ISSN 0280-6819

KAROLINSKA INSTITUTET

TECHNICAL AUDIOLOGY

Report TA127
Oct 1993

Efficient adaptive methods for measurements of speech reception thresholds in quiet and in noise

Björn Hagerman and Catharina Kinnefors

ISSN 0280-6819
TA127, Oct 1993

CONTENTS page

KAROLINSKA INSTITUTET
Teknisk Audiologi
Dept. of Technical Audiology

ISSN 0280-6819
TA127, Oct 1993

Efficient adaptive methods for measurements of speech reception threshold in quiet and in noise

Björn Hagerman and Catharina Kinnefors[1]

[1]Dept. of Audiology, Academic Hospital, Uppsala, Sweden

ABSTRACT

An adaptive method for measurement of speech reception threshold in noise (S/N-threshold) was evaluated regarding reliability and learning effect. A corresponding method was also tested for speech reception threshold in quiet (SRT). The methods are designed for use with a speech and noise material developed by Hagerman (1982a). Ten normal-hearing and 40 hearing-impaired subjects were tested. The results showed that the test-retest reliability of the S/N-threshold was much better with the new, adaptive, method than with the old one which was not adaptive. With the new method a S/N-threshold can for most patients be measured within two minutes (after the initial short training) with a reliability, i.e. standard deviation, of 0.78 dB and an almost negligible learning effect. It was also concluded that the adaptive method designed for the sentences in quiet gives at least as reliable SRT result as the old spondée method (including learning effects) but in shorter time.

This work was supported by the Swedish Council for Social Research

Mailing address:
Teknisk Audiologi / KI
Box 70025
S-100 44 STOCKHOLM
Sweden

Visiting address:
Teknisk Audiologi
Lindstedtsvägen 7
Kungliga Tekniska Högskolan
Stockholm

Telephone:
Nat.: 08-411 66 60
Int.: +46 8 411 66 60

Fax:
Nat.: 08-791 71 30
Int.: +46 8 791 71 30

INTRODUCTION

The ability of hearing-impaired patients to recognize speech may be divided into three parts, i.e. the speech reception threshold in quiet (SRT), the speech reception threshold in noise (S/N-threshold) and some speech recognition score at a comfortable speech level for monosyllabic words. Results of the two first seem to be rather independent of each other (Smoorenburg, 1992). Also the two latter are not highly correlated (Hagerman, 1984). Therefore it is a need for efficient clinical methods for measurements of all these three variables.

In the following **SRT** symbolises the **speech reception threshold in quiet** and **S/N-threshold** symbolises the **speech reception threshold in noise**.

Hagerman (1982a) developed a speech and noise material for measurement of the S/N-threshold. It consists of eleven lists, each with ten sentences of five words, e.g. *Peter kept seven old buttons, Karin has four black baskets*. The homogeinity of the lists is very good since they all use exactly the same recording of words. New lists were made from the recorded one by combining the words differently in the computer. The noise has the same average spectrum as the speech. On the whole it is possible to get very fast and reliable S/N-thresholds with this material. For normal hearing subjects the standard deviation for repeated S/N-thresholds was 0.44 dB using two lists taking less than four minutes (Hagerman, 1982a).

However, the threshold method recommended to the clinics, and earlier used to evaluate this speech material, might be further improved. In that method a whole list was used for each signal-to-noise ratio, which was changed in 3-dB steps in order to find the S/N-threshold (defined at 50% recognition) by interpolation. Since each word is scored separately it would be possible to use a sophisticated, adaptive, stepping procedure for each sentence.

A couple of such procedures were tested by computer simulation and one of them was chosen for a clinical study. Test of a corresponding adaptive procedure for measuring the SRT with the same speech material was also included in the study.

The aim of the study was to examine the test-retest reliability of these adaptive threshold procedures for hearing-impaired subjects at a clinic.

METHODS

Adaptive methods evaluated

S/N-threshold

For each sentence of the actual lists (Hagerman, 1982a) a score between zero and five correctly repeated words is possible. Depending on the score obtained for a certain sentence the S/N-ratio should be changed in such a way for the next sentence in the threshold procedure so as to obtain a score as close to the 50%-threshold as possible. For a sentence of five words, each correct word represents a recognition score of 20%. Since the slope of the psychometric function is 25%/dB for normal-hearing subjects (Hagerman, 1982a) and somewhat less steep for hearing-impaired subjects (Hagerman, 1984), a change of about 1 dB for each correct word should be appropriate to reach the threshold from one sentence to the next.

The 50%-threshold represents 2.5 words correct out of five. Since many impairments will give maximum recognition scores less than 100% also at the best speech level and S/N, the psychometric function will often be steepest at a score lower than 50%. From that point of view it is advantageous to focus on a sligthly lower score than 50% (Hagerman, 1979). For the reasons above the following stepping scheme was used:

Nos. of correct words	0	1	2	3	4	5
Change of noise level, dB	-2	-1	0	+1	+2	+3

However, due to some training effects (Hagerman, 1982a), a whole training list was used at the start of the session. The following S/N-ratios were used for the training list.

Sentence No.	1	2	3	4	5	6
S/N dB	+20	+10	+5	0	-5	-8

After sentence no. 6 of the training list, or earlier, as soon as two correct answers or less were obtained, adaptive stepping according to the first scheme was used. For subsequent lists the adaptive scheme was used and the starting level chosen according to the score obtained at the last sentence of the preceding list. Thresholds for five consecutive full lists were calculated according to the method described later.

Speech reception threshold in quiet (SRT)

For SRT the steepness of the psychometric function is only about half of the steepness of the function in noise (Hagerman, 1982a). Therefore the following stepping scheme was used for SRT:

Nos. of correct words	0	1	2	3	4	5
Change of speech level, dB	+4	+2	0	-2	-4	-6

No particular training was used for measurements of SRT, since the demand of very high reliability is less pronounced here. At the start two sentences were presented about 20 dB above the pure tone average threshold (PTA) of 500, 1000 and 2000 Hz. After that the stepping followed the adaptive scheme. Altogether three full lists were always run. However, only six sentences were used for calculating each of three consecutive thresholds. The calculation of the first threshold started with the first sentence where a score of two or less correct answers were obtained. On average over the 50 subjects 7.6 sentences were required to get down to a sentence score of two words or less, i.e. before the first threshold calculation started.

Calculation of thresholds

The calculation of thresholds was done for each list according to Plomp and Mimpen (1979). That is, the S/N-threshold was defined as the average of the S/N-ratios used for the actual sentences (or speech levels for SRT). However, since the score of a sentence influences the level of the **next** sentence, the S/N-ratios for sentence number 2 to 11 were actually used, where the last sentence (No. 11) was not necessarily presented. In the following example a speech level of 40 dB was used:

Sentence No.	1	2	3	4	5	6	7	8	9	10
Noise level, dB	48	47	48	49	47	47	49	48	46	47
S/N dB	-8	-7	-8	-9	-7	-7	-9	-8	-6	-7
Nos. of corr. words	1	3	3	0	2	4	1	0	3	1

The score of one correct word at the last sentence implies a S/N-ratio of -6 dB for the next sentence. Thus the threshold result is

(-7-8-9-7-7-9-8-6-7-6)/10 = -7.4 dB

The SRT was calculated correspondingly, but averaging only six levels for each threshold.

SRT for spondées

For comparison the SRT was also measured with **spondées** according to SAME (1983). In that procedure a couple of words are presented about 20 dB above the PTA. The level is reduced 5 dB for each word until difficulties to recognize arises. Then 10 words are presented at that level. The next level is chosen 5 dB higher or lower to get two scores for 10 words at each side of the 50% score. The 50%-threshold is interpolated from these two scores.

Subjects

Ten normal-hearing and 40 hearing-impaired subjects participated in the study. The normal-hearing subjects (NH) had thresholds not exceeding 10 dB HL in the range 125-8000 Hz with one exception who had threshold values of 15 dB at 4 and 8 kHz. The hearing-impaired subjects (HI) had sensorineural hearing losses confirmed by bone conduction measurements. Almost all audiograms were sloping towards high frequencies. Six were diagnosed as noise induced impairments. No ear needing contralateral masking was tested. The means and the standard deviations of the thresholds are shown in Table I.

Table I. Means and standard deviations for the pure tone thresholds of the hearing-impaired subjects (N=39, subject No. 46 excluded).

Freq	.25	.5	1	2	3	4	6	8	kHz
Mean	13.7	16.0	21.3	37.9	51.4	62.2	65.6	62.7	dB
SD	10.6	13.3	16.0	18.2	15.9	15.2	17.7	18.8	dB

Equipment

The experiment was run in an ordinary sound-proof room normally used for ordinary clinical audiometry. The speech material was presented from a compact disc player through an audiometer (Interacoustic, AC30). The speech and the noise (when used) were mixed in the audiometer. The subject listened monaurally through the headphone (TDH-39 with MX 41AR cushions). The calibration of the equipment was checked before the experiment.

The accuracy of the attenuators is crucial to ensure good repeatability of the measurements of S/N-threshold. However, the adaptive method (using many levels) should be less sensitive to such problems compared to the earlier method using interpolation between two levels. Within any decade of the attenuator that was used in the investigation, the relative error was not larger than 0.1 dB.

However, there was a problem regarding the absolute level due to a functional error of the audiometer. It was shown after the experiment that the absolute level was raised approximately 6 dB in the right channel and approximately 3 dB in the left channel when the two channel were mixed, as they are when measuring S/N-threshold. This does'nt matter since the comfortable level was used in the measurement of S/N-threshold and the absolute level is of no particular interest for this case. However, it precludes the comparison between absolute values of SRT with spondées and SRT with sentences, since it is not known whether a mixed setting of the channels was used or not for any of those measurements.

Procedure

The subjects were separated into two groups, A and B. Group A started with SRT, first with spondées and then with the sentences. After an intermission of at least five minutes, the testing continued with S/N-threshold. Group B was tested in reversed order (see Table II).

Table II. Presentation order (TR = training list).

Measurement	List No.		Order No.	
	Grp A	Grp B	Grp A	Grp B
SRT spondées	spondée	spondée	1	9
SRT sentences	TR	9	2	10
SRT sentences	2	10	3	11
SRT sentences	3	11	4	12
Assessm. MCL	4	TR	5	1
Pause			6	8
S/N-training	5	2	7	2
S/N-threshold	6	3	8	3
S/N-threshold	7	4	9	4
S/N-threshold	9	5	10	5
S/N-threshold	10	6	11	6
S/N-threshold	11	7	12	7

As seen, the two groups got different lists for the different tasks.

The S/N-threshold was measured at a comfortable level of speech. This level was estimated using one sentence list. Then the speech level was fixed and the S/N-ratio was varied by varying the noise level according to the method specified above.

Each subject listened to ten sentence lists altogether, three for measurement of SRT, one for assessment of comfortable level, one for training in noise, and five for measurement of S/N-threshold.

RESULTS AND DISCUSSION

The data were analysed by means of the statistical computer program STATGRAPHICS.

S/N-threshold

Normal-hearing subjects

Group B started with the S/N-threshold. Therefore, the normal-hearing subjects in group B showed a **learning effect** for that measurement of 0.13 dB per list. That is, the S/N-threshold got on average 0.13 dB better for each list. Group A was measured with three sentence lists in quiet, before the S/N-thresholds. The corresponding value for the normal-hearing in group A was -0.16 dB per list, i.e. rather a fatigue effect since they got this test as the later part of the experiment. These values represent the slopes of the linear regression lines calculated (with inverted sign, since a better performance is related to a lower threshold value). The learning effect of 0.13 dB per list fits reasonably well with unpublished data by the author that showed a linear slope of 0.09 dB per list for 15 lists presented to 17 normal-hearing subjects. It is, however, only half of the value found in an earlier study (Hagerman, 1982a), where the learning effect was 0.26 dB per list for the four first lists.

Earlier results with the old threshold method (Hagerman, 1982a) showed a **mean value** for the first threshold of -7.1 dB for normal-hearing subjects. This was obtained after one training list and two lists to assess the threshold.

The corresponding result for the new method was -7.8 dB (N=10, group A and B). This result fits very well with computer simulations that indicated about a 0.6 dB better threshold for the new, adaptive method compared to the old one (mainly due to threshold detection around 40% rather than 50%). However, one normal-hearing subject in group B showed a pathological threshold value of about -3 dB. On the other hand, half of the subjects (group A) had some extra learning during the measurement of SRT.

The **test-retest reliability** was obtained from the error term of the one way analysis of variance, which was 0.39. The standard deviation of repeated measurements should then be $\sqrt{0.39} = 0.62$. To be compared with the standard deviation of the old method this figure has to be divided by $\sqrt{2}$ since the old method used **two** lists for a threshold. $0.62/\sqrt{2} = 0.44$ which is exactly the same as for the old method (Hagerman, 1982a). However, in the old result the learning effect is outbalanced, since it was calculated as the standard deviation of the difference between the two first thresholds. In that case the learning effect is obtained as the **mean** value of the difference. In the present result the learning effect is included. Therefore, the new method definitely gives more reliable results for normal hearing subjects.

Hearing impaired subjects

As for the clinical result with the old method (Hagerman, 1984) it was for many reasons appropriate to divide the patients into **two groups**. The **better group** had S/N-thresholds < 0 dB (N=33) and the **worse group** had S/N-thresholds > 0 dB (N=6).

There was one hearing impaired subject (No. 46) with a diagnosis of retrocochlear hearing loss. His five S/N-thresholds varied between +4.3 and +9.5 dB. The high variability was probably due to a very flattened intelligibility curve, since his speech recognition score on a 50 word PB-list was only 60%. The method of measuring S/N-threshold is not well suited for patients with such low speech recognition scores in quiet. Therefore this subject was omitted from the analyses of S/N-thresholds.

The **learning effect** for the better group was on average 0.07 dB per list among the patients starting with threshold in noise, group B, but only 0.04 dB per list for those who heard some lists without noise first, group A.

However, for the worse group the learning effect was as big as 0.5 dB per list (N=6, group A plus group B). This shows that the learning effect is almost negligible for patients with S/N-thresholds < 0 dB. However, for patients with S/N-thresholds > 0 dB the learning effect is substantial, which is in accordance with Hagerman (1984, Fig. 8).

The **test-retest reliability** in terms of the error term of the one way analysis of variance was 0.62 for the better group. From that value the standard deviation of repeated measurements was calculated and compared to the corresponding value obtained with the old method (Hagerman, 1984). See Table II. As a matter of fact, for hearing-impaired subjects the new method gives almost as good reliability for thresholds obtained with one list only as the old method gave for thresholds obtained with two lists. Still the learning effect is included in the results for the new method but not for the old. Therefore, the new method definitely gives more reliable results for all patients as well as for normal-hearing subjects.

Table II. Test-retest reliability for S/N-threshold measurement using one or two lists. The variances are equal to the error term of the one way analysis of variance. The standard deviations (SD) are calculated from the variances.

Group	**normal-hearing**	**hearing-impaired** better	worse
N	10	33	6
Variance	0.39	0.62	1.50
SD using 1 list	0.62	0.78	1.23
SD using 2 lists	0.44	0.55	0.87
SD using 2 lists, old	0.44	0.71	1.1

Prediction of the reliability of an individual result

The reliability is to some degree related to the threshold result itself. This is seen in Figure 1. However with this new threshold method there should be another way to predict the variability. Since the threshold value is based upon the mean of ten attenuator values, one for each sentence, the threshold variability should be reflected in the variability of the attenuator settings. More specifically, the standard deviation of repeated threshold results should be predicted by the standard error of the mean of the attenuator values, i.e. the standard deviation of the attenuator values divided by the square root of 10. However, doing this, a slightly worse prediction than the prediction from absolute threshold shown in Fig. 1 was obtained. One reason for that is probably that the learning effect shows a rather big variability between subjects. Another reason may be that the learning effect is not completely linear with time. Therefore it might not be accounted for to the right degree in just one list.

Significance limits when comparing two results

In comparing two results obtained (e.g. comparing the two ears), it is of interest to know whether the difference between the results is significant, or whether it depends on hazard only. The significance limits were obtained by multiplying the standard deviations in Table II with the square root of two and with 1.64 ($p<0.01$, two-sided test and normal distribution presupposed). See Table III.

Table III. Required difference in dB between two S/N-thresholds for significance (~~p<0.01~~, two-sided test and normal distribution presupposed).
p<0,1

Mean threshold	< 0 dB	> 0 dB
1 list per threshold	1.8	2.9
2 list per threshold	1.3	2.0

SRT

Absolute thresholds

Due to the uncertainty of the absolute calibration at certain settings (see above) it was not possible to get an exact relation between the absolute values of the thresholds with spondées and with sentences. The difference between the mean spondée threshold and the mean sentence threshold for the whole group (N=50) was 4.7 dB. It was expected to lie between 3 and 6 dB, if the two channels of the audiometer were mixed during the measurements, which they probably were. The mean threshold for normal-hearing subjects is 22 dB SPL for spondées and it was 21.6 dB for the sentences with the old method (Hagerman, 1982a). With the new method the score defining the threshold is shifted from 50% to 40%. With a slope of the psychometric function of 10%/dB (Hagerman, 1982a) the threshold would change to about a 1 dB better value. Thus the new method is expected to give thresholds that are 1.4 dB better than that for the spondées.

Learning effect

The learning effect when measuring SRT using the sentences is shown in Table IV and in Figure 2 for the various groups. It was expected to be greatest in group A, which started with SRT. However, this was not the case for the hearing-impaired subjects. The improvement from the first to the second threshold in quiet, was largest for the hearing-impaired subjects in group B (5.7 dB). The reason may be that group B got a temporary threshold shift from the initial measurements of S/N-threshold with a gradual release during the following measurements in quiet. Normally the SRT is measured before loading the ear with any noise. Therefore the result of group A is more interesting. The 3.0 dB learning effect of the two first thresholds for the hearing-impaired subjects in group A can be compared with the corresponding value of 1.8 dB for spondées shown by Hagerman (1982b). The difference is probably due to some learning of the limited word material (cf. Hagerman, 1984, Fig. 1).

Table IV. Mean values (dB) of SRT for consecutive measurements showing the learning effect for various groups. These results are also illustrated in Figure 2.

	Normal hearing		Hearing-impaired	
	grp A	grp B	grp A	grp B
N	5	5	20	20
Threshold No. 1	3.5	1.7	19.6	13.2
Threshold No. 2	1.5	1.3	16.6	7.5
Threshold No. 3	1.5	0.1	15.1	6.8

Reliability

The test-retest reliability was obtained from the error term of the analysis of variance. Since the learning effect was not negligible in this case, one analysis was also made with the threshold number as a covariate in order to outbalance the mean learning effect. The results are shown in Table V.

Table V. Standard deviation (dB) of repeated measurements of SRT using six sentences obtained as the square root of the error variance from the analysis of variance. In the last row, the learning effect was outbalanced by using the order number of measurement as covariate in the analysis.

	Normal hearing		Hearing-impaired	
	grp A	grp B	grp A	grp B
N	5	5	20	20
Learning included	1.81	1.32	3.17	4.87
Learn. outbalanced	1.59	1.09	2.26	3.69

Also here the most relevant results are those obtained for group A for the same reasons as above. According to Hagerman (1982b), the reliability for ordinary spondée thresholds varies between 2.6 and 4.0 dB for various degrees of hearing impairment. This is for only two thresholds without the learning effect included. The learning effect will probably be better outbalanced using only two thresholds rather than three, since the learning is highly nonlinear with time as shown in Figure 2. This was checked for the hearing-impaired subjects in group A using only the two first thresholds, which gave a standard deviation for repeated measurements of 1.62 dB. Even if the sum of the learning effect and the standard deviation is taken as the error of measurement, the new adaptive method with sentences is at least as good as the old spondée method. The sum is in both cases around 4.6 dB. Furthermore, the old method takes on average 3 minutes and 10 seconds (on average 6+24 words are needed, Hagerman, 1982b), while the new one takes only 2 minutes and 35 seconds (only 14 sentences were used on average, including training). Therefore the new method would be preferred.

Check of psychogenic hearing loss is an important use of SRT measurements comparing the result with the pure tone average threshold. It might be more effective with the sentence material. Remembering five words at a time needs more attention than only one word. Therefore it is probably more difficult to simulate a hearing loss at a SRT measurement with sentences.

CONCLUSIONS

1. The test-retest reliability of the S/N-threshold was much better with the new, adaptive, method than with the old one which was not adaptive.

2. With the new method a S/N-threshold can for most patients be measured within two minutes (after the initial short training) with a reliability, i.e. standard deviation, of 0.78 dB and an almost negligible learning effect.

3. Using a similar method in quiet the sentences give at least as reliable SRT result as the old spondée method (including learning effects) but in shorter time.

ACKNOWLEDGMENTS

The authors wish to thank Ann-Cathrine Lindblad for helpful comments.

This work was supported by the Swedish Council for Social Research.

REFERENCES

Hagerman, B. 1979. Reliability in the determination of speech reception threshold (SRT). *Scand Audiol 8,* 195-202.

Hagerman, B. 1982a. Sentences for testing speech intelligibility in noise. *Scand Audiol 11,* 79-87.

Hagerman, B. 1982b. Measurement of speech reception threshold. *Scand Audiol 11,* 191-193.

Hagerman, B. 1984. Clinical measurements of speech reception threshold in noise. *Scand Audiol 13,* 57-63.

Plomp, R. & Mimpen, AM. 1979. Improving the reliability of testing the speech reception threshold for sentences. *Audiology 18,* 43-52.

SAME 1983. Metodbok i praktisk hörselmätning. LIC förlag, Solna. In English: Manual of Practical Audiometry, Volume 1, ed. Stig Arlinger, Whurr Publishers Ltd, London, 1989.

Smoorenburg, GF. 1992. Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram. *J Acoust Soc Am 91(1),* 421-437.

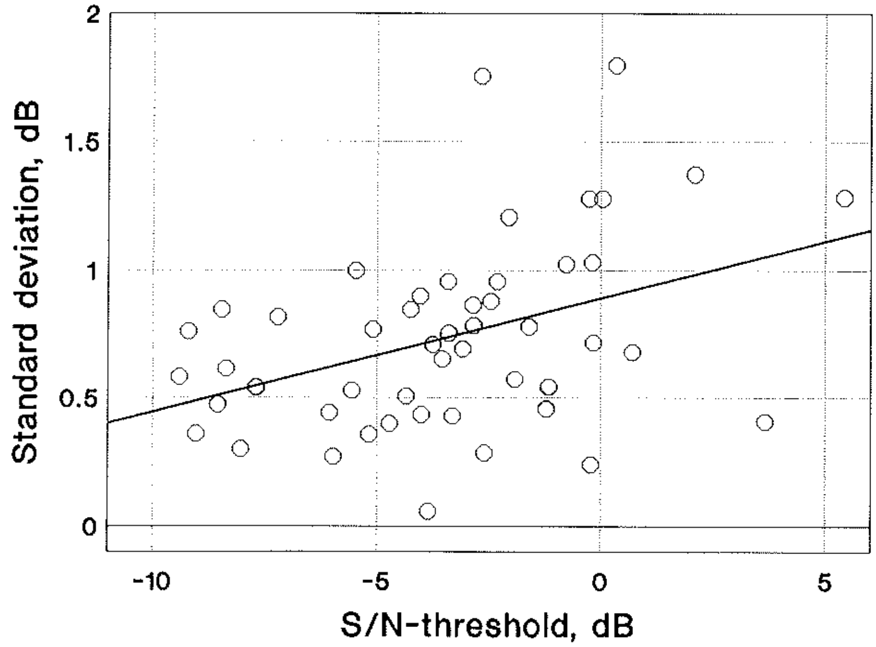

Figure 1. Relation between S/N-threshold (average for each subject) and standard deviation of five consecutive measurements. Regression line with intercept 0.89 dB and slope 0.0447. N=49.

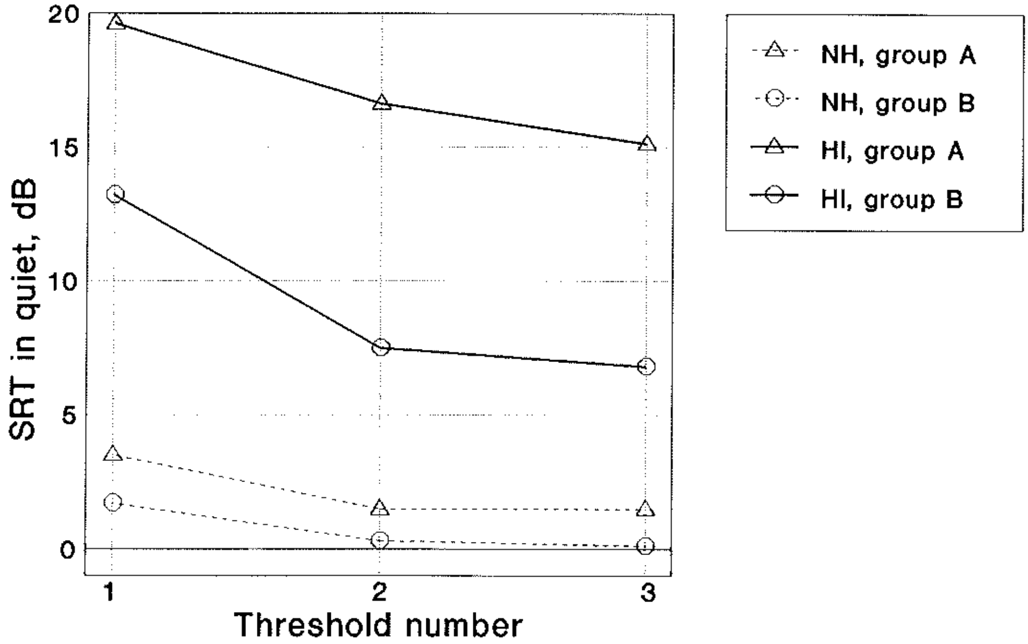

Figure 2. SRT as a function of No. of measurement showing the learning effect for various subject groups.