard deviation of the threshold differences was 1.0 dB and 1.6 dB repectively. From the last two values the standard deviation of repeated measurements on the same patient was calculated to be 0.71 dB and 1.1 dB respectively. For normal hearing subjects the corresponding value was only 0.44 dB (Hagerman, 1982). Since the accuracy of the threshold measurement clearly depends on the slope of the intelligibility curve (Hagerman, 1979), the increase from 0.71 to 1.1 may well be explained by the decrease of the mean slope from 14.0 to 9.0 %/dB for the two groups respectively (averaged over discrimination values from both threshold measurements). The corresponding value for the normal hearing subjects was 13.9 %/dB, and thus the increase in the standard deviation from 0.44 to 0.71, when the better patients are compared with the normal hearing subjects, can not be explained this way. It might simply be caused by a slightly fluctuating gain in the equipment at the clinics, probably not present in the laboratory equipment used for the normal hearing subjects.

Influence of speech level

For 10 patients (10 ears) both the PB-discrimination and the S/N-threshold was measured at two speech levels separated 7 dB as mentioned above. The difference between the two PB-results and the two S/N-results was then calculated respectively. The mean and standard deviation of this difference was 2.8 % and 7.6 % for the PB-results. This means that the PB-result deteriorated somewhat (however not significantly), when the speech level was 7 dB lower than the one preferred. The mean and standard deviation of the S/N-threshold difference was 1.5 dB and 1.4 dB respectively, meaning that the result in noise improved significantly at the lower level. Part of this improvement, however, might be due to learning effects, since the lower level was always tested after the test at the normal level. Although not significantly shown here, the hypothesis, that the optimum speech level is different for speech without noise and speech disturbed by noise, is worth testing in a future investigation. Chung & Mack (1979) also showed that the optimum level for word discrimination in noise was lower than expected on hearing impaired subjects.

Difference between the ears

Of the 9 patients that were measured on both ears there were three, who showed significantly different S/N-thresholds on the two ears. In two of these cases, however, the PB-discrimination scores also differed much in the same direction, and thus no extra information was achieved by the S/N-threshold measurement here. In the remaining case, however, the tone thresholds were quite similar for both ears. The PB-discrimination scores were 92 % and 94 % but the S/N-thresholds were -3.3 dB and 0.1 dB respectively. Here the better threshold was measured first, and thus any learning effect could not contribute to this difference.

Subjective ratings

Fig. 9 shows the subjective ratings of the ability to recognize speech in noisy environments plotted against the S/N-threshold results. The expected relation is present only as a trend, while there is a high variability, the reason of which is discussed below.

DISCUSSION

The purpose of the first experiment was to try out the sentence material as a discrimination test without noise. The advantage would be that a sentence list takes only about a third of the time required by a PB-word list. Thus it might be tempting to replace the PB-word lists with the new sentence lists at the clinics, although this was not the main purpose with the new speech material. Two drawbacks were found however. One was the big difference between the two tests, found for some patients. This is perhaps not serious in the long run though, when one is used to interpret the results from the new speech material. The other drawback was the learning effect, which is more serious. For PB-words this effect is negligible (Hagerman, 1976), which means that the particular speech material seems to be responsible for the big learning effect in the present study. According to Hagerman (1982) it was expected to be very small, but the reasoning there is probably not valid.

The speech reception threshold in noise gives a new and important aspect of the hearing impairment, since it is not highly correlated to the ordinary audiometric measures. It seems to be especially well suited to give further information about hearing impairments with very high discrimination scores without noise. Although the reliability was lower for patients in the clinical situation than for normals on the laboratory, the reliability was still quite as high as for Plomp & Mimpen's (1979b) speech material using normal hearing subjects. By using tapes with speech and noise mixed on the same channel with suitable S/N-ratios, the reliability might be further enhanced.

The number of patients tested at two speech levels and at both ears respectively, were not as many as planned. Therefore these parts might only be looked upon as pilot studies.

A desirable feature of a speech reception threshold test in noise, would be to reflect truly the patients ability to recognize speech disturbed by noise in real situations. This is true also for most of the possible applications mentioned in the introduction. The result of the subjective ratings was not too encouraging in this respect, since there was a high variability around the expected trend. One reason of this variability is probably the many different listening situations the patient will meet, which are impossible to include in one test. Influence of the ear not tested is also possible of course. A more elaborate method to measure the subjective ability would certainly give a better correlation, but the most important reason of the variability is perhaps the different lip-reading abilities of the patients. Therefore a similar test with lipreading included is a future desire. To make computer edited video-taped sentence lists in the corresponding way as the present material, is however probably almost impossible. Thus it remains to find another way of making several lists with equal difficulties, preferably regarding both auditive, visual and audio-visual information. Since there is no simple way of changing the lipreading difficulty of an item in a video-taped recording, this will be a heavy task to fulfil.

CONCLUSIONS

1. The material and method described, of testing the speech reception threshold in noise, is fitted for clinical use. The threshold measurement is fast and reliable and the learning effect is small.

2. Since no ordinary audiometric test showed a high correlation to the S/N-threshold, this measure gives a new and important aspect of the hearing impairment and is proposed to be incorporated in the clinical test battery.

ACKNOWLEDGEMENTS

I wish to thank the experimenters, Karin Eriksson, Anette Lagergren, Anne Larsen and Gun Nyquist for their careful work and kind cooperation.
This work was supported by the National Swedish Board for Technical Development.

REFERENCES

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

Chung, D.Y. & Mack, B. 1979. The effect of masking by noise on word discrimination scores in listeners with normal hearing and with noise-induced hearing loss. Scand Audiol 8, 139.

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.

Hagerman, B. 1979. Reliability in the determination of speech reception threshold (SRT). Scand Audiol 8, 195.

Hagerman, B. 1982. Sentences for testing speech intelligibility in noise. Scand Audiol 11, 79.

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. 1979a. Speech-reception threshold for sentences as a function of age and noise level. J Acoust Soc Amer 66(5), 1333.

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

Risberg, A. & Agelfors, E. 1978. Information extraction and information processing in speechreading. Dept. of Speech Comm., Royal Inst. Technology, Stockholm, STL-QSPR 2-3, 62.

Smoorenburg, G.F., de Laat, J.A.P.M. and Plomp, R. 1982. The effect of noise-induced hearing loss on the intelligibility of speech in noise. Hearing & Hearing Prophylaxis, (ed. H. M. Borchgrevink) Scand Audiol Suppl. 16, 123.

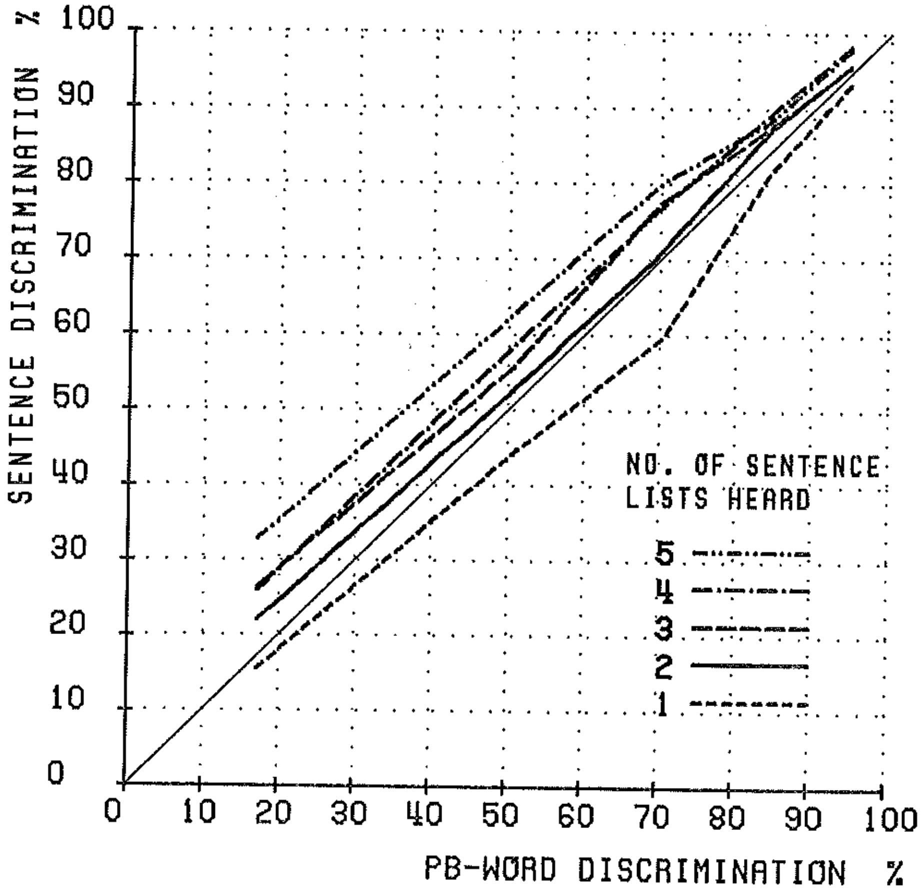

Fig 1. Sentence discrimination without noise as a function of PB-word discrimination. Mean values for 5 different discrimination classes. Parameter is No. of sentence lists heard. Data from Table I.

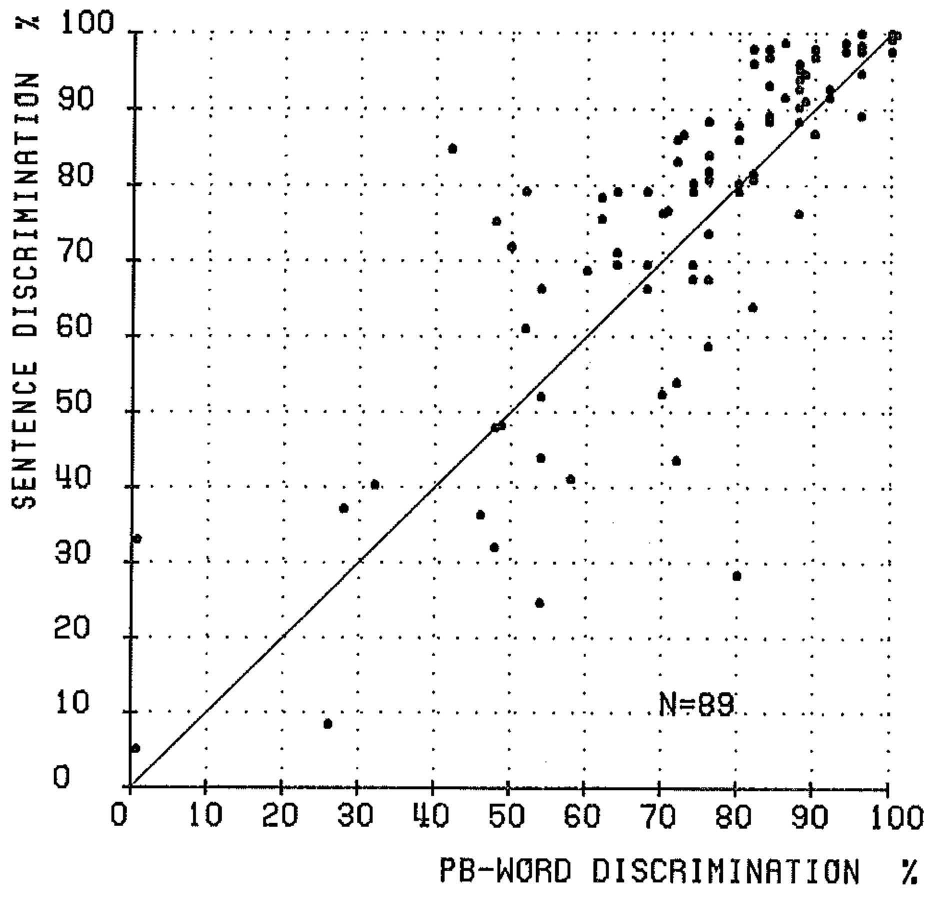

Fig 2. Mean discrimination score of 5 sentence lists for each patient plotted against the PB-word discrimination score, showing inter-subject variability.

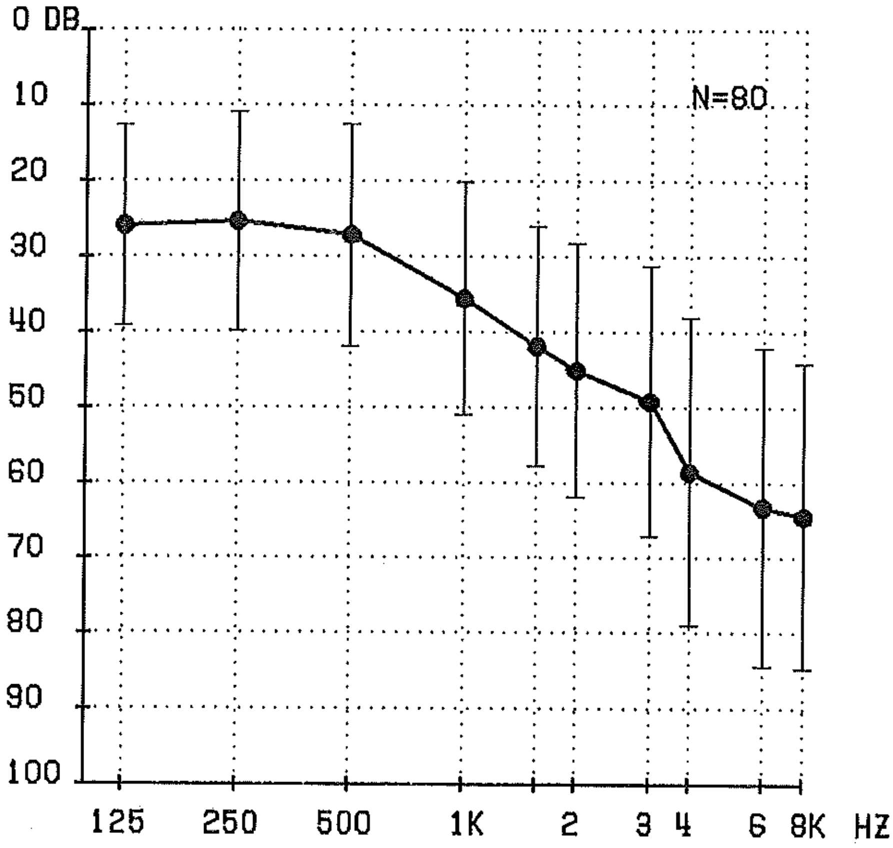

Fig 3. Mean audiogram (±SD) for the 80 ears with sensorineural impairment.

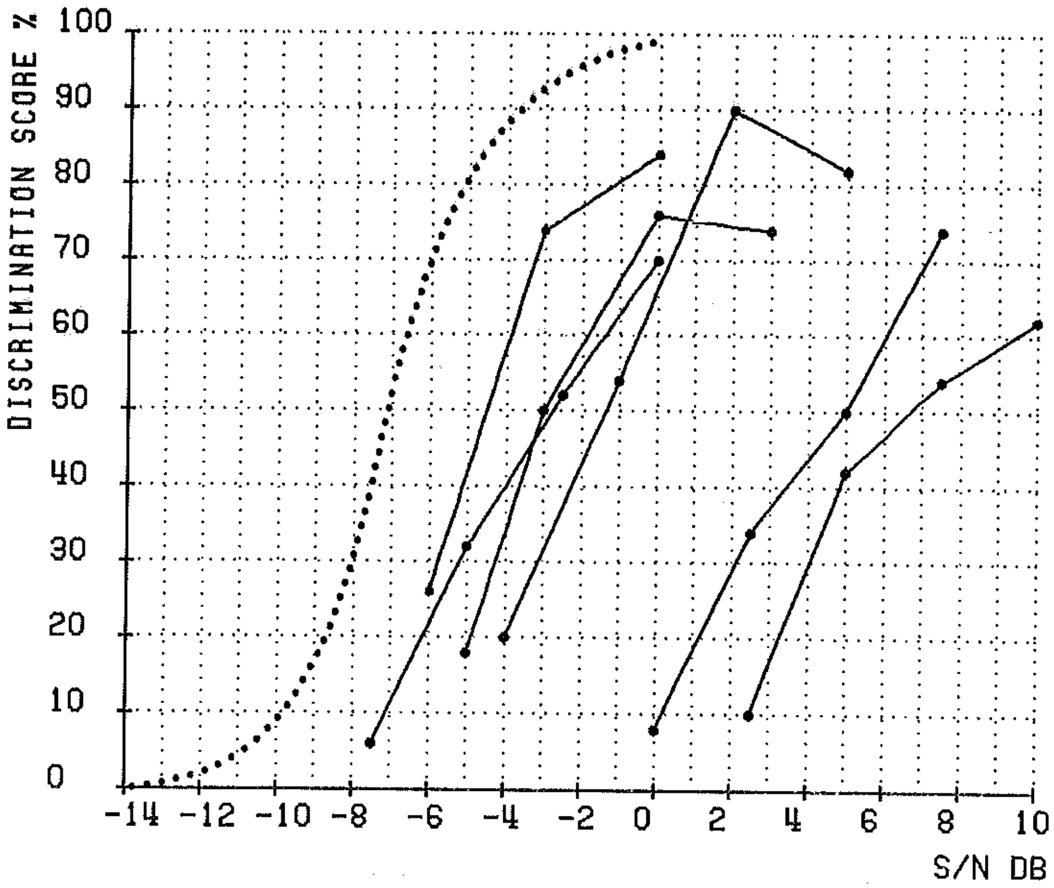

Fig 4. Intelligibility curves for sentences in noise of six extra patients. Dotted line is the average curve for normals (Hagerman, 1982).

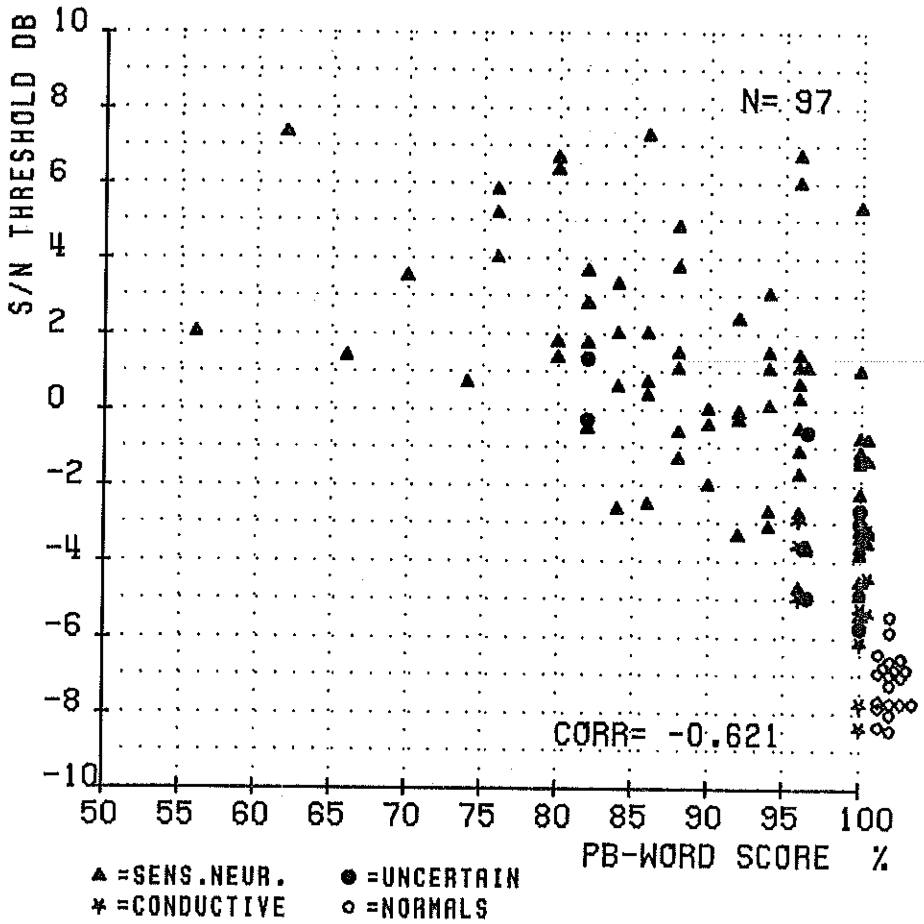

Fig 5. Scatter diagram of the relation between the PB-discrimination score without noise and the S/N-threshold for sentences. Different symbols for different diagnoses as shown by the figure. Normals from Hagerman (1982).

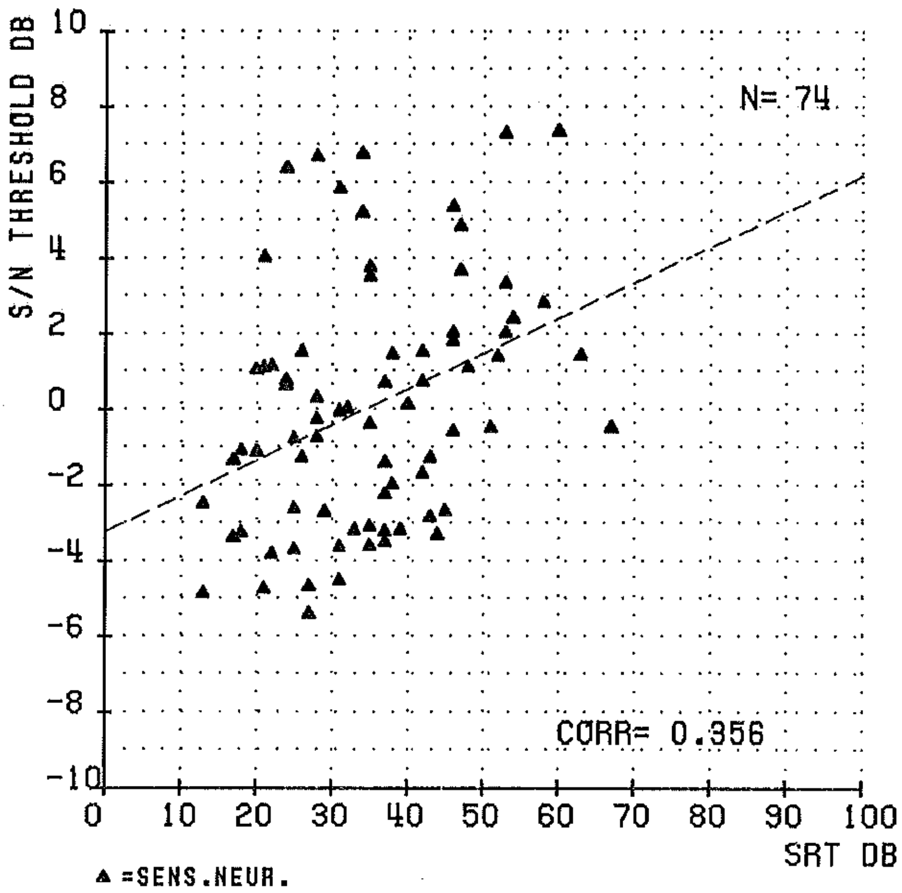

Fig 6. Scatter diagram of the relation between the speech reception threshold for spondees (SRT) without noise and the S/N-threshold for sentences. Only sensorineural impairments.

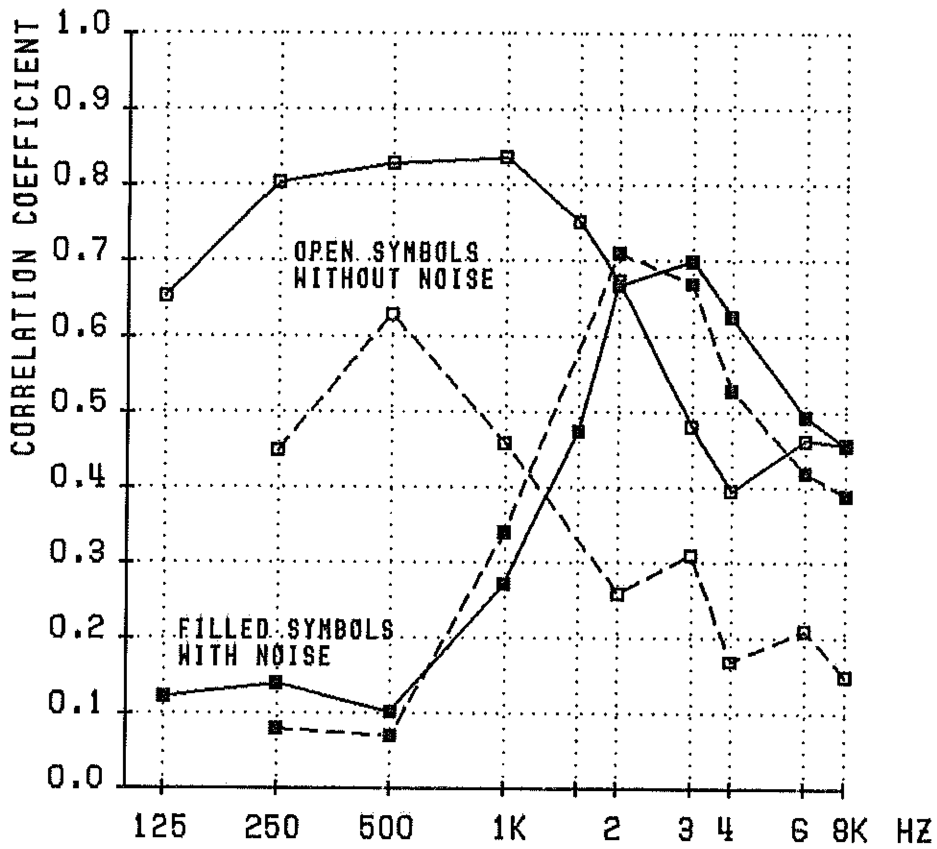

Fig 7. Open symbols show correlation between speech reception threshold without noise and tone thresholds at different frequencies. Filled symbols show correlation between S/N-threshold and tone thresholds. Continuous lines refer to the present investigation and broken lines to Smoorenburg et al. (1982). (With permission).

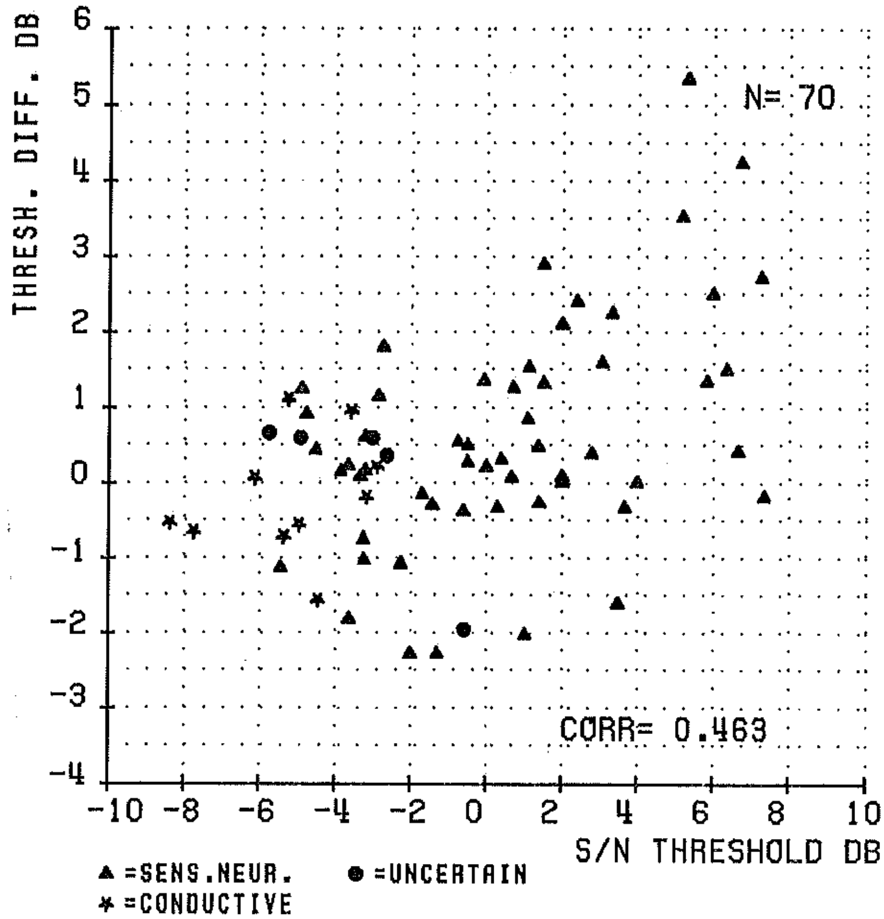

Fig 8. Scatter diagram of the relation between the first S/N-threshold, and the difference between the first and the second S/N-threshold. This shows the reliability and the learning effect. Different symbols for different diagnoses as shown by the figure.

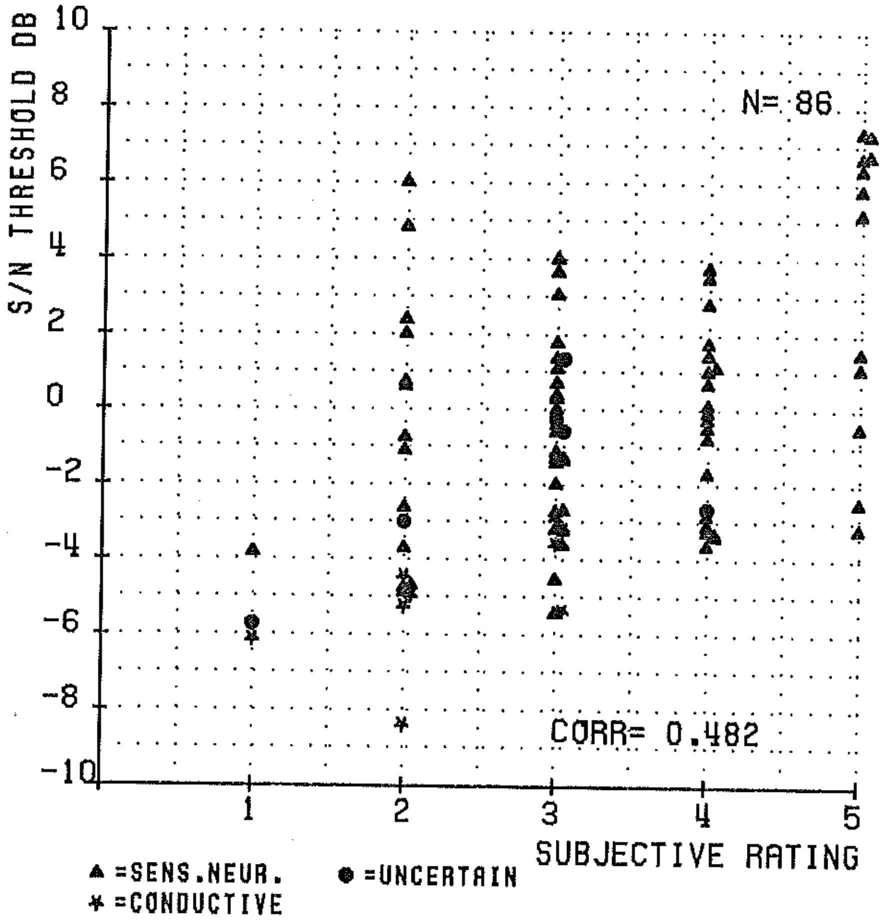

Fig 9. Scatter diagram of the relation between the subjective ability to recognize speech in noisy environments, and the S/N-threshold. Different symbols for different diagnoses as shown by the figure.

Discr. class	0-39%	40-58%	60-78%	80-88%	90-100%
N	5	14	28	26	16
PB-words %	17.2	50.6	70.6	84.4	95.1
Sentences % No. of lists heard					
1	15.6	44.1	59.6	81.5	93.3
2	22.0	52.6	71.1	87.2	95.8
3	26.4	55.6	77.6	87.5	95.6
4	26.0	58.3	77.1	89.3	98.3
5	32.8	62.6	80.2	88.4	97.9

Table I. Mean values of PB-word discrimination scores and sentence list scores without noise for different discrimination classes and different number of sentence lists heard, showing the learning effect.