ASPECTS OF TONE SENSATION

A Psychophysical Study

Reinier Plomp

Institute for Perception TNO, Soesterberg

and

Faculty of Medicine, Free University, Amsterdam

1976

ACADEMIC PRESS

London New York San Francisco

A Subsidiary of Harcourt Brace Jovanovich, Publishers

CHAPTER 1

THE EAR AS A FREQUENCY ANALYZER

We do not need a laboratory to convince us that our hearing organ is capable of analyzing sounds. We have no difficulties in discriminating sound waves from different sources, even though these waves are seemingly inextricably superimposed before reaching the ear. The most impressive demonstration of this analyzing power is the way in which we distinguish the tones produced by the musical instruments of an orchestra. Apparently, the superposition of the various periodic sound waves produced by the violins, flutes, oboes, etc., is no obstacle to hearing the notes individually.

At first sight, a rather simple model might appear adequate to describe the way in which the superimposed sound waves are perceived and analyzed by the auditory mechanism. We could suppose that the ear was provided with a large number of overlapping band-pass filters tuned to different frequencies from low to high. A one-to-one relationship between filter frequency and perceived pitch would explain how simultaneous tones are discriminated. This elegant model has played an important part in the history of hearing theory. Its greatest promotor was the eminent German scientist H. L. F. von Helmholtz. He published the theory in 1863 in his book "Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik". Von Helmholtz's comparison of the ear's frequency-analyzing mechanism with the strings of a piano appealed to the imagination and the book was soon considered a "classic". Its impact on the development of hearing theory during the last century has been immense.

Looking more carefully, however, this model appears to give too crude a description of the analyzing processes in the hearing system. Physics has shown that a typical musical tone consists of a large number of harmonics with frequency ratios 1 : 2 : 3 : 4 : 5 . . . Comparing the ear with a frequency analyzer suggests that any periodic sound wave should be heard as a sum of sinusoidal components, or partials, with their corresponding pitches. We do not, however, distinguish these individual harmonics when listening to a concert. The sound of each instrument results in the sensation of a single note with a single definite pitch equal to the pitch of the fundamental and a specific timbre depending upon the relative amplitudes of the harmonics. The frequency ranges of the harmonics of two tones may overlap considerably, but still we hear only two tones. Apparently, the ear is able to analyze a compound sound into complex rather than into simple tones. We do not have a physical model for such a filter system that separates the harmonics belonging to different tones but does not separate the harmonics belonging to the same tone.

Does this imply that we should abandon the frequency analyzer as a model of the ear's analyzing process? This would be necessary if, indeed, the ear were not capable of distinguishing the individual harmonics of a complex tone. Identification of the partials may be nearly impossible in listening to tones in a musical context, but it is possible under more favourable conditions. In this chapter the experimental evidence concerning the ear's frequency-analyzing power and its limits will be treated. A discussion of the imperfections of this model as a full description of the way in which the organ of hearing discriminates simultaneous tones can better be postponed until the final chapter of this study. Earlier chapters on loudness, timbre, and pitch as attributes of complex sounds will show that auditory analysis is followed by a process of synthesis.

1. Identification of the partials of a complex sound

The concept that the ear can be compared with a frequency analyzer was accurately formulated for the first time by the German physicist Ohm in 1843. His "definition of tone", nowadays known as "Ohm's acoustical law", states that a tone with frequency f is only heard if the complex sound contains $\sin(2\pi ft + \phi)$ as a component (Ohm, 1843). In von Helmholtz's phrasing, this definition reads: "Every motion of the air which corresponds to a composite mass of musical tones, is, according to Ohm's law, capable of being analysed into a sum of simple pendular vibrations, and to each such single simple vibration corresponds a simple tone, sensible to the ear, and having a pitch determined by the periodic time of the corresponding motion of the air." (von Helmholtz, 1863; English translation, p. 33.) We should notice that Ohm's definition does not exclude the possibility of analyzing power being limited, which, although admitted by von Helmholtz elsewhere, would appear to be excluded by his quoted definition.

Von Helmholtz was quite aware of the difficulty of distinguishing the harmonics of a complex tone and recommended, in order to direct the attention to a particular partial, to listen first to a tone of the same pitch as the partial. How important this advice is can be concluded from experiments bv Thurlow and Rawlings and by Pollack. Thurlow and Rawlings

(1959) presented various one-, two-, and three-tone stimuli to subjects, who had to judge how many tones were present. Pollack (1964) requested his subjects to decide whether a probe tone, presented after the stimulus, was or was not a member of the tone combination. In both cases the results were so poor that the authors questioned the validity of Ohm's acoustical law.

A more positive result was obtained in a series of experiments (Plomp, 1964; Plomp and Mimpen, 1968) in which von Helmholtz's advice to direct the attention to the partial to be distinguished was taken into account. In order to decide whether the subject did actually hear the component, the following procedure was used. The subject had at his disposal a switch with three positions. In the middle position he heard the complex stimulus

$$\sum_{n=1}^{12} \cos 2\pi n f_0 t. \tag{1}$$

In the positions 1 and 3 the subject heard probe tones, one with frequency nf_0 (n = integer), the other one with frequency $(n \pm \frac{1}{2})f_0$. The subject was allowed to switch freely from one position to another. He had to judge which of the two probe tones was also present in the complex tone. There were equal chances that either condition 1 or 3 contained the component nf_0, and that the frequency of the other probe tone was either $(n - \frac{1}{2})f_0$ or $(n + \frac{1}{2})f_0$. The chance of a correct response in this so-called two-alternative forced-choice procedure varied between 0·5 and 1, corresponding to 0% and 100% identifiability of the partial, respectively. The complex signal given by Eq. (1) was presented monaurally by headphone at a sensation level of about 60 dB.

In these experiments the fundamental frequency was varied between 44 Hz and 2000 Hz. For each f_0 the percentage of correct responses diminished monotonically for increasing n. This percentage was plotted as a function of n and fitted by a smooth curve of which the 75% point was taken as the limit of the number of distinguishable harmonics. The results for six intensively trained subjects are reproduced in Fig. 1. In this diagram 44 Hz is not included because none of the subjects scored more than 75% correct responses for the second and higher harmonics.

From these data the minimal frequency separation required to identify a partial out of a complex can be calculated. As an example, let us take $f_0 = 500$ Hz. The diagram shows that, on the average, the first five harmonics, up to 2500 Hz, could be identified. This means that at 2500 Hz a frequency separation of 500 Hz is required. Similarly, this limit can be determined for other frequencies and the result is plotted in Fig. 2 (open circles). The curve gives the best fit to the data points.

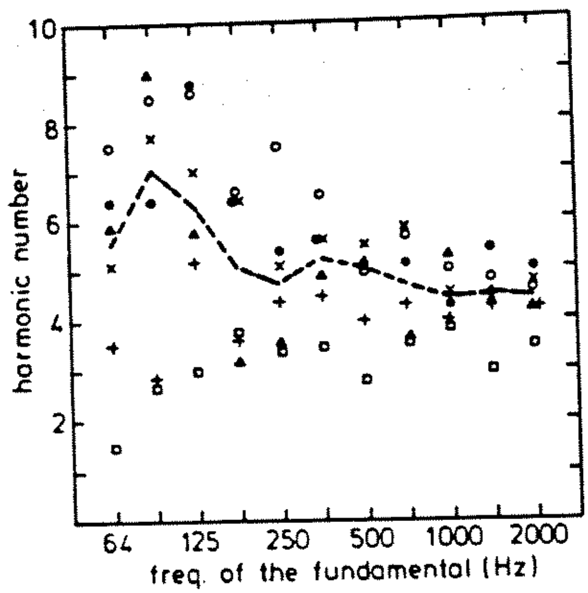

FIG. 1 Number of harmonics identified by six subjects as a function of the frequency of the fundamental. The dashed line represents the median. The complex tone contained 12 harmonics. (From Plomp and Mimpen, 1968.)

In order to verify whether the harmonic relationship of the partials had any effect on the results, the measurement was repeated for series of partials which were not harmonically related. Two of the subjects participated in this experiment and their average data are also reproduced in Fig. 2 (crosses). Additional data were published by Soderquist (1970), who adopted the same procedure as described above. His eight subjects, including four professional musicians, listened to two inharmonic series of 12 partials distributed evenly over the frequency range between about 230 Hz and 2200 Hz. The results, averaged over the two stimuli and the subjects within each group, are given for the nonmusicians (square symbol) and musicians (triangular symbol) separately.

We may conclude from Fig. 2 that the limit of the ear's ability to identify the individual components of a complex sound does not depend upon whether these partials are harmonically related or not. The spread of the data points in Fig. 1 and the distance between the two points adopted from Soderquist in Fig. 2 indicate that there are substantial differences among subjects in their ability to identify partials. At low frequencies this limit, expressed as the minimum frequency separation required for identification, has a constant value of about 60 Hz, at middle and higher frequencies

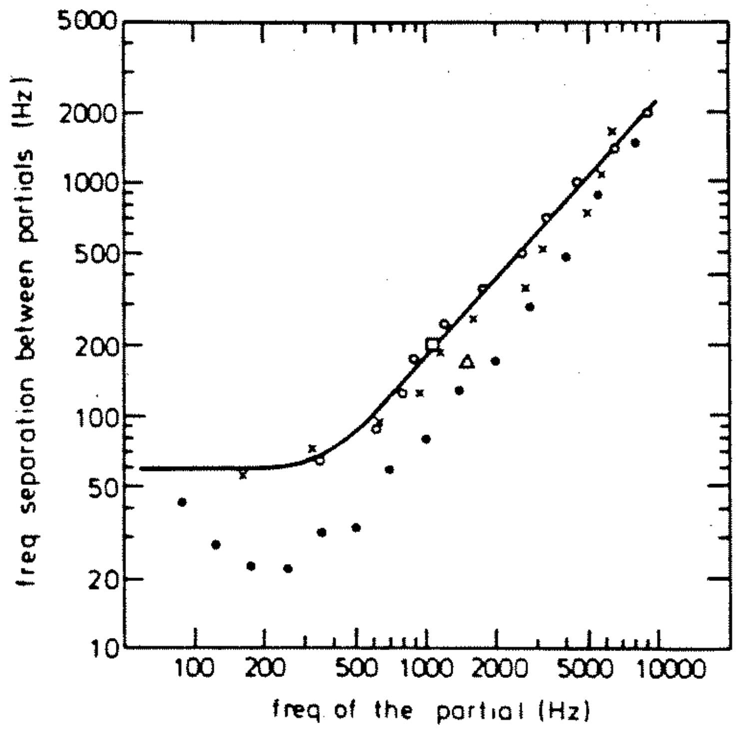

FIG. 2 Frequency separation between partials of a complex sound required for correct identification. The open circles, fitted with the curve, are from a harmonic tone complex (six subjects), the crosses from an inharmonic tone complex (two subjects). The solid points are from stimuli consisting of only two components (same two subjects). (Data from Plomp, 1964; Plomp and Mimpen, 1968.) The open square and triangle represent the average data of a similar experiment with inharmonic tone complexes for four nonmusicians and four musicians, respectively, by Soderquist (1970).

it is a nearly constant percentage of frequency of 15 to 20%. In musical terms the partials of a complex stimulus can be identified if their frequency separation exceeds a value between a minor and a major third.

It should be mentioned that, by applying a different procedure, some investigators have concluded that more harmonics are distinguishable than represented in Fig. 1. Schouten (1940a) suggested drawing the listener's attention to a particular harmonic of a complex tone by eliminating and reintroducing this partial repeatedly. Cardozo (1967, 1968), Duifhuis (1970, 1972) and Gibson (1971) adopted this technique, with minor modifications, for investigating the number of identifiable harmonics.

They found for low fundamental frequencies values two to three times higher than those given above. This discrepancy is due to the difference in procedure. The procedure used for obtaining the data presented in Fig. 1 is similar to the way in which the limit of spatial resolution of the eye is measured by means of a steady-state pattern of light and dark bars. Varying the amplitude of a particular harmonic in the auditory stimulus is essentially equivalent to flickering of a particular bar in the visual field. These fluctuations may be perceptible as differences in the amount of stimulation even though the separate components of the steady-state pattern are not resolved by the sense organ. In the author's opinion, the use of a steady-state stimulus is preferable as a procedure for defining the limit of the analyzing power of the system both in vision and in audition.

In view of the limited number of identifiable harmonics it may seem paradoxical that eliminating a single harmonic with a high harmonic number ($n > 16$) from a train of periodic pulses makes this same harmonic audible. Duifhuis (1970, 1971, 1972) who studied this intriguing phenomenon, showed it to be a consequence of the limited frequency-resolving power of the ear. Such a damped filter system responds separately to the successive pulses, with "silent" intervals between the individual pulse responses, as long as the pulse rate is low compared to the filter bandwidth. Elimination of an upper harmonic of the stimulus actually means that this component is added in these silent intervals and may become audible as a pure tone.

So far, all stimuli considered have been multitone complexes. The lower limit is a sound containing only two partials. For this case as well, the question of the minimal frequency separation required to identify the two pitches can be asked. Thurlow and Bernstein (1957) and Terhardt (1968) presented subjects with pairs of simultaneous tones over a range of frequency differences and asked them whether one or two pitches were audible. However, a more objective procedure, using probe tones, would be preferable. The solid symbols in Fig. 2 represent average results for two subjects in a forced-choice procedure similar as used for the multitone stimuli (Plomp, 1964). The frequency of one probe tone coincided with one component of the stimulus, whereas the frequency of the other probe tone was either $\frac{1}{4}\Delta f$ below the higher or $\frac{1}{4}\Delta f$ above the lower component of the stimulus, with $\Delta f =$ frequency difference between the components. The results are essentially the same as found by the other authors mentioned. The diagram indicates that, particularly at low frequencies, the individual tones of a pair can be identified for substantially smaller frequency separations than the partials of a complex tone. This appears also to be true for the lowest partial of complex sounds containing four components (Moore, 1973). These findings suggest that the slope of the excitation pattern (see

Chapter 1.4), not masked by a neighbouring component, contributes to the audibility of a partial.

2. Masking pattern and critical bandwidth

One sound can prevent another sound from becoming audible. This masking effect has been used for a long time as the main technique for studying auditory frequency analysis. The magnitude of the effect is expressed in the *masked threshold*, defined as the minimal sound-pressure level of a sinusoidal probe tone required to detect this tone in the presence of a masking stimulus. The dependence of the masked threshold upon the frequency of the probe tone gives the *masking pattern* which is highest near the stimulus frequencies. We may expect that the masking pattern of a simple-tone stimulus reflects the frequency characteristic of the auditory filter.

A. THE MASKING PATTERN OF SIMPLE TONES AND NARROW NOISE BANDS

The first data on the masking pattern of simple tones were published in 1924 in a classic article by Wegel and Lane. They found, however, that the method is not ideal. For probe-tone frequencies near the masking tone, the two tones interfere in the form of *beats*. Similar phenomena occur around the small multiples of the frequency of the masker. Furthermore, over a certain range above the masker frequency, the masked threshold appears to be determined by the detection of *combination tones*, due to nonlinear interaction between masker and probe tone, rather than of the probe tone itself.

These complications are illustrated in Fig. 3 (Egan and Hake, 1950; additional data were published by Ehmer, 1959a; Small, 1959). In one experiment Egan and Hake used a masker of 400 Hz. The variable probe tone was pulsed with 0·7 sec on and 0·7 sec off and rise and fall times of 0·1 sec to avoid clicks. The subjects were instructed to adjust the sound-pressure level of the probe tone to its detectability threshold, using 'detection of anything" as a criterion. The stimuli were presented monaurally by headphone. The solid points in the diagram represent the average results in dB SL of five subjects for a masker of 80 dB SPL. The dips at 400, 800, and 1200 Hz are due to beats and the broad minimum between 400 and 800 Hz is due to combination tones. An extensive study on the role of combination tones in the masking pattern of simple tones was published by Greenwood (1971). We will return to the role and origin of combination tones and beats in Chapters 2 and 3.

Several investigators have attempted to avoid these disturbing factors by substituting a narrow band of noise for either the masker or the probe

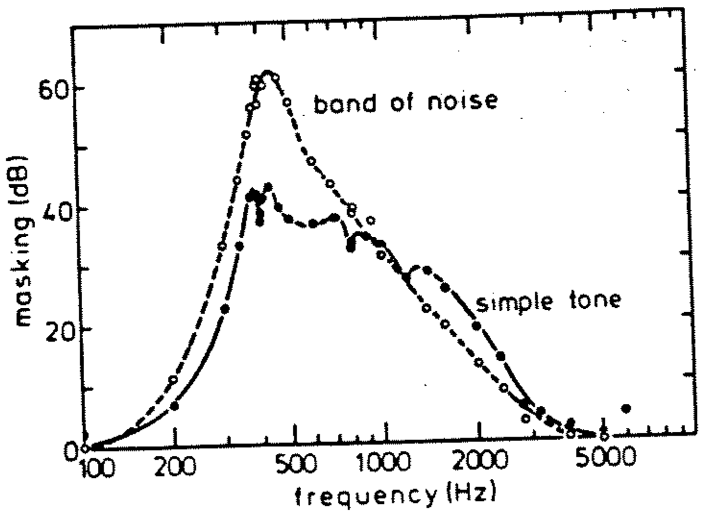

FIG. 3 Masking patterns in dB SL for a simple tone of 400 Hz (dots) and a 90-Hz band of noise (circles) with a centre frequency of 410 Hz, both at 80 dB SPL (average of five subjects). (From Egan and Hake, 1950.)

tone. Egan and Hake repeated their experiment using a 90-Hz wide band of noise with a centre frequency of 410 Hz as a masker. The masking pattern for a noise band of 80 dB SPL, averaged over the same five subjects, is also reproduced in Fig. 3. The authors pointed out that even then the probe tone occasionally "beats" with the randomly fluctuating stimulus. All subjects reported that, around 410 Hz, the probe tone was heard near the masked threshold as a "buzz" or "rattle". The authors estimated that the peak of the masking pattern is 4·2 dB lower than it would have been without beats. It is remarkable that the dashed curve does not seem to be significantly affected by audible combination products of the band of noise and the probe tone. Greenwood (1971) has shown that these products cannot be eliminated by using a noise stimulus.

Other studies in which noise bands were used as a masker include Bilger and Hirsh (1956), Chistovich (1957), Ehmer (1959b), and Maiwald (1967). They confirm that the low-frequency slope of the masking pattern is considerably steeper than the high-frequency slope. Extremely steep low-frequency slopes of more than 100 dB per octave have been found, demonstrating the high selectivity of the auditory filter (see Rodenburg *et al.*, 1974, for estimates of the slope derived from curves published by several authors, showing that substantial differences appear to exist). Whereas this slope is rather independent of the level of the masking stimulus, the high-frequency slope becomes flatter and flatter for higher sound-pressure levels. Applying very short probe-tone bursts in order to avoid beats, Vogten (1974) found that the peak of the masking pattern does not necessarily coincide with the frequency of the tonal masker. The origin of this phenomenon is not clear.

The interference of masker and probe tone can be avoided by presenting the probe tone after rather than during the masker. This procedure makes use of the fact that the masked threshold induced by a stimulus does not drop suddenly to the hearing threshold after the stimulus is switched off. As this post-stimulatory masking, nowadays usually called *forward masking*, disappears rapidly, it is necessary to use short probe-tone bursts immediately after the masker. In the past mainly the temporal aspects of forward masking have been studied. No systematic data on the masking patterns of various stimuli are available.

B. CRITICAL BANDWIDTH

A characteristic of primary importance for any filter system is its bandwidth. The masking pattern of a narrow band of noise reveals the slopes of the auditory filter but is not very appropriate for estimating the bandwidth. Zwicker (1954) applied a particular procedure for investigating this bandwidth. He studied the masking effect of two simple tones with frequencies f_1 and f_2 on a narrow band of noise with a centre frequency of $\frac{1}{2}(f_1 + f_2)$. The masked threshold was measured with the Békésy up-and-down tracking technique. This technique makes use of an attenuator, the attenuation of which decreases automatically as long as a push-button is pressed and increases as long as it is released. The subject is required to press the button if the probe sound, in this case the narrow band of noise, is audible and to release the button if the sound is not audible. The position of the attenuator is continuously recorded and this recording permits an accurate estimate of the masked threshold.

The solid curve in Fig. 4 represents the results for a masker consisting of two tones of 50 dB SPL at equal frequency distances below and above 570 Hz. The probe sound was a band of noise with a bandwidth of 30 Hz centred on 570 Hz. The diagram shows that the masked threshold is constant for frequency separations below about 130 Hz, but decreases progressively beyond that value. The other two curves demonstrate that such a critical frequency difference is not found for a masker consisting of either the lower or the higher frequency component separately. Similar measurements were performed for other frequencies and the resulting critical frequency differences for two subjects are plotted in Fig. 5. It appeared that this difference does not depend upon sound-pressure level.

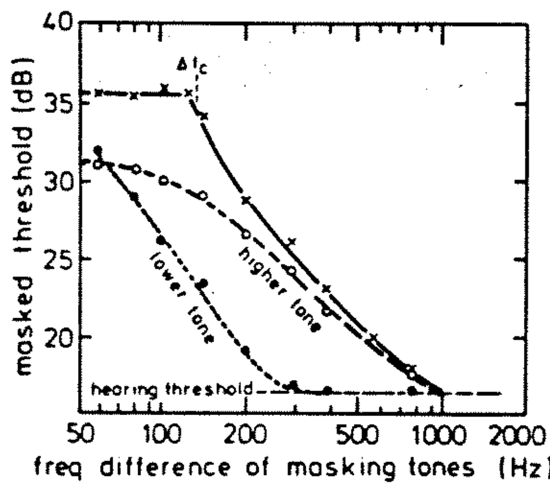

FIG. 4 Masked threshold in dB SPL of a noise band 30 Hz wide with a centre frequency of 570 Hz as a function of the frequency difference between two simple tones of 50 dB SPL at equal distances below and above 570 Hz (solid curve; one subject). The dashed curves give the masked thresholds for the simple tones separately. The value of Δf_c represents critical bandwidth. (From Zwicker, 1954.)

The curve in Fig. 5 can be considered as an estimate of the bandwidth of the auditory filter, usually called the *critical bandwidth*, as a function of frequency. Since the introduction of this term by Fletcher (1940), critical bandwidth has been found to explain a large variety of auditory phenomena. It is now accepted as a basic characteristic of hearing and we will refer to it again and again in the course of this study.

Zwicker's experiment has never been systematically repeated for a larger group of subjects. In his extensive study on masking Greenwood (1961a, 1961b) included some experiments similar to Zwicker's and found at frequencies above 1000 Hz somewhat smaller critical-bandwidth estimates. An investigation by Green (1965) with a simple tone as the probe sound resulted in wider critical frequency separations than found by Zwicker and Greenwood for a narrow band of noise. Glattke and Small (1967) used the same paradigm in forward-masking experiments. Unfortunately, they applied frequency separations too small to obtain significant results.

Finally, the question of whether the width of the critical band depends upon time should be mentioned. Whereas several investigations (Scholl, 1962; Elliott, 1965, 1967; Green, 1969; Srinivasan, 1971) seemed to suggest that critical bandwidth starts wide and becomes narrower after the masker is switched on, more recent studies (Zwicker and Fastl, 1972; Fastl, 1974a, 1974b) do not confirm that view.

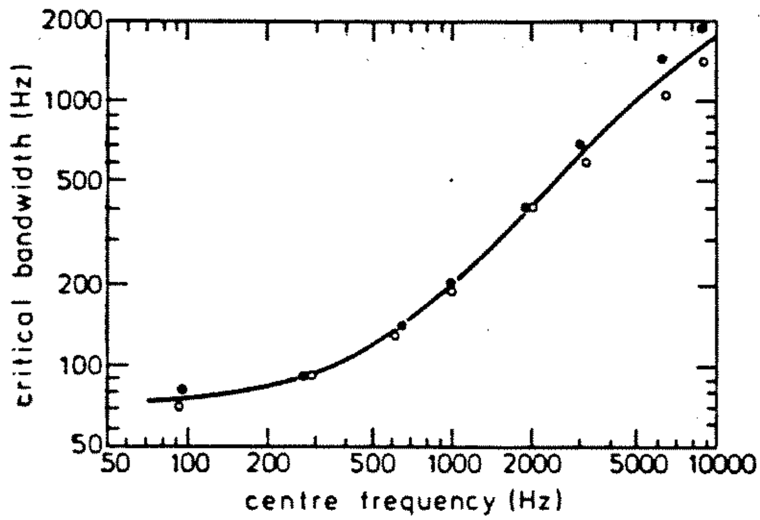

FIG. 5 Critical bandwidth, measured as in Fig. 4, as a function of frequency for two subjects. (Adapted from Zwicker, 1954.)

The concept of critical bands should not be understood as if the ear were provided with a fixed set of adjacent auditory filters. It refers merely to the fact that any device discriminating neighbouring components is characterized by its resolving power.

3. Pulsation threshold and lateral suppression

Recently, an alternative technique for exploring the selectivity of the auditory filter system has been proposed. The new measure of the ear's response to a sound obtained by this method will be referred to as the *pulsation threshold*. It is free from the main disadvantages of the preceding masking procedures by making use of a nonsimultaneous probe tone which can be substantially longer in duration than the brief tone bursts required in forward masking. Since the pulsation threshold reveals properties of the system not found in direct masking, its interpretation and relation to the masked threshold will be treated at some length.

A. AUDITORY INDUCTION AND THE PULSATION THRESHOLD

The new procedure stems from an auditory phenomenon discovered in three different places apparently independently.

In 1957 Thurlow published a short note on what he called "an auditory figure-ground effect". In exploring the effects obtained with two simple tones alternately sounding he observed that, under certain conditions, the more intense tone is heard as clearly intermittent and the less intense appears to sound continuously, in spite of the fact that it is physically absent during the bursts of the louder tone. In several subsequent studies (Thurlow and Elfner, 1959; Thurlow and Marten, 1962; Elfner and Caskey, 1965; Elfner and Homick, 1966, 1967; Elfner, 1969, 1971) the phenomenon was further explored.

Working from a quite different point of departure, another group of investigators rediscovered the effect. Warren and associates (Warren, 1970a; Warren and Warren, 1970; Warren and Obusek, 1971) observed that when an extraneous sound, such as a cough or tone, completely replaces a speech sound in a recorded sentence, listeners believe they hear the missing sound. Later on they became aware that this ability to perceptually synthesize missing phonemes can be considered as a special case of a much broader auditory phenomenon, described by them as *auditory induction*. In one of their experiments Warren *et al.* (1972) alternated a tone of 1000 Hz at 80 dB SPL with a weaker probe tone of varying frequency, each tone burst lasting 300 msec. They requested subjects to adjust the intensity of the probe tone to the highest level at which this tone seemed continuous. The general shape of the resulting curve resembled the masking pattern measured for the same subjects. From this and several further experiments the authors drew up the following rule: "If there is contextual evidence that a sound may be present at a given time, and if the peripheral units stimulated by a louder sound include those which would be stimulated by the anticipated fainter sound, then the fainter sound may be heard as present."

At the same time Houtgast (1972a) studied the suggested occurrence of "Mach bands" in hearing by investigating the detectability threshold of short probe-tone bursts presented in the gaps between repeated bursts of the masking noise band. He noticed that, at low levels relative to the masker, the bursts of the probe tone are not heard individually but as a tone continuing through the noise bursts. At high levels, however, the tone bursts are heard as a pulsating tone. From his experiments Houtgast concluded that the *pulsation threshold*, defined as the maximal level at which the probe tone still sounds continuous, is an appropriate measure for studying the ear's filter characteristic. In order to avoid misunderstandings, it must be emphasized that the pulsation threshold refers to the sensation of *pulsation* and does not refer to the detectability of the probe tone as such. Below pulsation threshold the tone is audible as a continuous tone rather than as a repetition of tone bursts.

Fig. 6 represents the frequency dependence of the pulsation threshold for nine different tonal stimuli (Houtgast, 1973). Each stimulus, with constant frequency and level given by the triangular symbols in the diagram,

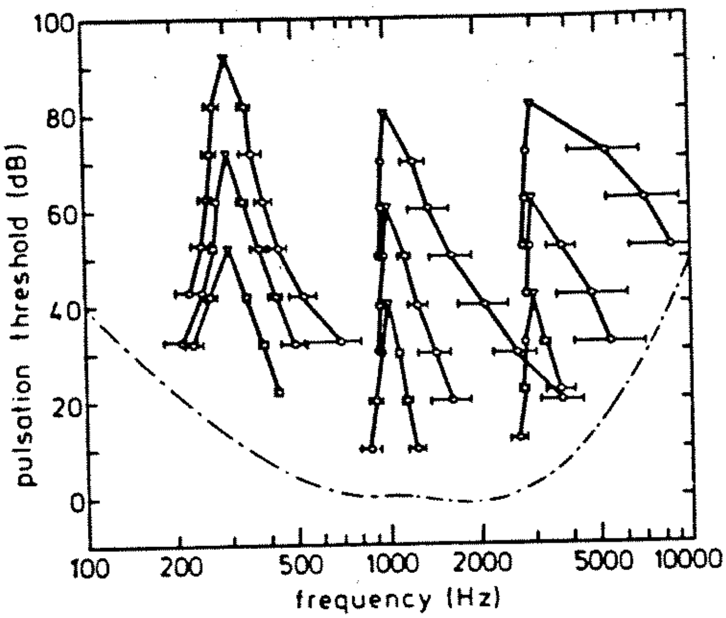

FIG. 6 Pulsation threshold as a function of frequency of the probe tone for tonal stimuli with frequency and level given by the triangular symbols. The horizontal dashes give the standard deviations for 12 ears in 7 subjects. The dashed curve represents the hearing threshold. (From Houtgast, 1973.)

was alternated four times per sec with the probe tone at various levels in steps of 10 dB. To avoid transition clicks, the tone bursts had a rise time and a decay time of 15 msec. The tones were presented monaurally by headphone at fixed sound-pressure levels. The subject was required to adjust the *frequency* of the probe-tone bursts until maximally different from the stimulus frequency without losing the impression of continuity. The horizontal dashes give the standard deviations for 12 ears in 7 subjects. The asymmetric shape of the curves resembles the masking pattern of a narrow band of noise (Fig. 3). The rather flat high-frequency slopes for stimuli of 3000 Hz approximate the slopes of the 1000-Hz curves if the pulsation threshold is replotted relative to the hearing threshold (dashed line). It is clear from this diagram that pulsation threshold, not influenced by combination tones and beats, represents an attractive measure for demonstrating the frequency-analyzing power of the hearing mechanism. It has been verified (Verschuure *et al.*, 1976a) that, without silent gaps between the alternate probe-tone and the stimulus bursts, the pulsation threshold is almost identical for burst durations between 100 and 200 msec.

For durations longer than this optimal range, the continuity effect loses its significance more and more, for shorter durations the bursts suffer more and more from spectral spread of energy.

The extent to which the harmonics of a complex tone are resolved by the ear is illustrated in Fig. 7 (Houtgast, 1974a). The first seven or eight harmonics of the stimulus with a fundamental of 250 Hz manifest themselves as peaks in the pulsation-threshold pattern.

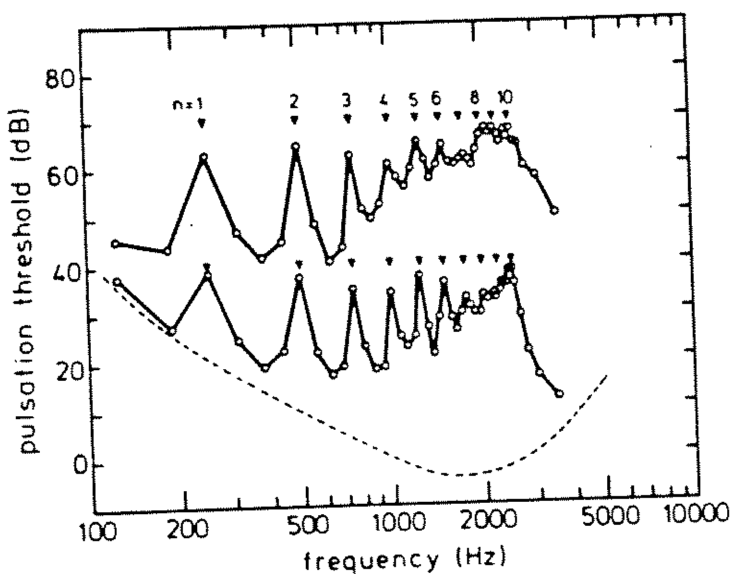

FIG. 7 Pulsation-threshold patterns of two complex tones consisting of the first ten harmonics of 250 Hz (average data of three subjects). The sound-pressure levels of the harmonics are given by the triangular symbols. The dashed curve represents the hearing threshold. (From Houtgast, 1974a.)

Whereas the general shape of the pulsation-threshold pattern for a single pure tone does not differ essentially from the masking pattern, the curves of Fig. 7 show an unexpected peculiarity. For some components the pulsation threshold of a probe tone coinciding in frequency with the harmonic is substantially lower than the latter's sound-pressure level (difference between triangular and circular symbols). With this we have encountered what may be an auditory manifestation of the general neural process of *lateral inhibition*. As, however, we should not exclude the possibility that (nonlinear) cochlear mechanics plays a part in this effect (see Chapter 2.4), we will refer to it by the more neutral term of *lateral suppression*. It is an important constituent of the ear's filter characteristic and we will consider it in more detail.

B. LATERAL SUPPRESSION

Until recently no direct psychophysical evidence for lateral suppression in hearing had been found. Indications that it might exist were provided by von Békésy (1963a) who observed that in an octave band of noise distinct pitches corresponding to the lower and upper cutoff frequencies of the bandfilter can be heard. For a pass band of 400–800 Hz subjects were able to match a probe tone rather accurately to these cutoff frequencies. Von Békésy considered this as demonstrating that, like the Mach bands in vision (Ratliff, 1965), the edges of a band of noise are emphasized by contrast phenomena. This would mean that lateral suppression plays a part in hearing just as it does in other senses. More extensive experiments by other investigators (Small and Daniloff, 1967; Rakowski, 1968; Rainbolt and Schubert, 1968; Fastl, 1971; Glave, 1973) showed under which conditions the "edge pitch" is most easily observed.

This rather indirect support for possible lateral suppression in hearing led Carterette *et al.* (1969, 1970) to trace whether similar edge effects are present in the masking pattern of noise bands. Their affirmative conclusion was strongly criticized by Greenwood (1970) on the basis of his own extensive data on masking and by Rainbolt and Small (1972; see also Small, 1975) who replicated the experiments with negative results.

Apparently, if Mach bands in hearing do exist, simultaneous masking is not the right method for revealing these bands. Houtgast (1971, 1972a) considered nonsimultaneous masking more appropriate to answer this question. Briefly formulated, his argument ran as follows: if not only the masking noise band but the probe tone, too, is subjected to the suppression process, the masking contour will not show up the effect; hence, one should use a nonsimultaneous rather than a simultaneous probe tone to obtain a positive result.

In order to verify this assumption Houtgast performed experiments in which he compared the results for a probe tone presented either simultaneously with the masking stimulus or nonsimultaneously during pauses between stimulus bursts.

The first experiment (Houtgast, 1972a) dealt with the question, raised by Carterette and his co-workers, whether Mach bands in hearing exist or not. The stimulus consisted of high-pass or low-pass filtered noise. With a nonsimultaneous probe tone, presented as brief tone bursts during 50-msec gaps in the stimulus, the forward-masking pattern was characterized by a distinct peak near the cutoff frequency of the noise band. Subsequent experiments (Houtgast, 1974a) showed a similar effect in the pattern of the pulsation threshold. This effect was absent in direct masking, demonstrating that a nonsimultaneous probe tone is essential for revealing the existence of an edge-enhancement mechanism in audition.

If such a mechanism does exist, one should expect that a tone is able to suppress to a certain extent another simultaneous tone in its neighbourhood. Similar effects have been demonstrated electrophysiologically at the level of the cochlear nerve (Sachs and Kiang, 1968; Arthur *et al.*, 1971). A preliminary experiment (Houtgast, 1971, 1972a) showed that the suppression of one tone by another can be also demonstrated psychophysically. In a subsequent experiment (Houtgast, 1973, 1974a) the measurement was repeated with five different experimental paradigms, indicated in the panels of the left column of Fig. 8. The stimulus (shaded areas) consisted of a constant weaker tone with $f_1 = 1000$ Hz and $L_1 = 40$ dB SL and a stronger suppressing tone with f_2 varying from 1000 up to 2500 Hz and $L_2 = L_1 + 20$ dB. The probe tone with $f_p = f_1 = 1000$ Hz was either presented during the stimulus (direct masking, procedures 1 and 2), immediately after the stimulus (forward masking, procedures 3 and 4), or alternately with the stimulus (pulsation threshold, procedure 5). The bursts of the stimulus and the probe tone were switched on and off smoothly with rise and fall times of about 20 msec to avoid clicks. In the upper four diagrams of the right column the curves represent the detectability threshold of the probe tone, in the lowest diagram the curve represents the pulsation threshold. In procedure 5 the stimulus component f_1 is always audible, except for f_2 very near to f_1, and the pulsation threshold is the level at which the alternating probe tone and the weaker tone of the stimulus, both fixed at 1000 Hz, are heard as a single continuous tone. Thus in this case the interpretation of the pulsation threshold is much simpler than for cases in which the stimulus does not contain a component coinciding in frequency with the probe tone. The diagrams in the right column show that, for direct masking, the detectability threshold of the probe tone increases monotonically for f_2 approaching 1000 Hz. However, when the probe tone is not simultaneous with the stimulus, the curves first drop to a distinct minimum at about 1150 Hz. This means that, for a stimulus of 1000 Hz, both the detectability threshold in forward masking and the pulsation threshold at 1000 Hz are considerably reduced if a loud tone of 1150 Hz is added. As the lowest diagram indicates, the reduction of a 40-dB tone can be up to 20 dB.

Fig. 9 illustrates how lateral suppression of a tone of 40 dB at 1000 Hz depends upon the level of the suppressing tone. The asymmetry of the curves demonstrates that a tone higher in frequency suppresses a lower tone substantially more strongly than conversely.

We may conclude from these experiments that the suppression of one

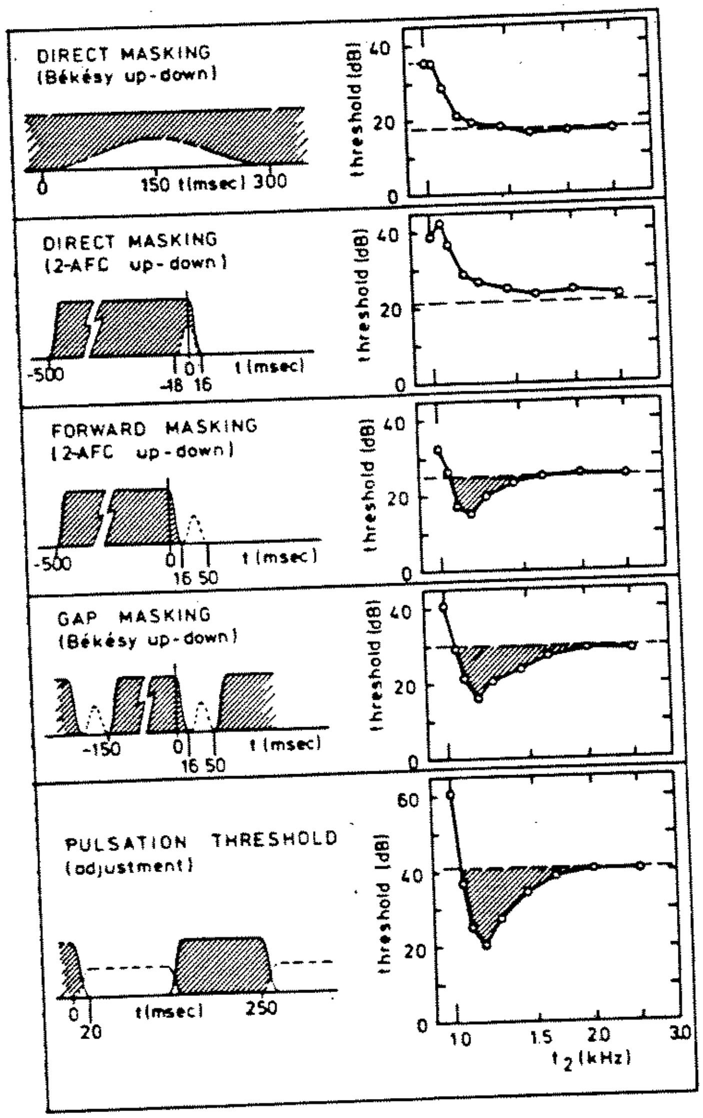

FIG. 8 Suppression of a tone of 1000 Hz at 40 dB SL by a 20 dB stronger tone of frequency f_2, varying from 1000 up to 2500 Hz. In the left column the shaded areas represent the two-tone stimulus and the dashed lines the 1000-Hz probe tone. The corresponding diagrams show the results for the five experimental procedures (average of two subjects). (Adapted from Houtgast, 1974a.)

tone by another becomes manifest if the probe tone and the stimulus are not simultaneous. In this respect the pulsation threshold appears to be similar to the detectability threshold in forward masking. Additional experimental evidence confirming this conclusion was recently provided by Shannon (1976). I will return to the intriguing question of why lateral suppression is absent in direct masking in Chapter 1.4.

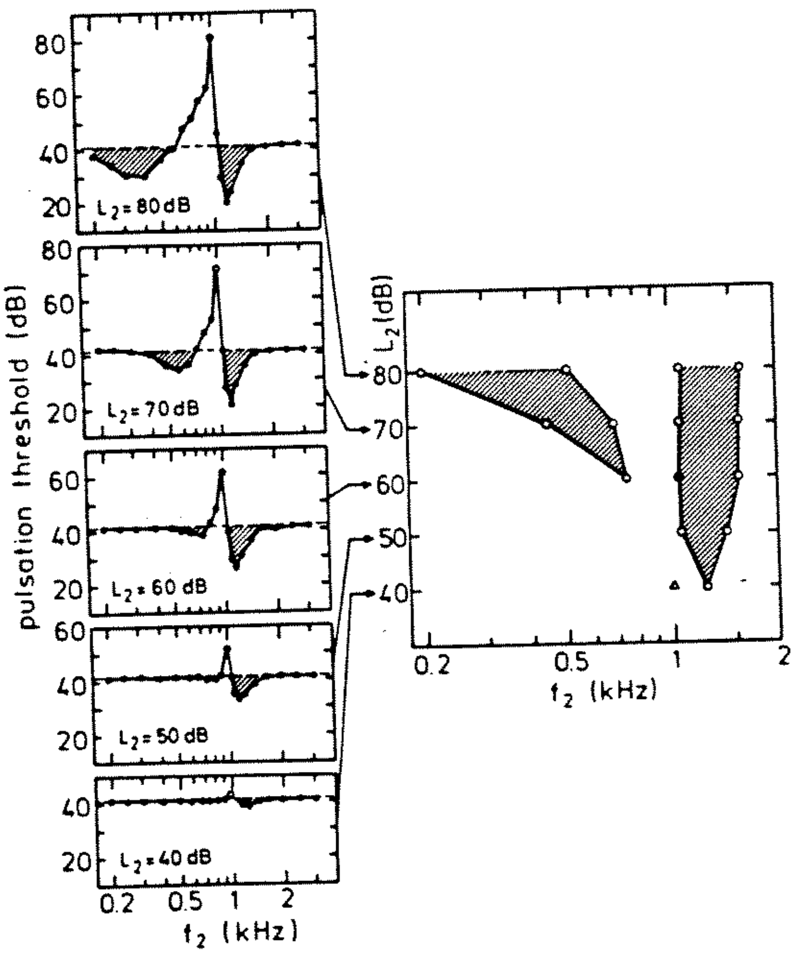

FIG. 9 Dependence of the amount of lateral suppression upon the frequency (abscissa) and level (parameter) of the suppressing tone (average data of seven subjects). The suppressed tone was 1000 Hz at 40 dB SL. The dashed areas in the right diagram give the frequency and level ranges over which more than 3 dB suppression was found. (From Houtgast, 1974a.)

C. SHAPE AND BANDWIDTH OF THE AUDITORY FILTER

Both the edge effect of noise bands and the tone-on-tone suppression suggest that the ear's frequency-analyzing mechanism is much more complicated than a simple filter model would suggest. Houtgast (1972b, 1974a) determined the filter characteristics for a simultaneous and for a nonsimultaneous probe tone using "rippled", or "comb-filtered", noise as the stimulus. The upper part of Fig. 10 shows the sinusoidally shaped power spectra of stimuli with an intensity difference between the peaks and valleys of 11 dB. For a linear filter, the output intensity is given by this noise spectrum weighted according to the frequency response of the filter. Assuming that the auditory filters are linear (which holds only as a first-order approximation), the response characteristic of a particular filter can be calculated from the variation of the overall intensity of the noise stimulus for a progressively increasing modulation frequency, required to keep the intensity at the output of the filter constant. As a measure of this constancy the detectability threshold of a simultaneous probe tone of constant intensity, or its pulsation threshold for a nonsimultaneous probe tone, can be taken. It suffices to measure the threshold difference for successive peaks and valleys, and for successive maximal positive and negative slopes, coinciding with the probe-tone frequency (conditions marked as a–d in Fig. 10). As an example, the middle diagrams of Fig. 10 give the results for a probe tone of 1000 Hz, and the lower diagrams the filter characteristics calculated from these data.

It is apparent that the two procedures give essentially different frequency-response curves for the auditory filter. The direct-masking data reveal a relatively "simple" filter with a bandwidth of about 210 Hz, the pulsation-threshold data a substantially sharper filter with a negative part at the high-frequency side. This negative part implies that the output of the 1000-Hz filter will be reduced by adding a strong tone of about 1150 Hz, precisely what we have seen in Fig. 8. The absence of a second negative minimum at the low-frequency side, required to explain the less pronounced suppression effect of a lower tone (Fig. 9) may be due to the fact that in the calculations a linear system was assumed.

The bandwidth difference between the two experimental conditions was further explored with a somewhat simplified procedure in which bandwidth was calculated from the threshold differences at peaks and valleys for only four modulation frequencies. Fig. 11 presents the results for five probe-tone frequencies and five subjects. The bandwidths derived from the pulsation-threshold data are roughy half the bandwidths derived from the masking data.

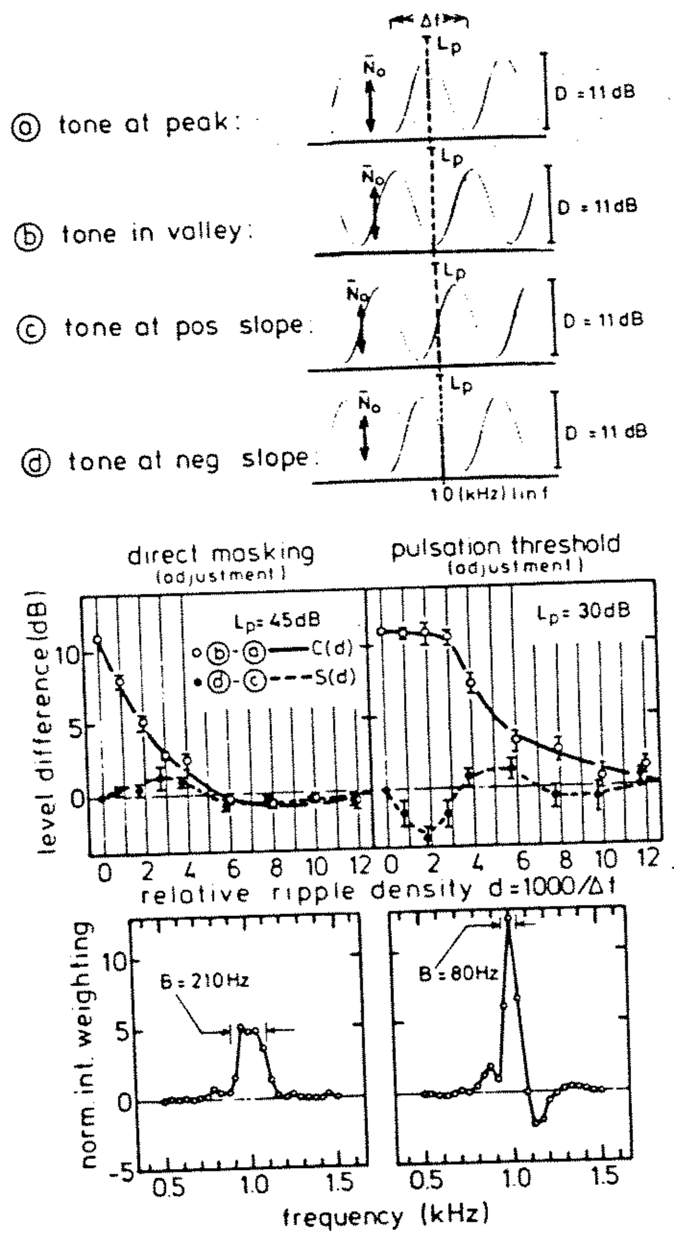

FIG. 10 Illustration of how the frequency-response characteristic of the 1000-Hz auditory filter can be determined using rippled-noise stimuli. The upper part shows the sinusoidally modulated noise spectra for one modulation frequency. The solid curve of the left diagram in the middle

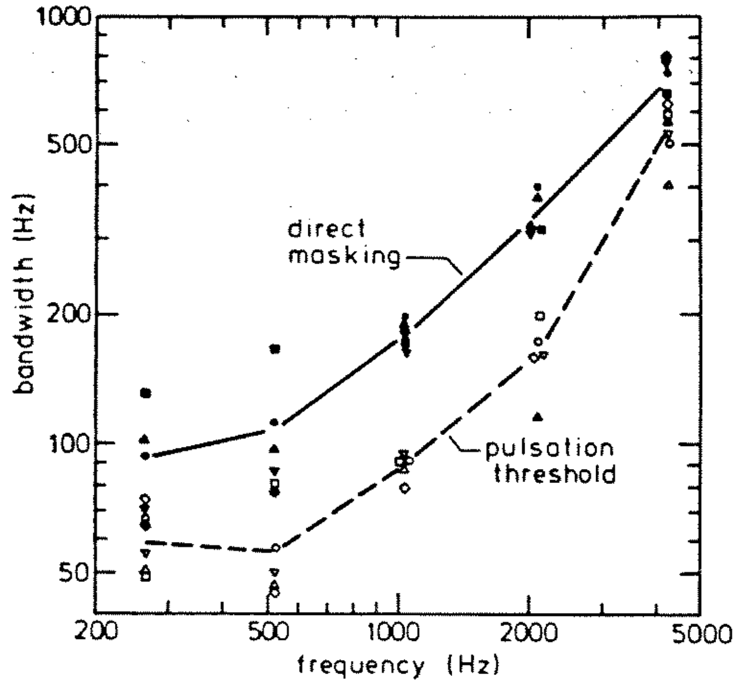

Fig. 11 Bandwidth of the auditory filter as a function of frequency, calculated from (1) the detectability threshold in direct masking (solid symbols), and (2) pulsation threshold (open symbols), with rippled noise as the stimulus (five subjects). The curves connect the average values. (Adapted from Houtgast, 1974a, with use of the individual data.)

4. Discussion

In the previous sections various procedures for investigating auditory frequency analysis were treated. The experimental results clearly indicate

---

represents the difference in noise levels required to mask a simultaneous probe tone fixed at 1000 Hz and 45 dB coinciding with the peak and the valley, respectively, of the noise, plotted as a function of the number of peaks in the frequency range 0–1000 Hz (condition b minus condition a). The dashed curve is the level difference for probe tones coinciding with the maximally positive and negative slopes, respectively (condition d minus condition c). Similarly, the right diagram represents the differences in noise level required to perceive a nonsimultaneous probe tone fixed at 1000 Hz and 30 dB as a continuous tone (data of one subject; the vertical dashes give the standard deviations). The lower diagrams represent the filter characteristics calculated from these data. (Adapted from Houtgast, 1974a.)

that such an analyzing mechanism does exist. However, they also demonstrate that we cannot speak unequivocally about "the" frequency-resolving power of the ear. The use of a probe tone nonsimultaneously presented with the stimulus reveals the effect of lateral suppression, not found for a simultaneous probe tone. In order to arrive at a coherent picture of the ear as a frequency analyzer we need an explanation for this different behaviour. It seems to me that a clearer view on the role of the probe tone in the experiments may help in understanding this difference.

The implications of using probe tones for exploring the properties of the auditory filter are not always fully realized by investigators applying them. Because the probe tone is subjected to the same processes as the stimulus, one has to be very careful in interpreting the resulting data. This is evident if adding a probe tone to the stimulus results in audible interference phenomena such as combination tones and beats, but also holds when no interference seems to occur.

By investigating the masking (or pulsation) pattern of a stimulus with the aid of a probe tone we attempt to scan the spatial distribution of neural activity along a scale on to which frequency is projected (say, along the cochlear partition). This activity pattern will be called the *excitation pattern*. We do not need to know by which physiological parameters this pattern is characterized if we use the physical variables of a probe tone scanning the excitation pattern as a measure of it. However, since the probe tone is subjected to the same imperfect frequency analysis, it is not simply true that the (detectability or pulsation) threshold of the probe tone as a function of frequency reflects the excitation pattern of the stimulus. An essential condition for obtaining a reliable picture of the excitation pattern of the stimulus is that the corresponding pattern for the probe tone is substantially sharper than for the stimulus so that the threshold of the probe tone is determined by its peak rather than by one of its shoulders or tails.

Fig. 4 may indicate that this condition is not generally fulfilled. The diagram shows that the masked threshold for two simultaneous tones is more elevated than one would expect from the thresholds for the tones separately. A possible explanation is that, due to the width of the excitation pattern of the probe sound, the condition for detecting this sound is most favourable at a place whose frequency is slightly off the nominal value. This effect may occur if the probe tone is situated somewhere on the tail of a one-tone masker, but can be neglected if it is situated between the components of the two-tone masker.

Let us now return to our main problem: why a nonsimultaneous probe tone does, and a simultaneous probe tone does not reveal any effect of lateral suppression. The explanation follows Houtgast's most recent exposition (Houtgast, 1974a). We consider, successively, the experimental findings on the detectability threshold in forward masking, the pulsation threshold for a probe tone sounding alternately with the stimulus, and the detectability threshold in direct masking.

(*a*) *Forward masking*. Whatever the origin of forward masking may be, it is reasonable to assume that, if all other parameters are kept constant, the threshold measured with this procedure is monotonically related to the excitation induced during stimulation. This implies that, since forward masking reveals the effect of lateral suppression, we may accept that the excitation pattern of the stimulus, too, includes this effect.

(*b*) *Pulsation threshold.* In all experiments considered, the shape of the pulsation-threshold curve as a function of frequency resembles the detectability-threshold curve in forward masking. Therefore, Houtgast's (1971, 1972a) hypothesis, proposed independently of the similar rule put forward by Warren *et al.* (1972) and quoted above, seems to be justified: "When a tone and a stimulus S are alternated (alternation cycle about 4 Hz) the tone is perceived as being continuous when the transition from S to tone causes no (perceptible) increase of nervous activity in any frequency region. The pulsation threshold, thus, is the highest level of the tone at which this condition still holds." If this hypothesis is correct, and if the condition of a relatively sharp excitation pattern of the probe tone is fulfilled, then the pulsation-threshold pattern is a direct measure of the excitation pattern induced by the stimulus, expressed in the physical parameters of the probe tone. Since the slope of the curves at the low-frequency side in Fig. 6 appears to increase with sound-pressure level, the condition of a very locally operating probe tone is violated here. As Verschuure *et al.* (1976b) have shown, the extremely steep slopes at high levels may be understood as a consequence of this phenomenon. Houtgast (1974d) found that for narrow bands of noise the low-frequency slope of the pulsation-threshold pattern is steeper than that of the masking pattern, as might be expected from the different filter characteristics of Fig. 10. Fastl (1974a), however, did not observe this difference. This discrepancy in experimental results remains unexplained.

(*c*) *Direct masking*. A very simple model can account for the fact that, unlike the previous procedures, direct masking does not demonstrate lateral suppression. If, in input terms, lateral suppression results from a reduction of the gain factor in an adjacent frequency region, it will affect both the masker and the probe tone in this region, leaving the signal-to-noise ratio ("probe-to-masker" ratio) unchanged. Since the detectability threshold in direct masking depends almost exclusively upon signal-to-noise ratio, lateral suppression will have no effect on the detectability threshold of the probe tone. Hence, a simultaneous probe tone is unsuitable for revealing lateral suppression. The masking pattern obtained with such a probe tone reflects the excitation pattern without the effect of lateral suppression. The finding that lateral suppression does manifest itself by using a nonsimultaneous probe tone implies that its effect disappears (almost) immediately after the suppressing sound has been switched off.

This explanation should be considered as a tentative and rather general model. It does not discriminate between a mechanical or neural origin of lateral suppression.

Only after this discussion of the seeming contradiction between the data obtained with a simultaneous and a nonsimultaneous probe tone, does it make sense to compare the bandwidth estimates for the various procedures described in this chapter. The main results of Figs. 2, 5, and 11 are replotted in Fig. 12.

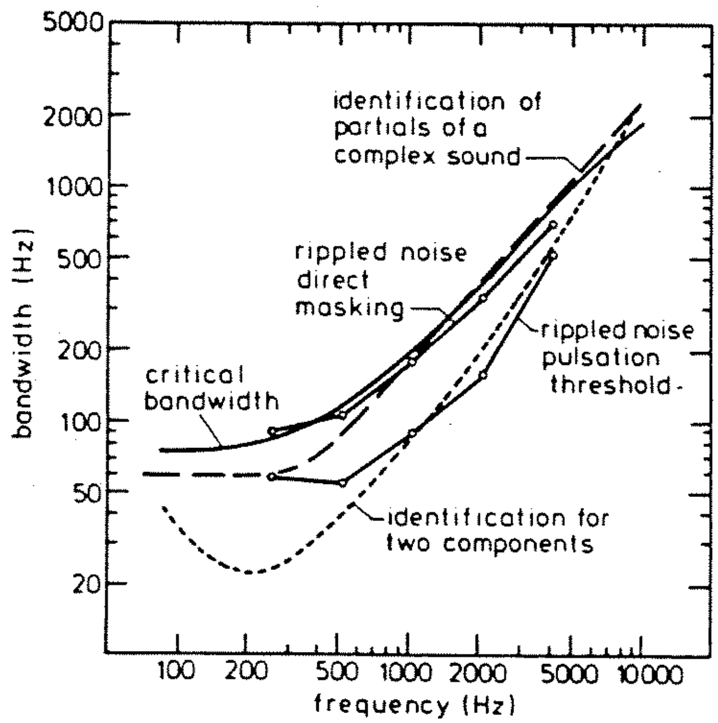

FIG. 12 Comparison of the various bandwidth estimates of the auditory filter from Figs. 2 (dashed curves), 5 (solid curve), and 11 (open symbols).

We see from this diagram that the bandwidth values calculated by Houtgast from his direct-masking measurements for rippled noise agree excellently with the critical-bandwidth data found by Zwicker also in direct masking. The bandwidth estimates based upon the pulsation threshold for rippled noise are about half these values. Our ability to identify a particular component of a multitone stimulus depends upon the peak-to-valley ratio required for distinguishing the corresponding peak in the excitation pattern of the stimulus. In view of this and other uncertain factors, we may conclude that the data points do not differ more than can be expected from quite different experimental procedures.

5. Conclusions

The main findings arrived at in this chapter can be summarized as follows:

(1) For complex sounds with equal amplitudes of the components, the ear is able to identify these partials as long as their frequencies are separated by more than 15 to 20% (first 5 to 7 harmonics of a complex tone), with a minimal frequency distance of about 60 Hz.

(2) This limit agrees well numerically with critical-band measures based on the minimal frequency difference of two tones giving separate peaks in the masking pattern with a simultaneously presented probe sound.

(3) The pulsation threshold, equal to the highest level at which probe-tone bursts still sound as a continuous tone when the probe tone is alternated with the stimulus, is an appropriate tool for scanning the excitation pattern of a simple or complex tone.

(4) A probe tone presented nonsimultaneously with the stimulus (forward-masking threshold and pulsation threshold) reveals lateral suppression indicative of lateral inhibitory interactions in the auditory system similar to those giving rise to Mach bands in vision.

(5) The existence of lateral suppression implies that we cannot speak unequivocally about "the" auditory bandwidth. The bandwidth estimates calculated from experiments with a nonsimultaneous probe tone are about half the critical bandwidths as found for a simultaneous probe tone.