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

Report TA113
June 1986

RELATIONS BETWEEN SPEECH INTELLIGIBILITY AND PSYCHOACOUSTICAL MODULATION TRANSFER FUNCTION (PMTF)

Björn Hagerman, Åke Olofsson and Ann-Cathrine Lindblad

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

ISSN 0280-6819
Report TA113
June 1986

RELATIONS BETWEEN SPEECH INTELLIGIBILITY AND PSYCHOACOUSTICAL MODULATION TRANSFER FUNCTION (PMTF)

Björn Hagerman, Åke Olofsson and Ann-Cathrine Lindblad

ABSTRACT

The modulation transfer function, MTF, has proved to be a powerful measure for predicting speech intelligibility in speech transmission channels. We extended it to include the ear, by measuring the psycho-acoustical MTF, i.e. the PMTF.

Tone thresholds of 11 normally hearing and 20 hearing impaired subjects were measured in presence of unmodulated and intensity modulated noise. The octave frequencies from 500 to 4000 Hz were used. The noise was octave filtered around the frequency of the probe tone. Six modulation frequencies from 1 to 50 Hz were used. From these results the PMTFs were calculated, as well as the corresponding psycho-acoustical speech transmission indices, i.e. the PSTIs. The subjects' speech discrimination scores and speech reception thresholds in noise were also measured.

A correlation coefficient of 0.85 between the speech discrimination score and the PSTI was obtained. For the speech reception threshold in noise and the PSTI the correlation was 0.71. The first of these two figures is promising, but our method needs some improvement since it gave some problems due to fatigue effects.

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

INTRODUCTION

As an alternative to speech intelligibility tests objective methods have been used for evaluating the quality of speech transmission channels. French and Steinberg (1947) derived a measure called the Articulation Index (AI). It is based on the speech levels in different frequency bands. Peutz (1971) developed a method to evaluate the influence of room acoustics on the speech intelligibility. Steeneken and Houtgast extended the AI into the time-domain by including intensity modulations and by measuring the remaining modulations after the transmission channel. The reason was that the amplitude modulations in speech carry important information and therefore should be taken into account. The resulting speech transmission index (STI) predicts speech intelligibility well for many different types of distortion, for example reverberation, for which the AI method was not particularly well suited (Steeneken & Houtgast, 1980).

The STI calculated from the modulation transfer function (MTF) seems so powerful that it is tempting to extend it to be a tool for predicting the listener's ability to recognize speech. Then it would be possible to assess the speech discrimination ability of a hearing-impaired listener without using speech stimuli, e.g. for foreign patients with language difficulties. Perhaps it would also be possible to use the method for hearing aid fitting, using gain values which best preserve the modulations for each frequency band.

If a noise carrier is fully intensity modulated with a sinusoid and received by the ear, the neural representation of the signal will not truly reflect the original. Backward and forward masking will fill in the silent valleys to some extent, reducing the modulation index from 1 to a lower value, depending on the modulation frequency. The depth of the valley can be estimated by measuring the pure tone threshold in presence of such a modulated noise signal. Similarly the peak can be estimated by measuring the tone threshold in presence of the same, but now unmodulated, noise signal. From these two values a corresponding modulation index and thus the psychoacoustical modulation transfer function (PMTF) can be calculated.

The aim of this investigation was to try out a method of measuring the PMTF, simple enough to be used clinically, and to compare the resulting psychoacoustical STI (PSTI) with the speech discrimination score without noise and with the speech reception threshold in noise.

METHODS

Subjects

The measurements were performed on 11 normally hearing and 20 hearing impaired subjects with sensorineural impairment. Most of the normally hearing subjects were chosen among the staff at our department. Their pure tone thresholds were 20 dB HL or better from 250 to 4000 Hz (one exception, 25 dB at 4000 Hz). Among the hearing impaired subjects only those were accepeted, whose threshold differences between the two ears were not too big, in order to avoid the need of contralateral masking. Not to risk that the required sound levels were not available from the equipment, subjects with severe hearing impairment were excluded.

Speech reception threshold (SRT) in noise

For measuring the SRT in noise we used the speech material developed by Hagerman (1982). Each list consists of 10 sentences with 5 words in each sentence. The noise has the same average spectrum as the speech and is slightly amplitude modulated with a low-frequency noise centred around 2 Hz. The modulation index is only .09 (RMS).

First the most comfortable level was assessed with some sentences without noise. Then a training list was presented at this level with the noise raised step by step after each sentence. From the result of the training list a signal-to-noise ratio, S/N, for the first test list was chosen to get a score around 50%. For the next list the noise level was changed 3 dB in a direction to get the two results at each side of 50%. Then the 50% threshold was interpolated from the two results, expressed as a signal-to-noise ratio. In order to achieve a very reliable estimate two thresholds were measured and their average was used as the final threshold value, called the S/N-threshold.

Speech discrimination

The speech discrimination was measured by using the phonetically balanced (PB) lists of monosyllabic words commonly used at the audiological clinics in Sweden. They were recorded at our department in 1966. For the speech intelligibility measurement two 50-word lists (No. 6 and No. 5) were run at the most comfortable level, which was earlier established with the sentences. The PB-discrimination score was thus based on 100 words altogether.

Psychoacoustical Modulation Transfer Function, PMTF

The PMTF refers to the ability of the ear to transfer intensity modulations with different modulation frequencies in the frequency range 0-50 Hz. This ability can be different for different carrier frequencies in the audio frequency range. It can be measured by assessment of the tone threshold in the presence of intensity modulated masking noise. We used a Bekesy audiometer for the threshold measurement of the probe tone. Unfortunately we had to use a continuous probe tone when measuring the threshold in a noise modulated with a low frequency. Otherwise the pulse frequency and the modulation frequency interfered with each other. Thus the probe tone was always continuous with the exception that we also measured two thresholds in unmodulated noise, using both a pulsed and a continuous probe tone. The reason to use also a pulsed tone in the unmodulated noise was that we had observed some fatigue effects in a pilot study preceeding this experiment. The thresholds obtained with a continuous tone in unmodulated noise are called CU-thresholds and those with a pulsed tone in unmodulated noise are called PU-thresholds in the following.

The probe tone frequencies 500, 1000, 2000 and 4000 Hz were used. For each of these, 8 thresholds were measured, namely with the intensity of the masking noise sinusoidally modulated with 1, 2, 5, 10, 20, 50 Hz and in unmodulated noise and finally with a pulsed probe tone in unmodulated noise. The noise was octave filtered around the frequency of the probe tone. For half of the subjects the order of the modulation frequencies was the one mentioned above and for the rest of the subjects the order was reversed. In both groups, however, the two thresholds in continuous noise were always the last measurements for each probe tone frequency. The threshold value was calculated as the mean of the last 12 reversals out of 14. A silent period of at least 10 seconds was inserted between each threshold measurement.

Since the aim of the study was to compare the PMTF-result with the results of the speech tests, the level of the intensity modulated noise was adapted to the speech level in each of the four octaves 500-4000 Hz. The two speech materials were analyzed with a B&K 1615 octave filter and a B&K 2305 level recorder. The writing speed of the pen was 160 mm/s. The mean of the 10 highest peaks for the two materials in each octave was then used as the peak level of the intensity modulated noise in the PMTF-test. Thus the noise levels were set to -5, -9, -12 and -15 dB below the calibration tone on the speech tape for the frequencies 500, 1000, 2000 and 4000 Hz repectively. In this way the noise level to be set for each subject was determined from the most comfortable level used for the speech tests.

From the threshold G in unmodulated noise and the threshold H in modulated noise the modulation transfer function, E, is calculated in the following way:

$X=(G-H)/10$ but if $G<H$ then $X=0$

$D=10**X$ intensity ratio
$**$ denotes exponentiation

$M=(D-1)/(D+1)$ modulation index

$E=10*\log(M)$

This is done for all combinations of carrier and modulation frequencies. For each carrier frequency a modulation transfer function is obtained, as a function of the modulation frequency.

Using the equations above, a speech transmission index can also be calculated:

$S=10*\log(M/(1-M))$ apparent signal to noise ratio

$T=(S+15)/30$ T is limited to $0<T<1$, however

With appropriate weigths (see RESULTS) the different T-values for different carrier and modulation frequencies are summed up to a psychoacoustical speech transmission index, PSTI.

Equipment

The patient listened to the test signals through TDH-39 earphones with MX-41/AR cushions sitting in a sound-proof room (*). Some parts of the equipment were built at our department, denoted here with an (*). So were the power amplifier (*) and the attenuator (*) driving the earphone.

For the speech tests a Revox B77 tape recorder was used with the speech on one channel and the noise on the other. The output signals from the tape recorder were then routed to two attenuators (Hewlett & Packard) and fed to a mixer (*) which made it possible to choose the decided signal-to-noise ratio.
For the PMTF-test a Demlar Bekesy-audiometer delivered the probe tone. The masking noise was taken from a white noise generator (*), routed through an octave filter (B&K 1613) and a programmable attenuator (*), which was controlled by a micro computer (Luxor ABC 80).

Procedure

The measurements were performed individually during two sessions on different days. Both sessions lasted about 45 minutes and included a short intermission. The first session started by measuring the Bekesy pure tone threshold curve, followed by an assessment of the most comfortable level for sentences without noise. Then the two S/N-thresholds were measured. After the intermission the two PB-lists were run, followed by simple measurements of three psycho-acoustical tuning curves, the results of which are not presented here. In the second session all the PMTF-measurements were carried out.

RESULTS

The hearing impaired subjects were separated into two groups considering the slope of the pure tone audiogram. The flat group includes 8 subjects with threshold differences less than or equal to 20 dB in the range 500 to 4000 Hz. The remaining 12 hearing impaired subjects are included in the steep group. The normally hearing subjects form a third group. Figure 1 shows the mean pure tone thresholds of the two hearing impaired groups.

The calculation of the PMTF is based on the difference between the thresholds in modulated noise and in unmodulated noise. It is therefore interesting to study these differences shown in Figures 2-4. The dashed line denotes 500 Hz carrier frequency and for the other lines the number of dots between the dashes denotes number of kHz of the carrier. The 0 dB base line refers to the CU-threshold. At the right end of the figures the differences between the CU- and the PU-thresholds are indicated. For the normal group (Fig. 2) it seems as if the ability of the normal ear to distinguish a tone in modulated noise is much better at higher frequencies of the tone. This is, however, probably wrong. The reason of this result is rather that the fatigue effect was more accentuated at higher frequencies for the continuous tone in unmodulated than in modulated noise. This was seen on the threshold curves of some subjects and was not present for the threshold of the pulsed tone. If the PU-threshold had been used as a reference, the 4000 Hz curve would instead have been worse than the other three.

Fig. 3 shows the corresponding curves for the steep group. The 500 Hz curve is very similar to that of the normal group. For increasing frequencies the results are increasingly worse. The results of the flat group (Fig. 4) do not show as big differences between the different frequencies. The larger hearing loss at 4 kHz for the steep group is reflected as a lower 4 kHz-curve in Fig. 3 compared to the 4 kHz-curve for the flat group in Fig. 4. A corresponding relation can also be seen for 0.5, 1 and 2 kHz. Note also that the PU-thresholds for all groups and probe tone frequencies are better than the CU-threshold (CU-PU>0). The difference increases with increasing frequency.

We were interested in using the threshold difference for calculating the PMTF and the PSTI, and to find the correlation between the PSTI-value and the speech intelligibility measures. If the CU-threshold should be used as a reference when calculating the threshold difference, the result would get too high due to the fatigue problem mentioned above. One could perhaps use some fixed value for each frequency; the same for all subjects. Since the different subjects use different bias in their listening strategy, this way of solving the problem does not seem very promising. Instead we chose to use the PU-threshold as the reference, but with a constant added, in order to get the positive threshold differences necessary for calculating the logarithm. So please note, that the CU-threshold used as the base line in Figs. 2-4 is not used as a reference threshold from now on.

The PMTF was calculated for each subject and for each of the four carrier frequencies 500-4000 Hz. As a reference we used the PU-threshold adding a constant of 7 dB. Then a PSTI-value was calculated, first with equal weights on the four octave frequencies and then with various weights to get higher correlations to the two speech intelligibility scores. (The 7 dB constant was also chosen to get high correlations.) The resulting correlation coefficients are shown in Table I. The weights in the second row are corresponding to those used in the original STI-measure (Steeneken & Houtgast, 1980). They are modified, however, to our range 500-4000 Hz instead of the original 125-8000 Hz. In the two lowest rows are shown the highest correlations obtained for the two different speech intelligibility measures. Scattergrams corresponding to these data are shown in Figs. 5 and 6.

The values of the PMTFs used above to calculate the PSTI:s are shown in Figs. 7-9 for the three groups of subjects. The higher hearing loss at high frequencies for the steep group and at low frequencies for the flat group are reflected also in these figures. From these results, however, it is difficult to draw any conclusions about the cut-off frequency or the slope of the "filter" of the ear that degrades modulations.

DISCUSSION

Attempts have been made before to find a psychoacoustical measure which is highly correlated to various measures of speech intelligibility (Dreschler & Plomp, 1980; Dreschler & Plomp, 1985; Festen & Plomp, 1983). The interesting point is not the SRT (without noise), where good correlations are achieved with suitable linear combinations of the tone thresholds at different frequencies (Noble, 1973). More important are different abilities at higher hearing levels, well above threshold, such as the maximum speech discrimination score for monosyllabic words, where correlations to pure tone thresholds are worse (Noble, 1973), and the SRT in noise. We obtained a high correlation between the PSTI-measure and the speech discrimination score (.85), but are somewhat puzzled about the weighting factors needed. Results for 500 and 4000 Hz were not used at all and 1000 Hz was much more important (.7) than 2000 Hz (.3). If more subjects with bad speech discrimination scores had been used, perhaps other weighting factors migth have been optimal. However, also with equal weighting factors the correlation coefficient was high (.79), keeping in mind the high variability of speech discrimination scores (Hagerman, 1976).

The correlation coefficient .71 obtained by us between PSTI and SRT in noise is of the same order as the corresponding value, .75, between the SRT in noise and the mean critical ratio at 500, 1000 and 2000 Hz, obtained by Dreschler & Plomp (1985).
It may be astonishing that the PSTI-measure gives a lower correlation to the S/N-threshold than to the PB-discrimination score, although the PSTI-measure is based on a threshold in noise. This measure, however, is based on a fully modulated noise and should reflect the masking effects of adjacent phonemes on each other - that is, it is more or less a simulation of speech. In the measurement of the S/N-threshold we go one level further. There is the real speech and we add noise as an extra disturbance. In our experiment this noise was only slightly modulated and does not correspond at all to the fully modulated noise used for the PMTF-measurement. Furthermore, for hearing impaired subjects the speech discrimination score is not very highly correlated to the S/N-threshold (Hagerman, 1984).

Due to fatigue problems, it was necessary to use a pulsed tone when measuring the reference threshold in an unmodulated noise. On the other hand we had to use a continuous tone for the threshold in modulated noise, in order not to confuse the patient with two different pulsating signals. This is unsatisfactory since it made it necessary to add a constant to the reference threshold. There is no theoretical ground for such a constant. Another problem is that the critical ratio is not as good for hearing impaired persons as for normally hearing. That is, their threshold in unmodulated broad band noise is worse, which means that their reference threshold for the PMTF is worse. This fact gives them a better PMTF unless the threshold in modulated noise is influenced to the same extent. (This problem is probably related to the fatigue problem.) However, a bad critical ratio presumably degrades the ability to discriminate speech. Instead of measuring the reference threshold for each patient, one might use a standard reference threshold obtained for normally hearing subjects. That should solve the reference problem but when we tried this the correlation got worse. The reason might be that the different subjects had different biases in their threshold concepts. This bias will be neutralized if two thresholds from the same subject are subtracted, but it may cause problems when the patient's threshold is subtracted from a standard value.

The method of measuring the PMTF used by Ahlstrom (1984) on normally hearing subjects might be worth trying also on hearing impaired. He measured the thresholds of short tone pips placed either in the peak or in the trough of the modulated noise and used the difference of these two thresholds. That made the threshold measurement in unmodulated noise superfluous and should have decreased the fatigue problem. This method, however, requires somewhat more sophisticated equipment.

CONCLUSIONS

1. There is a high correlation between the PSTI-measure derived from the PMTF and the PB-discrimination score.
2. The correlation between the PSTI and the S/N-threshold is only moderate.
3. The method used in this investigation has some drawbacks, mainly due to fatigue effects, and should be modified.
4. The PMTF seems to be an important measure of the ability of the ear to analyze speech sounds.

ACKNOWLEDGEMENTS

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

REFERENCES

Ahlstrom, C. 1984. Psychoacoustical modulation transfer functions and speech recognition in listeners with normal hearing. Thesis. Vanderbilt University, Nashville, Tennessee.

Dreschler, W.A. & Plomp, R. 1980. Relation between psychophysical data and speech perception for hearing-impaired subjects. I. J Acoust Soc Am 68(6), 1608-1615.

Dreschler, W.A. & Plomp, R. 1985. Relation between psychophysical data and speech perception for hearing-impaired subjects. II. J Acoust Soc Am 78(4), 1261-1270.

Festen, J.M. & Plomp, R. 1983. Relations between auditory functions in impaired hearing. J Acoust Soc Am 73(2), 652-662.

French, N.R. & Steinberg, J.C. 1947. Factors governing the intelligibility of speech sounds. J Acoust Soc Am 19, 90-119.

Hagerman, B. 1976. Reliability in the determination of speech discrimination. Scand Audiol 5, 219-228.

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

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

Noble, W.G. 1973. Pure-tone acuity, speech-hearing ability and deafness in acoustic trauma. Audiology 12, 291-315.

Peutz, V.M.A. 1971. Articulation loss of consonants as a criterion for speech transmission in a room. J Audio Eng Soc 19, 915-919.

Steeneken, H.J.M. & Houtgast, T. 1980. A physical method for measuring speech-transmission quality. J Acoust Soc Am 67, 318-326.

PSTI-weights					Correlation coefficients	
	frequency, kHz					
	.5	1	2	4	PB-score	S/N-threshold
equal	.25	.25	.25	.25	.79	-.66
orig STI	.195	.195	.318	.292	.77	-.65
optimal	.0	.7	.3	.0	.85	
optimal	.0	.7	.0	.3		-.71

Table I. Correlations between the PSTI-value and the two speech intelligibility measures. The PSTI-value is calculated with various weights for the octave frequencies, but with equal weights for all the modulation frequencies.

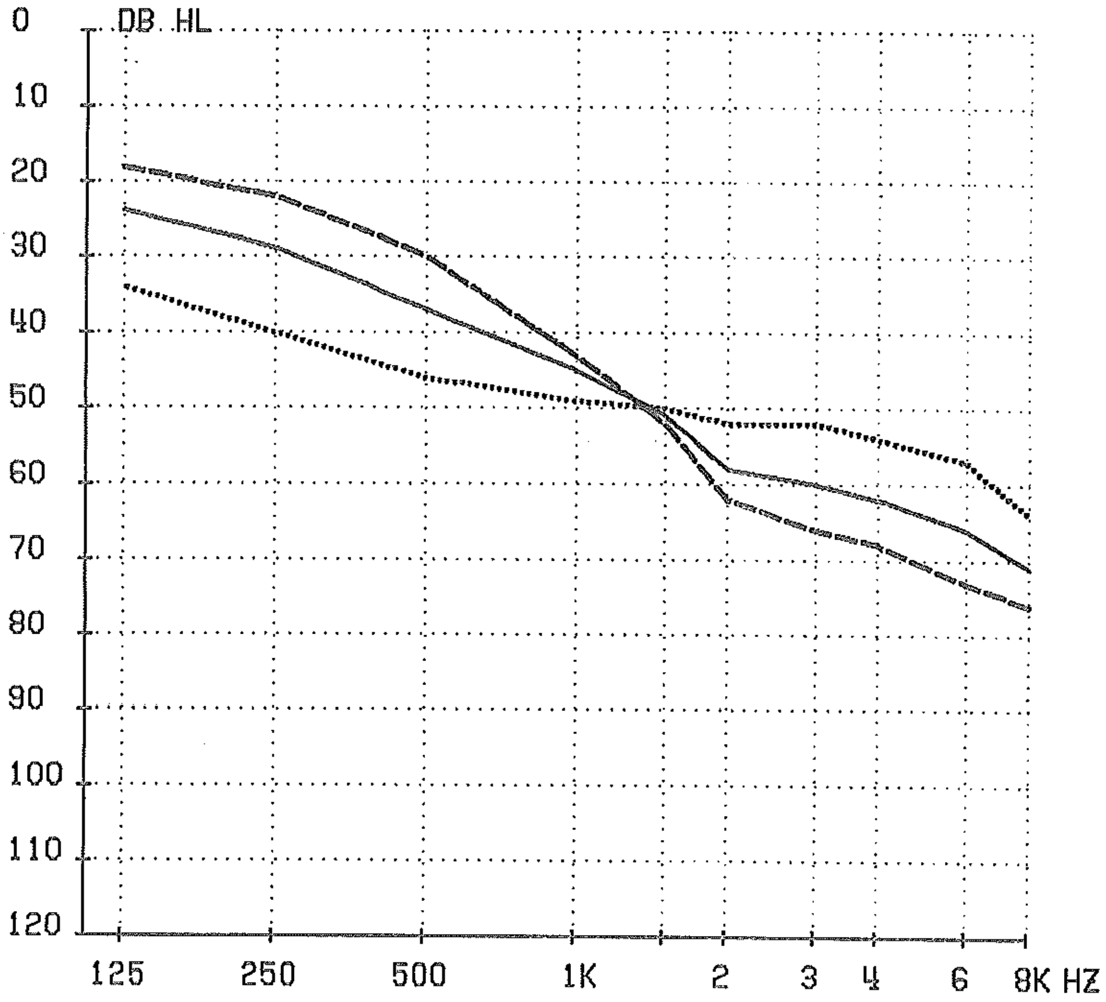

Fig. 1. Mean pure tone thresholds of the hearing impaired subjects. All the hearing impaired subjects, N=20, continuous line. Steep group, N=12, dashed line. Flat group, N=8, dotted line.

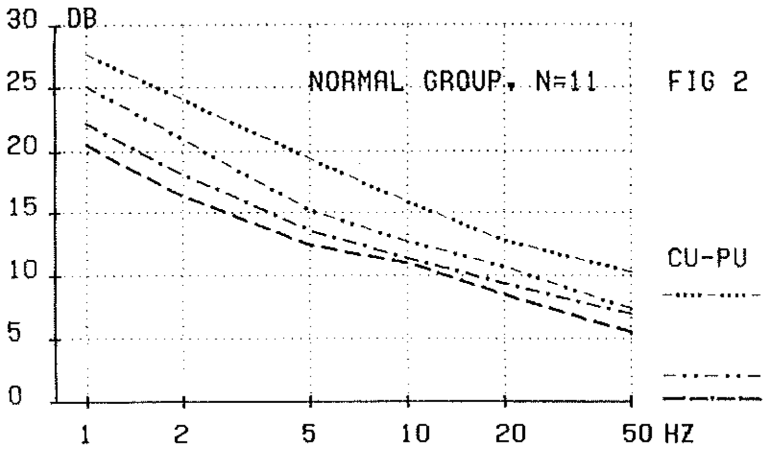

FIG 2

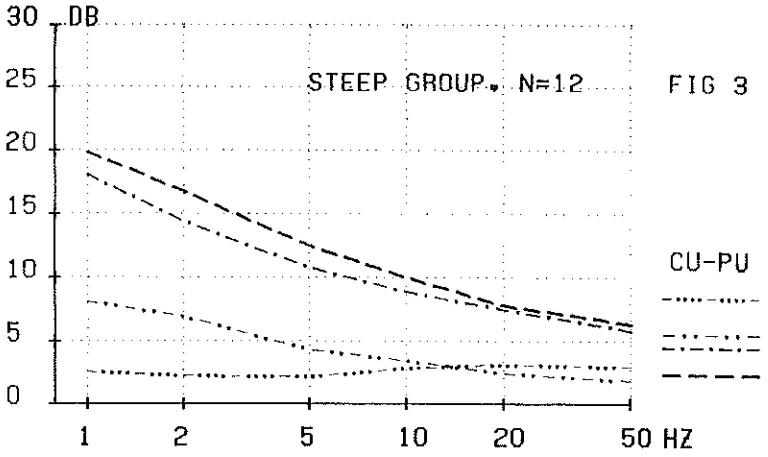

FIG 3

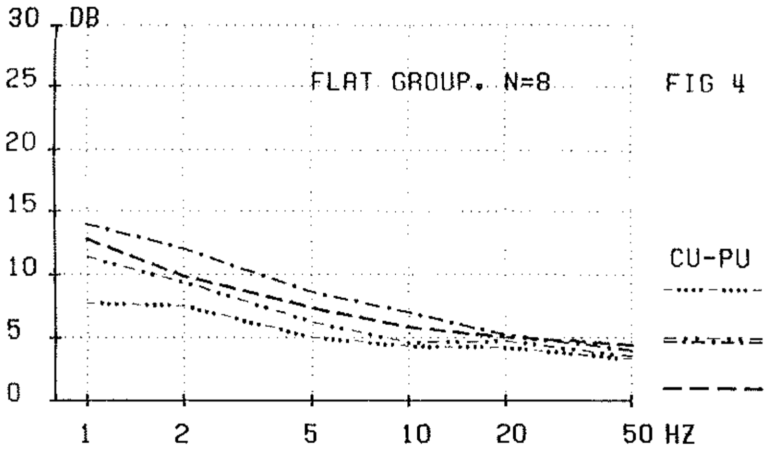

FIG 4

Fig 2-4. Differences between thresholds in continuous noise and in modulated noise for the three groups of subjects. The 0 dB base line represents the CU-thresholds.
- - - - - - 500 Hz probe tone.
-.-.-.-.- 1 kHz probe tone.
-..-..-..- 2 kHz probe tone.
-....-....- 4 kHz probe tone.
The lines to the right in the figures show the differences between the CU- and the PU-thresholds.

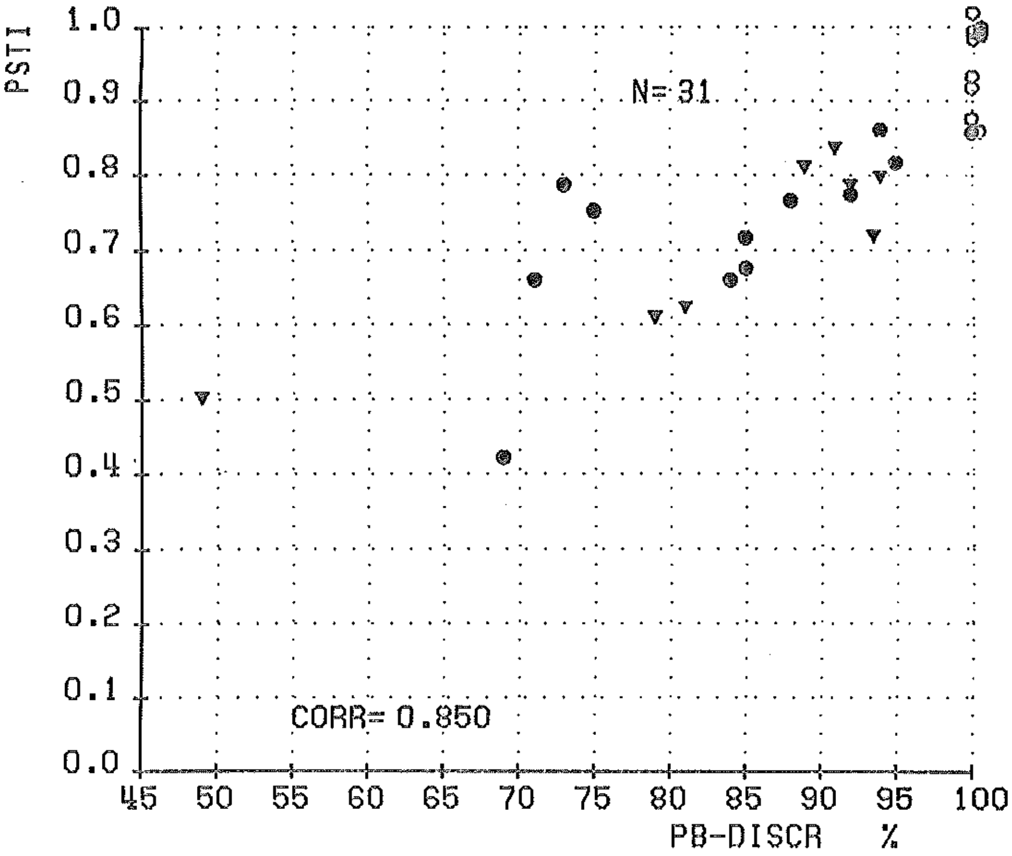

Fig. 5. Scattergram of the relation between
PB-discrimination score and PSTI.
Open circles - normally hearing group.
Filled circles - steep group.
Filled triangles - flat group.
All the normally hearing subjects are assumed to have a discrimination score of 100%.

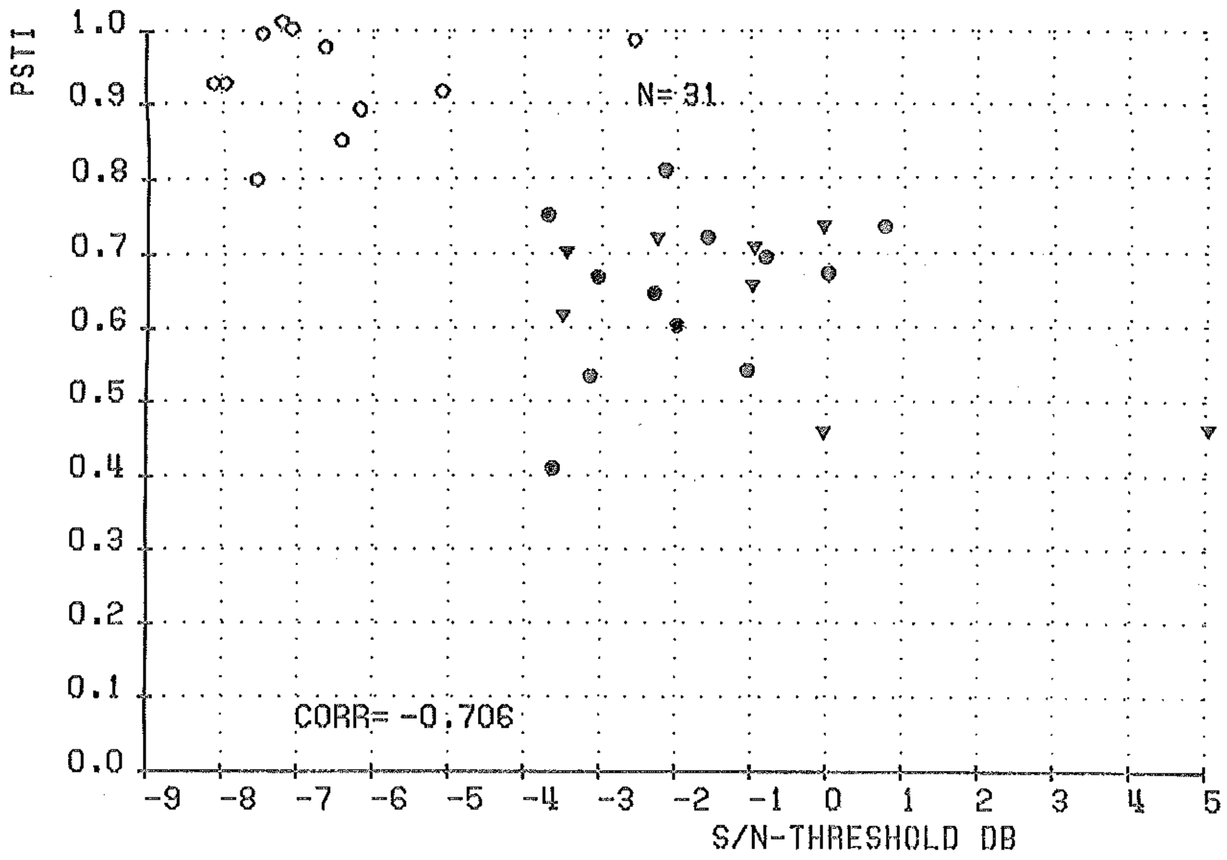

Fig. 6. Scattergram of the relation between S/N-threshold and PSTI.
Open circles - normally hearing group.
Filled circles - steep group.
Filled triangles - flat group.

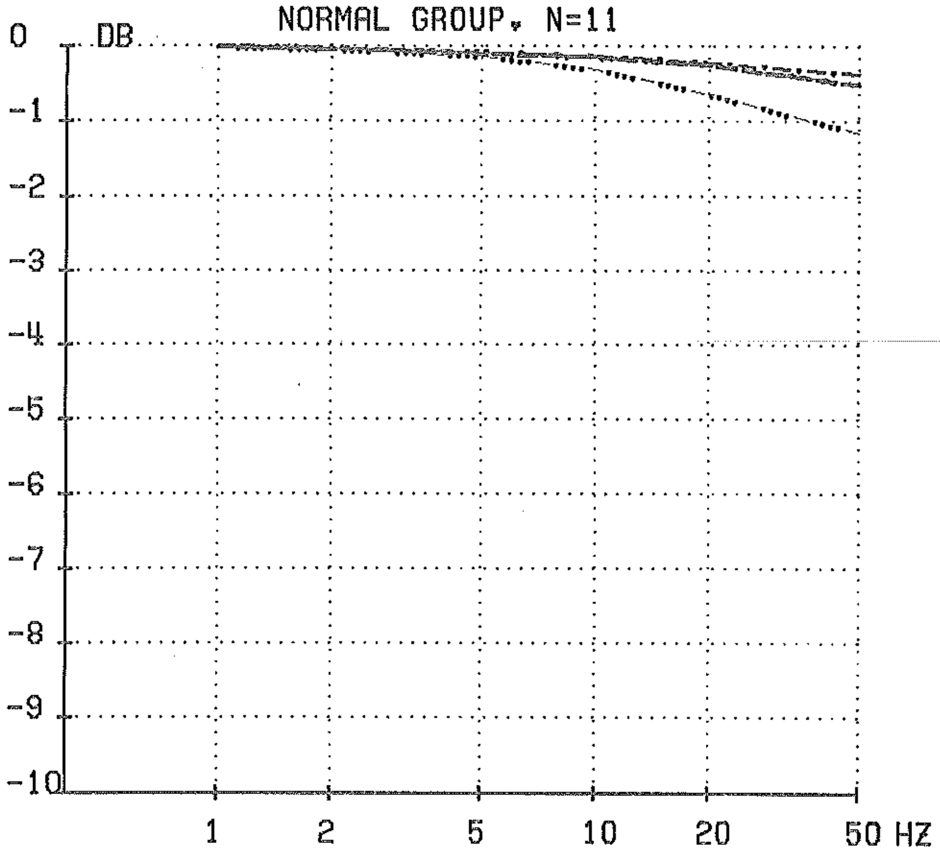

Fig. 7. Mean psychoacoustical modulation transfer functions (PMTF:s) for the normally hearing group (N=11), calculated with the reference level at the PU-threshold + 7 dB.
- - - - - 500 Hz probe tone.
-.-.-.-.- 1 kHz probe tone.
-..-..-..- 2 kHz probe tone.
-....-....- 4 kHz probe tone.

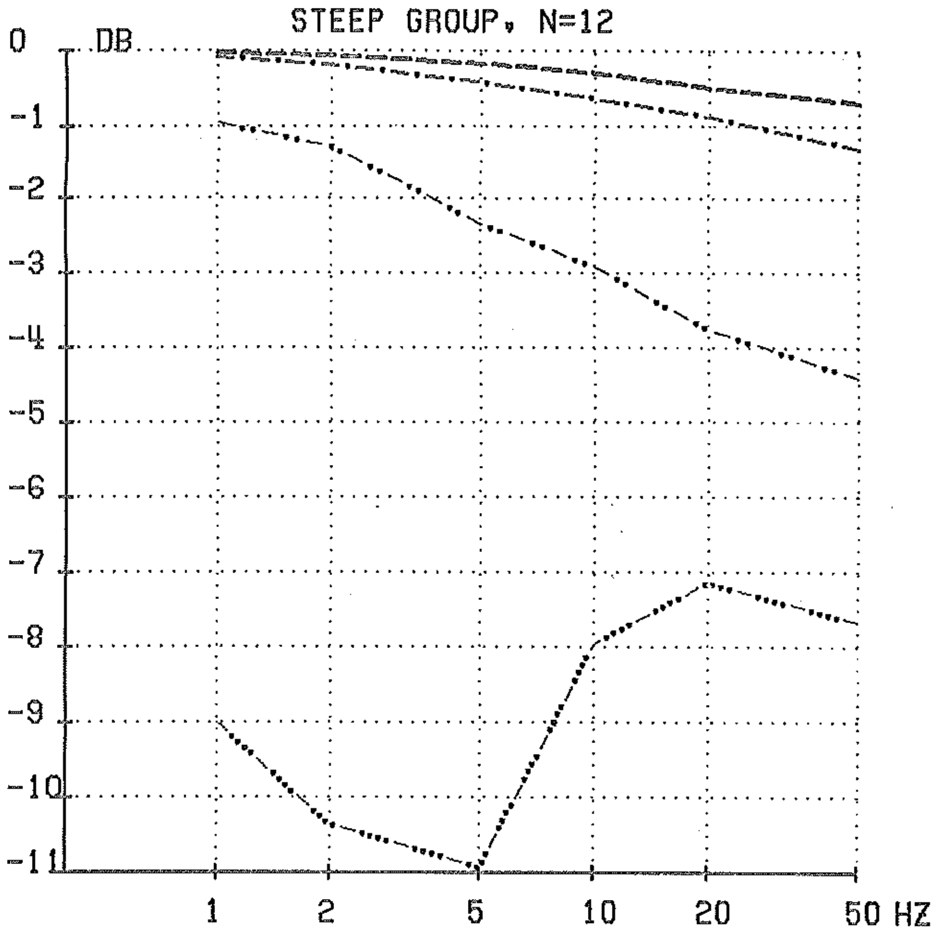

Fig. 8. Mean psychoacoustical modulation transfer functions (PMTF:s) for the steep group (N=12), calculated with the reference level at the PU-threshold + 7 dB.

- - - - - - 500 Hz probe tone.
-.-.-.-.- 1 kHz probe tone.
-..-..-..- 2 kHz probe tone.
-...-...- 4 kHz probe tone.

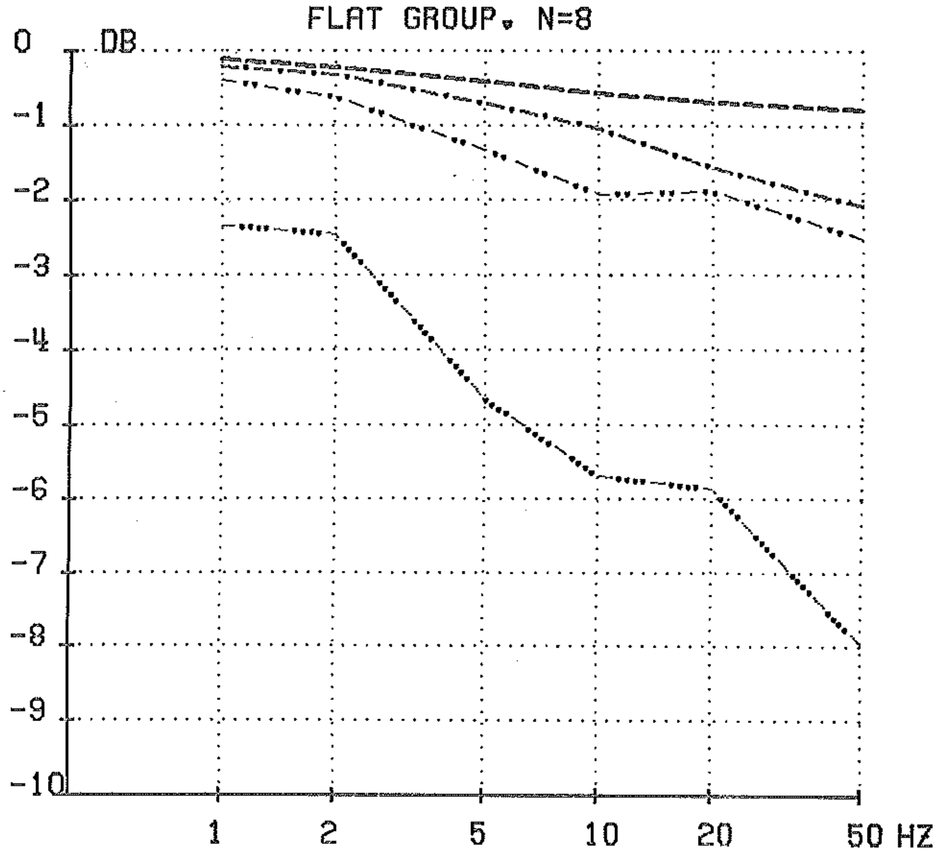

Fig. 9. Mean psychoacoustical modulation transfer functions (PMTF:s) for the flat group (N=8), calculated with the reference level at the PU-threshold + 7 dB.
- - - - - - 500 Hz probe tone.
-.-.-.-.- 1 kHz probe tone.
-..-..-..- 2 kHz probe tone.
-....-....- 4 kHz probe tone.