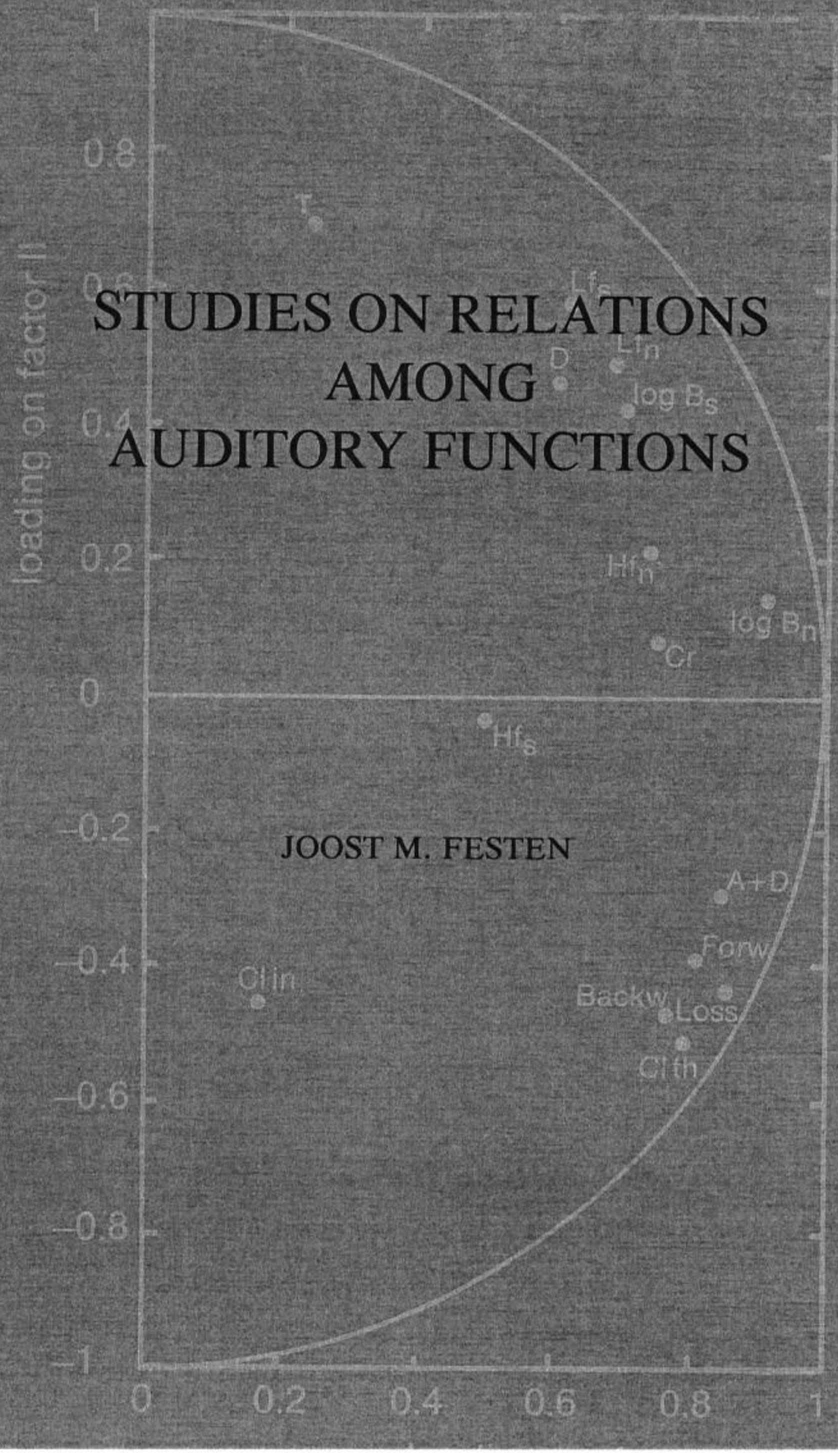
STUDIES ON RELATIONS AMONG AUDITORY FUNCTIONS

JOOST M. FESTEN



STUDIES ON RELATIONS AMONG AUDITORY FUNCTIONS
JOOST M. FESTEN

VRIJE UNIVERSITEIT TE AMSTERDAM

STUDIES ON RELATIONS AMONG AUDITORY FUNCTIONS

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van
doctor in de wiskunde en natuurwetenschappen
aan de Vrije Universiteit te Amsterdam,
op gezag van de rector magnificus dr. H. Verheul,
hoogleraar in de faculteit der wiskunde en natuurwetenschappen,
in het openbaar te verdedigen
op vrijdag 21 januari 1983 te 13.30 uur
in het hoofdgebouw der universiteit, De Boelelaan 1105

door

JUSTINUS MARIA FESTEN

geboren te Utrecht

van Kralingen's drukkerij & uitgeverij
Loenen aan de Vecht

Promotor: Prof. dr. ir. R. Plomp
Referent: Prof. dr. R.J. Ritsma

VOORWOORD

De kiem voor het onderzoek, waarvan dit proefschrift verslag doet, werd gelegd in een subsidie-aanvrage uit 1976 door Reinier Plomp en Guido Smoorenburg voor een zwaartepuntsproject bij ZWO. Daar aan deze subsidievorm een voortijdig einde kwam, werd dit project als zodanig niet gefinancierd, al kwamen in de jaren daarna delen ervan voor subsidie in aanmerking. Het door mij verrichtte onderzoek werd gefinancierd door de Vrije Universiteit en uitgevoerd in de vakgroep Keel-neus-oorheelkunde en Audiologie.

Aan het totstandkomen van dit proefschrift en aan mijn wetenschappelijke vorming hebben velen een bijdrage geleverd. In de eerste plaats natuurlijk mijn promotor Reinier Plomp, die na het leggen van de kiem alle levensfasen van dit onderzoek heeft meegemaakt en begeleid. Daarnaast ook Tammo Houtgast en Guido Smoorenburg, van wie ik veel geleerd heb, vooral in de periode waarin ik regelmatig naar Soesterberg ging. Bij Theo Kapteyn kon ik altijd terecht met vragen ten aanzien van slechthorendheid en audiometrie. Vanzelfsprekend hebben ook mijn naaste kollega's, Wout Dreschler, Toine Duquesnoy, Gerrit Bloothooft en Jan de Laat, door voortdurende discussies mijn ontwikkeling beïnvloed. Een zeer tastbare bijdrage aan het onderzoek is geleverd door de technici van de afdeling en in het bijzonder door Cor de Boer, die naast de zorg voor het computersysteem vele electronische schakelingen realiseerde.

In de slotfase is Prof. dr. R.J. Ritsma zo vriendelijk geweest als referent enige kritische kanttekeningen bij het manuscript te plaatsen. De uiterlijke vormgeving van dit proefschrift en de figuren zijn tot stand gekomen in nauwe samenwerking met Philip van Kralingen.

Tenslotte heeft mijn gezin, en vooral Maddy, op tal van punten bijgedragen en niet in het minst door mij de tijd te gunnen.

Aan allen ben ik dank verschuldigd.

aan Maddy, Rogier en Michiel

CONTENTS

CHAPTER 1

INTRODUCTION

The aim of auditory research is to describe and understand processes in hearing. As system analysis is a powerful scientific tool, researchers are inclined to describe the auditory system in terms of this analysis as an input-output system. The input to the system are sound waves to be expressed in physical terms like sound pressure and frequency. The description of the output depends on the experimental technique. In psychoacoustics the output is the sensation of sound and can be expressed in perceptual terms like loudness and pitch. In neurophysiology the output is the activity of haircells, auditory nerve fibers, etc., and is again expressed in physical terms like amplidude and firing rate. Whereas psychoacoustics is an overall approach in which the auditory system itself is regarded as a black box, physiology is analytical and studies subsystems within the auditory pathway. The present study is of the overall type and uses psychoacoustical techniques.

A possible way of measuring sensation is to express the sensation in perceptual terms. For instance, the subject can be asked to indicate how much louder, or higher in pitch, a signal is with respect to a reference signal. However, this appears to be an extremely difficult task, and the results are strongly dependent on details in the experimental procedure (cf. Warren, 1970). More reliable results are obtained in experiments in which input parameters are used as a reference. In this case the results are expressed in terms of properties of the input signal, for instance the level in decibels of a sound just masking another sound or the just distinguishable frequency difference in Hertz between two tones. These, and many other, quantities can be measured as a function of the physical parameters of the input signal and the result of such a measurement will be called an auditory function.

Traditionally in auditory research, in each investigation attention is focused on only one phenomenon. In order to explore this phenomenon, one or a few very closely related auditory functions are determined for a group of subjects. For another auditory function, in general, a new investigation is started with another group of subjects involved. Thus, the attention is directed to one auditory function for which the dependence on physical parameters is investigated. These studies, with normal-hearing subjects and also with hearing-impaired subjects, are of vital importance for hearing theory, but they also have serious shortcomings; they have taught us a lot about individual auditory functions, but they give little information about relations between various aspects of the hearing mechanism. This gap in our knowledge is particularly apparent if we try to understand the background of hearing impairment. In this respect it is insufficient to know variations in auditory bandwidth for one group of subjects, tone-on-tone masking for a second group, and forward masking for again another group. Instead, we need to have information about all these, and even more, auditory functions for the *same* group of subjects in order to study their relationships. Whereas the first can be described as a "vertical" approach, this alternative way of studying auditory functions can be considered to represent a "horizontal" approach. Of course, in such a broad study auditory functions cannot be measured as extensively as was done in studies focused on individual auditory functions. It is inevitable to reduce the number of measurement conditions very strongly, for instance, by restricting all measurements to a single signal frequency. This can be considered as a drawback of a "horizontal" study but both kinds of study, "horizontal" and "vertical", have their advantages and supplement each other.

The rationale for studying relations among auditory functions is the hypothesis that although we can measure many functions, the processes governing hearing can be described with only a few basic functions. Measured auditory functions are only combinations of these basic functions. For instance, frequency resolution of the auditory system as a function of frequency could be such a basic function. Frequency discrimination, critical bandwidth, and the steepness of masking curves are possibly all only derivatives of that single function. Theoretically the hypothesis can be tested by changing one of the parameters in the auditory system and investigating its influence on the various auditory functions. In psychoacoustics this is only possible to a very limited extent, for instance by introducing a temporal threshold shift by exposing subjects to noise. However, the same goal can be reached by using the differences in auditory functions existing among

subjects. Large differences are found among hearing-impaired subjects, but also among normal-hearing subjects differences in auditory functions can be demonstrated.

Literature on this approach of auditory research is very scarce. Relations between auditory abilities have been studied in an attempt to find optimum criteria in the selection of applicants for tasks involving auditory requirements. Seashore developed widely applied tests on musical abilities to select students for musical training (Harris, 1957) and Harris himself applied Seashore's tests along with other tests in a study on the selection of sonar operators. Results were poor, in spite of good test-retest reliabilities that were achieved. Most of the tests applied by Harris were of a complex nature and involved central processing. In direct connection with relations among auditory tests is a study by Elliott *et al.* (1966) on discrimination performance of normal-hearing subjects. However, factor analysis on the data showed only one general factor, interpreted as "discrimination factor". Most probably, this factor was the result of the influence of subject motivation or effort in all tests.

More research has been spent on the area of the relations between auditory tests and speech perception. For normal-hearing subjects (university students) Hanley (1956) conducted an experiment using 32 tests covering a very broad range of auditory traits. A large number of tests had a strong bearing on cognition, which was reflected in the denomination of the eight factors emerging from a factor analysis. Hanley hypothesized that speech perception could be related to "verbal facility", but she was not able to demonstrate this relation and concluded that her subject population was too homogeneous with respect to verbal facility. For a less homogeneous group of navy recruits, Solomon *et al.* (1960) repeated a large part of Hanley's test battery supplemented with some extra tests. The results were in reasonable agreement with those of Hanley's study and some relation between speech perception and verbal facility was found. Furthermore, Solomon showed by factor analysis that tests using filtered, reverberant, interrupted, clipped, or noise masked speech all loaded on one single factor, indicating that the ability to understand distorted speech is a single capability irrespective of the kind of distortion.

For hearing-impaired subjects nearly all earlier studies are devoted to an optimum prediction of speech-discrimination scores from other auditory parameters. Mullins and Bangs (1957) studied maximum speech discrimination in quiet and found it to be predominantly determined by the amount of hearing loss at high frequencies and by the slope of the hearing loss (a greater handicap in case of a steeper audiogram). Speech discrimination at 40 dB above the speech-reception threshold both in quiet and in noise was studied by Ross *et al.* (1965). Besides pure-tone thresholds and speech-discrimination tests, they included tests on intensity-difference limens and frequency-difference limens at various frequencies and also a test on aural overload. Although speech discrimination appeared to be related only to the extent and configuration of the hearing loss, they found that speech-discrimination scores in noise and in quiet depend on different parameters. Regarding the other audiometric data, low, but significant, correlations were found between the frequency-difference limen and audiometric loss at various frequencies; the difference limen for intensity was not consistently related to any other measure. In recent years there is a growing interest in the background of hearing loss. Especially much attention has been given to the deterioration of frequency resolution and its relation with the audiogram (Pick *et al.*, 1977; Wightman *et al.*, 1977; Zwicker and Schorn, 1978; Hoekstra, 1979). A relation between frequency resolution and the speech-reception threshold in noise was demonstrated by Leshowitz (1977) and Horst (1982). In other studies on the relations between speech perception and audiometric data even more parameters on peripheral signal processing have been incorporated. Tylor *et al.* (1980) measured temporal integration and frequency resolution along with speech-discrimination scores in noise for subjects suffering from noise-induced hearing loss. They found that both frequency analysis and temporal analysis are impaired in the region of hearing loss and both are correlated with poor speech discrimination in noise. Dreschler and Plomp (1980) tried to find intermediate stages between the perception of tones and speech. They studied speech-reception thresholds, vowel perception, and frequency resolution for hearing-impaired subjects and found distortions in the perception of vowels that could be attributed to reduced frequency resolution.

The present study is only indirectly related to speech perception. The main issue is the relation among auditory functions per se. The second chapter is based on a preliminary study (Festen *et al.*, 1977) and shows that reliable differences in auditory functions can be measured even for normal-hearing subjects. Chapter 3 deals with relations among auditory functions in normal-hearing subjects. This chapter is based on a study with a battery of 12 tests on 50 subjects and represents the text of a paper "Relations between auditory functions in normal hearing" (Festen and Plomp, 1981). Chapter 4 enters, as

an intermezzo, into some problems of the measurement technique and utilizes data from the previous and the following chapter. In Chapter 5 the relations among some auditory functions for hearing-impaired subjects are explored. This chapter is based on a study with 22 sensorineurally hearing-impaired subjects with moderate losses and represents the text of a paper "Relations between auditory functions in impaired hearing" (Festen and Plomp, 1983). Finally,in Chapter 6 results of the various experimental chapters are confronted with each other and the approach chosen is evaluated. This study ends with a summary in which a recapitulation of the experiments is given together with the main conclusions.

CHAPTER 2

INTERINDIVIDUAL DIFFERENCES OF AUDITORY FUNCTIONS: A PILOT STUDY *

Summary

This chapter shortly describes the apparatus used in these studies and deals with a pilot experiment that was conducted to determine the feasibility of measuring correlations between short tests. Incorporated tests are: the absolute threshold and masked thresholds for a tone and a click, the just-noticeable difference (JND) in the phase of one component of a complex tone plus JNDs in frequency and intensity, the perception of the low pitch of a complex tone, the auditory bandwidth measured with comb-filtered noise and tuning-curve slopes both measured in direct masking and with the pulsation-threshold technique, temporal resolution measured with intensity-modulated noise, cubic-difference-tone strenght, and lateral suppression. All tests were measured twice for a measuring frequency of 1000 Hz and 10 normal-hearing subjects participated in the experiment. For nearly all tests there was a good correlation between the results of the test and the retest which shows that the measurement error is small as compared to the interindividual differences.

2.1. INTRODUCTION

In order to use interindividual differences to provide insight into the interdependence of auditory functions and into the underlying properties of the hearing system, auditory functions have to be stable, at least during the experiment, and they have to be measured accurately. The main objective of this preliminar study is to see whether we can measure reliable interindividual differences, using short tests for a number of interesting auditory functions. Eight kinds of test were chosen, most of which received much attention in recent years and for which firm arguments exist concerning their relationships (Houtgast, 1974; Smoorenburg, 1972; Zwicker, 1970). Each test was performed twice on separate days. The differences between test and retest results reflect the combined effects of measurement error (accuracy) and day-to-day

* *Based on a paper in: Psychophysics and Physiology of Hearing, (1977).*

variations (stability) of the auditory functions. Ten normal-hearing subjects participated in the experiment and, to limit the number of measurement conditions, in all tests our attention was focused on the frequency region around 1000 Hz. Before going into details on the individual tests we will discuss the main lines of the apparatus used both for the experiment described in this chapter and for the experiments presented in the subsequent chapters.

2.2. APPARATUS

In a study as described here special demands are put upon the apparatus. Many auditory functions have to be measured for a group of subjects and for each subject we have only a very limited amount of time. This calls for an apparatus which can be quickly converted from one test to another and is suitable for quite diverse measurements. Traditionally, for each experiment pieces of equipment are put together in order to satisfy the needs for one specific case. This method is unsuitable here because it is impossible to convert the apparatus both quickly and reliably from one measurement to another, and the equipment is too expensive to have separate set-ups for each experiment. Besides flexible in its interconnections, the apparatus should also be flexible with respect to the signals. It should be possible to generate all kinds of signal, from pure tones to complex tones and from white noise to, for instance, comb-filtered noise. It should also be easy to change parameters of the signals, like frequency and level of the individual components. Last, but not least, because we are interested in relatively small interindividual differences, the apparatus should allow accurate measurement techniques like an adaptive two-alternative forced-choice procedure (2AFC) in which signal parameters in each new trial are based on the responses given so far, and the ultimate threshold is calculated from all responses.

A block diagram of the apparatus that meets the requirements listed above is given in Fig. 2.1. The diagram can be divided in two parts, at the left an analog part for the audio signals and at the right a digital part with a minicomputer for the control. Centrally we find the audio equipment for generation of the signals and timing of the stimuli. Generally this equipment has two kinds of input and output terminals, one for analog signals and one for digital signals needed for control. All analog terminals and part of the digital terminals are connected to a patch panel shown at the left in the block diagram. This panel is a matrix of connectors which can be linked in any desired fashion by means of prewired patch boards. On the right in the block diagram

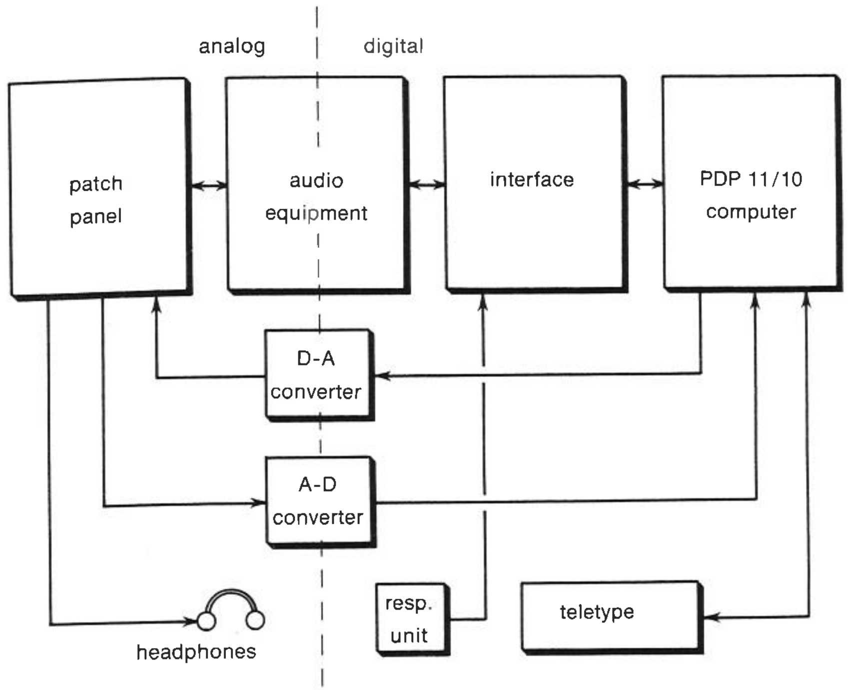


Fig. 2.1. Block diagram of the apparatus.

we find a minicomputer (PDP 11/10) linked with the audio equipment via an interface. The computer performs the following operations:

(1) It controls the measurement. Signal parameters as frequency, level, and duration are adjusted, subject responses are read, and new parameters are chosen on the basis of these responses.

(2) It calculates waveforms of complex signals which are stored in a function generator for reproduction independent of the computer.

(3) It generates non-periodic signals through the D-A converter. For instance, comb-filtered noise is produced by sampling of white noise, digital filtering, and generation via the D-A converter.

(4) Finally, data are stored in the computer during the experiments for later processing.

The operation of the apparatus can be illustrated with the help of a simple example. Suppose, we wish to determine the masked threshold of a tone in noise. On a patch board a noise generator, a bandfilter, an analog gate, an attenuator, and a mixer are linked in series. Apart, a second chain is build of a sine-wave generator, an analog gate and an attenuator. The signal from the second chain is fed into the same mixer and its output is linked to a headphone (Beyer DT 48) via a headphone amplifier. At the start of the experiment the sine-wave generator and both attenuators are adjusted by the computer. The timer, triggering the analog gates, is adjusted to produce a tone burst simultaneously with one of two successive noise bursts. Which noise burst contains the stimulus is chosen randomly by the computer. The subject's task is to detect the stimulus and to indicate whether it was in the first or the second noise burst. The signal level in the next trial is based on the current signal level and the subject's response (see Chapter 4). The signal level for all trials and the corresponding responses are stored for later calculations.

2.3. EXPERIMENTS

Eight tests were presented in a random order to each of the 10 subjects. All stimuli were presented monaurally, to the same ear of the subject. Typically each subject had two sessions in a week. The eight tests are described below.

2.3.1. Threshold

Thresholds were measured in a 2AFC adaptive procedure, with an initial level well above threshold. After each response the subject was provided with a visual feedback. The level of the test signal was raised by 2 dB after each false response and lowered by 2 dB after three successive correct responses. The mean signal level during 40 trials after the first 4 false responses was taken as the final threshold. Three different thresholds were measured three times over in the following order:
(1) masked threshold of a 1000-Hz pure tone with a duration of 180 ms in a continous pink noise with a level of 48 dB per one-third octave;
(2) masked threshold of a pulse pair with alternating polarity and a time distance of 20 ms, filtered by an octave filter centered at 1000 Hz, in the same pink noise;
(3) absolute threshold of a 1000-Hz pure tone, again with a duration of 180 ms.
The results of this test are the median values for each of the three thresholds. They will be referred to as *tone in noise thr.*, *pulse in noise thr.*, and *absolute thr.*, respectively.

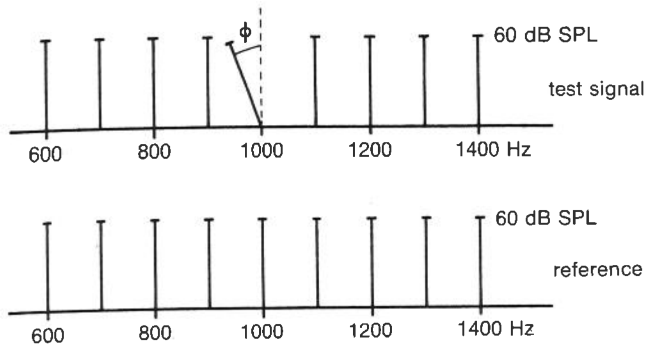


Fig. 2.2. Schematic presentation of the signals applied to measure the JND in phase. All components are in sine-phase except the 1000-Hz component of the test signal.

2.3.2. Phase-difference limen

The just-noticeable difference (JND) in phase was measured in a 2AFC adaptive procedure. The signal was a complex tone with a fundamental frequency of 100 Hz containing the 6th up to the 14th harmonic; see Fig. 2.2. All harmonics had the same level of 60 dB SPL. Each trial consisted of two stimulus pairs (stimulus duration 200 ms). The first stimulus of each pair was a reference tone in which all components were in sine phase. The second stimulus of either the first or the second pair contained the same complex tone, the other the complex tone with the 10th harmonic shifted in phase. A pink-noise floor with a level of 39 dB per one-third octave was presented continuously. The adaptive procedure started with a phase shift of 90 degrees. For each step the phase difference was multiplied by $\sqrt{2}$ after an incorrect response and divided by $\sqrt{2}$ after three successive correct responses. Again the subject was provided with a visual feedback after each response. The phase-difference limen was the mean shift in about 70 trials. In one session this limen was measured three times over. The result of this test is the median value of these three limens.

2.3.3. Intensity- and frequency-difference limen

The just-noticeable difference in intensity (ΔI) and in frequency (Δf) for a tone of 1000 Hz and 60 dB SPL were measured in one test. The same adaptive 2AFC procedure was used as in the previous test. In the first series of the test Δf was measured and in the second series ΔI. These two series were repeated once. In the ΔI-part of the test the intensity of the test stimulus was always below the intensity of the reference and started at a relative level of -8 dB. In the Δf-part the test tone had a higher frequency than the reference and started at 1200 Hz. For each step the intensity difference or frequency difference between signal and reference was multiplied or divided by $\sqrt{2}$. Throughout the test there was a pink-noise floor of 39 dB per one-third octave. In both cases the final difference limen was the mean difference over about 70 trials. The result of this test is the mean ΔI and the mean Δf.

2.3.4. Low pitch of a complex tone

The prominence of the low pitch of a complex tone was measured as a function of the harmonic number, n, of the middle component in the test signal. The subject had to judge the pitch jump between two stimuli. The first stimulus was the reference and had a duration of 500 ms. The second stimulus, presented after an interval of 100 ms, was the test signal and had a duration of only 100 ms to prevent the subjects from analytic listening. The test signal consisted of the $(n - 1)$th, n-th, and $(n + 1)$th harmonic at equal levels of 50 dB SPL as indicated in Fig. 2.3. The n-th harmonic was about 1000 Hz with small random fluctuations between the stimulus pairs (σ = 3%). In this test we used n = 4, 6, 8, 10, and 12. The reference signal, also with components of 50 dB SPL, contained the $(n - 4)$th, $(n - 3)$th, $(n - 2)$th, $(n + 2)$th, $(n + 3)$th,

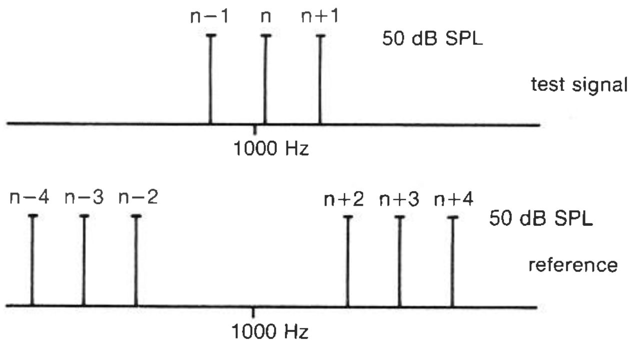


Fig. 2.3. Schematic presentation of the signals applied to measure low pitch. The test signal contains three harmonics, the middle one of which is about 1000 Hz with a small random fluctuation from trial to trial (σ = 3%). The reference is composed of 6 harmonics flanking the harmonics of the test signal. The fundamental of the reference is shifted 3% up or down in frequency with respect to the test signal.

and $(n + 4)$th harmonic of a fundamental shifted 3% up or down in frequency with respect to the test signal. Only for n = 4 the lowest harmonic was replaced by the $(n + 5)$th. Also in this test the subject was informed about his performance. One series consisted of 20 trials with the same harmonic number. The total session contained 25 series, five for each value of n. During the test there was a continuous pink-noise floor of 33 dB per one-third octave. Typically, the percentage of correct responses increases with decreasing harmonic number and reaches a saturation level for n = 4 and 6. The mean score for n = 4 and n = 6 will be referred to as *low-pitch score* and the value of n for which the subject achieved a score halfway the just defined low-pitch score and chance was called *n max for low pitch*.

2.3.5. Auditory bandwidth

An estimate of the bandwidth at 1000 Hz was obtained by means of a comb-filtered noise masker and a 1000-Hz probe tone of 40 dB SPL. The noise had a modulation depth of 28 dB and was low-pass filtered with a cut-off frequency of 3000 Hz. Four maskers were used in a random order, with peak spacings of 1000, 500, 250, and 125 Hz. For each peak spacing two adjustments were made, one with a peak and one with a trough at 1000 Hz and in one session each condition was measured twice. Two measurement techniques were used, simultaneous masking and the pulsation-threshold technique, and thresholds were reached by adjusting the level of the masker. In the pulsation-threshold technique masker and probe signal are presented in alternation, and for a sufficient masker level the signal is heard as continuously. This nonsimultaneous technique shows a greater frequency selectivity than simultaneous techniques, attributed to the influence of lateral suppression (cf. Houtgast, 1974). Both in simultaneous masking and for the pulsation threshold, the masker was presented four times per second. For the pulsation threshold the probe was presented in the gaps between the masker bursts and in the simultaneous masking test the probe was presented within every second burst of the masker. From the threshold differences between peak and trough the bandwidth of a Gaussian-shaped filter is calculated that accounts best for the data (cf. Houtgast, 1977). This bandwidth is used as an estimate of the auditory bandwidth. The results will be referred to as *bandwidth simult.* and *bandwidth pulsation.*

2.3.6. Time window of the auditory system

The internal time window was measured with signals which may be interpreted as the time-domain analog of the signals used in the previous test. Instead of an intensity modulation in frequency we used a masker that was intensity modulated in time and instead of a probe, well defined in frequency, we used a probe, well defined in time. The masker was pink noise with a modulation depth of 20 dB and was presented continuously. Modulation frequencies of 8, 16, 32, and 64 Hz were used. The probe was a pulse pair filtered through an octave filter centered at 1000 Hz and was repeated once every second. The two pulses had opposite polarity and their distance was fixed at 125 ms. The probe had a constant level of 87 dB SPL peak and the threshold was determined by adjusting the level of the noise. For each modulation frequency two thresholds were measured, one for the pulses coinciding with the peaks in the noise envelope and the other for the pulses coinciding with the troughs. In one session all conditions were measured twice. The threshold differences between peak and trough are described as resulting from a time window following $e^{-t/\tau}$. The parameter τ of this function will be used as the result of this test and gives a measure of the temporal resolution of the auditory system.

2.3.7. Slopes of the auditory filter

Measurements related to the slopes of the auditory filter at 1000 Hz were carried out both in simultaneous masking and with the pulsation threshold technique. The paradigms used, were the same as described in section 2.3.5. The probe was a 1000-Hz pure tone with a fixed level of 40 dB SPL. The masker was band-pass noise, 200 Hz wide. In order to obtain a masker with much steeper skirts than the auditory filter, the band-pass noise was produced by multiplying the center frequency f_c by low-pass noise with a cut-off frequency of 100 Hz and a slope of 48 dB/oct. This resulted in masker skirts of 48 dB per 100 Hz. Centre frequencies f_c of 700, 1000, and 1200 Hz were used. In one session the threshold of the 1000-Hz probe was determined four times for each value of f_c in a random order. Since the noise band was 200 Hz wide, the slopes calculated from these thresholds are no perfect estimates of the slopes of the masking pattern. The low- and high-frequency slopes calculated from pulsation thresholds and masking thresholds will be referred to by *LF slope pulsation*, *LF slope simult.*, *HF slope pulsation*, and *HF slope simult.*, respectively.

2.3.8. Nonlinearity

The strength of the cubic difference tone $2f_1 - f_2$ and lateral suppression at 1000 Hz were measured as a function of the frequency separation between the components f_1 and f_2. Both phenomena depend strongly on frequency separation (cf. Goldstein (1967) for the cubic difference tone and Houtgast (1974) for lateral suppression). Frequency ratios f_2/f_1 of 1.1, 1.2, 1.3, 1.4, and 1.5 were used in a random order. For each frequency ratio the test consisted of three parts. In part one the strength of the cubic difference tone (CDT) at 1000 Hz was measured with the method of cancellation (see Fig. 2.4). The primaries f_1 and f_2 had a level of 70 and 60 dB SPL, respectively. In part two of the test, again the CDT level at 1000 Hz was measured, but now with the pulsation-threshold technique. As masker the same signal was used as in part 1 and the probe tone was given a phase equal to the phase of the CDT just found in the cancellation experiment. In part three lateral suppression was determined by measuring the pulsation threshold at 1000 Hz for a two-component masker which was composed of a tone at 1000 Hz with a level of 50 dB SPL and a suppressor tone of 70 dB SPL at the frequency of f_1.

The level of the CDT sharply dropped with increasing frequency ratio between f_1 and f_2; for f_2/f_1 = 1.5 only a very weak CDT was found. For a frequency ratio f_2/f_1 of 1.1 the cancellation in part one of the test was not possible for all the subjects, due to masking. Therefore, only the remaining ratios were used. The average cancellation and pulsation level for f_2/f_1 = 1.2, 1.3, and 1.4 are called *CDT level canc.* and *CDT level pulsation*, respectively. From the results of part three the mean thresholds are calculated for

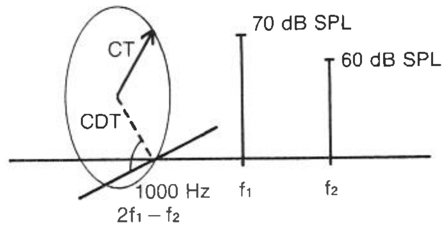


Fig. 2.4. Schematic presentation of the signals applied in the cancellation of the cubic difference tone (CDT). Phase and amplitude of the cancellation tone (CT) are adjusted by the subject to cancel the internally generated CDT. Shown is a vector addition of CDT and CT for equal amplitudes.

the two conditions f_2/f_1 = 1.1 and 1.2, and for the two conditions f_2/f_1 = 1.4 and 1.5. Because lateral suppression is strongest for frequency ratios between 1.1 and 1.2 and is almost absent for ratios of 1.4 and 1.5, the difference between the two means will be referred to as *suppression*.

2.4. RESULTS

Table 2.1 gives the test-retest correlation coefficients for all measurements. These coefficients, r_{tr}, are a measure for the reliability of the differences among subjects. Additionally Table 2.1 gives the mean, the standard

Test	measuring technique	r_{tr}	average score	σ between subjects	standard error	dimension
1.1 Tone in noise thr.	2AFC	-0.09	48.5	0.0	1.2	dB SPL
1.2 Pulse in noise thr.	2AFC	0.90	86.4	2.0	0.7	dB SPL peak
1.3 Absolute thr.	2AFC	0.83	0.8	4.6	1.6	dB SPL
2 JND in phase	2AFC	0.93	38.0	16.9	3.7	degr.
3.1 JND in intensity	2AFC	0.82	3.0	0.8	0.3	dB
3.2 JND in frequency	2AFC	0.85	4.6	1.4	0.5	Hz
4.1 Low-pitch score	2AFC	0.96	86.3	14.0	2.0	%
4.2 N max for low pitch	2AFC	0.56	8.6	0.4	0.3	-
5.1 Bandwidth simult.	adj.	0.85	258.	21.4	7.0	Hz
5.2 Bandwidth pulsation	adj.	0.64	122.	15.5	8.4	Hz
6 Temporal window	adj.	0.50	13.5	2.3	1.8	ms
7.1 LF slope pulsation	adj.	0.67	48.	7.2	3.9	dB/oct
7.2 LF slope simult.	adj.	0.58	45.	6.6	4.0	dB/oct
7.3 HF slope pulsation	adj.	0.81	135.	27.0	11.0	dB/oct
7.4 HF slope simult.	adj.	0.51	35.	8.3	6.2	dB/oct
8.1 CDT level canc.	adj.	0.15	32.5	3.3	6.3	dB SPL
8.2 CDT level pulsation	adj.	0.48	24.7	2.1	1.6	dB SPL
8.3 Suppression	adj.	0.80	15.7	3.6	1.4	dB

Table 2.1. Summary of the results of tests listed in the first column. The second column gives the measuring technique (2AFC or adjustment). The third column shows the correlation between test and retest, and the fourth column gives the average over 10 subjects. The standard deviation between subjects and the standard error are shown in the fifth and sixth column, respectively.

Test	Kind of score	1.3	2	3.1	3.2	4.1	4.2	5.1	5.2	6	7.1	7.2	7.3	7.4	8.2	8.3
1.2 Pulse in noise thr.	level	0.46	-0.05	0.46	0.62	-0.02	-0.13	0.36	0.48	-0.12	-0.70	-0.40	-0.33	-0.21	0.02	0.08
1.3 Absolute thr.	level		0.26	-0.09	0.11	-0.12	-0.30	-0.37	-0.23	0.00	-0.07	0.24	-0.51	-0.22	0.26	-0.33
2 JND in phase	phase diff.			0.32	-0.18	-0.81	-0.21	-0.67	-0.27	0.09	0.36	0.11	-0.33	-0.44	0.02	-0.40
3.1 JND in intensity	level diff.				0.26	-0.18	0.15	0.23	0.54	0.51	-0.41	-0.32	-0.55	-0.64	0.31	0.00
3.2 JND in frequency	frequency diff.					-0.17	-0.53	0.71	0.41	-0.03	-0.51	-0.16	-0.18	0.18	0.00	0.04
4.1 Low-pitch score	percentage						0.25	0.38	0.14	0.09	-0.26	-0.03	0.30	0.19	0.13	0.29
4.2 N max for low pitch	harm. number							-0.17	0.24	0.29	-0.22	-0.46	-0.10	-0.23	-0.12	-0.05
5.1 Bandwidth simult.	width								0.63	0.05	-0.55	-0.19	0.10	0.27	0.12	0.42
5.2 Bandwidth pulsation	width									0.40	-0.83	-0.66	-0.20	0.14	0.26	-0.03
6 Temporal window	width										-0.30	-0.16	-0.58	-0.24	0.51	-0.54
7.1 LF slope pulsation	slope											0.77	0.15	-0.13	0.05	0.23
7.2 LF slope simult.	slope												-0.12	-0.23	0.38	0.33
7.3 HF slope pulsation	slope													0.59	-0.64	0.23
7.4 HF slope simult.	slope														-0.22	-0.12
8.2 CDT level pulsation	strength															0.12
8.3 Suppression	strength															

Table 2.2. Matrix of correlation coefficients between the results of the tests. The coefficients that are significant at a level of 5% are underlined. With each test the kind of score is indicated for a correct interpretation of the sign of the correlations.

deviation between subjects, and the standard error for all tests. The table shows that the reliability of the experiments 1.1 and 8.1 is very low; therefore, these experiments have been excluded in further calculations. Table 2.2 gives the matrix of correlation coefficients between all other tests. Because these correlations are based on the test results of only 10 subjects the matrix is not suitable for an analysis of principal components (cf. Dziuban and Shirkey, 1974). However, a number of isolated high correlations are found, indicating that, possibly, the tests in question have a basic property of the auditory system in common. The highest correlations we found, can be described as follows.

(1) A large auditory bandwidth accompanies a shallow low-frequency slope of the psychophysical tuning curve, both measured using the pulsation-threshold technique. For this correlation, which can be understood with a simple filter model, a scatter diagram is shown in Fig. 2.5 panel (a). The corresponding relationship in simultaneous masking was not found.

(2) Subjects with a high phase sensitivity also show a high low-pitch score

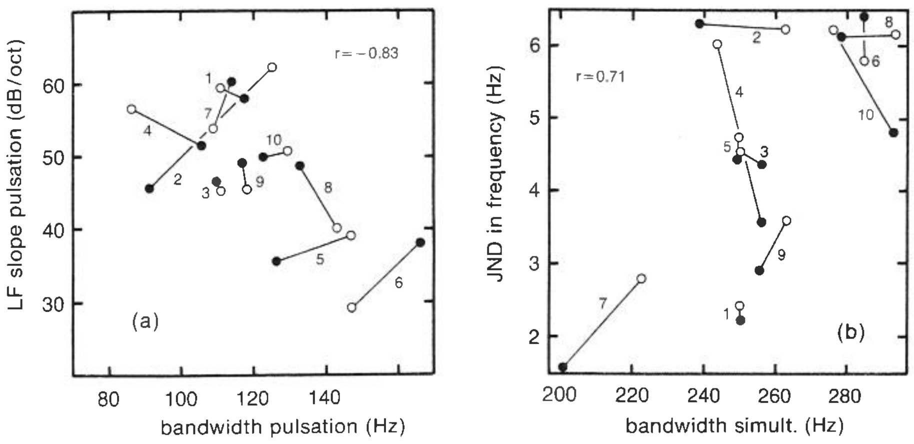


Fig. 2.5. Scatter diagrams for tests with a high correlation. The test and retest data are indicated by filled and open circles, respectively. The numbers refer to the individual subjects. r Represents the correlation coefficient for the mean values of test and retest. Panel (a) shows the relationship between bandwidth and low-frequency slope of the psychophysical tuning curve, both measured when using the pulsation-threshold technique. Panel (b) shows the relationship between the bandwidth measured in simultaneous masking and the JND in frequency.

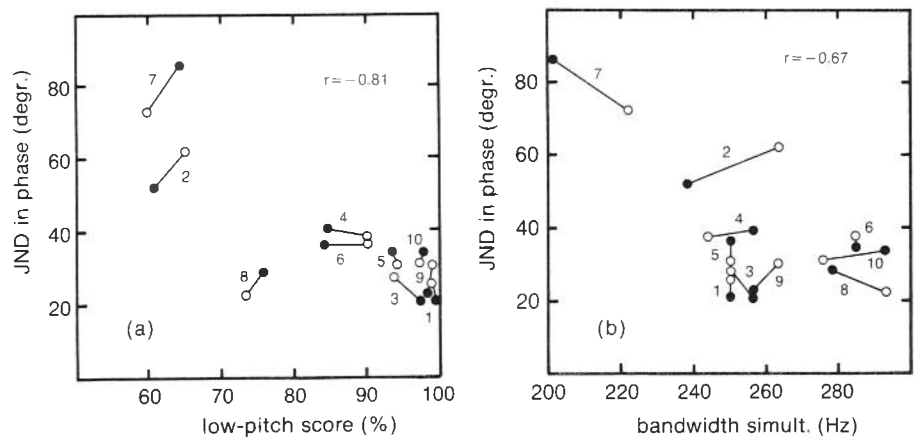


Fig. 2.6. Scatter diagrams as in Fig. 2.5. Panel (a) shows the relationship between the low-pitch score and the JND in phase. Panel (b) shows the relationship between the bandwidth measured in simultaneous masking and again the JND in phase.

as well as a large auditory bandwidth measured in simultaneous masking (see Fig. 2.6). Both phase sensitivity and low-pitch score are reported to be related to the strength of combination tones (cf. Smoorenburg, 1972; Buunen *et al.*, 1974). However, in the present study no correlation with the level of the cubic difference tone could be demonstrated.

(3) A steep low-frequency slope of the psychophysical tuning curve, measured in simultaneous masking, is connected with a steep low-frequency slope when measured with the pulsation-threshold technique.

(4) A large auditory bandwidth, measured in simultaneous masking, is accompanied with a large just-noticeable difference in frequency (see Fig. 2.5 panel b). A relation between the JND in frequency and the steepness of the high-frequency slope of the tuning curve (corresponding to the steep edge of the masker) as proposed by Zwicker (1970) was not found.

(5) Subjects with a low threshold for a pulse in noise have a steep low-frequency slope of the psychophysical tuning curve.

2.5. DISCUSSION

Table 2.1 shows that reliable differences in auditory functions among individual subjects can be measured with tests taking only 15 to 30 minutes. Only for the measurements 1.1 and 8.1 the test-retest correlations are very poor. For the Tone in noise thr. (1.1) this is apparently due to very similar thresholds among the subjects. As for the CDT level canc. (8.1) we may conclude from the large standard error (6.3 dB), that cancellation is a difficult task for untrained subjects.

By and large, the 2AFC tests gave more reliable results than the tests in which adjustment was used as a measuring technique as shown by the test-retest correlations in Table 2.1. Also, because results of a two-alternative forced-choice procedure are criterion free (see Chapter 4), the application of 2AFC tests precludes correlations that are based solely on systematic differences in criterion among subjects. Another possible source of spurious correlations is the influence of training and motivation in various tests. To eliminate these effects from the test results, we should, whenever possible, use test scores that are based on differences between threshold determinations, like the slopes of the psychophysical tuning curve or the bandwidth from the comb-filtered noise experiment.

To optimize conditions for finding correlations, we have to reduce measurement error as far as possible. Therefore, no randomisation of the sequence of tests and measurement conditions should be applied, because systematic effects do not influence correlation coefficients, but they constitute an extra source of variance adding to the measurement error when randomisation is applied indeed. Furthermore, tests should be carefully tuned to each other; it is not enough to adopt tests in the battery just because those tests by themselfs give interesting results, but we should hypothesize a relation with some of the other tests in the battery. Finally, the significance of correlations is strongly influenced by the number of subjects. In this respect the number of 10 subjects used in this experiment must be considered as very low.

CHAPTER 3

RELATIONS BETWEEN AUDITORY FUNCTIONS IN NORMAL HEARING *

Summary

The relations among a number of auditory functions were studied by concentrating on small interindividual differences in these functions. For this purpose a battery of 12 tests was applied to 50 normal-hearing subjects. The tests included absolute threshold, auditory bandwidth measured with comb-filtered noise in direct and in forward masking, psychophysical tuning curve both in direct and in forward masking, temporal resolution measured with intensity-modulated noise, forward- and backward-masking curves, cubic-difference-tone strength, and lateral suppression. In all cases the test frequency was 1000 Hz. Among the relations found, are (1) a positive correlation between the shift of the steep edge of the tuning curve, away from the probe frequency, and the width of the auditory filter as measured with comb-filtered noise, (2) an inverse relation between the width of the tuning curve and the width of the temporal window, and (3) a positive correlation between the width of the auditory filter and the strength of the cubic difference tone. Low correlations among tests were not caused by poor test reliability.

3.1. INTRODUCTION

Properties of human hearing are commonly described in the form of auditory functions showing the relations between parameters of the stimulus and properties of the corresponding sensation. Examples are absolute and masked thresholds as a function of frequency, and the level of combination tones as a function of frequency and level of the primary tones. In the past many auditory functions have been determined and used as materials for developing hearing theory. In such a theory the auditory system is described in terms of basic properties, like frequency resolution and nonlinearity. A typical characteristic of this approach is that the theory is derived from data representing the behavior of the "ideal" test subject, i.e., the average of the subjects actually used.

* *Paper published in: J. Acoust. Soc. Am., 1981, (70), 356-369.*

In the present study we take quite the opposite view: it may be possible to explore the interindividual differences, giving us information about the relations between auditory functions. For instance, a difference in frequency-resolving power may be expected to affect both the bandwidths of the auditory filters and their slopes, suggesting that we should find a relation between the critical bandwidth and the slopes of the psychophysical tuning curve. This approach will only succeed if the interindividual differences at the level of the basic properties are not obscured by effects not included in this study.

Few studies in the literature are related to our approach. More than twenty years ago Harris (1957) gave a review of research into what he called primary auditory abilities. All studies which he mentioned were directed to rather central auditory processes, whereas our main interest is in peripheral processing. Harris strongly emphasized the need for tests having a high test-retest reliability in order to find correlations between them. Closely allied to our approach is an extensive study by Elliott *et al.* (1966) on the discrimination performance of a group of normal-hearing subjects. They were interested both in the range of normal performance and in the relations between the auditory tests. However, a factor analysis showed that besides a small general discrimination factor, the various discrimination tests tend to be relatively independent of each other.

From a preliminary study on a wide variety of experiments (Festen *et al.*, 1977) it became clear that relations among auditory functions, if traced by means of interindividual differences, can only be demonstrated with highly reliable measurements. Accordingly, much effort was spent to reduce unwanted data variance as much as possible.

This study deals with three important properties of the peripheral hearing organ: frequency resolution, time resolution, and nonlinearity, as well as with the absolute threshold. To obtain data suitable for factor analysis, we included several tests related to each of the three hypothetical underlying properties. To test frequency resolution we measured the auditory bandwidth with comb-filtered noise and the slopes of the psychophysical tuning curve, both in direct masking and in forward masking. To test temporal resolution, we measured the width of a temporal window, which is the time-domain analog of the auditory filter, and the slopes of forward and backward masking curves. To test nonlinearity, we measured the strength of the cubic difference tone for an optimal frequency ratio between the primaries and also the degree of lateral suppression for one suppressor condition.

Because the tested properties may vary essentially with signal frequency, all tests were administered at a single frequency, 1000 Hz. To save time, tests were as simple as possible and contained a minimum number of conditions. They were performed at the same sound-pressure level for all subjects.

3.2. METHODS

3.2.1. Procedure

We used an adaptive two-alternative-forced-choise (2AFC) procedure (Levitt, 1971) with visual feedback. In the various tests, threshold was reached by varying different signal parameters such as probe-signal intensity, masker frequency, and time delay between masker and probe. For example, in an experiment with a constant probe signal and a variable masker intensity, each run of 2AFC trials consisted of three stages. The first stage started well above threshold and after each correct response the masker level was raised by a fixed amount until the first false response. In the second stage the masker intensity was decreased after each false response and raised after two successive correct responses. This stage was introduced to provide a good starting point for the last stage and was terminated after the fourth false response. Stage three consisted of a constant number of trials (20 for the simultaneous-masking experiments and 30 for nonsimultaneous masking) and the mean level of the successive trials was adopted as the final estimate of the threshold. In this stage the masker level was raised after three successive correct responses which procedure converges to a detectability chance of 79%. The step size in the stages 2 and 3 was 2 dB for the direct masking and 3 dB for the forward-masking conditions.

3.2.2. Difference scores

An important source of unwanted variance is the difference in training and alertness among the subjects, causing different thresholds for audiologically similar subjects. We can get around this difficulty by focusing our attention on properties that can be described by the difference between two threshold values. Examples are the threshold difference between peaks and troughs of comb-filtered noise, and the slopes of the psychophysical tuning curve which are calculated from the difference in threshold for two masker frequencies divided by their frequency separation. To minimize the influence of training and alertness on the "difference score" the two measurements always followed each other immediately. A disadvantage of these scores is that the error variance is double that from the raw data, and so it becomes especially important to keep measurement error as small as possible.

3.2.3. Experimental design

Finally, the sequence of tests in the whole experiment may contribute to the reduction of measurement error. A randomization of the test sequence over subjects eliminates systematic errors in the mean test results but causes error variance in the interindividual differences as a consequence of sequence effects. However, in this study we are not in the first place interested in optimum average test results but rather in optimum interindividual differences, and for this reason all subjects were tested according to the same schedule. Testing took place in four morning sessions on four successive days. Half of the tests were carried out in the first session and the other half in the second session. The third and the fourth session were exact replications of the first two sessions, thereby permitting us to estimate training effects and test reliability. The coefficient of reliability (r_{tt}) is defined as unity minus the proportion of error variance, or alternatively as the proportion of "true" variance in a test. It can be estimated from the correlation coefficient between test and retest (r_{tr}) by applying the formula of Spearman and Brown:

$$r_{tt} = 2r_{tr} \;/\; (1+r_{tr}) \qquad (3.1)$$

(cf. Guilford, 1954; Nunnally, 1967).

On the average a test block lasted for a quarter of an hour, after which the subject had a break of the same duration. In the breaks for one subject another subject was tested.

3.2.4. Subjects

To apply factor analysis the number of subjects has to be at least a few times greater than the number of tests. The number of subjects also affects the significance of the individual correlation coefficients. Accordingly, we

tested 50 subjects, which means that correlation coefficients exceeding ±0.36 are significant ($p < 0.01$). The subjects were tested at one ear. At the test frequency all subjects were within 15 dB of normal hearing. Their age ranged from 17 to 31.

3.2.5. Apparatus

A PDP-11/10 computer controlled the generation of the signals, their presentation, and the adaptive procedure and stored responses. The tonal signals were stored in two revolving memories of 512 time samples with 16 bits of resolution. The noise signals were produced by a noise generator (Wandel und Goltermann RG-1); time delay, attenuation, and summation necessary for the comb-filtered noise were performed by the computer. The stimuli were gated with cosine-squared onset and termination with rise and fall times of 15 ms, and were presented via an electro-dynamic earphone (Beyer DT 48).

3.3. EXPERIMENTS

Figure 3.1 schematically presents this study's twelve experiments, each of which is discussed in the following sections.

3.3.1. Absolute threshold (1000 Hz)

Each session started with a determination of the absolute threshold for 1000 Hz. In the 2AFC procedure the visually indicated observation periods of 500 ms each were separated by 300 ms; the signal duration was 200 ms. A measurement block consisted of three runs. Stage 3 of each run contained 20 trials and the step size was 2 dB.

Results and discussion

The mean threshold is 5.9 dB SPL and the standard deviation between subjects is 4.6 dB. An analysis of variance showed that the interindividual differences are highly significant and constitute the greater part of the total variance. The interaction between *subjects* and *sessions* is also significant, although there is no main effect of the latter source. The interaction may result from influences such as fluctuations in alertness, training effects, or even physical health. Apparently these influences are significant for the individual subjects but not for the average results. Such influences are

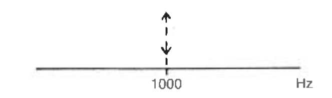
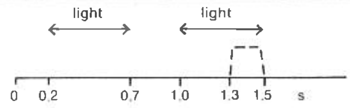
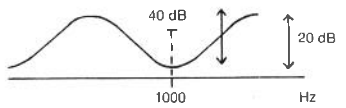
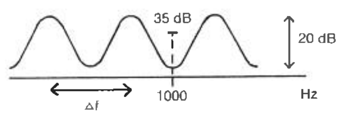
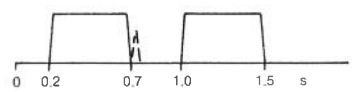
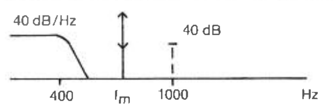
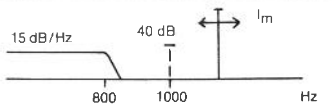
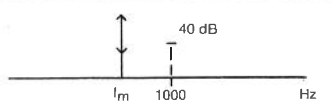
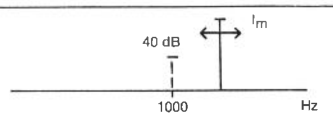
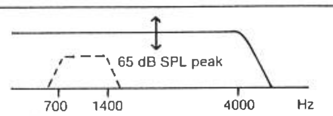
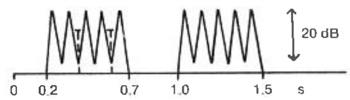
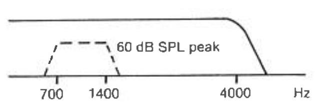
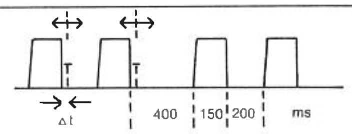
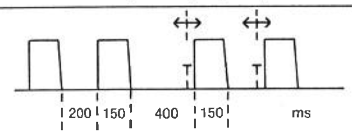
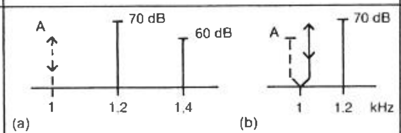
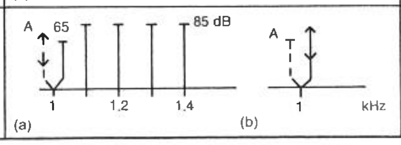
Experiment	Spectrum	Temporal structure	Conditions	Run length and Step size
1 Absolute Threshold				20 trials 2 dB
2 Bandwidth simultaneous masking			peak and trough for Δf = 1000, 667, 500 Hz	20 trials 2 dB
3 Bandwidth nonsimultaneous masking			peak and trough for Δf = 333, 250, 167 Hz	30 trials 3 dB
4 Shallow Edge simultaneous masking		as Experiment 2	f_m = 550, 950 Hz	20 trials 2 dB
5 Steep Edge simultaneous masking		as Experiment 2	l_m = 70, 90 dB	20 trials 4% of Δf
6 Shallow Edge nonsimultaneous masking		as Experiment 3	f_m = 750, 950 Hz	30 trials 3 dB
7 Steep Edge nonsimultaneous masking		as Experiment 3	l_m = 70, 90 dB	30 trials 10% of Δt
8 Temporal Window			peak and trough for f_{mod} = 10, 15, 20 Hz	20 trials 2 dB
9 Forward Masking			l_m = 35, 55 dB/Hz	30 trials 10% of Δt
10 Backward Masking	as Experiment 9		l_m = 35, 55 dB/Hz	30 trials 20% of Δt
11 Cubic Difference Tone		as Experiment 3		30 trials 3 dB
12 Suppression		as Experiment 3		30 trials 3 dB

Fig. 3.1. Schematic representation of 12 experiments. In the first column the experiments are labeled. The second and third columns show the spectral and temporal structures of the signals, respectively. The probe signals are dashed and the maskers are fully drawn. For all experiments the dependent variable is indicated with an arrow. The fourth column gives the measurement conditions and the fifth column shows for the adaptive procedure the step size and run length.

minimized in all other experiments where difference scores are used. The standard deviation of the individual runs, pooled over subjects, is 2.8 dB. If we use the results of the first two sessions as test and the results of the last two sessions as retest, the reliability coefficient r_{tt} = 0.91. The average threshold and its spread over subjects are in good agreement with data by Zwicker and Heinz (1955), who measured thresholds at different frequencies for a group of 100 students, and are also comparable to the values found by Mrass and Diestel (1959) for 70 normal-hearing subjects between 18 and 30 years of age.

3.3.2. Auditory bandwidth

The auditory spectral resolution was derived from the "internal" peak-to-trough ratio of comb-filtered noise as a function of its fineness of peak spacing. This method has been used both in electrophysiology (Wilson and Evans, 1971) and in psychophysics (Houtgast, 1974) where the threshold difference between a peak and a trough is determined for a probe tone of constant frequency. If we assume that the internal filter acts as an intensity-weighting function and that its shape is Gaussian on a linear frequency scale (cf. Patterson, 1976), than a simple expression can be found for the threshold difference between the peaks and troughs as a function of peak spacing. The smooth curves in Fig. 3.2 are calculated under these assumptions for a masker with a peak-to-trough ratio of 20 dB. The parameter of the curves is the equivalent rectangular bandwidth B.

The bandwidth was measured under both simultaneous masking and nonsimultaneous masking, which yield different values for frequency resolution. Houtgast (1974) ascribed this difference to lateral suppression, which is investigated explicitly in experiment 12.

The spectral and temporal characteristics of the signals are given in Fig. 3.1 lines 2 and 3. For optimum discrimination between subjects the measurements were performed for peak spacings of 1000, 667, and 500 Hz in simultaneous masking and 333, 250, and 167 Hz in nonsimultaneous masking. In nonsimultaneous masking, the 1000-Hz probe tone had a cosine-squared envelope with rise and fall times of 15 ms and no steady state. It was presented immediately after the masker. The peak level was 35 dB SPL and the bandwidth of this short tone was about 50 Hz.

In each measurement block six thresholds were determined (peak and trough conditions for three peak spacings). Under each masking condition the bandwidth was determined twice for both the test and the retest.

a. Results

The threshold-level difference between peak and trough as a function of peak spacing is given in Fig. 3.2. To give an impression of the variability, subjects are divided in five subgroups of ten subjects each on the basis of their mean threshold differences. Because of the low correlation between the bandwidths in direct masking and in forward masking (r = 0.17), the subgroups in these two experiments were composed independently. The five heavy lines on the left give the threshold differences in simultaneous or direct masking for these five groups. The standard deviation of the threshold differences, pooled over subjects and peak spacings, is 2.1 dB. By means of a least-squares estimation the bandwidth B of a Gaussian filter was calculated for the results of the test and for the retest of each subject. The reliability coefficient for these bandwidths is 0.44. Results of the same kind were obtained with forward masking. Here the pooled standard deviation of the individual differences is 4.2 dB. The curves on the right in Fig. 3.2 show the threshold differences

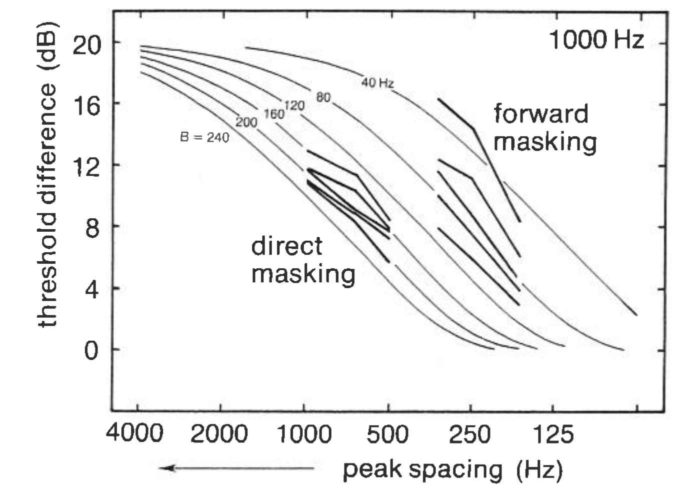


Fig. 3.2. Threshold-level difference for a 1000-Hz probe tone between peak and trough of comb-filtered noise as a function of peak spacing. The curves on the left give the data in direct masking for five subgroups of ten subjects. The curves on the right give the data in forward masking, also for five subgroups; the composition of the subgroups in direct and in forward masking is not the same. The smooth curves represent calculated threshold differences for a Gaussian-shaped filter.

in forward masking. Bandwidths were again calculated under the assumption of a Gaussian shape; the mean bandwidth is 72 Hz. The reliability coefficient of these bandwidths (0.59) is not significantly higher than the reliability of the direct-masking bandwidths.

b. Discussion

The mean width of the auditory filter found in direct masking is 177 Hz which is in good agreement with the traditional values of the critical bandwidth according to Zwicker (1954) and Scharf (1970). The mean bandwidth and variability are comparable to data by Houtgast (1974) on five subjects.

The reliability coefficient for the bandwidth data in direct masking is rather low, as is also indicated by the small fraction of the variance that is accounted for by the factor *subjects* in an analysis of variance. This low reliability is considered to be detrimental to the correlations between this "direct-masking" bandwidth and the other tests.

The bandwidths measured in nonsimultaneous masking are in good agreement with the pulsation-threshold data by Houtgast (1977). An analysis of variance on the forward-masking threshold differences showed a significant interaction between *subjects* and *peak spacings*. This effect can also be seen in Fig. 3.2. Subjects with a large average threshold difference have a steeper curve, and thus a greater effect of peak spacing, than subjects with a small threshold difference. These differences cannot be accounted for by the smooth curves calculated from a Gaussian-shaped filter. A possible explanation for this effect is that narrow filters have steeper slopes than a Gaussian filter, or even have negative-going parts, resulting in a sharper decrease of the threshold difference as a function of peak spacing.

3.3.3. Psychophysical tuning curve in simultaneous masking

Another measure of auditory frequency resolution is provided by pure-tone masking, introduced by Wegel and Lane (1924) with the measurement of the extent of masking caused by a fixed masker. Many investigators (Moore, 1978; Vogten, 1978a and b; Zwicker, 1974) have recently studied pure-tone masking for a fixed probe tone and measured either masker level as a function of frequency or, because of the steepness of the curves, masker frequency as a function of level. The resulting "iso-response" curves bear resemblance to neurophysiological "iso-rate" curves or tuning curves. To limit the number of measurements, the tuning curves were approximated by four points, two on the

steep edge and two on the shallow edge. A schematic representation of the signals is given in Fig. 3.1.

The shallow edge was approximated by determining the levels at which masker frequencies of 950 and 550 Hz just masked a 40-dB, 1000-Hz signal. In order to avoid detection of combination tones a continuous low-pass noise was present with a spectral density of 40 dB per Hz and a cutoff frequency of 400 Hz. The slope of the shallow edge is the difference score. In test and retest the two masker levels were determined three times over within one measurement block.

To measure the other edge of the tuning curve the frequency of a fixed-level masker was adjusted so that the signal was just masked. Because of the steepness of this slope a constant masker frequency could easily have led to unacceptably high masker levels. Now the masker levels were 70 dB and 90 dB SPL. Low-pass noise (15 dB per Hz and cutoff frequency of 800 Hz) was added to mask the distortion products. To make sure that the masker frequency did not reach a value of 1000 Hz, the step size in the adaptive procedure was a fixed percentage (4%) of the frequency difference between the masker and the probe tone.

a. Results

The filled symbols and solid lines in Fig. 3.3 give the results of both sets of measurements. For the shallow edge the mean values and the standard deviations over 50 subjects are indicated for both masker frequencies. The standard deviation within subjects is 1.7 dB. The correlation coefficient between the threshold data at 550 and 950 Hz is 0.51, indicating that, for the shallow edge, the subjects differed in their slopes as well as in their mean thresholds. For this reason no grouping of the data is shown. The average slope is 17 dB/oct with a standard deviation of 3.8 dB/oct and the reliability coefficient is 0.82.

For the steep-edge data all calculations were carried out on the logarithm of the frequency difference between probe and masker because these log (Δf) values are normally distributed for both masker levels, and because the standard deviation is independent of the mean value. The average slope is about 220 dB/oct. An analysis of variance of the raw thresholds showed that the proportion of variance accounted for by the factor *subjects* is seven times greater than the part accounted for by the interaction bewteen the factors *subjects* and *masker levels*. Thus the bulk of the between-subjects variance is caused by a parallel shift of the steep edge, as is also illustrated by the correlation

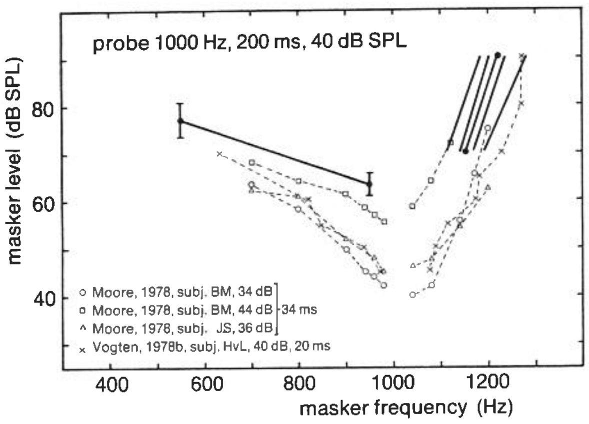


Fig. 3.3. Tuning curves in simultaneous masking. The solid symbols give the mean results for 50 subjects and the between-subjects standard deviations are indicated by bars. At the steep edge the mean results for five equally large subgroups are connected by straight lines. Combination tones were masked by low-pass noise, for the shallow edge 40 dB per Hz with a cutoff frequency of 400 Hz and for the steep edge 15 dB per Hz with a cutoff frequency of 800 Hz. For comparison some results from the literature are indicated with dashed lines and open symbols.

coefficient between the thresholds for the 70- and 90-dB masker ($r = 0.78$). To show this variability the subjects were grouped, on the basis of their mean log (Δf) value, in five subgroups of ten subjects each. The average thresholds for each group are connected by straight lines in Fig. 3.3. Despite the small fraction of the variance accounted for by differences in slope, the reliability coefficient of the slope values is 0.92. For each subject an estimation of Q_{10dB} was calculated by approximating the tuning curve with two straight lines. The average Q_{10dB} is 3.3 with a standard deviation between subjects of 0.7.

b. Discussion

The slopes of the psychophysical tuning curve found in our experiments are in good agreement with the results of Moore (1978) and Vogten (1978b), as may be seen in Fig. 3.3. The slope values summarized by Moore are steeper on the low-frequency side and shallower on the high-frequency side, but he based his slope values, more than we did, on a region near the tip of the tuning curve. Because we did not take into account the sharp tip of the tuning curve, our estimations of Q_{10dB} are also slightly smaller than the 3.6 to 4.9 given by

Moore (1978). The absolute levels of the tuning curves are not quite comparable because the duration of the probe tone was 200 ms in our case and 34 or 20 ms for the other two authors. This difference in duration can account for a vertical shift of 5 to 10 dB.

3.3.4. Psychophysical tuning curve in nonsimultaneous masking

Nearly all authors (Houtgast, 1973; Moore, 1978; Vogten, 1978b) found much steeper slopes for the tuning curve in nonsimultaneous masking than in simultaneous masking. {Only Rodenburg *et al.* (1974) found essentially similar slopes for the two conditions except that the tip of the tuning curve was sharper in nonsimultaneous masking.} Vogten (1978b) and Moore (1978) ascribe the differences largely to lateral suppression, which is only effective simultaneously with the masker. In our nonsimultaneous-masking experiment we used the technique of forward masking and again the tuning curve was approximated by four points (see Fig. 3.1).

For the shallow edge (Experiment 6) masker frequencies of 950 and 750 Hz were used and the step size in the adaptive procedure was 3 dB. For the steep edge (Experiment 7) masker frequency was the dependent variable; masker levels of 70 and 90 dB SPL were used. Because of the shallower psychometric function in forward masking the step size was, contrary to experiment 5, 10% of the frequency difference between masker and probe tone. In these experiments we used no low-pass background noise.

a. Results

The four filled symbols in Fig. 3.4 and the associated bars are the average thresholds and the between-subjects standard deviations. In forward masking large interindividual differences may be expected (see Experiment 9). This yields a set of tuning curves which have shifted parallel to each other along the vertical axis. To illustrate this behavior the subjects were divided into five subgroups differing mainly in this aspect of their tuning curve. In order to prevent the classification in subgroups from being contaminated with differences in the steepness of the slopes, the grouping was made on the basis of the masker level at the tip of the tuning curve. To this end for each subject the tuning curve was approximated with two straight lines and the minimal masker level required was calculated from the intersection. The five solid curves are the average results for the five subgroups obtained in this way.

On the shallow edge the standard deviation within subjects is 3.9 dB. The correlation coefficient between the results at 750 and 950 Hz is 0.73. The average slope is 63 dB/oct with a standard deviation of 13 dB/oct and the test-retest correlation coefficient of the slope data is 0.41, which gives a reliability coefficient r_{tt} = 0.58.

For the steep edge the standard deviation within subjects is about 20% of the frequency difference between probe and masker. The correlation coefficient between the results for the two masker conditions (70 and 90 dB SPL) is r = 0.85, which fits in with the parallel shift discussed above. The slopes on the high-frequency side of the tuning curve are very steep, averaging 380 dB/ oct. For the same reasons as discussed with Experiment 5, the steep-edge slope values that will be used in our further calculations are the differences between the log (Δf) values for the two masker conditions. The reliability coefficient of the steep slope is 0.80. The Q_{10dB} for the tuning curves in nonsimultaneous masking is 8.4, with a between-subjects standard deviation of 1.8.

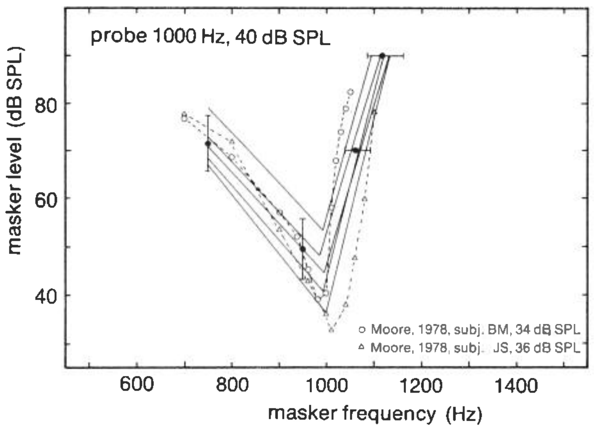


Fig. 3.4. Tuning curves in nonsimultaneous masking. The solid symbols and the associated bars are the average thresholds and the between-subjects standard deviations, respectively. The five fully-drawn lines give the mean results for subgroups of ten subjects, composed on the basis of the minimum masker level obtained by extrapolation from the four data points per subject. For comparison some results by Moore (1978) are given with open symbols.

b. Discussion

The major difference between the tuning curves in simultaneous and nonsimultaneous masking is their sharpness. Vogten (1978b) ascribes this difference largely to the influence of lateral suppression, which is only effective simultaneously with the masker. By and large the tuning curve in nonsimultaneous masking should correspond to neural tuning, and the tuning curve in simultaneous masking should correspond to neural tuning including the inhibition areas. Moore (1978), however, argues that psychophysical tuning should be sharper than neural tuning, because for one thing the observer can make use of pitch cues (differences between masker and probe), and for another the observer is able to shift his "auditory filter" in order to optimize the signal-to-noise ratio. For comparison some of the results obtained by Moore are shown in Fig. 3.4. The resemblance is good, but our straight lines are only approximations of the more extensively determined tuning curves. For this reason there is some discrepancy between the average value of Q_{10dB} found in this study and that found by Moore (1978) (8.4 and 11.2, respectively).

3.3.5. Temporal resolution

Green (1973) suggests that at least two time constants can be distinguished in the auditory system: (1) a relatively long integration time, involved, for example, in measuring the threshold for a tone burst as a function of its duration, and (2) a short integration time, involved, for example, in the detection of a gap in noise as a function of gap duration. The short integration time should reflect peripheral processes, and the longer one may be a parameter of a more central mechanism. Because our prime interest is in peripheral auditory processing, we sought to measure the short integration time. Analogous to the bandwidth in the spectral resolution we chose a description by means of an intensity-weighting function (time window), and used a measuring technique which is the time-domain analog of Experiment 2 (see Fig. 3.1). The masker was sinusoidally intensity-modulated white noise (peak-to-trough ratio of 20 dB), low-pass filtered (48 dB/oct) with a cutoff frequency of 4000 Hz. The difference in masked threshold between peak and trough was measured as a function of the modulation frequency. The probe was an 0.4-ms click, octave filtered with a central frequency of 1000 Hz to limit detection of the click to the frequency region under investigation. The time sequence shown in Fig. 3.1 is one trial of the 2AFC procedure for the trough condition with a modulation frequency of 10 Hz. The probe, which had a fixed level of 65 dB SPL peak, was presented twice within one observation period, with a fixed interval of 200 ms, and occurred in all conditions at the same time relative to the start of the masker. Masked thresholds were determined at the peaks and the troughs for three modulation frequencies: 10, 15, and 20 Hz. In each measurement block all conditions were measured once in a fixed order, with the peak condition first, followed immediately by the corresponding trough condition. Both test and retest consisted of two blocks.

a. Results

The subjects were divided into five subgroups on basis of their average threshold difference between peak and trough. Fig. 3.5 gives the threshold differences for these five groups as a function of modulation frequency. The smooth curves give the theoretical peak-to-trough difference for a window with a Gaussian shape. The data, however, show a sharper decrease in threshold difference as a function of modulation frequency, and could only be optimally described by a rectangular window, which is, for each measuring condition, shifted to the position that gives an optimum signal-to-noise ratio. However, such a window is very unrealistic; moreover, it would imply that the threshold difference reaches zero for modulation frequencies substantially lower than those found by Rodenburg (1977) and in our own laboratory. To account for threshold differences at high modulation frequencies we need a temporal window with a sharp tip as, for example, an exponential function, which, however, gives too shallow a decrease for the threshold difference in the region between 10 and 20 Hz.

For the data presented here the Gaussian window shape offers a reasonable compromise. This window was applied in a least-squares approximation to the data, in order to determine the temporal width τ. For the whole group the average width $\overline{\tau}$ is 19.6 ms with a between-subjects standard deviation of 4.1 ms. The reliability coefficient of the temporal widths is 0.84.

b. Discussion

The width of the window that fits the data depends strongly upon the window shape adopted in the calculations. We found an integration time of about 20 ms for a Gaussian window, 25 ms for a rectangular window, and 10 ms for an exponential window. The differences between the results of this experiment and the data presented earlier (Festen *et al.*, 1977), with τ = 13.5 ms, reflect such differences in the window definition. Although recently Buunen and van Valkenburg (1979) found an integration time of 25 ms for the detection of a

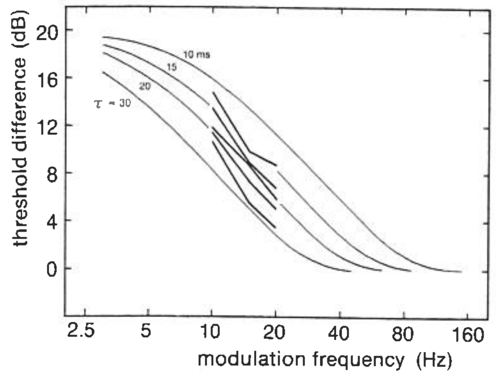


Fig. 3.5. Threshold-level difference, for octave-filtered clicks around 1000 Hz, between peak and trough of intensity-modulated noise as a function of modulation frequency. The modulation depth was 20 dB. The smooth curves give the theoretical threshold differences for a window with a Gaussian shape; the width of the window is the parameter of the curves. The five heavy lines give the results for five subgroups of ten subjects. The subgroups were composed on the basis of the mean threshold difference.

gap in noise, there are considerable differences between our results and the data by other investigators who used slightly different methods. Rodenburg (1977) used intensity-modulated noise and found, with a 4000-Hz probe, a transfer function with a cutoff frequency of 50 Hz (τ = 3 ms). Viemeister (1977) found a low-pass function with a cutoff frequency of 35 Hz (τ = 4.5 ms). He used an amplitude-modulated noise masker and a 500-Hz wide click probe centered at 1000 Hz. Green (1973) reported a number of experiments in which the minimum integration time was measured, and found 2-3 ms in nearly all tests. Only the detectability of a click in a rectangular burst of noise shows a breaking point between 10 and 20 ms.

In explaining these differences we have to realize that the time constant of a low-pass filter describing the modulation transfer depends not only on the cutoff frequency but also on the shape of the filter. Other differences must be caused by the use of amplitude-modulated rather than intensity-modulated noise in many of the experiments and by differences in click frequency. Fortunately these problems have only minor influence on the relations between the width of the temporal window and the results of other experiments, because only the interindividual differences in the integration time rather than the overall magnitude were used for further analysis.

3.3.6. Forward and backward masking

Another way to study temporal resolution is to measure the extent of time over which a burst of noise masks a preceding or following signal (cf. Pickett, 1959; Elliott, 1962; Plomp, 1964; and Wilson and Carhart, 1971). Under backward masking Pickett (1959) found masking to extend back 25 ms, almost independent of the masker level, and Elliott (1962) found 50 ms. Under forward masking Plomp (1964) could fit his data on a logarithmic time scale by straight lines which reach the absolute threshold after 200 ms, independent of masker level. Wilson and Carhart (1971), who measured forward and backward masking as well as the combined effect, also found a maximum of 200 ms in forward masking. An extensive review of temporal masking was given by Duifhuis (1973).

In this experiment, we again used a constant probe. The masking curves obtained in this way are essentially the time-domain analogs of the tuning curves of the Experiments 4 and 5. As in Experiment 8, the probe was a 0.4-ms click, octave-filtered with a central frequency of 1000 Hz. The masker was a burst of noise, low-pass filtered (48 dB/oct), with a cutoff frequency of 4000 Hz (see Fig. 3.1). Because steep masking patterns are to be expected, the time interval Δt between masker and probe was used as the dependent variable. The step size in the adaptive procedure was 10% of Δt. The small Δt values obtained in backward masking necessitated a step size of 20% after the first false response. In each block the masked threshold was measured three times over for two spectral densities (35 and 55 dB per Hz), with the lower level always coming first.

Fig. 3.1 also shows the time sequence of the 2AFC procedure for both forward and backward masking. The masker was presented twice in each observation period, and in one of the observation periods both masker bursts were followed by octave-filtered clicks, with the same interval relative to the masker. This procedure was adopted because it appeared that the subject could discriminate more accurately between the probe and the random fluctuations of the noise near the end of the masker than with a conventional 2AFC. In the backward-masking experiment a comparable procedure was used. The masker was switched within 1 ms.

a. Results

Since in the adaptive procedure the step size was a constant fraction of the time between the probe and the masker, all calculations were performed on

the log (Δt) values. The results are given in Fig. 3.6 as a function of the time interval between the probe and the masker. The probe was presented at t = 0, so the forward-masking data, in which the masker preceded the probe, are shown in the left-hand panel and the backward-masking data are shown in the right-hand panel. The filled circles are the mean results of all subjects. The standard deviation, pooled over subjects and masker levels, is 7% of the interval between probe and masker in forward masking and 9.5% in backward masking. For both experiments there is a strong correlation between the results at the two masker levels (0.79 in forward masking and 0.85 in backward masking).

This shows that the larger part of the differences between subjects is caused by a parallel shift of the masking curves, as illustrated by the results of the five subgroups which were selected on the basis of the masking interval averaged over the two masker levels. Because of the only moderate correlation between the mean results in backward and forward masking (r = 0.50), the five subgroups in these two experiments were composed independently.

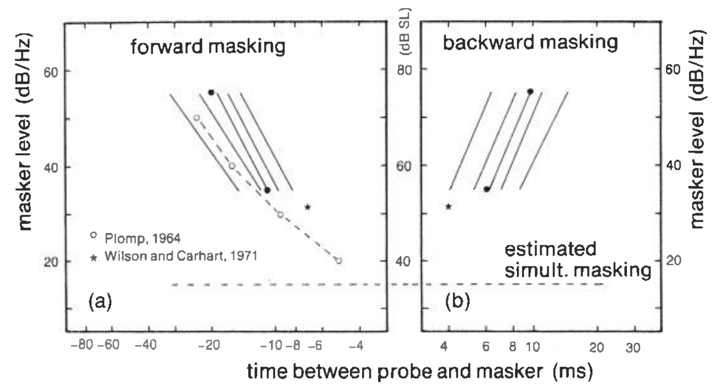


Fig. 3.6. Level of wideband noise just masking an octave-filtered click probe (1000 Hz) as a function of the time between masker and probe; panel (a) for forward masking and panel (b) for backward masking. The filled symbols represent the mean results for 50 subjects. The fully-drawn lines give the results for five subgroups of 10 subjects; the subgroups were composed separately for the forward and the backward masking on basis of the mean Δt-values. The open circles give results by Plomp (1964) and the asterisks represent data by Wilson and Carhart (1971).

Whereas the temporal locations are different, the slopes in forward and backward masking are similar. The average slope in forward masking is 21.6 dB per doubling of the time interval between masker and probe, with a between-subjects standard deviation of 4 dB, and in backward masking the slope is 25.3 dB per doubling of Δt, with a standard deviation of 5 dB. The reliability coefficient of these slopes is 0.76 in forward masking and 0.64 in backward masking.

b. Discussion

By estimating the sensation level of our signals some of the data from the literature can be plotted in Fig. 3.6. The ordinate between the two panels of Fig. 3.6 gives the estimated sensation level of the masker and the dashed horizontal line gives the estimated masker level needed for simultaneous masking of the probe. This level was derived from the results of Experiment 8, in which the same probe was used in simultaneous masking.

Plomp (1964) determined the length of the just-noticeable time interval between two bursts of noise as a function of the level of the second burst, for three levels of the first burst. The open circles in Fig. 3.6 give his results for a second burst of 35 dB SL. A noise burst at this level is just capable of masking a probe as used in our experiment. The asterisks in Fig. 3.6 are estimates based on the data by Wilson and Carhart (1971). In spite of the differences among the studies (Wilson and Carhart used a wideband probe) there is reasonably good agreement among the three sets of data.

The origin of nonsimultaneous masking is discussed extensively by Duifhuis (1973). In his theory backward masking and a short-term component of forward masking are consequences of peripheral frequency resolution. The long-term component of forward masking is considered to be related to neural adaptation. The transition from the short-term to the long-term component takes place at time intervals of about 20 ms, so that our data should mainly concern the steep short-term component. However, Elliott (1962) mentioned a more rapid neural transmission of more intense stimuli as a possible explanation for backward masking.

3.3.7. Cubic difference tone

The most important distortion products generated by the nonlinearity of the auditory system are the difference tone $f_2 - f_1$ and the members of the class $f_1 - n(f_2 - f_1)$ with $f_2 > f_1$ and n a small integer (Plomp, 1965). For

many years the scarce data available for the difference tone $f_2 - f_1$ indicated that this distortion product is generated by an "overload" type of nonlinearity, which can be described by a small quadratic term in the transfer function (Zwicker, 1955; Goldstein, 1967). Only recently has it been shown that properties of the difference tone are more complicated (cf. Hall, 1972; Humes, 1979).

For the more prominent combination tones $f_1 - n(f_2 - f_1)$ more extensive data are available (see Smoorenburg, 1972). The level of these distortion products relative to the level of the primary tones is nearly constant over a large range (20-70 dB SL) and is, under favorable conditions, only 10 dB lower than the primaries.

This motivated Goldstein (1967) to refer to an "essential" nonlinearity. Moreover, this nonlinearity seems tightly coupled to the frequency resolution: (1) the generation of combination tones is strongly favored by a small frequency separation between the primaries, suggesting that filtering precedes their origin, whereas (2) the combination tones act psychophysically as ordinary acoustic signal components, suggesting that they are already present in the cochlea. The strongest member of this class is the cubic difference tone (CDT) for which $n = 1$. In this study the level of the CDT was measured for one specific condition of the generating components. The levels, L_1 and L_2, of the primary tones were 70 and 60 dB SPL, respectively. The frequencies f_1 and f_2 were 1200 and 1400 Hz, so that the probe frequency, being $2f_1 - f_2$ was 1000 Hz as in all other experiments. A drawback of this choice is that the region where the CDT is generated is presumably a little beyond the frequency under investigation. This is not too serious provided that the properties of the auditory system do not change rapidly with frequency.

The level of the CDT is considered to be equal to the level of a reference masker (of frequency $2f_1 - f_2$) which gives the same amount of forward masking. The measuring procedure consisted of two stages as shown in Fig. 3.1. First the forward-masking-threshold level of a 1000-Hz probe was measured for the two-tone masker generating a CDT at the probe frequency, and secondly, the level of an acoustic component of frequency $2f_1 - f_2$ necessary to mask this probe was measured. A component with frequency f_1 and with the same level as used in the two-tone stimulus is added to the reference masker in order to make the reference maximally similar to the stimulus. In this way both masking effects and suppression effects caused by f_1 are present in the original stimulus as well as in the reference. Both in the test and in the retest the measurement of the CDT, including the measurement with the reference masker, was performed three times.

Results and discussion

The average CDT level is 44.9 dB SPL and the between-subjects standard deviation is 3.6 dB. The accuracy of the CDT level appears from the within-subjects standard deviation, which is 4.4 dB. The reliability coefficient of the CDT levels as estimated from the test-retest correlation is 0.71.

Zwicker (1955) determined the cancellation level of the CDT generated by f_1 = 1000 Hz and f_2 = 1200 Hz with $L_2 = L_1 - 10$ dB as a function of L_1. With seven subjects he found at L_1 = 70 dB SPL an average CDT level of 43 dB SPL, and a range of about 10 dB among the subjects. This is in good agreement with our data.

Both in the cancellation data by Zwicker and in our data effects of suppression do not play a role. We would have found, however, lower CDT levels due to lateral suppression by using a reference masker without the f_1 component (1200 Hz) (cf. Smoorenburg, 1974).

3.3.8. Suppression

In the final experiment a second manifestation of auditory nonlinearity was studied. In a nonsimultaneous masking experiment with masker and probe at the same frequency, the masked threshold may under certain conditions decrease when a second tone is added to the masker.

This reduction may be interpreted as a suppression of the first masker component by the second component, called the suppressor. Houtgast (1973) was the first to investigate this phenomenon. He used a wide variety of measurement techniques and concluded that a reduction in masked threshold only occurs under those conditions in which probe and masker are presented nonsimultaneously. He argued that in simultaneous masking both the masker and the probe are suppressed by the suppressor so that no effect can be measured. In conditions of nonsimultaneous masking the masker is suppressed and the probe is not, provided that suppression acts only instantaneously. In general weak signals are suppressed by stronger signals that are higher in frequency; only at high intensities does suppression also occur on the high-frequency side of the suppressor.

In the present study the suppression of a 1000-Hz tone of 65 dB SPL was measured in forward masking. In order to introduce strong suppression effects the suppressor was not a pure tone but a complex tone which contained four partials of 85 dB SPL each as shown in Fig. 3.1. All masker components were added in sine phase. A drawback of this multicomponent suppressor is the generation of combination tones, which influence the measured suppression. The procedure is the same as the one used in Experiment 11 for the CDT. The level of the suppressed tone is defined as the level of a reference masker that gives the same amount of forward masking. The complete measurement, i.e., the determination of the probe level and subsequently the determination of the reference masker level, was repeated three times over in the test and in the retest.

Results

The average level of the suppressed 1000-Hz component was 42.5 dB SPL, with an input level of 65 dB SPL, giving an average suppression of 22.5 dB. The between-subjects standard deviation was 11.5 dB, but presumably this value was caused partly by interindividual differences in level and phase of the combination tones. Like the measurement of the CDT this test also had a high accuracy. The within-subjects standard deviation was 5 dB, giving a standard error of 2 dB. As a consequence of the small standard error and the large differences between subjects, the reliability of this test is very high: r_{tt} = 0.96.

3.4. RELATIONS AMONG THE TESTS

In the previous section we described the 12 tests of this study and we discussed the results and their reliability. However, we were not interested primarily in the individual tests, but rather in the relations among tests. In this section we will discuss these relations in terms of the correlation coefficients between the test results and we will show how those coefficients are affected by measurement error. The statistical analysis was performed by applying the SPSS package of computer programs (Nie *et al.*, 1975).

3.4.1. Difference scores

Before studying the relations among the test results, we have to (1) verify whether we succeeded in canceling the effects of training by using difference scores, and (2) check whether the reliability of the tests justifies looking for relationships.

a. Effects of training

An effect of training in the test results would have been better (lower) thresholds in the retest than in the test, and should be revealed in an analysis of variance as an effect of *sessions*. Interindividual differences in the effect of training should be reflected in the interaction of *subjects* and *sessions*. Another important constituent of this interaction are day-to-day variations in the auditory functions. As effects of training will, in a first approximation, have the same influence on both thresholds that constitute the difference score, they will be canceled in the difference scores, but day-to-day variations in the auditory functions may manifest themselves in the raw scores as well as in the difference scores.

Table 3.1 shows some of the results of an analysis of variance on each of the 12 tests for the raw data and for the difference scores. The second column gives the average difference between test and retest for the raw data. The plus sign, which is present for all these differences, indicates that the threshold was better in the retest than in the test. Clearly, those differences are caused by training, and although only a small part of the total variance in the raw data (column 3), they are significant for most of the tests. In the difference scores, however, the *session* effect is not significant for any test (column 8), which demonstrates that the average effect of training was eliminated.

Experiments	Raw data					Difference scores			
	Sessions			*Sessions×subj.*		*Sessions*		*Sessions×subj.*	
	Mean difference	Variance (%)	*p*	Variance (%)	*p*	Variance (%)	*p*	Variance (%)	*p*
1 Absolute threshold	+0.4 dB	<0.1	0.56	15	<0.001				
2 Bandwidth simult.	+0.6 dB	0.3	0.004	1.6	<0.001	<0.1	0.51	2.2	0.02
3 Bandwidth nonsimult.	+0.7 dB	<0.1	0.21	3.7	<0.001	<0.1	0.56	4.4	<0.001
4 Shallow edge simult.	+0.5 dB	<0.1	0.04	0.9	<0.001	<0.1	0.96	7.9	<0.001
5 Steep edge simult.	+4%	0.5	<0.001	0.7	<0.001	<0.1	0.59	3.0	<0.001
6 Shallow edge nonsimult.	+2.0 dB	0.5	0.001	2.0	<0.001	0.6	0.13	15.3	<0.001
7 Steep edge nonsimult.	+9.5%	0.7	0.002	2.3	<0.001	0.1	0.23	1.5	0.24
8 Temporal window	+1.4 dB	1.3	<0.001	1.3	<0.001	<0.1	0.12	<0.1	0.44
9 Forward masking	+10%	1.0	<0.001	3.3	<0.001	<0.1	0.60	5.4	0.009
10 Backward masking	+18.5%	3.6	<0.001	4.4	<0.001	<0.1	0.26	6.5	0.008
11 Cubic difference tone	+0.4 dB	<0.1	0.31	2.2	0.03	0.4	0.10	2.4	0.17
12 Suppression	+0.6 dB	<0.1	0.15	2.0	<0.001	<0.1	0.14	0.7	0.14

Table 3.1. Results of 23 analyses of variance (for each of 12 tests and for both raw data and difference scores) for the mean effect of *sessions* and the interaction of *sessions* and *subjects*. The variances are given as percentages of the total variance, which is much smaller for the difference scores than for the raw data.

As mentioned before, in the raw data the interaction of *sessions* and *subjects* is made up of two parts: (1) differences in training effects among the subjects, and (2) day-to-day variations in auditory functions. Since the average effect of training has been eliminated in the difference scores, it is reasonable to assume that this also holds for the individual subjects and that if a significant interaction remains in the difference scores this must be due to day-to-day variations in the auditory function.

b. Reliability of the tests

The degree to which a correlation coefficient is affected by measurement error is given by the equation

$$r_\infty = r_{xy} \,/\, (r_{xx} r_{yy})^{0.5} \;, \qquad (3.2)$$

where r_∞ is the correlation coefficient between the true components in the

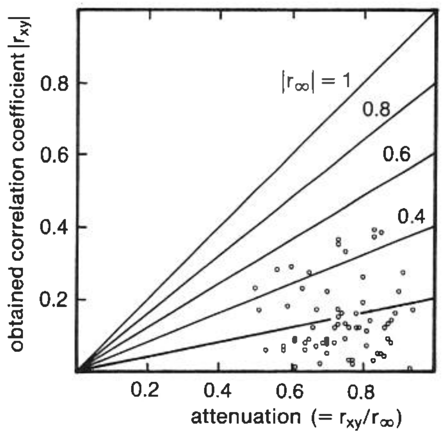


Fig. 3.7. Scatter diagram of the obtained correlation coefficients versus their attenuation coefficients. The correlations corrected for attenuation can be gathered from the sloping lines.

tests X and Y, or between the average results of the tests X and Y if both had an infinite duration, r_{xy} is the obtained correlation coefficient and r_{xx} and r_{yy} are the coefficients of reliability of X and Y, respectively (cf. Guilford, 1954). The *attenuation* of correlation coefficients, given by the denominator in Eq. 3.2, can be calculated for each pair of tests and since we also know the obtained correlations, from the next section, we are able to estimate the true correlation. For this purpose a scatter diagram of the obtained correlations versus their corresponding attenuations is given in Fig. 3.7. In this diagram the true, or corrected, correlations are represented by straight lines at different angles to the abscissa. We see that the attenuation coefficients are close to unity (caused by high test reliabilities) and, consequently, after correction for attenuation most correlations are still very low; none of the corrected correlations is greater than 0.5. If the obtained correlations had been low because of measurement error, a correction for attenuation should have recovered the correlations. Fig. 3.7 shows that test reliability did not severely limit the obtained correlations.

c. Correlations

The obtained coefficients of correlation between the difference scores (including the absolute threshold) are given in Table 3.2. A factor analysis of this matrix of correlations would show up underlying factors, representing basic properties of the auditory system. However, it turned out that this correlation matrix is not appropriate for factor analysis. Dziuban and Shirkey (1974) recommended three tests to determine whether or not a correlation matrix is suitable for factor analysis. First of all, Bartlett's test of sphericity which tests whether the correlation matrix comes from a multivariate normal population in which the variables of interest are independent. This test is regarded as a first selection and for our matrix of correlations the hypothesis of independence is rejected (p = 0.005). But the matrix does not pass the criterion of the other two procedures, testing its psychometric adequacy. Despite this finding we performed the analysis but found no comprehensible structure.

Although factor analysis has to be abandoned, we found a number of interesting correlations that are not accidental as shown by Bartlett's test. For instance, there is a significant correlation (r = -0.29) between the steepness of the shallow edge of the tuning curve and the width of the auditory filter, both measured in simultaneous masking. This correlation may be interpreted as a causal relation in which the shallow edge is one of the determinants of the

Experiment	Kind of score	2	3	4	5	6	7	8	9	10	11	12
1 Absolute threshold	level	0.23	0.17	-0.04	0.27	-0.36	-0.38	0.08	0.37	0.12	0.26	0.00
2 Bandwidth simult.	width		0.17	-0.29	-0.08	-0.23	-0.18	-0.09	-0.06	-0.06	0.28	0.27
3 Bandwidth nonsimult.	width			-0.05	0.16	-0.07	0.08	-0.08	0.07	0.01	-0.09	-0.13
4 Shallow edge simult.	slope				-0.13	0.09	-0.07	0.39	0.12	0.08	0.03	-0.16
5 Steep edge simult.	$(slope)^{-1}$					-0.15	-0.12	-0.14	0.05	-0.03	-0.11	-0.17
6 Shallow edge nonsimult.	slope						-0.02	-0.22	-0.12	-0.09	-0.06	0.33
7 Steep edge nonsimult.	$(slope)^{-1}$							0.14	-0.12	0.12	-0.09	-0.08
8 Temporal window	width								0.19	0.35	-0.02	-0.22
9 Forward masking	$(slope)^{-1}$									0.08	0.13	0.05
10 Backward masking	$(slope)^{-1}$										-0.17	-0.16
11 Cubic difference tone	strength											-0.03
12 Suppression	strength											

Table 3.2. Matrix of correlation coefficients between the difference scores (including the absolute threshold). With each of the experiments the kind of score is indicated for a correct interpretation of the sign of the correlations. Underlined values are significant at 1% level; dashed lines indicate a 5% level of significance.

bandwidth. There is a positive correlation (r = 0.28) between the strength of the cubic difference tone and the auditory bandwidth, which is in line with the general finding that the generation of combination tones is favored by a strong interaction between the primaries. Regarding nonlinearity, there is also a positive correlation (r = 0.33) between the strength of lateral suppression and the steepness of the shallow edge of the tuning curve measured in nonsimultaneous masking, suggesting a sharpening of the frequency selectivity due to suppression.

With respect to temporal resolution, there are a few significant correlations. There is a positive correlation (r = 0.39) between the width of the temporal window and the steepness of the shallow edge of the tuning curve. Furthermore, there is also a positive correlation (r = 0.35) between the reciprocal of the slope of the backward-masking curve and the width of the temporal window, which means that a steep slope goes with a narrow window. Finally, the absolute threshold is significantly correlated with the forward-masking slope and with each of the slopes of the tuning curve in nonsimultaneous masking. For these correlations we have no simple explanation.

3.4.2. Some additional scores

Besides difference scores, other interesting scores can be calculated from the raw data. As is shown in the discussion of the individual experiments, these additional scores may represent large interindividual differences. In this section the following scores are studied.

(1) For the tuning curve in simultaneous masking the mean threshold level for the two masker frequencies on the shallow edge (score 4a) and the mean threshold frequency for the two masker levels on the steep edge (score 5a). In simultaneous masking there are large differences among the subjects in the frequency position of the steep edge of the tuning curve, see Fig. 3.3. For the shallow edge there exists a similar effect, although less marked and therefore not indicated in Fig. 3.3. The same effect can be seen in the tuning curve measured in nonsimultaneous masking (Fig. 3.4), but it is not relevant here because it only leads to a trivial correlation with the forward-masking data.

(2) For the tuning curve in simultaneous masking as well as in nonsimultaneous masking the Q_{10dB} was calculated (score 4,5, and 6,7, respectively). Here we expected a strong correlation with the auditory bandwidth because both Q_{10dB} and the bandwidth as measured in the Experiments 2 and 3 are related to the tip of the auditory filter.

(3) Just as we found most of the interindividual differences in a parallel shift of the steep edge of the tuning curve, the majority of the differences among subjects in backward and forward masking are also caused by parallel shifts (see Fig. 3.6). For this reason the mean Δt between masker and probe in forward and in backward masking were incorporated in our study on relations between auditory tests.

The correlations between these additional scores and the difference scores and those among the additional scores are combined in the matrix of Table 3.3. Correlations between scores obtained from the same data are omitted. For instance, correlating Q_{10dB} with any other measure of the tuning curve is not allowed because both parameters are based on the same data, so that the error parts of the scores would coincide. The problem with the additional scores representing the parallel shift of a masking curve is that these scores are confounded with differences in training and alertness among the subjects. For

the correlations between the "shift" scores and the difference scores this problem is not too serious because the confounding effects are only present in one of the two tests and cannot cause spurious correlations. However, the correlations among the "shift" scores mutually are questionable. High correlations may very well be the result of effects of training and alertness in both tests, as, for instance, with respect to the mean Δt in forward masking and in backward masking ($r = 0.50$).

Relations among the scores

Apart from the problems discussed above, the additional scores introduce a number of interesting correlations. In simultaneous masking the parallel

	Experiment	Kind of score	4a	5a	9a	10a	4,5	6,7
1	Absolute threshold	level	-0.06	0.29	0.23	0.06	-0.06	-0.16
2	Bandwidth simult.	width	-0.25	0.38	0.03	-0.03	-0.26	-0.05
3	Bandwidth nonsimult.	width	0.01	0.14	-0.07	0.02	-0.08	-0.12
4	Shallow edge simult.	slope		0.06	0.24	0.50		0.11
5	Steep edge simult.	$(\text{slope})^{-1}$	-0.03		-0.09	-0.23		-0.20
6	Shallow edge nonsimult.	slope	0.22	-0.33	0.00	0.00	0.11	
7	Steep edge nonsimult.	$(\text{slope})^{-1}$	0.08	-0.22	-0.08	-0.03	-0.04	
8	Temporal window	width	0.11	0.18	0.49	0.53	0.39	-0.36
9	Forward masking	$(\text{slope})^{-1}$	0.00	0.10		0.17	0.13	-0.02
10	Backward masking	$(\text{slope})^{-1}$	0.05	-0.05	0.28		0.08	-0.22
11	Cubic difference tone	strength	0.13	0.19	0.04	-0.05	0.07	-0.03
12	Suppression	strength	0.07	-0.22	-0.21	-0.09	-0.11	0.48
4a	Shallow edge simult.	mean level		-0.30	-0.25	0.07		0.26
5a	Steep edge simult.	mean freq.			0.50	0.34		-0.27
9a	Forward masking	mean Δt				0.50	0.23	-0.19
10a	Backward masking	mean Δt					0.52	-0.09
4,5	Q_{10dB} simult.	sharpness						0.15
6,7	Q_{10dB} nonsimult.	sharpness						

Table 3.3. Matrix of correlation coefficients introduced by the additional scores. Correlations between scores obtained from the same data are omitted. Underlined values are significant at 1% level; dashed lines indicate a 5% level of significance.

shift of the steep edge, away from the probe frequency, correlates ($r = 0.38$) with the bandwidth of the auditory filter. This indicates once more that the width of the tuning curve is related to the width of the auditory filter measured with comb-filtered noise, as expected. This conclusion is supported by the negative correlation between the sharpness of the tuning curve (Q_{10dB}) and the auditory bandwidth in simultaneous masking ($r = -0.26$). On the other hand, only a very small part of the variance ($r^2 = 0.15$) in the width of the auditory filter is accounted for by the parallel shift in the steep edge of the tuning curve and, moreover, this relation is not found in nonsimultaneous masking.

The possible relation between suppression and the shift of the steep edge, which was discussed with Experiment 6, could not be demonstrated (the correlation -0.22 is small and has the wrong sign). With respect to suppression, however, there appears to be a rather high correlation with Q_{10dB} in nonsimultaneous masking ($r = 0.48$). This correlation may be understood as a sharpening of the auditory filter as a result of suppression. The correlation between suppression and the slope of the shallow edge of the tuning curve in nonsimultaneous masking ($r = 0.33$) may be understood in the same way.

Finally, some high correlations are related to temporal resolution. The width of the temporal window is correlated with the parallel shift away from the masker in both the forward-masking curve ($r = 0.49$) and in the backward-masking curve ($r = 0.53$). Additionally, this shift of the backward-masking curve away from the masker is correlated with good frequency resolution as measured by the slope of the shallow edge in simultaneous masking ($r = 0.50$) and by Q_{10dB} in simultaneous masking ($r = 0.52$). These relations show a trade-off between temporal resolution, on the one hand, and frequency resolution, on the other hand, which is in agreement with the theory proposed by Duifhuis (1973).

3.5. DISCUSSION

In the previous sections it was shown that the reliability of the tests was high enough to reveal correlations. For the pair of tests with the lowest reliabilities the attenuation was 0.5 and, consequently, even a true correlation as low as 0.56 would lead to a significant obtained correlation (5% level, two tailed). However, since few significant correlations among the tests were found, we may be tempted to conclude that most of the auditory functions as studied here are independent. Or to put it in another way, the characteristics of the auditory system in the region of 1000 Hz cannot be

described by only a few parameters. One can find two reasons why this conclusion may be somewhat premature.

First, two auditory functions that are in fact dependent will be highly correlated only when all other influences are held constant. For this reason we strove to eliminate confounding variables like training, alertness, and test sequence. There could be, however, other variables influencing the auditory functions and obscuring the correlations. But even in a rather homogeneous group of subjects, as in this study, such influences are not very likely. Perhaps this question can be answered in a sequel to this study with hearing-impaired subjects.

Second, we have to consider to what degree our tests reflect the auditory functions. In order to keep measuring time within reasonable limits, we had to accept a rather restricted number of conditions in each test. We cannot exclude the possibility that the auditory functions are not smooth enough to be represented by only a few data points. This problem does not affect the reliability of the tests because it is not caused by measurement error, but it can affect the correlations between the tests.

3.6. CONCLUSIONS

The most important conclusions of this study are:

(1) The results of the individual tests are in good agreement with the existing literature;

(2) A rigid experimental design with a test and a retest for each experiment is necessary to estimate all sources of variance and to minimize the influence of systematic errors;

(3) The reliability of the tests is high enough to show correlations between the tests, notwithstanding the use of untrained subjects;

(4) The strength of the cubic difference tone is favored by a wide auditory filter;

(5) There is a trade-off between frequency resolution on the one hand and temporal resolution on the other hand;

(6) Most of the investigated auditory functions appear to be independent of each other.

CHAPTER 4

STATISTICAL INTERMEZZO ON THRESHOLD MEASUREMENTS

Summary

In this chapter the measurement procedure used in these studies is investigated utilizing the obtained data. A short introduction of the interpretation of the psychophysical threshold is given and the need of forced-choice procedures to obtain criterion-free results is discussed. The measurement procedure is divided in two phases. In the first phase observations are made according to an adaptive procedure and in the second phase the threshold is estimated from the obtained data. The parameters of the adaptive procedure are discussed and three procedures for estimating the threshold are outlined. The differences found between these procedures are small and constant over subjects. The second part of the chapter is devoted to measurement error in the thresholds. By using an analysis of variance, measurement error is estimated for various tests as a function of the number of trials in the adaptive procedure. For runs of up to 30 trials in the measuring stage no signs are found of a reduced alertness towards the end of the run. In addition, the accuracy of thresholds from different estimating techniques appears to be very much alike. Finally, because the start of the adaptive procedure provides misleading information, the measurement error is studied as a function of the number of leading trials which are disregarded in estimating the threshold.

4.1. INTRODUCTION

For the threshold determinations presented in the previous chapter a number of parameters had to be chosen, either based on information from the literature or on our own experience. In this chapter we will try to evaluate these procedures using the obtained data. The threshold determinations can be divided in two phases, the first in which observations are made according to a certain procedure and the second in which the threshold is estimated from the obtained data.

The goal of the measurement procedure in phase one is to select the experimental parameters of each trial, such that the subject's responses convey a maximum of information about the position of the threshold. Trials with too high or too low stimulus levels evoke highly predictable responses and supply only marginal information. So, trials should be placed near the anticipated threshold. Because we have only limited *a priori* information about the position of the threshold, the results on previous trials should be used to select new test levels. The most simple form of an adaptive procedure is the "Simple Up-Down" or "staircase" method for estimating the 50% level, as introduced by Dixon and Mood (1948) for testing the sensitivity of explosives to shock. The application of this procedure, and of deduced procedures to estimate other points of the psychometric function, is widely spread in psychoacoustics; for instance, the sequential up-and-down method (Cardozo, 1966), PEST (Parameter Estimation by Sequential Testing) (Taylor and Creelman, 1967), UDTR (Up-Down Transformed Response) (Levitt, 1971), and BUDTIF (Block-Up-and-Down, Two-Interval, Forced-choice) (Cambell, 1974). In this study we used the UDTR procedure with some modifications. The procedure started well above threshold to help the subject in directing his attention to the stimulus. But, in order to prevent a waste of many trials at inadequate test levels, the first stage of each test sequence or *run* was a starting procedure which efficiently guides the stimulus level to the threshold region.

An important parameter of the adaptive procedure is the step size. If, in the starting procedure, the step size is too small, it takes many trials to reach the threshold. On the other hand, if the steps are too large, the procedure may easily pass over the threshold region to conditions for which the stimulus is completely inaudible. After the threshold has been reached, the procedure should track this threshold during a number of trials. If in this stage steps are too large, there will be too many trials at "uninteresting" test levels. If, on the contrary, steps are too small, the procedure is not able to follow spontaneous fluctuations in the subject's threshold. In the original UDTR procedure the step size is variable, but in our experiments always a constant step was used throughout the test. The step size was adapted to the steepness of the psychometric function and amounts about 15% of the total transition from no detection to perfect detection. When starting above threshold, this step size gives, on the average, the first false response at about the target percentage of 75% in the psychometric function for a 2AFC task.

Another important parameter of the procedure is the run length. Statistically, the error variance will be reduced by a factor of two for each doubling of the number of trials, with an exception for very short runs. Because for

each new run a number of trials is wasted in the starting stage, it would be preferable to spend as many trials in one run as needed for the desired accuracy. However, subjects may become less alert with increasing run length and, as a result, accuracy will grow more gradually or will even diminish, depending on the amount of reduction in alertness. For this reason fairly short runs were used. To study the possible reduction of alertness we will calculate the error variance as a function of the number of trials (see Section 4.3.2.).

Regarding phase two of the threshold determinations, there are several methods to estimate the threshold from the obtained data. A very simple method, in which the average of the stimulus level over all trials after the starting stage is taken, was applied in the previous chapter. Here, we will compare, on the basis of error variance, this simple method with a method in which the average stimulus level of the reversals in the sequence is used (Levitt, 1971), and with the more sophisticated method of "maximum likelihood".

Often Monte-Carlo simulations have been used to optimize the parameters of measurement procedures. These simulations have two great advantages: first of all, they give quick results, and secondly, all parameters determining the adaptive procedure can easily be changed. This gives the opportunity to study also sub-optimum procedures. However, the Monte-Carlo technique, being purely statistical, is unrealistic because it cannot account for peculiarities of subjects like lapses and fatique. In fact results of Monte-Carlo simulations constitute an upper bound for the efficiency of adaptive procedures in case subjects always take optimum decisions. For more realistic simulations the Monte-Carlo technique can study at the most recovery from lapses, but their frequency of occurrence must be implemented. Also the influence of fatigue can only be modeled if we know its course and its effect on the process of making decisions. The only way to overcome these restrictions is to study adaptive procedures in practice.

In the first paragraph of this chapter some details of threshold determinations are presented. The first section of this paragraph deals with the measurement procedure and gives a brief introduction to modern opinions on the psychophysical threshold, to illustrate the need for forced-choice procedures. In the second section a few threshold-estimating techniques are described and the differences among their results will be discussed. The last section of this paragraph is devoted to the steepness of the psychometric function which can be estimated with the maximum-likelihood procedure. The results are used to evaluate the step size chosen for the adaptive procedure. Measurement error in threshold determinations as a function of the run length in the adaptive procedure and for different threshold-estimating techniques is discussed in the second paragraph.

4.2. PSYCHOACOUSTIC THRESHOLD

4.2.1. The measurement procedure

Before discussing the measurement of thresholds in detail, we will shortly dwell upon the nature of psychoacoustic thresholds. In the classical way of thinking there is a fixed, physiologically determined, threshold and we only measure a gradual transition between no detection and perfect detection because of subject failures. Since the late fifties, however, this conception is abandoned for a number of reasons, one of which is that statistically significant differences in threshold were found for different measurement procedures. The alternative "detection theory" (Green, 1960) starts from the detection of signals in noise, where the threshold is determined by the signal-to-noise ratio. For each observation the subject has two alternative hypotheses: signal plus noise or noise alone; he has to determine the likelihood of both and subsequently apply some decision rule for choosing. In this decision rule the subject incorporates the values and costs of both decisions. He may use quite different criteria like: "I don't want to miss any signal" or "I don't want to give whatever incorrect responses". Thus, according to detection theory, the result of such an experiment is determined by the statistical properties of signal and noise and by the criterion of the subject. The criteria used by different subjects need not be the same and even for individual subjects the criterion may shift, which is a trade-off between false positives and false negatives. The discovered dependence of threshold upon the measurement procedure is a consequence of differences in criterion imposed by the procedure. Criterion-free thresholds can be measured when we prevent from asking the subject whether there was a signal or not, which is accomplished in a so-called "forced-choice procedure". In this case the noise is presented in two or more intervals and the signal is added in only one interval. For each interval the subject has to determine the likelihood of both hypotheses and thereupon he must pick the interval for which the ratio of the two likelihoods is most in favor of the signal. When using such a procedure, the percentage of correct responses as a function of signal level (the psychometric function) depends, apart from on the number of alternatives, only upon the characteristics of

signal and noise. If the underlying distributions of noise and noise plus signal are Gaussian, the psychometric function can be described by a cumulative normal curve when plotted against signal energy (Green and Swets, 1973). Usually the threshold is defined as the test level at which the psychometric function reaches some target probability.

As mentioned in the introduction, the conditioning of trials in the region of the ultimate threshold is of crucial importance; responses will be nearly always correct for high test levels, or at chance level for low test levels. An example of the adaptive measurement procedure used in this study is shown in Fig. 4.1(a), and has been discussed in the previous chapter. The stages 1 and 2 provide a starting procedure in which the test level is quickly adapted to the threshold region. In stage 3 the threshold is tracked with a constant number of trials. The test level is lowered in this stage only after three successive correct responses and raised after each incorrect response. With this procedure the chances for stepping upward and downward are equal if the probability on a correct response is 0.794 for an individual trial. When lowering the test level after two correct responses, as used in stage 2, this probability is 0.707. Although both probabilities are equally distant from the midpoint of the psychometric function for a 2AFC task, we preferred the first because more correct responses may help to keep up the motivation of the subject. The step size is constant, and adjusted to an *a priori* estimate of the steepness of the psychometric function.

4.2.2. Operational definitions of the threshold

Taking the procedures discussed so far as a tool for optimum placing of observations, we can now concentrate on estimating the threshold from the obtained data. A statistically based treatment of the data is offered by the maximum-likelihood method. This method is illustrated in Fig. 4.1. For each test level i the number of trials N_i and the number of correct responses K_i are calculated, and next a psychometric function is fitted such that the probability for the measured data is optimum. With this method all trials are used, also those in the first two stages of the run. As discussed earlier, the form of the psychometric function can be described very well as cumulative normal, thus only mean and sigma have to be estimated. However, computations can be grossly simplified when using an approximation to the cumulative normal curve by a logistic curve

$$p(x) = \frac{1}{n} \left(1 + \frac{n - 1}{EXP\ \{-(x - M)\ /\ c.S\}} \right), \qquad (4.1)$$

where p(x) is the probability of a correct response, x is the stimulus level, and n is the number of alternative responses in the forced choice; the parameters M and S are the mean and spread of the psychometric function. With the constant c the logistic curve is adapted to the cumulative normal curve. For c = 0.4431, as used here, the steepness at x = M is the same for both functions, making S comparable to sigma. If, for each test level, correct responses are binomially distributed, the probability of K_i correct responses out of N_i trials is given by

$$P_i = \frac{N_i!}{K_i!.(N_i - K_i)!} \ .\ p_i^{K_i} \ .\ (1 - p_i)^{N_i - K_i} \ . \qquad (4.2)$$

When the chances at different test levels are independent, the compound probability for the complete set of responses is the product of P_i over all levels i. To estimate mean and spread of the psychometric function we have to maximize this probability, which is equivalent to finding the maximum of the logarithm of the probability without multiplication constant,

$$P^* = \sum_i K_i.\ln(p_i) + (N_i - K_i).\ln(1 - p_i) \ . \qquad (4.3)$$

The maximum of this function may be found, using an iterative search procedure like the method of "steepest descent".

An alternative and much simpler estimate of the threshold from a sequence of up-down data is the average of reversals in the sequence, as proposed by Levitt (1971). In order to reduce bias in the threshold estimate an even number of reversals should be used. Apart from the simplicity, this method is also relatively insensitive to the accidental course of the adaptive procedure. A third, also very simple, threshold estimate is the average of stimulus levels of all individual trials in stage 3. Trials in the starting procedure (stages 1 and 2) are excluded from the average because the threshold region may not yet have been reached. An advantage of this estimate above the average of reversals seems that more of the acquired information is used. The data in the previous chapter were calculated, using the latter simple procedure.

The various threshold-estimation procedures yield different thresholds. For instance, the mean from the maximum-likelihood procedure with n = 2 estimates the stimulus level for which the chance on a correct response, p(x), is 0.75, whereas the average of reversals in a procedure as shown in Fig. 4.1 estimates p(x) = 0.79. Table 4.1 shows average thresholds resulting from dif-

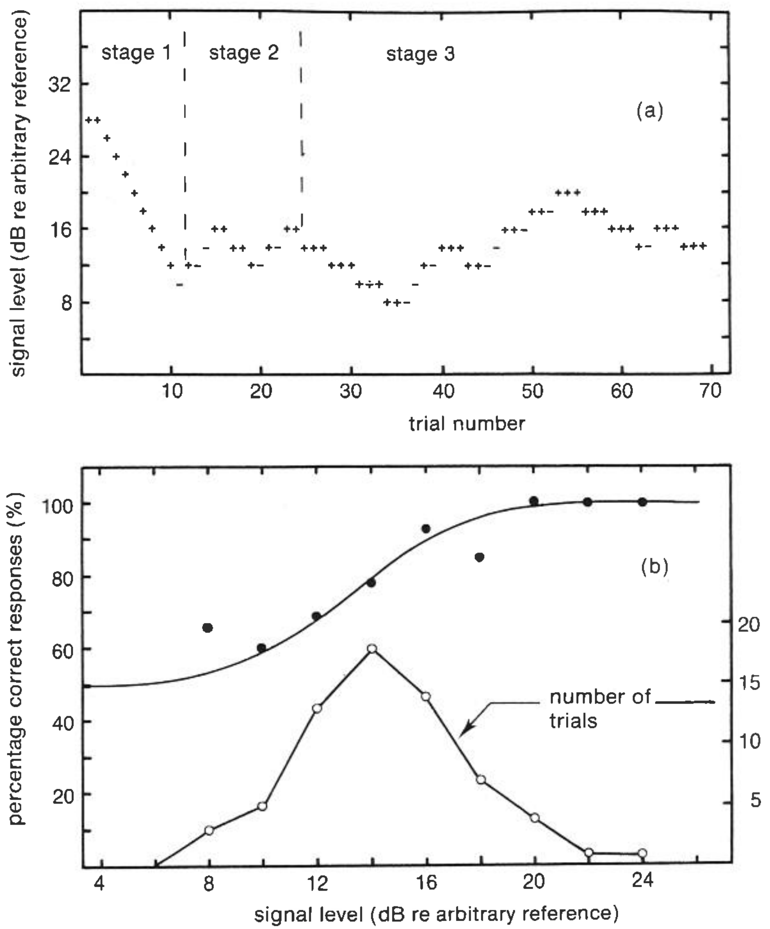


Fig. 4.1. Illustration of the method of maximum likelihood. Panel (a) gives an example of the course of the signal level within a run (+ for a correct response, - for an incorrect response). Panel (b) gives the percentage of correct responses from Panel (a) as a function of signal level (closed dots). The smooth line gives the psychometric function fitted to these data by optimizing Eq. (4.3). The number of trials for each signal level is given by open symbols.

Test	mean from max. likelihood	average of reversals	average of signal levels
Absolute threshold	5.45 dB	5.57 dB	5.91 dB
Bandwidth simult.	18.63 dB	18.42 dB	18.13 dB
Bandwidth nonsimult.	25.88 dB	25.33 dB	24.81 dB
Shallow edge simult.	70.73 dB	70.59 dB	70.24 dB
Shallow edge nonsimult.	61.68 dB	61.06 dB	60.53 dB

Table 4.1. Threshold estimates averaged over 50 subjects, for 5 tests and 3 estimation procedures.

ferent estimation procedures for five tests from the previous chapter. Note that the probe-signal level was varied for the absolute threshold and the masker level for the other four tests. The differences among estimates are up to 0.5 dB in simultaneous masking and up to 1 dB in nonsimultaneous masking. These differences are not subject dependent for no significant interaction between *threshold-estimating procedures* and *subjects* was found. So, the differences among estimation procedures are merely constants that do not obscure auditory differences among subjects. Concerning the differences among subjects, we may just as well use a simple threshold-estimating procedure instead of an advanced but laborious one. However, also the measurement error accompanying these methods should be considered, but, as will be shown in paragraph 4.3, these are very much alike.

4.2.3. Steepness of the psychometric function

Recently, fitting of the psychometric function by maximum-likelihood estimation was studied in Monte-Carlo simulation by Hall (1981). He used a PEST adaptive procedure, with minor modifications, for the placing of observations in a 4AFC task and concluded that the method was highly efficient for estimating the mean. However, from these simulations it appeared that the estimate of the spread (S), determining the steepness of the psychometric function, is severely biased. The magnitude of this bias was determined by Hall by comparing the spread values from the method of maximum likelihood with the original spread used in the Monte-Carlo simulations. For the PEST procedure which he used the estimated spread was between 0.7 and 0.9 of the true spread. This bias is caused by the adaptive procedure which selects test levels with the

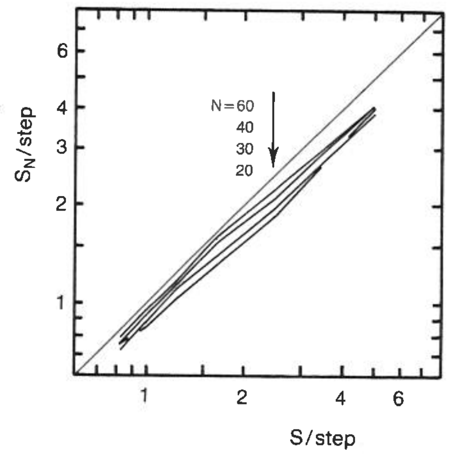


Fig. 4.2. Spread of the psychometric function estimated with the method of maximum likelihood as a function of the real spread, for runs with 20, 30, 40, and 60 trials. Each condition is based on 1000 simulations.

aid of previous responses. To concentrate trials effectively in a narrow region of the psychometric function, the adaptive procedure disregards levels that are judged as "too high" or "too low" on the basis of very short sequences. For instance, if we have three successive correct responses the corresponding test level is judged as "too high" and will be lowered. However, the chance on a correct response will presumably not be 100%. A step in the other direction is made after each incorrect response. This means that we leave the test level after collecting a minimum of information. As a consequence of this selection of test levels we obtain a transition from no detection to perfect detection which is sharper than reality. Spread values may be corrected for this bias, but for our adaptive procedure, used in a 2AFC task, new Monte-Carlo simulations are needed. The results of these simulations for different step sizes and run lengths are shown in Fig. 4.2, and the corrections are applied to the spread estimates of different tests. Table 4.2 gives the results for the tests used with normal-hearing subjects and Table 4.3 gives the corresponding results for the tests from the next chapter with hearing-impaired subjects. From these tables it can be seen that, for all tests, there is a reasonable relation between the step size in the adaptive procedure and the spread of the psychometric function. The spread is smaller for hearing-impaired subjects than for normal-hearing subjects in comparable tests, issuing in a little larger relative step size for the first group. To determine an optimum step size from our data is not possible because only one step size was used.

Test	S_N	step size	N	S	dependent variable
1 Absolute threshold	3.0 dB	2 dB	20	3.8 dB	signal level
2 Bandwidth simult.	2.6 dB	2 dB	20	3.2 dB	masker level
3 Bandwidth nonsimult.	6.3 dB	3 dB	30	7.9 dB	masker level
4 Shallow edge simult.	2.2 dB	2 dB	20	2.6 dB	masker level
5 Steep edge simult.	7.6 %	4 %	20	8.0 %	frequency diff. masker-signal
6 Shallow edge nonsimult.	6.3 dB	3 dB	30	7.9 dB	masker level
7 Steep edge nonsimult.	28 %	10 %	30	35 %	frequency diff. masker-signal
8 Temporal window	2.7 dB	2 dB	20	3.4 dB	masker level
9 Forward masking	27 %	10 %	30	35 %	time between masker and signal
10 Backward masking	41 %	20 %	30	52 %	time between masker and signal
11 Cubic difference tone	4.5 dB	3 dB	30	5.4 dB	signal level
11a Cubic difference tone	7.0 dB	3 dB	30	8.8 dB	masker level
12 Suppression	4.5 dB	3 dB	30	5.4 dB	signal level
12a Suppression	7.9 dB	3 dB	30	10 dB	masker level

Table 4.2. Spread of the psychometric function for various tests with normal-hearing subjects. The second column gives maximum-likelihood estimates, and the fifth column gives estimates corrected for bias. The fourth column shows the run length. The parameter which is varied in order to reach the threshold is listed in the last column.

4.3. MEASUREMENT ERROR

Sources of error variance may be divided in two main classes: variations within tests and variations between tests. The major source of error variance within tests is resulting from the statistical character of the threshold. At each stimulus level there is a certain chance on a correct response. In the long run the proportion of correct responses is expected to be equal to this chance, but the actual fraction will spread. Other sources of error within tests are: guessing for the correct interval in the 2AFC, and fluctuations within the subjects such as lapses and fatigue.

Error variance between tests occurs because two tests never mirror each other perfectly. For instance, a systematic error may occur because the retest is not administered at the same time of the day as the test. Not only tests, but also subjects may change, due to learning effects, changes in motivation, and day-to-day variations in auditory thresholds. Effects of this class of errors were discussed in the previous chapter with the introduction of difference scores. Here we will concentrate on the error variance within tests, and evaluate the development of measurement error during the course of the adaptive threshold-seeking procedure of 2AFC trials.

4.3.1. Estimation of measurement error

Measurement error can be estimated from repeated measurements after all systematic effects, like, for instance, learning effects are isolated. A convenient way to perform this isolation is offered by an analysis of variance (ANOVA). In such an analysis, without replications, or when replications are introduced as a separate factor, the "mean square" of the highest order interaction is an upper limit to the error variance. This mean square is composed of two parts, namely, the true mean square of this interaction and the error variance. When there are a large number of factors in the experimental design, it is likely that there is no true mean square for the highest order interaction and only the error variance remains. Fig. 4.3(a) shows, for the absolute threshold (Experiment 1 of Chapter 2), the distribution of the total variance

Test	S_N	step size	N	S	dependent variable
1 Audiometric loss	2.2 dB	2 dB	20	2.7 dB	signal level
2 Bandwidth simult.	2.7 dB	2 dB	40	2.9 dB	signal level
3 Critical ratio	2.4 dB	2 dB	20	2.9 dB	signal level
4 Bandwidth nonsimult.	3.6 dB	3 dB	60	3.9 dB	masker level
5 PTC low-frequency edge simult.	3.1 dB	2 dB	40	3.3 dB	masker level
6 PTC high-frequency edge simult.	18 Hz	10 Hz	40	20 Hz	masker freq. and level
7 PTC low-frequency edge nonsimult.	4.4 dB	3 dB	60	4.6 dB	masker level
8 PTC high-frequency edge nonsimult.	20 Hz	10 Hz	60	22 Hz	masker freq. and level
9 Temporal window	2.5 dB	2 dB	40	2.7 dB	signal level
10 Forward masking	5.5 dB	3 dB	60	5.9 dB	signal level
11 Backward masking	6.3 dB	3 dB	60	6.9 dB	signal level
12 Click threshold	2.7 dB	3 dB	60	2.8 dB	signal level
13 Click in noise	2.6 dB	3 dB	60	2.7 dB	signal level

Table 4.3. Spread of the psychometric function for 13 tests with hearing-impaired subjects. Various columns as in Table 4.2.

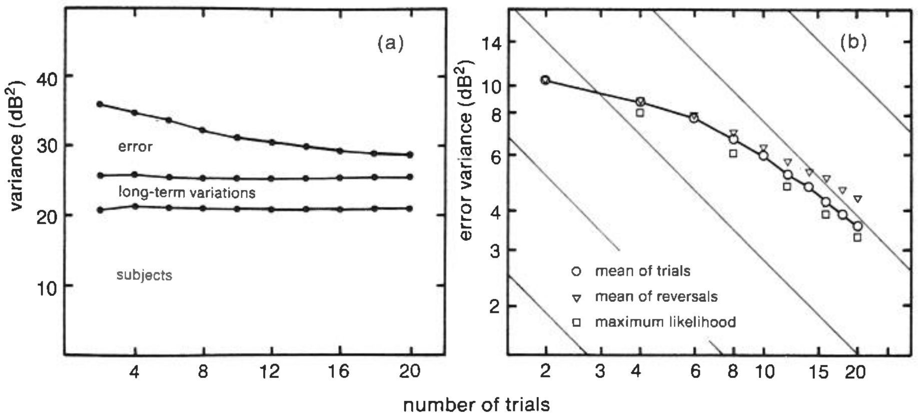


Fig. 4.3. Panel (a): Partition of the total variance in the "Threshold" experiment of Chapter 3 over the sources *subjects, long-term variations,* and *error* as a function of the number of trials in stage 3 that is considered. The data shown are for threshold estimates from the average stimulus level over individual trials; the other threshold-estimating procedures give the same results, except for the error part of the variance.
Panel (b): Error variance from panel (a) on a log-log scale for three threshold-estimating procedures.

over the sources *subjects*, *long-term variations*, and *error* as a function of the number of trials in stage 3 that is considered. For estimating variance components from ANOVA, we used a method described by Vaughan and Corballis (1969). The analysis has been performed for threshold estimates from three procedures: the maximum-likelihood procedure, the average stimulus level of all trials in stage 3, and the mean stimulus level for an even number (N_{rev}) of reversals in stage 3, which degenerates to the mean stimulus level if $N_{rev} < 2$. The variance accounted for by differences among subjects and the day-to-day variations in thresholds are constant as a function of the considered number of trials; only the "error" part of the total variance gradually diminishes. The three threshold-estimating procedures give the same results, except for the error variance which is shown in detail in Fig. 4.3(b) on a log-log scale. The differences are only very small. The maximum-likelihood procedure shows the least error variance, but in this procedure also trials in the stages 1 and 2 of the run are used. For the other two procedures the average stimulus level of individual trials gives the smaller error variance. Also for other tests it appeared that there are only very slight differences in error variance between different threshold estimates. For this reason we

will use only the average stimulus level of all individual trials in the further analysis, and the optimum condition will be compared with the results from the maximum-likelihood procedure.

4.3.2. Measurement error and considered trials

a. The run length

Our main interest now is the decay of measurement error as a function of run length, as shown in detail in Fig. 4.3(b). Purely statistically, the error should decrease by a factor of two for each doubling of the number of trials, as represented by the straight lines. This holds when the individual trials are independent. For the data used, the individual trials are coupled via the adaptive procedure with a limited step size and the error can diminish only slowly. When increasing the number of trials the influence of the coupling between successive trials on the possible test levels becomes less important and in the long run the logarithm of the error variance will diminish proportionally to the logarithm of the number of trials. However, if the run is very long, alertness of the subjects will be reduced and the decrease of error becomes less sharp. For studying this phenomenon, the total error is not very sensitive and a better indication is given by the error of only a few trials as function of their position within the run (see Fig. 4.4 and 4.5). This may be called the *differential error variance* because it monitors the development of error variance.

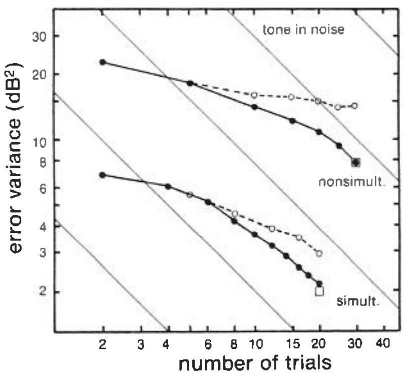


Fig. 4.4. Error variance of threshold estimates in the "Bandwidth" experiments of Chapter 3 as a function of the considered number of trials for normal-hearing subjects. Open circles represent the error variance in estimates based on 5 trials as a function of their position within the run. The result of the maximum-likelihood method is shown by a square.

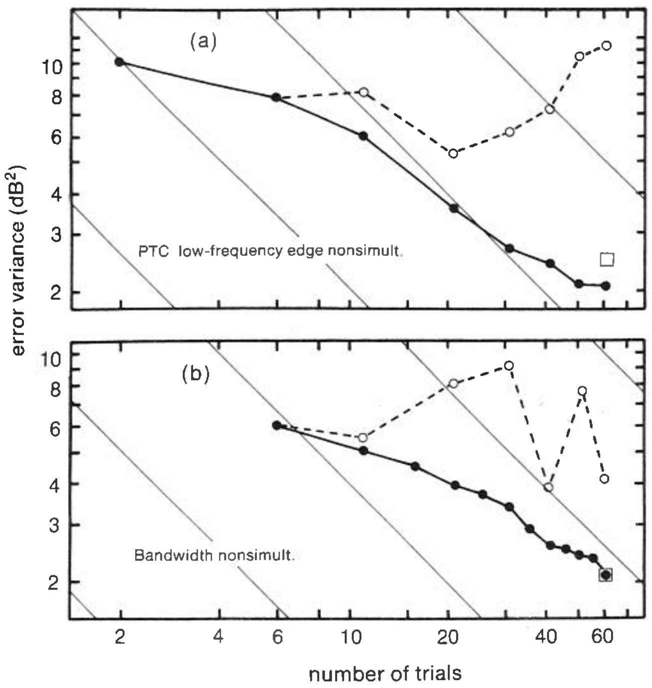


Fig. 4.5. Error variance of threshold estimates as a function of the considered number of trials for two tests (panels (a) and (b)) with hearing-impaired subjects. The open circles represent the differential error variance, as in Fig. 4.4, and the squares give the result of the maximum-likelihood method.

As long as there is only a slight reduction in the efficiency of individual trials, the accuracy can be further increased by extending the run length. If, however, the efficiency drops too much it will be better to start a new run, in spite of the large number of trials (about 20) that is required as a starting procedure in the stages 1 and 2. As can be seen in Fig. 4.4 the point were alertness of subjects starts to reduce, is by no means reached in runs with 20 or 30 trials in stage 3. In fact the differential error variance is still decreasing towards the end of the run. Considerablely longer runs (up to 60 trials in stage 3) were used in the study with hearing-impaired subjects that will be discussed in the next chapter. Fig. 4.5 gives the error variance for two tests from that study. The steepness of these curves is less than would be expected on a statistical basis, for which there may be two reasons. First of all, there is an indication that the differential error variance is

growing towards the end of the run, but this is difficult to see because the curves are strongly fluctuating due to the small number of subjects (22). Secondly, together with longer runs we used less replications, giving a more simple experimental design. As a consequence, the highest order interaction in the ANOVA may contain more than only error variance. Because the non-error part of the variance will not decrease as a function of the number of trials, the total reduction rate may be too shallow.

b. The starting procedure

When using the average of stimulus levels as an estimate of the threshold, it is clear that the starting procedure at the beginning of each run will provide misleading information. The number of trials at the start of a run which has to be omitted in order to arrive at minimum error variance, is the final question we want to deal with in this chapter. For this purpose we have evaluated the error variance for two tests as a function of the number of trials, and for various starting points (see Fig. 4.6). For different curves the aver-

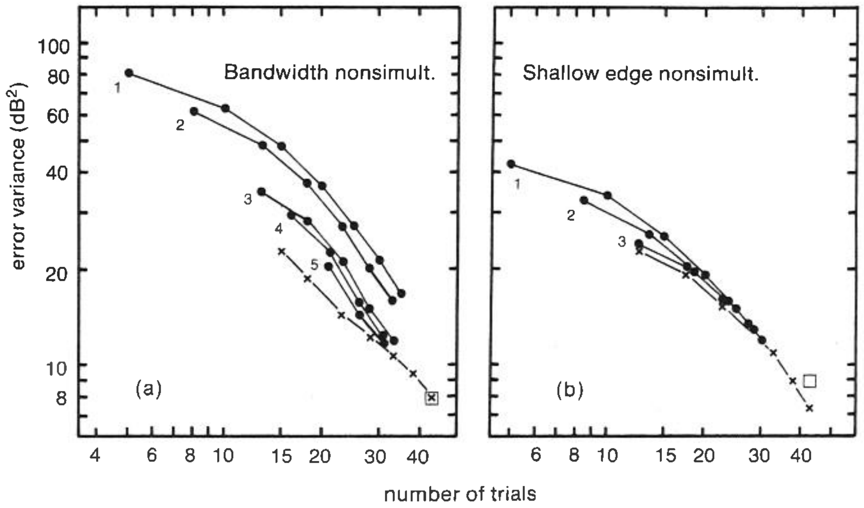


Fig. 4.6. Error variance of threshold estimates as a function of the number of trials spent since the first reversal in the adaptive procedure. Panel (a) for "Bandwidth nonsimult." and panel (b) for "Shallow edge nonsimult.". Parameter of the curves with closed circles is the reversal number in the adaptive procedure that is used as a starting point for averaging. The first point of each curve is based on 5 trials. In panel (b) the curves 4 and 5 are not shown because they fall on top of the displayed curves. The error variance obtained for only stage 3 is indicated with crosses and the maximum-likelihood result is given by a square.

aging starts after a different number of reversals, indicated with the parameter. In contrast with the previous figures, error variance is not given as a function of the number of trials used to calculate the threshold but as a function of the total number of trials spent after the first reversal in the run. The first data point of each curve gives the error variance based on five trials, but the position of these data points along the abscisa is determined by the number of trials needed after the first reversal. For instance, the curve with number 1 starts at N = 5 because 5 trials are needed for averaging after the first reversal, the curve with number 2 (panel a) starts at N = 8 because again 5 trials are needed and on the average it takes 3 trials from the first to the second reversal. The error variance for data in stage 3, as used in all previous figures, is indicated with crosses. In both panels of Fig. 4.6 it can be seen that for a low number of trials the error variance reduces when a larger part of the starting procedure is omitted. However, as the number of trials grows these differences diminish, due to a reduced influence of the starting procedure. In panel (a) this reduction is slow. When incorporating trials preceding the third reversal (curves 1 and 2), the error variance of threshold estimates for N = 30 is still clearly higher than when omitting these trials (curves 3, 4, and 5). For threshold estimates based on trials after the third, fourth, or fifth reversal, the error variance is almost independent of the starting point. In panel (b) the influence of the starting procedure appears to be much smaller, and beyond N = 20 the error variance is independent of the omitted number of trials. The curves for the error variance of estimates from averaging after the fourth or fifth reversal are left out of panel (b) because they fall on top of the curves shown.

4.4. DISCUSSION

Not all questions put forward in the introduction can be answered by using the obtained data. For instance, in the procedure for placing of observations only one rule was used for increasing or lowering the stimulus level, which disables us to evaluate this rule. The same holds for the step size used in this procedure. We can only compare the step size used in various tests with the steepness of the psychometric function, as shown in the Tables 4.2 and 4.3. As can be seen in these tables, there is a reasonable relation between these two.

More may be said about the development of measurement error as a function of run length. It appeared that within runs of up to 30 trials in stage 3, the error variance decreased without any sign of reduction of alertness during the run. In fact, the differential error variance, giving the error for a fixed number of trials as a function of their position within the run, even diminishes towards the end of the run (see Fig. 4.4). On the basis of these findings we decided to use longer runs in the study with hearing-impaired subjects. The decrease of error variance in these longer runs is less sharp. It is, however, not clear whether this is the result of an artifact of the variance-estimation procedure or caused by a reduction of alertness during the run. For the results of error variance as a function of run length an important restriction has to be made. Error variance as a function of run length is calculated by considering only a part of the trials in a run of fixed length. This only gives a perfect prediction for the accuracy of shorter runs if measurement error depends only on the trial number within the run and is independent of the total run length. If, on the contrary, the subjects anticipate the length of the run and adapt their alertness, no effects would be found during the runs if the subjects manage to maintain this level of alertness.

When leaving out leading trials in estimating thresholds from the average stimulus level, there are two opposite effects on the error variance. On the one hand trials with a large contribution to the error are left out and on the other hand the total number of trials determining the threshold decreases. As shown in Fig. 4.6, there is a large number of trials for which these two effects cancel each other. The curves starting at the 3th, 4th, and 5th reversal fall on top of each other, which means that in this region the error variance is independent of the starting point and is only a function of the total number of trials. When leaving out more trials than all until the fifth reversal, variance estimates become unreliable because a large part of the runs is too short. Therefore the corresponding error variance could not be calculated and is not shown in Fig. 4.6, but it is clear that at some point the error will start to increase as a consequence of the reduced number of trials.

Regarding the techniques for the estimation of thresholds from the obtained data, we reached clear results. With the three procedures that were investigated different points of the psychometric function are estimated, but the differences involved are subject independent and can be considered as constants. The error variance accompanying estimates from the three procedures differs very little, and for complete runs the maximum-likelihood method is only slightly better than the average over all trials.

CHAPTER 5

RELATIONS BETWEEN AUDITORY FUNCTIONS IN IMPAIRED HEARING *

Summary

Relations between auditory functions, as expressed by coefficients of correlation, were studied for a group of 22 sensorineurally hearing-impaired subjects with moderate losses. In addition to the audiogram, we measured frequency resolution, temporal resolution, and speech reception in quiet and in noise. Frequency resolution was derived from masking with comb-filtered noise and from the psychophysical tuning curve, for both paradigms in simultaneous and in nonsimultaneous masking. The critical ratio was also determined. Temporal resolution was determined with intensity-modulated noise and from backward and forward masking. All tests were performed at 1000 Hz. Correlations among tests were gathered in a matrix and subjected to a principal-components analysis. It turned out that tests on frequency resolution cluster, and are approximately independent of audiometric loss. Furthermore, hearing loss for speech in noise is closely allied to frequency resolution, whereas hearing loss for speech in quiet is governed by audiometric loss.

5.1. INTRODUCTION

Knowledge about relations among auditory functions is of vital importance to our understanding of the auditory system. Usually, speculations on the relations between specific auditory functions are based upon theoretical considerations and physiological data. From many of these speculations a theoretical framework has been developed, in which the hearing process is described in terms of basic properties like frequency resolution and nonlinearity.

A more direct way of finding relations among auditory functions is by making use of interindividual differences in these functions and correlating the results from a number of tests for a group of subjects. Following this statistical approach, basic properties of the auditory system can be traced by applying a factor analysis to the matrix of correlation coefficients between

* *Paper accepted for publication in: J. Acoust. Soc. Am.*

tests. In such an approach it is first of all essential that the results of various tests should not be confounded with differences among the subjects which are not under investigation, like practice and motivation. A second condition for such an approach is the need for scores to have a high test-retest reliability in order to yield significant correlations. Both these requirements were met in a study on auditory functions in normal hearing by Festen and Plomp (1981); in that study frequency resolution, temporal resolution, and nonlinearity were investigated, but notwithstanding the reliability of the scores, only very few significant correlations were found. The present investigations are an extension of that study to sensorineurally hearing-impaired subjects. In particular for this group the factor-analytic approach is promising since it capitalizes on the heterogeneity of the hearing-impaired population, while heterogeneity presents problems in more traditional studies on hearing impairment. Because the interindividual differences between hearing-impaired subjects are much larger than those between normal-hearing subjects, interfering factors will have less influence on the differences in auditory thresholds, and we expect more clear-cut results in this case.

We determined frequency resolution from the masking produced by comb-filtered noise and from psychophysical tuning curves, both in simultaneous masking and in nonsimultaneous masking. Temporal resolution was measured with intensity-modulated noise, following a procedure which is the time-domain analog of bandwidth determination with comb-filtered noise. In addition, we measured backward and forward masking curves for a burst of noise. All these tests were administered at a single frequency, 1000 Hz. Finally, we measured speech-reception thresholds in quiet and in noise to see how auditory communication of the hearing impaired is related to the auditory functions.

The choice of signal levels constitutes a general problem in measurements on hearing-impaired subjects. On the one hand, the absolute threshold is raised considerably, but, on the other hand, the maximally tolerable signal level does not change or is even lowered. Especially the comparison of test results between subjects with different absolute thresholds is troublesome. For instance, there is ample evidence, both physiological and psychophysical, that frequency resolution becomes poorer with increasing signal level (Evans, 1977; Rhode, 1978; Scharf and Meiselman, 1977). Because hearing-impaired subjects require signals to be presented at high sound-pressure level, the deterioration of frequency resolution found for them (Wightman *et al.*, 1977; Hoekstra and Ritsma, 1977; Zwicker and Schorn, 1978; Florentine *et al.*, 1980; Ritsma *et al.*, 1980) may to some extend be the result of level effects.

Recently, some reports were published revealing level-independent tuning curves measured in forward masking for normal-hearing subjects when presenting the probe tone at a constant sensation level by adding background noise (Nelson, 1980; Green *et al.*, 1981). It has been argued that short probe tones easily give rise to substantial spectral "splatter" which may be detected at high sensation levels. However, it seems unlikely that these effects can account completely for poorer frequency resolution at higher levels, even when measured with short probe signals (Wightman and Raz, 1980). In fact, an appropriate comparison of auditory functions between normal-hearing and hearing-impaired subjects should be made at equal sound-pressure levels if the auditory function under investigation precedes the reduction of sensitivity, and at equal levels above the absolute threshold for the opposite case. As things now stand, however, an intermingling of sensitivity reduction and auditory functions seems to be most probable. On the one hand, this offers the opportunity to find differences in auditory functions among hearing-impaired subjects, on the other hand it hampers comparisons between subjects with different losses. In this study we have tried to perform our measurements in a narrow range of sound-pressure levels of the maskers that were used.

Another factor that may lead to erroneous interpretation of auditory functions in impaired hearing is the influence of absolute or quiet thresholds. Martin and Pickett (1970) showed that for sloping hearing loss the high frequency side of masking patterns may easily reach the absolute threshold and could wrongly be interpreted as an excessive upward spread of masking. With other auditory functions we should also be alert not to measure the absolute threshold when aiming at masked thresholds.

5.2. METHODS

5.2.1. Procedure

We used an adaptive two-alternative forced-choice (2AFC) procedure (Levitt, 1971) with visual feedback. In the various tests, threshold was reached by varying different signal parameters such as probe-signal level, masker level, or masker frequency. Details of the measurement procedure are described in the previous study (Chapter 3) and in Chapter 4; now, however, the ultimate thresholds are averages over 40 trials in simultaneous masking and 60 trials in nonsimultaneous masking.

Bias in the data resulting from differences in training and motivation among the subjects may cause different thresholds for audiologically similar subjects. In most tests in the present study this bias was compensated for by using measures derived from the difference between two threshold values. For instance, the auditory bandwidth was calculated from the threshold difference between peaks and troughs of comb-filtered noise, and the slopes of the psychophysical tuning curve were determined from differences in the thresholds obtained for various masker frequencies. To minimize the influence of training and motivation on the difference score, the two measurements always followed each other immediately.

To eliminate measurement errors caused by sequence effects, the order of tests and of conditions within tests were fixed for all subjects. This may cause biased average results, but it minimizes error variance in the interindividual differences, which are of prime interest in this study. After a training session, testing took place in four morning sessions. Half of the tests were carried out in the first session and the other half in the second session. In the last two sessions all tests were repeated. The proportion of error variance was estimated from the correlation between test and retest. On the average a test block lasted twenty minutes, after which the subject had a break of the same duration. In the breaks for one subject another subject was tested.

5.2.2. Subjects

Twenty-two sensorineurally hearing-impaired subjects were drawn from files of the university clinic; they were tested at their better ear. Subjects had to have a hearing loss of between 30 dB and 60 dB at 1000 Hz without air-bone gap. Audiograms ranged from essentially flat to sloping with increasing losses toward higher frequencies; the maximal slope was 18 dB/oct. Subjects suffering from Ménière's disease or having a tinnitus were excluded, but no further constraints on the etiology were used. Decisive in the selection of subjects was their performance in a pilot session. Subjects had to understand the measurement procedure and be able to produce stable results. Seven males and fifteen females participated in this study; their age ranged from 30 to 72 with an average of 58. They were payed for their cooperation.

5.2.3. Apparatus

A PDP-11/10 computer controlled the generation of signals, their presentation, and the adaptive procedure; and stored the responses. The tonal signals were stored in two revolving memories of 512 time samples with 16 bits of resolution. The noise maskers were produced by a noise generator (Wandel und Goltermann RG-1); time delay, attenuation, and summation necessary for the comb-filtered noise were carried out by the computer. The stimuli were gated with a cosine-squared onset and termination, with rise and fall times of 15 ms, and were presented via an electro-dynamic earphone (Beyer DT 48). Speech-reception thresholds were measured with short sentences recorded on a Revox A 77, together with masking noise. In this test a Scintrex MK-IV earphone was used.

5.3. EXPERIMENTS

In the following sections the individual tests are introduced briefly. Table 5.1 gives a summary of the results together with their accuracy.

5.3.1. Pure-tone audiometry

In the first session measurements started with the determination of a pure-tone audiogram in order to find the better ear and to confirm the sensorineural origin of the hearing loss. Thresholds were determined at octave frequencies from 250 Hz to 4000 Hz for both air conduction and bone conduction. By definition, the ear with the lower bone-conduction threshold averaged over 500, 1000, and 2000 Hz is considered to be the better ear. For the subjects selected, bone-conduction losses in the better ear were essentially equal to air-conduction losses. Bone-conduction thresholds have not been used further and all data presented in this study were obtained from the better ear.

Each session started with a detailed measurement of the auditory threshold in the 1000-Hz region. Thresholds were determined for 200-ms tones of 630, 800, 1000, 1260, and 1590 Hz. Subsequently the threshold for a short 1000-Hz tone pulse having a cosine-squared envelope (no steady state) with a total duration of 30 ms was measured. The latter threshold was used to select probe-tone levels for nonsimultaneous-masking experiments and for calculating temporal integration.

A principal-components analysis on the five thresholds in the 1000-Hz region showed that 94.5 % of the total variance can be accounted for in two dimensions. These dimensions can roughly be characterised as *mean audiometric loss* and *audiometric slope*. For further analysis we used the real mean loss and audiometric slope by which 92.9 % of the variance is covered. The slope of temporal integration at 1000 Hz, calculated from the threshold difference between the 200-ms and the 30-ms tone bursts, amounts to 6.5 dB/log-unit-of-time on the average.

5.3.2. Frequency resolution

From the multitude of paradigms used in studying auditory frequency resolution three were selected: auditory bandwidth derived from the masking effect of comb-filtered noise, psychophysical tuning curve, and critical ratio. As in the previous study on normal-hearing subjects, the first two paradigms were used both in simultaneous masking and in nonsimultaneous masking. A schematic representation of the measurements is given in Fig. 5.1. In the following sections we will give a brief outline of each test.

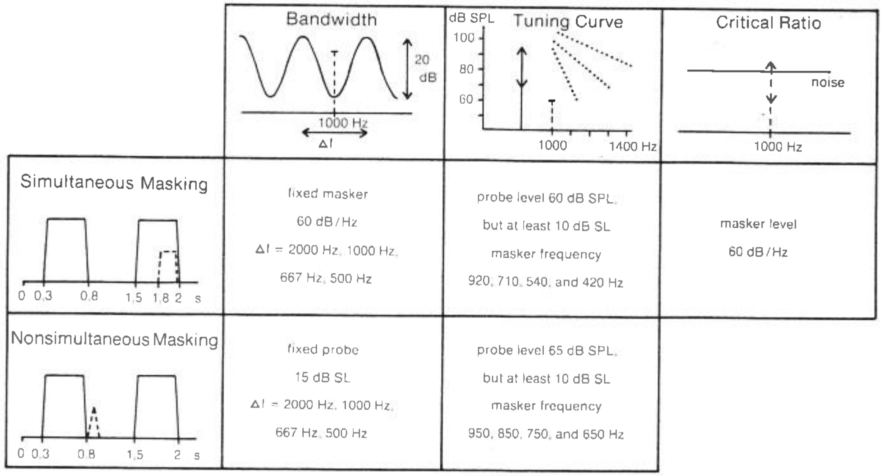


Fig. 5.1. Schematic representation of the experiments on frequency resolution. The columns give the spectra of the signals and the rows give the temporal structure. Probe signals are dashed and maskers are fully drawn. Measurement conditions are listed in the cells. Signals are gated with cosine-squared onset and termination, and rise and fall times of 15 ms. The envelope of the probe tone in nonsimultaneous masking is a Henning function with a total duration of 30 ms.

a. Bandwidth

The difference in masking between peak and trough of comb-filtered noise as a function of peak spacing gives a rather direct estimate of the auditory bandwidth. This method was introduced in psychoacoustics by Houtgast (1974), and was used for hearing-impaired subjects by Pick *et al.* (1977). Peak spacings were 500, 667, 1000, and 2000 Hz, and the peak-to-trough ratio was 20 dB. In simultaneous masking we used a constant masker with an average level of 60 dB per Hz. In order to prevent the results from being contaminated with differences in the decay of masking, a constant probe was used in forward masking. Because a constant sound-pressure level for the probe leads to unacceptably high masker levels for some subjects, we were forced to use a constant sensation level (SL) of about 15 dB in this test.

The differences in threshold-level between peak and trough averaged over subjects are given in Fig. 5.2. In the simultaneous-masking experiment the masked threshold in the trough exceeded the absolute threshold by more than 15 dB for all subjects, so true peak-to-trough differences were measured. The smooth curves show theoretical peak-to-trough differences for a Gaussian-shaped filter, with equivalent rectangular bandwidth as the parameter. For each sub-

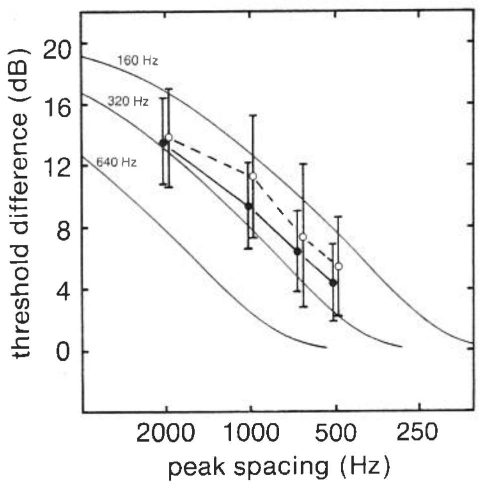


Fig. 5.2. Threshold-level difference for a 1000-Hz probe tone between peak and trough of comb-filtered noise as a function of peak spacing. Closed symbols are for simultaneous masking and open symbols for nonsimultaneous masking. The between-subjects standard deviation is indicated by vertical bars. The smooth curves give calculated threshold differences for a Gaussian-shaped filter.

ject the best-fitting bandwidth was calculated. These values appeared to be not normally distributed, but their logarithms show a symmetric distribution and were used in the further analysis.

Whereas in normal-hearing subjects the bandwidth in nonsimultaneous masking is about half the bandwidth in simultaneous masking, these bandwidths differ only slightly for the hearing-impaired subjects, see Fig. 5.2. The difference is only significant for a subset of the least hearing-impaired subjects. This finding is in harmony with observations by Wightman *et al.* (1977) and by Leshowitz and Lindstrom (1977) indicating that lateral suppression is greatly reduced by sensorineural hearing losses; more details on this subject will be given in section 5.4.1.(b).

The critical ratio, equal to the threshold level of a pure tone relative to the spectral density of a wide-band noise, is usually considered to be closely related to critical bandwidth. In this study the critical ratio was measured for only one condition. Spectral density of the noise masker was fixed at 60 dB per Hz and the level of a 1000-Hz probe tone was varied.

b. Psychophysical tuning curve

Recently, psychophysical tuning curves (PTCs) have been measured with hearing-impaired subjects in a number of studies (e.g. Hoekstra and Ritsma, 1977; Wightman *et al.*, 1977; Zwicker and Schorn, 1978; Tyler *et al.*, 1980). All studies reveal a broader tuning curve for impaired ears than for normal ears. However, because hearing-impaired subjects have to be tested at higher sound-pressure levels, part of the deterioration may simply be a level effect. In order to reduce this confounding, we tried a probe signal with a fixed sound-pressure level. No background noise was added to mask off-frequency probe-signal energy for in the spectrum of a pure-tone signal with a cosine-squared envelope secondary maxima are more than 30 dB below the level at the signal frequency.

In the simultaneous-masking experiment the probe-tone level was, generally, 60 dB SPL; however, for subjects with a high threshold this probe tone was too weak, so for them a level of 10 dB SL was used instead. The low-frequency edge of the tuning curve was determined from the masked thresholds at frequencies of 920, 710, 540, and 420 Hz. The high-frequency edge was measured at three points. In order to have the opportunity to measure very steep skirts and shallow skirts with the same procedure, masker level and frequency were varied together according to three fixed rules as indicated by dotted lines in Fig. 5.1. Step size in the adaptive procedure was 10 Hz and the corresponding num-

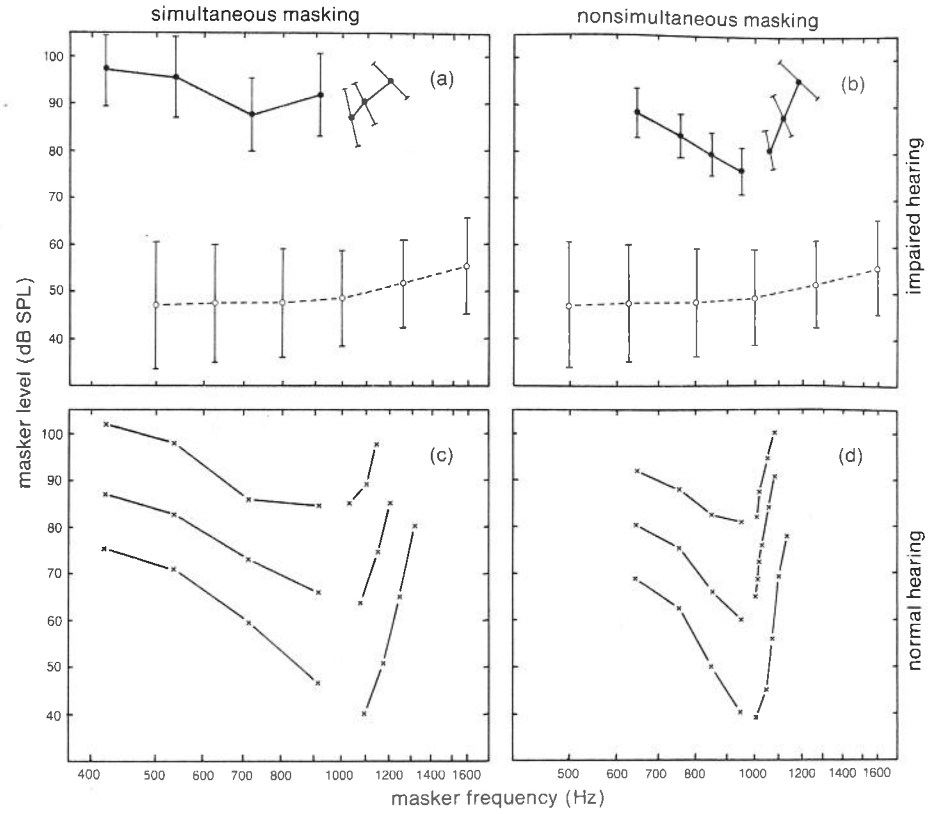


Fig. 5.3. Psychophysical tuning curves in simultaneous masking (panel a) and in nonsimultaneous masking (panel b). Open symbols represent the average pure-tone thresholds. The standard deviation between subjects is given by bars. Panels (c) and (d) give average psychophysical tuning curves for two normal-hearing subjects and for probe-tone levels of 20, 40, and 60 dB SPL, respectively.

ber of decibels. To avoid detection of combination tones, which may occur especially with the least impaired subjects, a continuous low-pass noise was present with a spectral density of 25 dB per Hz and a cut-off frequency of 800 Hz. For the tuning curves in nonsimultaneous masking the technique of forward masking was used (see Fig. 5.1). The short 1000-Hz probe tone was presented at 65 dB SPL or, if necessary, at a higher level corresponding to 10 dB SL. The low-frequency edge was measured at masker frequencies of 950, 850, 750, and 650 Hz. Three points on the high-frequency edge were again determined with the procedure in which masker level and frequency were varied together as for the simultaneous masking conditions (see Fig. 5.1).

The average results and the between-subjects standard deviations are given in Fig. 5.3, panels (a) and (b) for simultaneous and nonsimultaneous masking, respectively. For comparison corresponding tuning curves for normal-hearing subjects are reproduced in the panels (c) and (d). In simultaneous masking we had to use probe-tone levels in excess of 60 dB SPL for 10 subjects; the highest level was 75 dB SPL. Because this range of probe-tone levels is small, we expect no serious level effects among the tuning curves of different subjects. In nonsimultaneous masking we used a probe-tone level greater than 65 dB SPL for 11 subjects; here the highest level was 85 dB SPL. In comparing the results with those from other tests we use the slopes of the tuning curves and their sharpness. Slopes are calculated from a least-squares fit of a straight line. Because of the capricious shape of simultaneous-masking PTCs, it makes no sense to calculate Q_{10dB} here. As an alternative we introduced a tuning metric, namely the difference in level between the lowest threshold in the tuning curve and the average threshold of the two outermost points. For nonsimultaneous masking Q_{10dB} could be calculated from straight line approximations of the two edges.

Psychophysical tuning curves in simultaneous masking show two conspicuous differences between normal and impaired hearing. Firstly, subjects with sensorineural hearing losses have much broader PTCs than normal-hearing subjects. Even when comparing PTCs at corresponding sound-pressure levels, the slopes in normal hearing are about twice as steep. Secondly, for the hearing impaired many curves show a notch near the probe frequency. These W-shaped curves are seen in a number of studies (e.g. Leshowitz and Lindstrom, 1977; Hoekstra and Ritsma, 1977); according to Viemeister (1977) they may be introduced by beating of probe and masker or by a roughness sensation which stretches over a large frequency range due to deteriorated frequency resolution. In testing this hypothesis we accepted as a measure for W-shape the difference in masker level between 920 Hz and 710 Hz. Positive numbers correspond to a W-shape and negative numbers to a more or less normal tuning. The correlations will be discussed in section 5.4.2.

5.3.3. Temporal resolution

The ability of the listeners to resolve auditory events in the time domain was studied in three experiments. Firstly, we measured the width of the tempo-

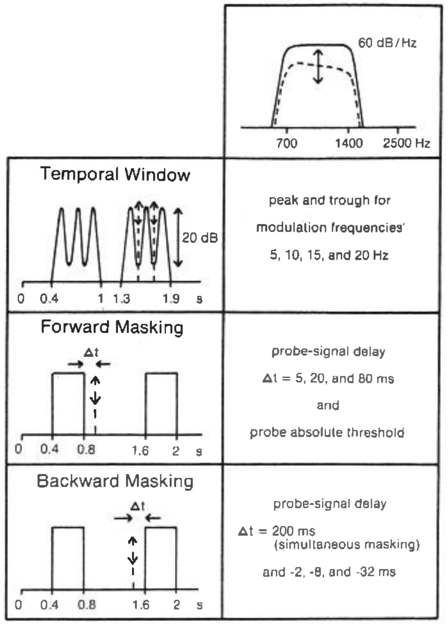


Fig. 5.4. Schematic representation of the experiments on temporal resolution. Probe signals are dashed and maskers are fully drawn. Each row represents a masking paradigm and the spectrum of the signals is given at the top.

ral window in a test representing the time-domain analog of the frequency-resolution measured with comb-filtered noise. Furthermore, the time courses of forward and backward masking were determined. A schematic survey of the experiments is given in Fig. 5.4.

a. Temporal window

The temporal window was determined by measuring the difference in masking between peak and trough of intensity-modulated noise. We used a fixed masker with an average level of 60 dB per Hz and a peak-to-trough ratio of 20 dB. The probe was a 0.4-ms click which was octave filtered, together with the masker, with a central frequency of 1000 Hz. The probe was presented twice within one observation period, with a fixed interval of 200 ms, and in all conditions at the same time relative to the start of the masker. Masked thresholds in peak and trough were determined for modulation frequencies of 5, 10, 15, and 20 Hz.

Threshold differences between peak and trough are given in Fig. 5.5. Thresholds in the trough condition exceeded the click's absolute threshold by at least 15 dB for all subjects, so a true representation of the internal modulation depth of the noise was obtained. Calculation of the width of the temporal window is similar to the calculation of the bandwidth. The smooth curves in Fig. 5.5 give the peak-to-trough differences for a temporal window with a Gaussian shape. Parameter of the curves is the equivalent rectangular width τ. Per subject the width of the window is calculated by a least-squares fit of this function to the threshold differences.

The close resemblance between theoretical peak-to-trough differences as a function of modulation frequency and the actual data suggests that the temporal-window shape is adequately described by the assumed function. However, any asymmetry in the window skirts is not detected with this measurement technique, because it only involves symmetric probe-signal positions. When comparing the results for hearing-impaired subjects with the results for normal-hearing subjects from Festen and Plomp (1981) (Chapter 3), the hearing-impaired subjects perform worse; their average time constant is nearly twice that of normal-hearing subjects. However, the tests are not fully comparable; as opposed to the present measurement, the data for the normal-hearing subjects were gathered using a constant probe and a wide-band masker.

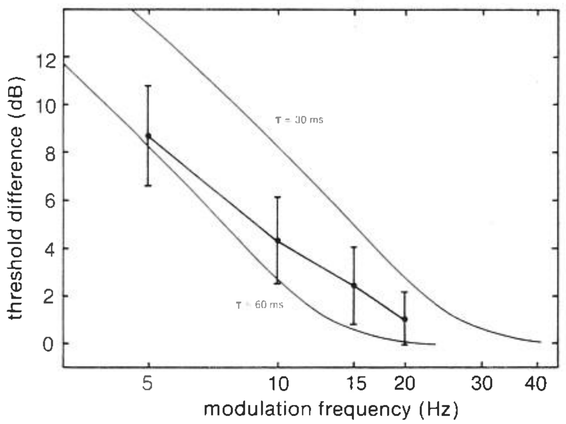


Fig. 5.5. Threshold-level difference, for octave-filtered clicks at 1000 Hz, between peak and trough of intensity-modulated noise as a function of modulation frequency. The between-subjects standard deviation is indicated by bars. The smooth curves give the theoretical threshold difference for a window with a Gaussian shape; the width of the window is the parameter of the curves.

b. Forward and backward masking

Another approach to temporal resolution, and one more frequently applied, is to measure the masking as a function of time for a signal preceding or following a burst of noise. Again a fixed masker of 60 dB per Hz was used. The probe signal was again a 0.4-ms click, and both masker and probe were octave filtered with a central frequency of 1000 Hz. The masker was switched on and off within 1 ms. In forward masking the threshold level for the probe was determined at 5, 20, and 80 ms after termination of the masker. In the same test block the absolute threshold for the short probe signal was measured. Backward masking was measured at intervals between probe and masker of 2, 8, and 32 ms. In this test block, additionally, the probe threshold in simultaneous masking was determined.

Whereas for normal-hearing subjects a straight line on a log-time scale gives a good description of forward-masking data, in this experiment a linear time scale seems more appropriate; Fig. 5.6 shows the results. Because of the strong effect of absolute threshold, the subjects have been split into four subgroups on the basis of their mean audiometric loss. The two middle groups contain 6 subjects each and the other groups 5 subjects. Panel (a) gives the results for backward masking and panel (b) for forward masking. For comparison, the average results of two normal-hearing subjects are given by means of dashed lines for masker levels of 20, 40, and 60 dB per Hz. Panels (c) and (d) give the same data as (a) and (b), but relative to the absolute threshold of the probe signal. Contrary to the results in terms of sound-pressure levels, in these "sensation-level" diagrams the uppermost curves represent the least impaired subjects. In the further analysis we used the slopes of forward and backward masking, calculated by means of a least-squares fit of a straight line. In addition, relations of the masked and quiet threshold of the click with other tests will be investigated.

Clear effects of sensation level are manifest in the results of forward and backward masking. For normal-hearing subjects the extent of masking effects before and after presentation of the masker is nearly independent of masker level: 160-200 ms in forward masking (cf. Plomp, 1964; Wilson and Carhart, 1971), and about 50 ms in backward masking (Elliott, 1962; Pickett, 1959). As a consequence, the slopes of backward and forward masking become steeper for higher masker levels, as seen in Fig. 5.6. Very shallow masking slopes are found for hearing-impaired subjects; as can be seen in the panels (a) and (b), the most impaired subjects show the shallowest slopes. To some extent this effect can be accounted for by the lower sensation level at which

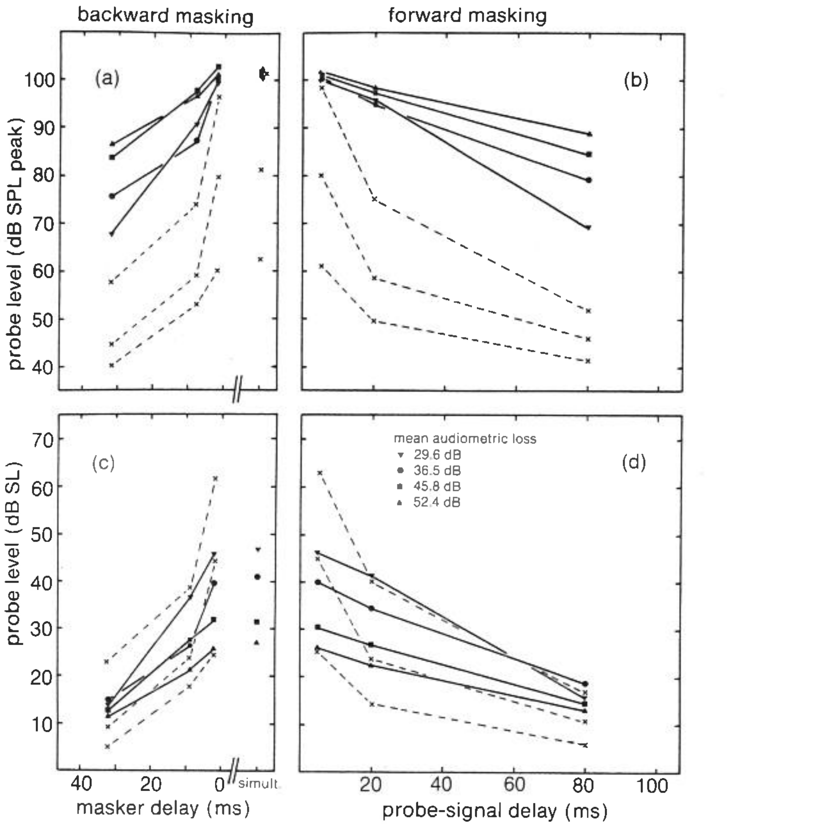


Fig. 5.6. Threshold for an octave-filtered click probe (1000 Hz) as a function of the time between masker and probe; panel (a) for backward masking and panel (b) for forward masking. The masker is an octave band of noise with a central frequency of 1000 Hz and a spectral density of 60 dB per Hz. The various symbols indicate subgroups with different mean losses. The dashed lines represent the average results for two normal-hearing subjects and for masker levels of 20, 40, and 60 dB per Hz, respectively. Panels (c) and (d) show the same results, but expressed in sensation level.

the impaired subjects receive the masker. In terms of sensation level (panels (c) and (d)) the effects of hearing loss are less dramatic. For normal-hearing subjects masking drops sharply immediately before and after the masker and more gradually at greater delays. According to Duifhuis (1973) forward masking is described by two time constants; the steep part of the curve is a consequence of peripheral frequency resolution and the shallow part is considered to be related to neural adaptation. For hearing-impaired subjects, however, a

gradual decay is found over the whole range. In terms of the model with two time constants, this means that the steep "frequency-resolution" part is swamped by the much stronger effects of adaptation. If an extrapolation to greater delay values is permissible, the masking curves both in forward and in backward masking reach the absolute threshold at about the same delay as found in normal hearing. These results are in agreement with those by Nelson and Turner (1980) who studied forward masking with a fixed-level 1000-Hz probe. They also found shallower masking curves for the hearing impaired, but regardless of hearing loss, no masking was found beyond 160 ms.

5.3.4. Speech reception

In the literature on sensorineural hearing loss there is overwhelming evidence for a reduced ability to discriminate speech when it is presented against a background of noise. In this study we incorporated an experiment on the speech-reception threshold as a function of noise level to see how speech reception is affected by basic parameters of the auditory system. For a more detailed study of possible intermediate stages between the perception of pure tones and the perception of speech, the reader is referred to Dreschler and Plomp (1983). Intelligibility of sentences was investigated by means of a test, developed by Plomp and Mimpen (1979), in which each threshold measurement is performed with a list of 13 short sentences. Speech-reception thresholds were measured for 70, 55, 40, and 25 dBA of interfering noise and in quiet. The noise had a spectrum equivalent to the long-term average spectrum of the sentences. Following the model by Plomp (1978), the results for hearing-impaired subjects can be described using two parameters: (1) the D parameter representing hearing loss for speech in noise and interpreted as a distortion term; (2) the (A+D) parameter representing hearing loss for speech in quiet and interpreted as resulting from attenuation (A) and distortion (D) together. In the further analysis these two parameters are used.

5.4. RELATIONS AMONG THE TESTS

5.4.1. Scores from independent measurements

a. Correlations

All tests described in the previous section were administered in test and retest, making it possible to calculate their reliability (see Table 5.1). The

Test	average	σ between subjects	standard error	dimension	r_{tt}
1 Mean audiometric loss	41.3	9.0	0.8	dB	0.99
2 Log bandwidth simult.	2.41	0.15	0.04	log Hz	0.95
3 Critical ratio	22.5	1.9	0.6	dB	0.91
4 Log bandwidth nonsimult.	2.35	0.23	0.05	log Hz	0.96
5 PTC low-frequency edge simult.	6.8	7.5	1.1	dB/oct	0.98
6 PTC high-frequency edge simult.	47.5	50.8	10.8	dB/oct	0.97
7 PTC low-frequency edge nonsimult.	23.0	10.6	3.1	dB/oct	0.92
8 PTC high-frequency edge nonsimult.	35.3	82.6	31.5	dB/oct	0.90
9 Temporal window	52.9	12.6	3.4	ms	0.93
10 Forward masking	0.26	0.12	0.02	dB/ms	0.96
11 Backward masking	0.71	0.25	0.07	dB/ms	0.93
12 Click threshold	65.3	8.8	1.2	dB	0.98
13 Click in noise	101.9	1.9	0.9	dB	0.84
14 Speech hearing loss in quiet	34.9	10.6	1.0	dB	0.99
15 Speech hearing loss in noise	3.8	3.2	0.5	dB	0.97
16 Mean audiometric slope	9.6	11.0	1.1	dB/oct	0.99
17 Temporal integration	6.5	2.1	1.5	dB/log t	0.67
18 Tuning metric simult.	13.2	4.5	1.5	dB	0.90
19 Q_{10dB} nonsimult.	3.0	1.3	0.3	-	0.94
20 W-shape	4.3	5.9	1.2	dB	0.96

Table 5.1. Results of the battery of tests. The second column gives the average over 22 subjects and the third and fourth columns give the between-subjects standard deviation and the standard error, respectively. The reliability of the scores is given in the last column.

coefficient of reliability (r_{tt}) is defined as unity minus the proportion of error variance or, alternatively, as the proportion of "true" variance in a test. It can be estimated from the correlation coefficient between test and retest (r_{tr}), by applying the formula of Spearman and Brown:

$$r_{tt} = 2r_{tr} / (1+r_{tr}) \qquad (5.1)$$

(cf. Nunnally, 1967). A high accuracy was reached; coefficients of reliability for nearly all scores are greater than 0.90. As a consequence correlations

among tests are virtually not affected by measurement error, and the correlations obtained may be regarded as "true" correlations.

The first 15 scores in Table 5.1 originate from mutually independent data, and the correlations among those scores are given in Table 5.2. Correlation coefficients greater than 0.54 or less than -0.54 are significant at a level of 1% and correlations exceeding ±0.42 are significant at a level of 5%. Several tests appear to be correlated with average hearing loss, all pointing to a less selective ear for a greater loss. We have discussed some of these relations with the results of the individual tests. Furthermore, a number of correlations among the scores refer to frequency resolution. For instance, in simultaneous masking the width of the auditory filter (score 2) is inversely related to the steepness of the low-frequency edge of the psychophysical tuning curve (score 5) (r = -0.64), as was also found for normal-hearing subjects (Festen and Plomp, 1981). In nonsimultaneous masking there is a similar relationship (r = -0.72). Both measures for the auditory bandwidth (scores 2 and 4) are positively correlated with critical ratio (3) (r = 0.64 and r = 0.68 for simultaneous and nonsimultaneous masking, respectively). However, the

Test	Kind of score	2 log B_s	3 Cr	4 log B_n	5 Lf_s	6 Hf_s	7 Lf_n	8 Hf_n	9 τ	10 Forw	11 Backw	12 CLth	13 CLin	14 A+D	15 D
1 Mean audiometric loss	loss	0.44	0.59	0.71	-0.32	-0.34	-0.37	-0.49	-0.18	-0.78	-0.80	0.95	0.34	0.83	0.38
2 Log bandwidth simult.	width		0.64	0.72	-0.64	-0.32	-0.63	-0.48	0.40	-0.24	-0.27	0.41	0.05	0.41	0.56
3 Critical ratio	level			0.68	-0.42	-0.16	-0.42	-0.46	0.22	-0.49	-0.49	0.52	0.29	0.53	0.63
4 Log bandwidth nonsimult.	width				-0.61	-0.40	-0.72	-0.66	0.26	-0.63	-0.59	0.69	-0.02	0.69	0.51
5 PTC low-freq. edge simult.	slope					0.24	0.66	0.46	-0.55	0.28	0.17	-0.22	0.08	-0.39	-0.60
6 PTC high-freq. edge simult.	slope						0.45	0.65	0.00	0.36	0.34	-0.33	-0.19	-0.30	-0.08
7 PTC low-freq. edge nonsimult.	slope							0.62	-0.42	0.35	0.23	-0.30	0.14	-0.42	-0.49
8 PTC high-freq. edge nonsimult.	slope								-0.19	0.53	0.44	-0.41	0.07	-0.46	-0.56
9 Temporal window	width									-0.02	-0.04	-0.27	-0.11	0.17	0.34
10 Forward masking	slope										0.91	-0.75	-0.16	-0.85	-0.32
11 Backward masking	slope											-0.79	-0.23	-0.81	-0.24
12 Click threshold	level												0.23	0.80	0.22
13 Click in noise	level													0.26	-0.14
14 Speech hearing loss in quiet	loss														0.31
15 Speech hearing loss in noise	loss														

Table 5.2. Matrix of correlation coefficients among 15 independently determined scores. With each test the kind of score is indicated for a correct interpretation of the sign of the correlations. Underlined values are significant at 1% level; dashed lines indicate the 5% level.

bandwidth measured in nonsimultaneous masking (4) is correlated with scores of the temporal resolution, whereas the bandwidth in simultaneous masking (2) is not. The width of the temporal window (9) and the threshold of a click in noise (13) show very few significant correlation coefficients with the other tests, indicating that they are largely independent.

b. Principal components

A description of the data with a minimum number of variables is achieved in a principal-components analysis (Harman, 1970). This analysis also offers a convenient survey of relations among tests summarised in a matrix of correlation coefficients. Normalising the results of the 15 tests given in Table 5.2, we can represent the data per subject as a point in a 15-dimensional Euclidian space. Each of the original dimensions contributes 6.7% to the total variance. Because a number of tests are correlated, it must be possible to find new dimensions explaining much more than 6.7% of the total variance. A principal-components analysis computes dimensions in the multi-dimensional cluster that explain as much as possible of the total variance. Or, to put it in another way, the analysis derives composites of the original tests in a way that maximizes the amount of variance explained. The dimension accounting for most of the variance is called *factor* 1, that with the second highest variance is called *factor* 2, etc. Mathematically the factors can be determined by computing the eigenvectors of the correlation matrix. The eigenvalue, being the length of the eigenvector squared, gives the variance in the corresponding dimension. The elements of the eigenvector are called *factor loadings* and represent the contributions of the individual tests to the factor. There are various criteria for the number of significant factors that can be determined. For instance, the percentage of variance accounted for is plotted as a function of the factor number, as in Fig. 5.7 (b), and the breaking point in this curve is interpreted as separation of significant and nonsignificant factors, or only factors with an eigenvalue greater than unity are considered to be significant.

Prior to applying a principal-components analysis the matrix of correlations was subjected to three tests, recommended by Dziuban and Shirkey (1974), and showed to be suitable for this analysis. In the analysis all slope scores have been inverted, so that lower numbers represent better hearing for all tests. Fig. 5.7 panel (a) shows the factor loadings for the first two dimensions. The squared distance of the tests from the origin represents the fraction of the test variance that is accounted for. Panel (b) gives the propor-

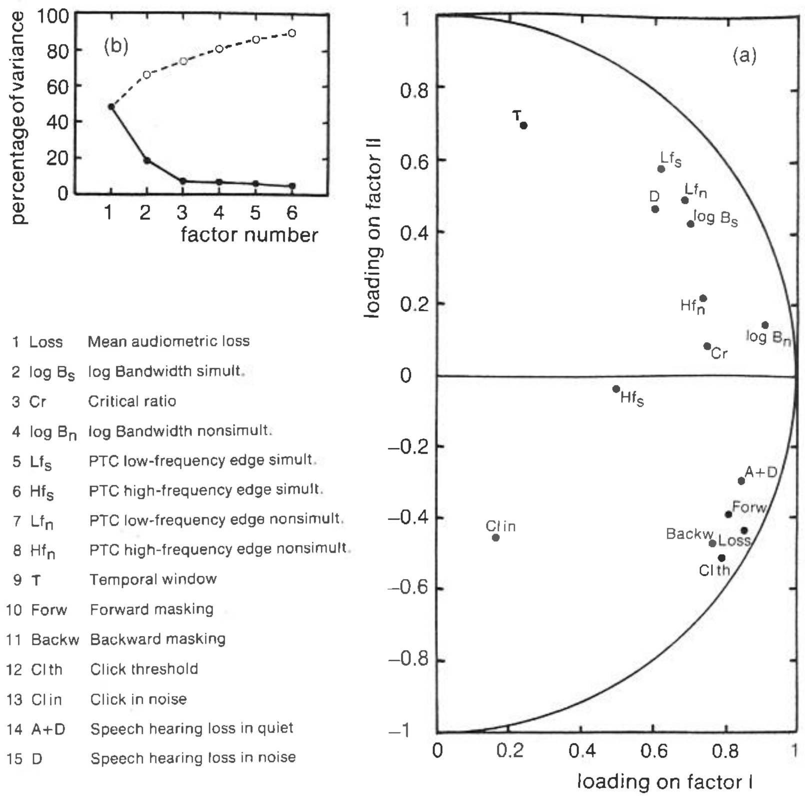


Fig. 5.7. Panel (a): Factor loadings on the first two dimensions from a principal-components analysis on 15 tests. The squared distance of the tests from the origin represents the fraction of the test variance that is accounted for. Panel (b): Proportion of the total variance accounted for by the individual factors and the cumulative proportion of variance (dashed line).

tion of the total variance accounted for by successive factors (fully drawn line) and the cumulative proportion of variance (dashed line). The first two factors explain 48.1 and 17.6% of the variance, respectively. For further extracted factors the corresponding variance decreases very gradually (8.0, 6.8, 5.7, and 2.8% for the factors 3 to 7, respectively), indicating that these factors are based largely on uncorrelated variance. Indeed the eigenvalues of the third and the fourth factor are greater than unity but this is only fractional. Because of the low subjects-to-variable ratio in this analysis, the stability of the derived principal components was investigated with separate analyses on the results of the test and the retest. These two analyses showed almost the same principal components as for the average results and thus showed configurations of tests which are highly comparable to the one in Fig. 5.7.

In the factor loadings of panel (a) two distinct clusters can be seen. The upper cluster contains hearing loss for speech in noise and frequency-selectivity scores, like the low-frequency edge of the PTC both in simultaneous and in nonsimultaneous masking and the bandwidth in simultaneous masking. The lower cluster contains scores related to the absolute threshold, like mean audiometric loss and hearing loss for speech in quiet, but also the slopes of forward and backward masking. The close correspondence between these slopes and absolute threshold is definitely caused by the sensation level effect in forward and backward masking as discussed in the section on temporal resolution. Seen from the origin the two clusters are not in perpendicular directions, which means that there is at least some relationship. In other words, there may be a general tendency of deteriorating frequency resolution with increasing hearing loss, but there is by no means a one-to-one relation.

Critical ratio and both bandwidth and high-frequency edge of the PTC in nonsimultaneous masking take up positions between the two clusters, indicating that they are related to both effects. According to Patterson (1976), the

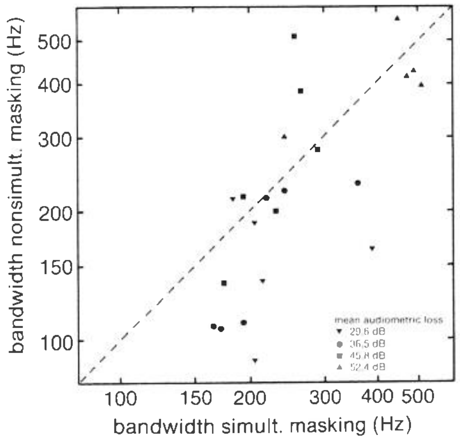


Fig. 5.8. Scatter diagram of the nonsimultaneous-masking versus the simultaneous-masking bandwidth. Individual subjects are indicated with symbols corresponding to categories of hearing loss as introduced in section 5.3.3. The dashed line represents equal bandwidths in simultaneous and in nonsimultaneous masking.

threshold of a tone in noise is determined by two subject-dependent quantities: the width of the auditory filter and a constant representing the efficiency of the system. It may be that this efficiency, not measured separately, also is a function of hearing loss. Nevertheless, there is a strong relation between critical ratio and auditory bandwidth. On the average critical ratio increases by 3.7 dB per octave of increase in bandwidth, which is close to the theoretical 3 dB/oct for an energy-detector model. In section 5.3.2 we saw that the average bandwidths in simultaneous and in nonsimultaneous masking are about the same, and we concluded that lateral suppression may be very vulnerable to hearing loss. An illustration can be found in Fig. 5.8, showing a scatter diagram of the bandwidths in simultaneous and nonsimultaneous masking. Individual subjects are indicated by means of symbols corresponding to the different categories of hearing loss as introduced in section 5.3.3. Although there is much dispersion, systematic differences between the two bandwidths seem to occur for the narrowest filters which are found in the least hearing-impaired subjects. Thus the bandwidth in nonsimultaneous masking is determined by frequency resolution, but also by lateral suppression as an additional loss-dependent factor and, as a consequence, the bandwidth in nonsimultaneous masking has a position between the two clusters in Fig. 5.7. A similar reasoning applies to the high-frequency edge of the PTC in nonsimultaneous masking.

The high-frequency edge of the PTC in simultaneous masking and the threshold of a click in noise are situated relatively close to the origin, implying that these scores are rather poorly represented in the two-dimensional subspace of the first two factors, and have only a weak relation with the other scores. The width of the temporal window is also a rather solitary score, only weakly related to frequency resolution and independent of hearing loss.

5.4.2. Related scores

The last 5 scores in Table 5.1 were not incorporated in the principal-components analysis, because these scores are based on the same data as scores already included. For instance, for the psychophysical tuning curve in nonsimultaneous masking we calculated both the slopes and Q_{10dB}. The correlation between these scores will be biased because error parts of the scores coincide. The correlations introduced by these additional scores are given in Table 5.3. Possibly biased correlations are given with italics.

The mean audiometric slope (score 16) and the temporal-integration slope (score 17) seem almost independent from the other tests. Although the correla-

Test	Kind of score	16	17	18	19	20
1 Mean audiometric loss	loss	*-0.86*	*-0.05*	-0.33	-0.51	0.30
2 Log bandwidth simult.	width	0.02	-0.19	-0.57	-0.66	0.54
3 Critical ratio	level	-0.42	-0.07	-0.32	-0.54	0.23
4 Log bandwidth nonsimult.	width	-0.02	-0.21	-0.66	-0.79	0.47
5 PTC low-freq. edge simult.	slope	-0.37	0.06	*0.59*	0.67	*-0.58*
6 PTC high-freq. edge simult.	slope	0.18	0.47	*0.74*	0.49	-0.19
7 PTC low-freq. edge nonsimult.	slope	-0.24	0.15	0.68	*0.95*	-0.63
8 PTC high-freq. edge nonsimult.	slope	0.17	0.59	0.68	*0.82*	-0.42
9 Temporal window	width	0.19	-0.04	-0.25	-0.41	0.49
10 Forward masking	slope	0.43	0.27	0.31	0.49	-0.16
11 Backward masking	slope	0.40	0.18	0.24	0.38	-0.15
12 Click threshold	level	-0.22	-0.09	-0.34	-0.42	0.19
13 Click in noise	level	-0.43	0.23	0.03	0.12	-0.25
14 Speech hearing loss in quiet	loss	-0.24	-0.08	-0.37	-0.54	0.40
15 Speech hearing loss in noise	loss	-0.01	-0.24	-0.25	-0.63	0.63
16 Mean audiometric slope	slope		*0.10*	-0.12	-0.07	0.43
17 Temporal integration	slope			0.32	0.30	-0.01
18 Tuning metric simult.	sharpness				0.68	*-0.47*
19 Q_{10dB} nonsimult.	sharpness					-0.66
20 W-shape	deformation					

Table 5.3. Matrix of correlation coefficients introduced by the related scores. Correlations between scores obtained from the same data are given in italics. Underlined values are significant at 1% level; dashed lines indicate the 5% level.

tions of both scores with mean audiometric loss are biased, this effect will be very small because of the high reliability of the mean audiometric loss. The absence of a significant correlation between slope (16) and loss (1) in the audiograms is most probably due to the selection of subjects in a limited range of losses. The absence of correlation between mean audiometric loss (1) and temporal integration (17) is surprising, because, firstly, temporal-integration curves for the hearing impaired are much flatter than the average 10 dB/log-unit-of-time for normal-hearing subjects, as has also been found in other studies (cf. Elliott, 1975) and, secondly, such a correlation was found

for noise-induced hearing loss by Tyler *et al.* (1980). On the other hand, according to Green (1973) temporal integration is the result of a time constant in a central process. If this is correct, we should not expect strong relations with scores of peripheral origin.

The tuning metric in simultaneous masking (18) and Q_{10dB} in nonsimultaneous masking (19) are closely related to bandwidth, both in simultaneous and in nonsimultaneous masking (2 and 4), and to other scores of frequency resolution. The W-shape metric (20) is also tightly coupled to frequency resolution and hearing loss for speech in noise (15). Besides, there is a positive relation between the W-shape (20) and the width of the temporal window (9) (r = 0.49), suggesting an explanation of this deformation of the psychophysical tuning curve in terms of temporal aspects like roughness perception.

As discussed before, the positions in Fig. 5.7 of the scores for forward and backward masking are brought about by their strong dependence on signal sensation level. In a further analysis of this confounding the influence of mean audiometric loss in these scores was eliminated by means of a linear regression. The rest scores were correlated with the scores from all other tests, but no statistically significant correlations were found. In particular the correlation coefficients with the width of the temporal window were only 0.32 and 0.37 for forward and backward masking, respectively.

5.5. DISCUSSION

For a good evaluation of the results of this study it may be worthwhile to dwell upon the aim of the approach applied in these and the previous experiments on normal-hearing subjects (Festen and Plomp, 1981). Hearing may be considered to consist of a number of processes localized along the auditory pathway from the inner ear up to the cortex. Each of these processes is determined by one or more parameters, which may be more or less relevant in hearing. The underlying hypothesis of our previous experiments was that, even in normal-hearing subjects, the parameters, or characteristics, of the individual processes will vary from subject to subject. The finding that the test reliability of the auditory functions measured was sufficiently high to reveal correlations among the scores confirmed that hypothesis. The fact that, nevertheless, only weak correlations were obtained, may indicate that in the auditory functions measured many parameters are involved, all varying among subjects.

In using hearing-impaired rather than normal-hearing subjects we have introduced much greater interindividual differences in auditory functions.

Consequently, the test reliability was much greater than in the previous study. However, this does not guarantee that correlations between auditory functions will be found. If in hearing impairment, just as in normal hearing, many auditory processes are involved with each their own parameters, we still have to expect weak correlations between the scores. The situation will be different if hearing impairment is primarily localized in specific hearing processes determined by a few parameters. In that case the interindividual differences in the auditory functions measured can be explained by a few parameters, which will show up in the data. Fig. 5.7 demonstrates that these interindividual differences can be described by two dimensions, revealing that the hearing impairments of the subjects used are largely determined by two parameters, one related to absolute threshold and the other to frequency resolution. The fact that only 65% of the total variance can be explained by these two factors indicates that several other parameters also play a part.

Of the relations found in the study on normal-hearing subjects, the correspondence between frequency resolution as measured with comb-filtered noise and the sharpness of the psychophysical tuning curve are confirmed in the present study. However, the trade-off between frequency resolution and temporal resolution was not found here. Perhaps this effect is overruled by the general deterioration of auditory functions with increasing hearing loss.

5.6. CONCLUSIONS

The main conclusions from this study are:

(1) A principal-components analysis shows two distinct clusters of tests, one for frequency resolution and one for scores related to audiometric loss. Scores within each cluster are mutually highly correlated but between the clusters only weak relations are found;

(2) For hearing-impaired subjects, the average bandwidth measured in nonsimultaneous masking is equal to the bandwidth measured in simultaneous masking, suggesting a deterioration of the suppression mechanism;

(3) Frequency resolution deduced from the psychophysical tuning curve is related to the auditory bandwidth as measured with comb-filtered noise;

(4) Hearing loss for speech in quiet is determined by mean audiometric loss, while hearing loss for speech in noise is related to frequency resolution;

(5) For hearing-impaired subjects forward and backward masking reach the absolute threshold within about the same time as for normal-hearing subjects.

CHAPTER 6

FINAL DISCUSSION

In the series of experiments presented here we have tried to find relations among various aspects of the hearing mechanism. In the theory on hearing numerous auditory functions are explained in terms of a rather limited set of auditory properties. The aim of these studies was to give experimental evidence for this theoretical framework by applying data reduction techniques to related test results and describing these results with a minimum number of parameters. In order to find relations, we used the differences in auditory functions as present among subjects. Due to these differences related tests will show correlations. Now, it is appropriate to look back in order to see what we have gained with our approach. In three experimental chapters we presented three studies applying a battery of tests on a group of subjects and each time we came up with a matrix of correlation coefficients.

The study in Chapter 2 was a pilot study carried out with a small group of 10 normal-hearing subjects and was intended to demonstrate the possibility of measuring interindividual differences with an accuracy sufficient to find correlations. Acceptable correlations between test and retest were found (Table 2.1), in particular for those tests in which a 2AFC procedure was used. This shows that interindividual differences are not swamped in measurement error. However, accuracy was not excellent and the utmost should be done to keep measurement error as low as possible, the more so as it was decided to use difference scores to reduce the influence of training and motivation. Because sequence effects constitute an extra source of variance, adding to the error variance when a randomisation of tests is applied, we chose for a fixed order of tests in the second and the third study. Admittedly, this introduces systematic sequence effects, but as far as these are constants they do not influence the correlations. Notwithstanding the high values of the correlation coefficients, only very few significant correlations were found in this pilot study.

In Chapter 3 we hypothesized three basic characteristics of peripheral hearing: frequency resolution, temporal resolution, and nonlinearity. A few tests on each of these characteristics were collected in a battery which was applied to 50 normal-hearing subjects. The matrix of correlations (Table 3.2) contained only a very few significant correlations and it was not possible to apply a principal-components analysis. A critical evaluation of potential reasons for these low correlations leads to the following conclusions. The reliability of the obtained scores was sufficient to reveal correlations, as is shown in Fig. 3.7. The applied test-retest reliability covers a range of attributes of the scores. A high reliability assures a favorable relation between error variance and variability among subjects, stability of scores over measurement days, and uniformity of training effects among subjects. In addition, the development of measurement error within measurement runs, as investigated in Chapter 4, shows no sign of a reduction of alertness towards the end of the runs. It is not to be expected that a larger number of subjects would have revealed better correlations. The number of subjects only determines the significance of correlation coefficients but not their value; for 50 subjects even correlations as low as 0.36 are significant at 1% level. Given all precautions that were taken in this experiment, we are left with only a few possible reasons for the discovered lack of correlations. In the first place, the simplification of auditory functions made in this study may be too crude. When describing a function with only a few points, incidental high or low values, which reproduce but are only of local significance in the investigated auditory function, may lead to stable but at the same time misleading scores. Secondly, the auditory system may be much more complicated than admitted in simple models used in the description of isolated auditory functions.

For a comparison of results from the different studies a number of correlation coefficients is collected in Table 6.1. From each of the three original matrices four pairs of tests were chosen showing high correlations and for which a corresponding pair of tests was found in at least one of the other matrices. These four correlations from each study are compared with the results from the other studies. The comparison between the results of Chapter 3 and the pilot study (Chapter 2) does not show much resemblance, for which there are a number of reasons. In the first place, not all tests from the pilot study were adopted in Chapter 3 and for comparable tests there are still

Study	Correlated tests		Correlation coefficients Ch. 2 (10 normal)	Ch. 3 (50 normal)	Ch. 5 (22 impaired)
Ch. 2	Bandwidth nonsimult.	LF slope nonsimult.	-0.83	-0.07	-0.72
	LF slope nonsimult.	LF slope simult.	0.77	0.09	0.66
	Pulse in noise thr.	LF slope nonsimult.	-0.70		0.14
	Bandwidth nonsimult.	LF slope simult.	-0.66	-0.05	-0.61
Ch. 3	Temporal window	Q_{10dB} simult.*		0.39	-0.25
	Temporal window	Shallow edge simult.	-0.16	0.39	-0.55
	Shallow edge nonsimult.	Absolute threshold	-0.07	-0.36	-0.37
	Temporal window	Q_{10dB} nonsimult.		-0.36	-0.41
Ch. 5	Bandwidth nonsimult.	Q_{10dB} nonsimult.		-0.12	-0.79
	Bandwidth nonsimult.	LF slope nonsimult.	-0.83	-0.07	-0.72
	Bandwidth nonsimult.	Bandwidth simult.	0.63	0.17	0.72
	Bandwidth nonsimult.	Absolute threshold	-0.23	0.17	0.71

* For Ch. 5 the Tuning metric was taken

Table 6.1. Four high correlations from each of the three studies (see Tables 2.2; 3.2 and 3.3; 5.2 and 5.3) compared with corresponding correlations from the other two studies. Pairs of tests showing a high correlation were only adopted if a corresponding pair of tests could be found in at least one of the other tables. Underlined values are significant at 1% level; dashed lines indicate a significance at 5% level.

considerably differences in the details as signal level, masker frequencies, and the choice which variable was the dependent and which the independent one. Secondly, there are indeed a few significant correlations in both matrices, but also correlations that are significantly different from zero have their confidence limits. Especially for correlations that just reach significance these confidence limits are quite remote from the observed correlations. Therefore, a significant correlation can indeed be used as strong indication for a relationship, and thus a true correlation different from zero, but we should be careful with conclusions about the tightness of these relationships.

In Chapter 5 a battery of tests was applied to hearing-impaired subjects, offering the advantage of large interindividual differences caused by the heterogeneity of this group. With these large differences, the simplification of auditory functions, introduced by measuring only a few points in order to keep the measuring time within reasonable limits, will have a much smaller influence on the test scores. In this final study we investigated the audiometric loss, frequency resolution, temporal resolution, and speech-reception thresholds in quiet and in noise. The matrix of correlation coefficients (Table 5.2) showed a lot of significant correlations and with a principal-components analysis it was possible to arrive at a simple configuration of tests, in which frequency resolution together with speech reception in noise was found to be largely independent of audiometric loss. However, it was not possible to identify temporal resolution as a separate factor and we are still far from a reduction in the number of parameters needed for the description of the hearing mechanism. A comparison of correlations found for hearing-impaired subjects with those for normal-hearing subjects from the previous studies is also given in Table 6.1. The comparison is difficult for the same reasons as mentioned above and also because hearing impairment may introduce relations which do not exist for normal-hearing subjects. For instance, in normal hearing there seems to be a trade-off between temporal resolution and frequency resolution, reflected in the correlation between Temporal window and Shallow edge simult. in Table 6.1. But if both these functions deteriorate with hearing loss, the trade-off is lost and the sign of the correlation coefficient may be inverted.

With regard to the generally low correlation coefficients found, it is not to be expected that longer measurement runs or more replications would have lead to more reliable results and higher correlations. Within runs accuracy will be limited by effects of fatigue, although we did not reach this limit as shown in Chapter 4, and when we spread the measurements over more days, accuracy is limited by day-to-day variations in auditory functions which is addressed in section 3.3.2. But, above all, it was shown that for the obtained correlations test reliability was not the limiting factor.

It is my strong opinion that although the enterprise undertaken was not completely successful, it was worhtwhile to try out and essential to test the coherence of hearing theory. For the time being, our general conclusion must be that we cannot describe important features of the peripheral auditory system with just a few basic parameters.

SUMMARY

In the analytical description of the human auditory process properties like frequency resolution, temporal resolution, and nonlinearity are expressed as input-output functions. Examples of such functions are: the bandwidth and the steepness of skirts of the auditory "filters" as a function of the center frequency or as a function of intensity, and the level of intermodulation products as a function of the intensity of the generating components. In psychoacoustics nowadays, these functions are usually measured as detection thresholds for a test tone in the presence of a masking signal. According to hearing theory many of these auditory functions have a common origin or are closely related. In this thesis an attempt is made to find experimental evidence for these relations, using psychophysical measurement techniques. The intention is to reduce the number of parameters needed in the description of the auditory process. The investigations are restricted to the frequency region of 1000 Hz and they make use of natural differences among subjects; relations are traced by calculating Pearson's product-moment correlations between test results. In order to arrive at reliable correlations various sources of error variance are reduced as far as possible.

After the introduction in Chapter 1, a pilot experiment with 10 normal-hearing subjects is presented in Chapter 2. This experiment included a variety of tests on properties like: the threshold of hearing, just noticeable differences in frequency and intensity, the perception of the pitch of complex tones, frequency resolution, temporal resolution, and the nonlinear distortion of the ear. All tests were conducted twice in order to estimate their reliability. It was found that small differences among subjects can be measured with sufficient accuracy to calculate reliable correlation coefficients. In particular, tests using a forced-choice procedure appeared to be more reliable than tests in which the subject had to adjust the detection threshold. The

latter thresholds may be affected by variations in the threshold criterion of the subject.

Based on the preliminary experiment, the third chapter deals with an experiment in which 50 normal-hearing subjects were tested on a battery of 12 carefully selected tests. These tests involve threshold of hearing, frequency resolution and temporal resolution of the ear, and the occurrence of nonlinear distortion. In this experiment only a few significant correlations were found. On the assumption that the applied simplification of auditory functions was justified, it was concluded that the measured auditory functions are largely independent of each other.

As an intermezzo, the fourth chapter considers properties of the measurement procedure used. The procedure is divided into two phases: the first, in which observations are made according to an efficient adaptive procedure, and the second, in which the threshold is estimated from the obtained results. The adaptive procedure of forced-choice trials used in making the observations is explained and three procedures to estimate the threshold are discussed. In the second part of this chapter the test results from the Chapters 3 and 5 are used to determine the measurement error as a function of the number of trials in the adaptive procedure and for the different methods of estimating the threshold It appears that the accuracy of various threshold estimations is very much alike and that within runs of up to 30 trials in the measuring stage no signs are found of a reduced accuracy towards the end of the run, as would be expected in the case of fatigue.

In Chapter 5 a slightly modified battery of tests is applied to a group of 22 sensorineurally hearing-impaired subjects. In addition to the audiogram, we measured frequency resolution, temporal resolution, and speech reception in quiet and in noise. This experiment showed a number of clear correlations between the tests, which were gathered in a matrix and subjected to a principal-components analysis. It turned out that 66% of the total variance could be accounted for in two dimensions. In this subspace two clusters of tests are found; the first cluster contains tests on frequency resolution and is approximately independent of the second cluster, which is determined by the audiometric loss. Furthermore, hearing loss for speech in noise is closely allied to frequency resolution, whereas hearing loss for speech in quiet is governed by audiometric loss. The clear results for sensorineurally hearing-impaired subjects as opposed to the results for normal-hearing subjects can be understood if we assume that hearing impairment affects only specific parameters of hearing.

Finally, in Chapter 6 the results of the various studies are confronted with each other and are evaluated in view of the aim to reduce the number of parameters needed in the description of the auditory system. It is concluded that, for the time being, we cannot describe important features of the peripheral auditory system with just a few parameters.

SAMENVATTING

In de analytische beschrijving van het auditieve proces bij de mens worden eigenschappen als frequentieresolutie, temporele resolutie en nietlineariteit uitgedrukt in de vorm van "input-output" functies. Voorbeelden hiervan zijn: de bandbreedte en de steilheid van de flanken van de auditieve "filters" als functie van de centrale frequentie of als functie van de intensiteit, en de sterkte van frequentiekomponenten tengevolge van intermodulatie, eveneens als functie van de intensiteit. In de psychoakoestiek worden deze functies tegenwoordig meestal gemeten in de vorm van maskeerdrempels voor een testtoon tegen een achtergrond van een maskerend geluidsignaal. Uit de theorie over het horen volgt dat er een functionele samenhang moet zijn tussen een groot aantal van de hierboven genoemde auditieve functies. In dit proefschrift wordt geprobeerd deze relaties experimenteel aan te tonen met behulp van psychofysische meetmethoden om zo te komen tot een reductie van het aantal parameters waarmee het gehoor beschreven wordt. Het onderzoek beperkt zich tot frequenties rond 1000 Hz en maakt gebruik van de natuurlijke verschillen tussen proefpersonen; relaties worden opgespoord door het berekenen van Pearson's produkt-momentcorrelaties tussen de resultaten van verschillende tests. Om betrouwbare correlaties te vinden, worden de verschillende te onderscheiden foutenbronnen zoveel mogelijk beperkt.

Na de inleiding in Hoofdstuk 1 wordt in Hoofdstuk 2 een verkennend experiment behandeld met 10 normaalhorende proefpersonen. Het experiment bestond uit een breed scala van tests die alle tweemaal werden uitgevoerd om een indruk te krijgen van de meetnauwkeurigheid. De tests hadden betrekking op de hoordrempel, het juist waarneembare verschil in frequentie en in intensiteit, de toonhoogtewaarneming in samengestelde signalen, het oplossend vermogen in frequentie en in tijd, en het optreden van door het oor geïntroduceerde nietlineaire vervorming. Aangetoond werd dat de kleine verschillen tussen proefpersonen nauwkeurig genoeg gemeten kunnen worden om betrouwbare correlaties te

berekenen. In het bijzonder bleken tests waarbij gebruik gemaakt werd van een gedwongen-keuzemethode nauwkeuriger resultaten op te leveren dan tests waarbij de proefpersoon de waarnemingsdrempel zelf moest instellen. In deze tweede procedure kunnen de drempels beinvloed worden door variaties in het detectiekriterium van de proefpersoon.

Voortbouwend op het voorafgaande experiment behandelt het derde hoofdstuk een experiment met een batterij van 12 zorgvuldig op elkaar afgestemde tests uitgevoerd met 50 normaalhorende proefpersonen. Naast de hoordrempel, hadden de tests betrekking op het oplossend vermogen van het gehoor in frequentie en in tijd, en op aspecten van de optredende nietlineaire vervorming. In dit experiment werden slechts een gering aantal significante correlaties gevonden. De conclusie was dat de gemeten auditieve functies grotendeels onafhankelijk van elkaar zijn, mits de toegepaste vereenvoudigingen, door slechts voor enkele condities te meten, gerechtvaardigd zijn.

Het vierde hoofdstuk behandelt, als een intermezzo, eigenschappen van de gebruikte meetmethode. De meetprocedure wordt onderverdeeld in twee fasen: de eerste, waarin observaties gedaan worden volgens een efficiente adaptieve procedure en de tweede, waarin de drempel geschat wordt op grond van de verkregen resultaten. De adaptieve procedure, met gedwongen keuzen, die gebruikt is voor het doen van de observaties wordt toegelicht en drie procedures voor het schatten van de drempel worden besproken. In het tweede deel van dit hoofdstuk wordt aan de hand van de testresultaten uit de Hoofdstukken 3 en 5 de meetnauwkeurigheid bepaald als functie van het aantal observaties in de adaptieve procedure en voor de verschillende methoden om de drempel te schatten. Het blijkt dat de nauwkeurigheid van de verschillende schattingsmethoden elkaar nauwelijks ontloopt en dat binnen reeksen tot 30 observaties niets te merken is van een afnemende nauwkeurigheid tegen het einde van de reeks, zoals te verwachten zou zijn bij vermoeidheid.

In Hoofdstuk 5 wordt een iets gewijzigde batterij tests toegepast op een groep van 22 slechthorende proefpersonen. Naast het audiogram, werden de frequentieresolutie, de temporele resolutie en het spraakverstaan in ruis en in stilte gemeten. Dit experiment leverde een aantal duidelijke correlaties op; de matrix van correlaties werd onderworpen aan een principale-komponentenanalyse. Er wordt aangetoond dat verschillende tests voor de frequentieresolutie met elkaar samenhangen en als groep onafhankelijk zijn van het gehoorverlies voor zuivere tonen. Bovendien blijkt dat het gehoorverlies voor spraak tegen een achtergrond van ruis nauw verbonden is met de frequentieresolutie, terwijl het gehoorverlies voor spraak in stilte bepaald wordt door het gehoorverlies voor zuivere tonen. De duidelijke resultaten bij slechthorende proefpersonen, in tegenstelling tot de resultaten bij normaalhorende proefpersonen, kunnen worden verklaard door aan te nemen dat de slechthorendheid slechts zeer specifieke parameters van het gehoor heeft aangetast.

Tenslotte worden in Hoofdstuk 6 de resultaten van de verschillende studies met elkaar vergeleken en wordt de balans opgemaakt in het licht van de doelstelling om te komen tot een reductie van het aantal parameters dat nodig is om het gehoor te beschrijven. De conclusie hiervan is, dat wij tot nu toe niet in staat zijn om belangrijke eigenschappen van het perifere gehoor te beschrijven met slechts enkele parameters.

GLOSSARY

Air-bone gap	difference between air-conduction loss measured with a headphone and bone-conduction loss measured with a vibrator on the mastoid.
Air-conduction loss	threshold of audition, when a headphone is used as stimulator, relative to the normal threshold of young listeners without ear pathology; the sound waves reach the inner ear via vibrations of the air stimulating the ear drum and the ossicular chain.
Analysis of variance (ANOVA)	statistical analysis, giving a decomposition of variance in the dependent variable due to various independent variables, and estimating the strength of effects introduced by the independent variables.
Auditory bandwidth	width of hypothetical filters in the ear representing its frequency selectivity.
Backward masking	masking of signals occurring (shortly) before the masking signal is switched on.
Bone-conduction loss	threshold of audition, when a vibrator on the mastiod is used as stimulator, relative to the normal threshold for young listeners without ear pathology; the sound waves reach the inner ear through vibrations of the bony surroundings of the inner ear.
Combination tones	tones generated under certain conditions within the ear when two tones are presented simultaneously.
Critical bandwidth	phenomenologically defined measure of the ear's frequency resolution; a number of properties of the ear

	show a sudden change when the bandwidth of the signal reaches this critical value, which is over a large range of frequencies about 1/3 octave.
Critical ratio	threshold in decibels of a tone relative to the spectral density of a wide-band noise just masking the tone.
Cubic difference tone (CDT)	combination tone with frequency $2f_1 - f_2$ where f_1 and f_2 ($f_1 < f_2$) are the frequencies of the tones stimulating the ear; the occurrence of this combination tone is often attributed to a cubic term in the transfer function.
Factor analysis	statistical analysis, describing the results for a group of subjects on a number of tests with a minimum number of variables being composites of the original tests.
Forward masking	masking of signals occurring after the masker is switched off.
Frequency discrimination	ability to distinguish frequency differences between two signals sounding in succession.
Interaction (in ANOVA)	part of the variance in the dependent variable that cannot be accounted for by a simple addition of the effects of the individual independent variables.
Lateral suppression	suppression of "weaker" parts in the stimulus pattern by adjacent "stronger" parts; in audition strong spectral parts suppress weaker parts.
Low pitch	pitch corresponding to the fundamental of a complex tone, even if the fundamental itself is absent.
Ménière's disease	combination of hearing loss, vertigo (dizziness), and tinnitus usually consisting of "attacks" and often ascribed to an increased fluid pressure in the membranous labyrinth (Newby, 1979).
Monte-Carlo method	simulation of a process governed by chance processes.
Nonsimultaneous masking	masking of a signal occurring shortly before or after the masking signal.

Principal-components analysis	kind of factor analysis in which all the variance in the tests is used to establish new dimensions, as opposed to "principal-factors" analysis in which per test a fraction of the variance may be unique (Harman, 1970).
Psychophysical tuning curve (PTC)	curve showing level as a function of frequency for a tone just capable of masking a weak tone of fixed level and frequency.
Psychometric function	detection chance as a function of some signal parameter (e.g. intensity).
Pulsation threshold	highest signal level for which, when using an alternation of masker and probe signal, the latter is sounding as continuous (signal and masker duration both about 125 ms); this signal level is considered to evoke a neural activity which is for no frequency larger than the activity evoked by the masker.
Q_{10dB}	filter quality, defined as its width, at a level 10 dB from the peak, divided by the centre frequency.
Sensation level (SL)	level in decibels above the detection threshold.
Spectral density	noise intensity in 1-Hz intervals in decibels relative to the intensity of a tone with a level of 20 μPa.
Speech-reception threshold	level in decibels at which 50% of speech is correctly understood.
Speech-discrimination score	percentage of correctly reproduced words.
Sound-pressure level	root-mean-square sound pressure re 20 μPa.
Temporal integration	decrease of the detection threshold with increasing duration for sounds shorter than about 200 ms.
Tinnitus	sound impression generated spontaneously within the ear.
Two-alternative forced choice (2AFC)	procedure for measuring detection thresholds in which the phenomenon to be detected is present in one of two observation intervals and the subject is forced to choose the right interval.

REFERENCES

BUUNEN, T.J.F., FESTEN, J.M., BILSEN, F.A., and VAN DEN BRINK, G. (1974). "Phase effects in a three-component signal", J. Acoust. Soc. Am. *55*, 297-303.

BUUNEN, T.J.F. and VAN VALKENBURG, D.A. (1979). "Auditory detection of a single gap in noise", J. Acoust. Soc. Am. *65*, 534-537.

CAMPBELL, R.A. (1974). "Modifications to the BUDTIF procedure", J. Acoust. Soc. Am. *56*, S57.

CARDOZO, B.L. (1966). "A sequential up-and-down method", IPO Annual Progress Report 1, 110-114.

DIXON, W.J. and MOOD, A.M. (1948). "A method for obtaining and analysing sensitivity data", J. Am. Stat. Ass. *43*, 109-126.

DRESCHLER, W.A. and PLOMP, R. (1980). "Relation between psychophysical data and speech perception for hearing-impaired subjects. I", J. Acoust. Soc. Am. *68*, 1608-1615.

DRESCHLER, W.A. and PLOMP, R. (1983). "Relation between psychophysical data and speech perception for hearing-impaired subjects. II", submitted for publication to J. Acoust. Soc. Am.

DUIFHUIS, H. (1973). "Consequences of peripheral frequency selectivity for nonsimultaneous masking", J. Acoust. Soc. Am. *54*, 1471-1488.

DZIUBAN, C.D. and SHIRKEY, E.C. (1974). "When is a correlation matrix appropriate for Factor Analysis", Psychol. Bull. *81*, 358-361.

ELLIOTT, D.N., RIACH, W.D., SHEPOSH, J.P., and TRAHIOTIS, C. (1966). "Discrimination performance of high school sophomores on a battery of auditory tests", Acta Oto-Laryngol., Suppl. 216.

ELLIOTT, L.L. (1962). "Backward masking: monotic and dichotic conditions", J. Acoust. Soc. Am. *34*, 1108-1115.

ELLIOTT, L.L. (1975). "Temporal and masking phenomena in persons with sensorineural hearing loss", Audiology *14*, 336-353.

EVANS, E.F. (1977). "Frequency selectivity at high signal levels of single units in cochlear nerve and cochlear nucleus", in *Psychophysics and Physiology of Hearing*, E.F. Evans and J.P. Wilson, Eds. (Academic Press, London), 185-192.

FESTEN, J.M. (1980). "Relations between auditory functions", in *Psychophysical, Physiological and Behavioural Studies in Hearing*, G. van den Brink and F. A. Bilsen, Eds. (Delft University Press), 450-457.

FESTEN, J.M., HOUTGAST, T., PLOMP, R., and SMOORENBURG, G.F. (1977). "Relations between interindividual differences of auditory functions", in *Psychophysics and Physiology of Hearing*, E.F. Evans and J.P. Wilson, Eds. (Academic Press, London), 311-319.

FESTEN, J.M., and PLOMP, R. (1981). "Relations between auditory functions in normal hearing", J. Acoust. Soc. Am. *70*, 356-369.

FESTEN, J.M., and PLOMP, R. (1983). "Relations between auditory functions in impaired hearing", accepted for publication in J. Acoust. Soc. Am.

FLORENTINE, M., BUUS, S., SCHARF, B., and ZWICKER, E. (1980). "Frequency selectivity in normal-hearing and hearing-impaired observers", J. Speech. Hear. Res. *23*, 646-669.

GOLDSTEIN, J.L. (1967). "Auditory nonlinearity", J. Acoust. Soc. Am. *41*, 676-689.

GREEN, D.M. (1960). "Psychoacoustics and detection theory", J. Acoust. Soc. Am. *32*, 1189-1203.

GREEN, D.M. (1973). "Minimum integration time", in *Basic Mechanisms in Hearing*, A.R. Møller, Ed. (Academic Press, New York), 829-843.

GREEN, D.M., and SWETS, J.A. (1973). *Signal detection theory and psychophysics*. (Krieger comp., Huntington, NY).

GREEN, D.M., SHELTON, B.R., PICARDI, M.C., and HAFTER, E.R. (1981). "Psychophysical tuning curves independent of signal level", J. Acoust. Soc. Am. *69*, 1758-1762.

GUILFORD, J.P. (1954). *Psychometric Methods*. (McGraw-Hill, New York).

HALL, J.L. (1972). "Auditory distortion products $f_2 - f_1$ and $2f_1 - f_2$", J. Acoust. Soc. Am. *51*, 1863-1871.

HALL, J.L. (1981). "Hybrid adaptive procedure for estimation of psychometric functions", J. Acoust. Soc. Am. *69*, 1763-1769.

HANLEY, C.N. (1956). "Factorial analysis of speech perception", J. Speech. Hear. Dis. *21*, 76-87.

HARMAN, H.H. (1970). *Modern factor analysis*. (The University of Chicago Press, Chicago).

HARRIS, J.D. (1957). "A search toward the primary auditory abilities", Med. Res. Lab. Memo. Rep. No 57-4 (U.S. Navel Submarine Base, New London).

HOEKSTRA, A. (1979). *Frequency discrimination and frequency analysis in hearing*, Ph. D. dissertation.

HOEKSTRA, A., and RITSMA, R.J. (1977). "Perceptive hearing loss and frequency selectivity", in *Psychophysics and Physiology of Hearing*, E.F. Evans and J.P. Wilson, Eds. (Academic Press, London), 263-271.

HORST, J.W. (1982). *Discrimination of complex signals in hearing*, Ph. D. dissertation.

HOUTGAST, T. (1973). "Psychophysical experiments on tuning curves and two-tone suppression", Acustica *29*, 168-179.

HOUTGAST, T. (1974). *Lateral Suppression in Hearing*, Ph. D. dissertation.

HOUTGAST, T. (1977). "Auditory-filter characteristics derived from direct-masking data and pulsation-threshold data with a rippled-noise masker", J. Acoust. Soc. Am. *62*, 409-415.

HUMES, L.E. (1979). "Perception of the simple difference tone ($f_2 - f_1$)", J. Acoust. Soc. Am. *66*, 1064-1074.

LESHOWITZ, B. (1977). "Speech intelligibility in noise for listeners with sensorineural hearing damage", IPO Annual Progress Report 12, 11-23.

LESHOWITZ, B.H., and LINDSTROM, R. (1977). "Measurement of nonlinearities in listeners with sensorineural hearing loss", in *Psychophysics and Physiology of Hearing*, E.F. Evans and J.P. Wilson, Eds. (Academic Press, London), 283-292.

LEVITT, H. (1971). "Transformed up-down methods in psychoacoustics", J. Acoust. Soc. Am. *49*, 467-477.

MARTIN, M.S., and PICKETT, J.M. (1970). "Sensorineural hearing loss and upward spread of masking", J. Speech. Hear. Res. *13*, 426-437.

MOORE, B.C.J. (1978). "Psychophysical tuning curves measured in simultaneous and forward masking", J. Acoust. Soc. Am. *63*, 524-532.

MRASS, H., and Diestel, H.G. (1959). "Bestimmung der Normalhörschwelle für reine Töne bei einohrigem Hören mit Hilfe eines Kopfhörers", Acustica *9*, 61-64.

MULLINS, C.J., and BANGS, J.L. (1957). "Relationships between speech discrimination and other audiometric data", Acta Otolaryngol. *47*, 149-157.

NELSON, D.A. (1980). Comment on a paper by D.L. Weber, D. Johnson-Davies, and R.D. Patterson in *Psychophysical, Physiological and Behavioural Studies in Hearing*, G. van den Brink and F.A. Bilsen, Eds. (Delft University Press), 116-117.

NELSON, D.A., and TURNER, C.W. (1980). "Decay of masking and frequency resolution in sensorineural hearing-impaired listeners", in *Psychophysical, Physiological and Behavioural Studies in Hearing*, G. van den Brink and F.A. Bilsen, Eds. (Delft University Press), 175-182.

NEWBY, H.A. (1979). *Audiology*. (Prentice-Hall, Inc., Englewood Cliffs, NJ).

NIE, N.H., HADLAI HULL, C., JENKINS, J.G., STEINBRENNER, K., and BENT, D.H. (1975). *Statistical Package for the Social Sciences*. (McGraw-Hill, New York).

NUNNALLY, J.C. (1967). *Psychometric Theory*. (McGraw-Hill, New York).

PATTERSON, R.D. (1976). "Auditory filter shapes derived with noise stimuli", J. Acoust. Soc. Am. *59*, 640-654.

PICK, G.F., EVANS, E.F., and WILSON, J.P. (1977). "Frequency resolution in patients with hearing loss of cochlear origin", in *Psychophysics and Physiology of Hearing*, E.F. Evans and J.P. Wilson, Eds. (Academic Press, London), 273-281.

PICKETT, J.M. (1959). "Backward masking", J. Acoust. Soc. Am. *31*, 1613-1615.

PLOMP, R. (1964). "Rate of decay of auditory sensation", J. Acoust. Soc. Am. *36*, 277-282.

PLOMP, R. (1965). "Detectability threshold for combination tones", J. Acoust. Soc. Am. *37*, 1110-1123.

PLOMP, R. (1978). "Auditory handicap of hearing impaired and the limited benefit of hearing aids", J. Acoust. Soc. Am. *63*, 533-549.

PLOMP, R., and MIMPEN, A.M. (1979). "Improving the reliability of testing the speech-reception threshold for sentences", Audiology *18*, 43-52.

RITSMA, R.J., WIT, H.P., and VAN DER LANS, W.P. (1980). "Relations between hearing loss, maximal word discrimination scores and width of psychophysical tuning curves", in *Psychophysical, Physiological and Behavioural Studies in Hearing*, G. van den Brink and F.A. Bilsen, Eds. (Delft University Press), 472-475.

RHODE, W.S. (1978). "Some observations on cochlear mechanics", J. Acoust. Soc. Am. *64*, 158-176.

RODENBURG, M. (1977). "Investigation of temporal effects with amplitude modulated signals", in *Psychophysics and Physiology of Hearing*, E.F. Evans and J.P. Wilson, Eds. (Academic Press, London), 429-437.

RODENBURG, M., VERSCHUURE, J., and BROCAAR, P.M. (1974). "Comparison of two masking methods", Acustica *31*, 99-106.

ROSS, M., HUNTINGTON, D.A., NEWBY, H.A., and DIXON, R.F. (1965). "Speech discrimination of hearing-impaired individuals in noise", J. Aud. Res. *5*, 47-72.

SCHARF, B. (1970). "Critical bands", in *Foundations of Modern Auditory Theory*, J.V. Tobias, Ed. (Academic Press, New York), 157-202.

SCHARF, B., and MEISELMAN, C.H. (1977). "Critical bandwidth at high intensities", in *Psychophysics and Physiology of Hearing*, E.F. Evans and J.P. Wilson, Eds. (Academic Press, London), 221-232.

SMOORENBURG, G.F. (1972). "Audibility region of combination tones", J. Acoust. Soc. Am. *52*, 603-614.

SMOORENBURG, G.F. (1974). "On the mechanisms of combination tone generation and lateral inhibition in hearing", in *Facts and Models in Hearing*, E. Zwicker and E. Terhardt, Eds. (Springer, Berlin), 332-343.

SOLOMON, L.N., WEBSTER, J.C., and CURTIS, J.F. (1960). "A factorial study of speech perception", J. Speech. Hear. Res. *3*, 101-107.

TAYLOR, M.M., and CREELMAN, C.D. (1967). "PEST: efficient estimates on probability functions", J. Acoust. Soc. Am. *41*, 782-787.

TYLER, R., FERNANDES, M., and WOOD, E.J. (1980). "Masking, temporal integration and speech intelligibility in individuals with noise-induced hearing loss", in *Disorders of Auditory Function, III*, I.G. Taylor, Ed. (Academic Press, London), 211-236.

VAUGHAN, G.M., and CORBALLIS, M.C. (1969). "Beyond tests of significance: estimating strength of effects in selected anova designs", Psych. Bull. *72*, 204-213.

VIEMEISTER, N.F. (1977). "Temporal factors in audition: a system analysis approach", in *Psychophysics and Physiology of Hearing*, E.F. Evans and J.P. Wilson, Eds. (Academic Press, London), 419-428.

VIEMEISTER, N.F. (1977). Comment on a paper by B.H. Leshowitz and R. Lindstrom in *Psychophysics and Physiology of Hearing*, E.F. Evans and J.P. Wilson, Eds. (Academic Press, London), 293.

VOGTEN, L.L.M. (1978a). "Simultaneous pure-tone masking: the dependence of masking asymmetries on intensity", J. Acoust. Soc. Am. *63*, 1509-1519.

VOGTEN, L.L.M. (1978b). "Low-level pure-tone masking: A comparison of tuning curves obtained with simultaneous and forward masking", J. Acoust. Soc. Am. *63*, 1520-1527.

WARREN, R.M. (1970). "Elimination of biases in loudness judgments for tones", J. Acoust. Soc. Am. *48*, 1397-1403.

WEGEL, R.L., and LANE, C.E. (1924). "The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear", Phys. Rev. *23*, 266-285.

WIGHTMAN, F.L., McGEE, T., and KRAMER, M.C. (1977). "Factors influencing frequency selectivity in normal and hearing-impaired listeners", in *Psychophysics and Physiology of Hearing*, E.F. Evans and J.P. Wilson, Eds. (Academic Press, London), 295-306.

WIGHTMAN, F.L., and RAZ, I. (1980). "Frequency selectivity as a function of sound pressure level", J. Acoust. Soc. Am. Suppl. 1, *67*, S8.

WILSON, R.H., and CARHART, R. (1971). "Forward and backward masking: interactions and additivity", J. Acoust. Soc. Am. *49*, 1254-1263.

WILSON, J.P., and EVANS, E.F. (1971). "Grating acuity of the ear: psychophysical and neurophysiological measures of frequency resolving power", 7th Int. Congr. Acoust., Budapest, 20 H 14.

ZWICKER, E. (1954). "Die Verdeckung von Schmalbandgeräuschen durch sinustöne", Acustica *4*, 415-420.

ZWICKER, E. (1955). "Der ungewöhnliche Amplitudengang der nichtlinearen Verzerrungen des Ohres", Acustica *5*, 67-74.

ZWICKER, E. (1970). "Masking and psychological excitation as consequences of the ears frequency analysis", in *Frequency Analysis and Periodicity Detection in Hearing*, R. Plomp and G.F. Smoorenburg, Eds. (Sijthoff, Leiden), 376-394.

ZWICKER, E. (1974). "On a psychoacoustical equivalent of tuning curves", in *Facts and Models in Hearing*, E. Zwicker and E. Terhardt, Eds. (Springer, Berlin), 132-141.

ZWICKER, E., and HEINZ, W. (1955). "Zur Häufigkeitsverteilung der menschlichen Hörschwelle", Acustica *5*, 75-80.

ZWICKER, E., and SCHORN, K. (1978). "Psychoacoustical tuning curves in audiology", Audiology *17*, 120-140.