Symposium
on
HEARING THEORY 1972

Introductions by

F.A. Bilsen
G. van den Brink
H. Duifhuis
E.F. Evans
H. Fastl
J.L. Goldstein
L.M. Grobben
T. Houtgast
P.I.M. Johannesma
H.R. de Jongh
P.J.J. Lamoré
A. Møller
B.C.J. Moore
A. Rakowski
M. Rodenburg
B. Scharf
M.R. Schroeder
C.R. Steele
E. Terhardt
I.C. Whitfield
F.L. Wightman
J.P. Wilson
E. Zwicker

Delft
Rotterdam
Eindhoven
Keele
München
Boston
Utrecht
Soesterberg
Nijmegen
Amsterdam
Rotterdam
Stockholm
Reading
Warszawa
Rotterdam
Boston
Göttingen
Stanford
München
Birmingham
San Diego
Keele
München

HELD AT THE
INSTITUTE FOR PERCEPTION RESEARCH - INSTITUUT VOOR PERCEPTIE ONDERZOEK

INSULINDELAAN 2 EINDHOVEN HOLLAND
Symposium Committee:

B.L. Cardozo, Eindhoven, Secretary
E. de Boer, Amsterdam
R. Plomp, Soesterberg
PREFACE

Hearing theory is a meeting place of researchers with various scientific backgrounds. A number of workers in this field enthousiastically agreed upon presenting preprint material for a small Symposium which was to be held 22 and 23 June 1972 in the Institute of Perception Research, Eindhoven.

In this material several themes are discussed, e.g.

- The mechanics of the cochlea
- Coding of frequency and time information in the auditory pathway
- Inferences from the frequency and time characteristics of masking
- Relations between binaural hearing and pitch perception
- The pitch of complex tones

This collection of 22 preprints is not to be regarded as formal proceedings of the Symposium. It is incomplete in that not all preprints were available at the moment of printing. Moreover, it is essentially incomplete because the discussions are lacking.

It is a pleasure to thank the authors for their willingness to prepare the preprints according to the Symposium format in a very brief period of time. The wholeharted cooperation of co workers at the IPO is gratefully acknowledged. A special word of thanks is adressed to Jan F. Schouten without whose support this Symposium would not have been held.

Ben L. Cardozo

Eindhoven, May 29th 1972
## CONTENTS

<table>
<thead>
<tr>
<th>Page</th>
<th>Author</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>F.A. Bilsen</td>
<td>Pitch of Dichotically Delayed Noise</td>
</tr>
<tr>
<td>9</td>
<td>G. van den Brink</td>
<td>The Influence of Fatigue upon the Pitch of Pure Tones and Complex Sounds</td>
</tr>
<tr>
<td>16</td>
<td>H. Duifhuis</td>
<td>Peripheral Aspects of Non-Simultaneous Masking</td>
</tr>
<tr>
<td>27</td>
<td>E.F. Evans</td>
<td>Does Frequency Sharpening Occur in the Cochlea?</td>
</tr>
<tr>
<td>35</td>
<td>H. Fastl</td>
<td>Temporal Effects in Masking</td>
</tr>
<tr>
<td>42</td>
<td>L.M. Grobben</td>
<td>Pitch and Power Spectra of Short Tone Pulses</td>
</tr>
<tr>
<td>50</td>
<td>T. Houtgast</td>
<td>Psychophysical Experiments on Grating Acuity</td>
</tr>
<tr>
<td>58</td>
<td>P.I.M. Johannesma</td>
<td>The Pre-Response Stimulus Ensemble of Neurons in the Cochlear Nucleus</td>
</tr>
<tr>
<td>70</td>
<td>H.R. de Jongh</td>
<td>About Coding in the VIIIth Nerve</td>
</tr>
<tr>
<td>78</td>
<td>L.J.J. Lamoré</td>
<td>Perception of Octave Complexes</td>
</tr>
<tr>
<td>90</td>
<td>A.R. Möller</td>
<td>Coding of AM and FM Sounds in Cochlear Nucleus</td>
</tr>
<tr>
<td>96</td>
<td>B.C.J. Moore</td>
<td>Audibility of Partial in a Complex Tone, in Relation to the Pitch of the Complex as a Whole</td>
</tr>
</tbody>
</table>
105 A Rakowski
Direct Comparison of Absolute and Relative Pitch

109 M. Rodenburg
Intensity Discrimination of Noise Bands as a Function of Bandwidth and Duration

115 B. Scharf
Frequency Selectivity and Sound Localization

123 M.R. Schroeder
An Integrable Model for the Basilar Membrane

135 C.R. Steele
Analysis of Fluid-Elastic Interaction in Cochlea

142 E. Terhardt
Frequency and Time Resolution of the Ear in Pitch Perception of Complex Tones

154 I.C. Whitfield
The Relation of the Medial Geniculate Body to the Tonotopic Organization of the Auditory Pathway

161 F.L. Wightman
Pitch as Auditory Pattern Recognition

172 J.P. Wilson and J.R. Johnstone
Capacitive Probe Measures of Basilar Membrane Vibration

182 E. Zwicker
Investigation of the Inner Ear of the Domestic Pig and the Squirrel Monkey with Special Regard to the Hydromechanics of the Cochlear Duct

186 J.L. Goldstein
Evidence from Aural Combination Tones and Musical Tones against Classical Periodicity Theory

209 List of Participants
PITCH OF DICHOTICALLY DELAYED NOISE

Frans A. Bilsen

Research Laboratory of Electronics, M.I.T., Cambridge, Mass.

1. Introduction.

Continuous noise (or any other appropriate sound) presented to one ear and the same noise delayed to the other ear give rise to the following sensations. For a delay shorter than (roughly spoken) 2 ms, the noises fuse and a single noise image is perceived whose position depends on the delay. As the delay $\tau$ increases beyond this bound, the noise image remains at one side of the head, but becomes more diffuse.

Recently, we have observed that, in addition to this increase in diffuseness, a faint but distinct pitch image corresponding to $1/\tau$ appears in the middle of the head.

In view of the close analogy that exists between this pitch phenomenon and monotic repetition pitch (MRP), produced by noise added to its delayed version in the same ear (Bilsen, 1970), we shall for convenience refer to the former as dichotic repetition pitch (DRP).

Although DRP is fainter than MRP, both have equal subjective pitch and timbre qualities. There is, however, a significant difference in existence region; MRP has been reported for $1 < \tau < 10$ ms, whereas DRP exists for roughly $\tau > 3$ ms.

2. Experiments

Pitch matching experiments by five subjects, using wide-band white noise as well as narrow-band white noise as basic stimuli, were performed to explore the characteristics of DRP in more detail. A subject, who heard the signal by headphones at a sensation level of about 25 dB (in a silent anechoic room), was free to follow two possible matching procedures. Either he was allowed to make an $\text{MRP}(\tau_m)$ equal to a $\text{DRP}(\tau_d)$ by adjusting the (monotic) delay $\tau_m$, for a fixed
PITCH OF DICHTOTICALLY DELAYED NOISE

(dichotic) delay $\tau_d$ (see Fig. 1). Or, in addition, he might use an MRP($\tau_o$) and a DRP($\tau_o$) as a fixed reference. In the latter case, he was matching a musical MRP interval against a musical DRP interval. Control experiments were performed with a pure tone (period $\tau_m$ and $\tau_o$) as a matching stimulus.

The results of individual pitch matchings are represented in Figs 2 and 3 for wide-band noise (white noise with high-cutoff 2000 Hz). When the (un)delayed noise is phase inverted, the pitch, DRP, deviates significantly from the pitch, DRP', for equal polarity of the undelayed and delayed noise. In general, two pitches can be perceived, one a little higher, the other a little lower than $1/\tau_d$ (ambiguity of pitch).

Measured points for narrow-band white noise (third octave with center frequency $f_0$) are represented in Fig.4 in normalized form. Here DRP($\tau_d, f_0$)/DRP($\tau_d, f_0$) is plotted as a function of $n (= f_0 \tau_d)$, for several values of $f_0$. Note that DRP($\tau_d, f_0$) is always equal to $1/\tau_d$.

With good approximation the results can be represented by the following empirical formulas (solid lines):

$$\text{DRP}_- (\tau_d) = 1/(\tau_d \pm 0.0008) ; \text{DRP}_- (\tau_d, f_0) = 1/(\tau_d \pm 1/(2f_0)).$$

The wide-band DRP values may be related to the narrow-band DRP values by assuming the existence of a dominant spectral region (c.f. Bilsen, 1970, for MRP). This region is found by equating the two expressions; thus,

$$f_0 (\text{dominant}) = 1/(2 \times 0.0008) = 625 \text{ Hz}.$$

It is noteworthy that this is approximately the frequency region for optimal binaural beats (Licklider et al., 1950).

Additional experiments with multiple-source dichotic stimuli (see Figs 1, 5 and 6) show that the DRP phenomena are subjectively similar and probably involve the same binaural mechanisms as the FP phenomena studied by Fourcin (1970). The principal new finding is that pitch can be evolved by a single dichotically-presented source.

3. Conclusions and speculations

Because DRP signals do not provide the cochlea with spectral information, given the essential independence of the two cochlea, timing information must be used in the creation of a central pattern of neural activity from which pitch is extracted.

Houtsma and Goldstein (1972) have supplied evidence that musical pitch of complex tones is mediated by a central processor operating on neural signals derived from those effective stimulus partials that are
PITCH OF DICHOTICALLY DELAYED NOISE

BILSEN

Tonotopically resolved. Thus, parsimony would require the neural activity pattern pitch is extracted from to resemble a "central spectrum".

Within the framework of binaural mechanisms that effectively add the cochlear outputs separately for resolved frequency bands, like those postulated by Durlach (1970) and Colburn (1969) in their models for binaural signal detection, the central spectrum should be a cosine-like function of frequency for $\text{DRP}_+$, or a sine-like function for $\text{DRP}_-$. Compare BMLD patterns for dichotically delayed noise.

In particular, this can explain the similarity between the narrow-band MRP behavior (Bilsen, 1970) and the corresponding DRP behavior as expressed by the empirical formulas.

Consideration of how pitch is extracted from the central spectrum leads to questions of place- or time-pattern processing, partly, like those that arise in monaural pitch (de Boer, 1956; Schouten et al., 1962; Ritsma, 1970; Whitfield, 1970; Bilsen, 1970).

Acknowledgements

The stimulating discussions with Julius Goldstein in particular, Nathaniel Durlach and Steven Colburn highly contributed to the initiation and progress of this research. My 7 months visit at M.I.T. was supported by the Netherlands Organisation for the Advancement of Pure Research (Z.W.O.).

Present address: Appl. Phys. Dept, Delft University of Technology.

Literature


Fig. 1 Stimulus configurations

Fig. 2 Wide-band DRP+ matches

Fig. 3 Wide-band DRP− matches

Fig. 4 Narrow-band DRP− matches

Fig. 5 Wide-band FP+ matches

Fig. 6 Narrow-band FP+ matches
THE INFLUENCE OF FATIGUE UPON THE PITCH OF PURE TONES AND COMPLEX SOUNDS

G. van den Brink

Dept. of Biological and Medical Physics, Medical Faculty Rotterdam; Lab. for Technical Physics, Technical University Delft, The Netherlands

Introduction

Certain kinds of hearing loss are accompanied by pitch changes. Even in cases which are not considered to be pathological, small irregularities in threshold and equal loudness audiograms, which are not more than a few decibels, are correlated with pitch irregularities (van den Brink, 1969). Not only for permanent deafness, but also in cases of temporary hearing losses, induced by exposure to a loud sound, are accompanied by pitch changes.

The purpose of the present experiments were to study the influence of a fatiguing signal upon the pitch of pure tones and to verify whether earlier findings (Van den Brink, 1971) about the link between the pitch of a complex signal and the pitches of its separate spectral components are valid also in the case of auditory fatigue.

We can be quite sure about the fact, that a temporary threshold shift caused by exposure to loud sounds is due to temporary inactivity of the most sensitive haircells in the organ of Corti. The measurement of pitch changes for pure tones and complex signals caused by fatigue, therefore, may enable us to decide whether the spectral components of residue-like signals (in our case harmonic AM signals) interacted already on or before the level of the organ of Corti.

Experimental set-up and procedure

The experimental set-up as well as the measuring procedure are similar to these applied in earlier experiments on binaural
INFLUENCE OF FATIGUE UPON PITCH

VAN DEN BRINK

Diplacusis (1969 and 1971). Detailed information about the set up has been shown before (1971); only the sequence of stimulus presentation is illustrated here in Fig. 1.

Fig. 1. Sequence of stimulus presentation.

With intervals of 1.6 sec. a fatiguing sound was presented periodically during 1.6 sec. in the subject's left ear. During the intervals two pairs of sound bursts were presented alternately to the two ears in a sequence left - right - left - right. The duration of each burst was 0.4 sec. They consisted either of pure tones or of harmonic AM signals. The fatiguing signal was always a pure tone. The after-effect of the fatiguing tone upon the pitch of the left ear signal (ref.) is very strong immediately after it is switched off; it decreases rapidly at first and more slowly later on. It was verified that fatigue was in a steady state after 30 to 45 seconds under our circumstances. The effect of fatigue upon pitch was unnoticeable after a period of about 10 minutes. The decision whether the test signal in the right ear was either equal or not to the signal in the left ear was always made at a moment 1.2 sec. after the end of the fatiguing signal. The frequency of the test signal (right) was carefully adjusted to a pitch that was equal to the pitch of the left ear signal at that moment. Matchings were made as a function of the frequency of the left ear signal. Although the effect of diplacusis was always superimposed on the effect of fatigue, this binaural matching procedure enabled us to compare a fatigued ear with an unfatigued ear. It has been verified that the fatiguing signal in the left ear had no
INFLUENCE OF FATIGUE UPON PITCH

VAN DEN BRINK 3

influence on pitch in the right ear.

Results
In the Figs 2 - 5 the curves indicated with a) represent data obtained with pure tones. The relative frequency difference \( f_r - f_1 / f_1 \) as is necessary for equal pitches in both ears is plotted as a function of the frequency of the pure tone signal in the left ear. The curves indicated with b) result from pitch matchings with harmonically amplitude modulated signals.

The value of \( \Delta f = \Delta g / g \) is plotted as a function of \( g; k = f/g \), \( f \) being the carrier frequency, \( g \) being the modulation frequency. The modulation depth was 100%. The curves indicated with c) are obtained by - rather arbitrarily - calculating the average matching value for the separate spectral components of the AM signal, giving the carrier component \( f = kg \) weight 2 and the components \( (k - 1)g \) and \( (k + 1)g \) weight 1. We are fully aware of the fact that this choice of 1 : 2 : 1 for the weights is rather arbitrary indeed. Spectral dominance as well as combination tones might be taken into account.

Fig. 2. a, b and c see text; \( k = 5 \); no fatiguing tone ref. and test signal 60 dB SPL.

Fig. 3. \( k = 5 \); fatiguing tone 700 Hz, 110 dB SPL. ref. and test signal 60 dB SPL.
The data calculated in this way, however, already show a remarkable agreement with the measured AM data in the case of no fatigue, provided that the spectral components of the signal are beyond 2000 Hz (1971). A rather severe hearing loss between 2000 and 3500 Hz in both ears of this subject may be the cause of the last restriction.

The agreement mentioned above is clearly shown in Fig. 2: there is a convincing correspondence between the curves b (measured) and c (calculated), except in the right part of the curves. Other data (1971) show that a lack of correspondence in the fine structure exists systematically in all measurements where the frequency of the spectral components exceeds 2000 Hz. The rough trend in the curves remains, however, also beyond 2000 Hz.

In the case of fatigue, the same phenomena exist, as is shown in Fig. 3. Due to a 700 Hz, 110 dB SPL fatiguing signal there is a systematic elevation of the pure tone diplacusis curve. Beyond about 900 Hz the test tone in the right ear had to be adjusted about 2% higher than without fatigue in order to have the same pitch as in the left ear. It is trivial that curve c) shows roughly the same elevation beyond 180 Hz, since it is calculated from curve a). The measured values for AM signals, however, show a similar trend, although the correspondence in the fine structure is not present.

Fig. 4 shows the results of a measurement with $k = 6$ and the same fatiguing signal, whereas Fig. 5 represents data obtained with a fatiguing signal with a frequency of 1000 Hz, 110 dB SPL and $k = 5$. Also in Figs 4 and 5 the elevation of the measured AM curves (b) (compared with the case of no fatigue) is about the same as in the calculated curves. In Fig. 5 there even is a striking correspondence in the fine structure which, however, may be due to a coincidence: There also is a correspondence with the fine structure of the pure tone curve. Particularly in this case the maxima in the pure tone curve are spaced such along the frequency scale that the $f - g$ and $f + g$ components coincide with maxima simultaneously with the carrier component.
INFLUENCE OF FATIGUE UPON PITCH

Fig. 4. k = 6; fatiguing tone 700 Hz, 110 dB SPL.

Fig. 5. k = 5; fatiguing tone 1000 Hz, 110 dB SPL.

Accuracy

During these measurements the impression grew that the accuracy of these matchings was less in the sloping parts than it was for frequencies where maxima or minima existed.

Fig. 6. Top: pure tone diplacusis; bottom: two times the "standard deviation."
In the upper part of Fig. 6 a pure tone diplacusis pattern between 1000 and 2000 Hz is given. This curve was obtained with 11 matchings per measuring point. The highest two and the lowest two of the 11 values were rejected. The measuring points give the averages of the median seven values. At the bottom of this figure we plotted the width of the frequency range as determined by these seven values, as an approximation of twice the standard deviation. This result confirms that, indeed, the measuring accuracy, is systematically depending upon the slope of the curve: the less steep the curve, the better the accuracy. The differences, however, are rather small. In this one experiment the average value of $\sigma$ is about 0.2%. The earlier described experiments, however, were carried out more carefully than this one, so that the accuracy can be estimated at 0.1 to 0.2%.

Conclusions
As usually is the case, the results of one experiment give inspiration for at least one following experiment. The results so far, show that the effect of fatigue upon pitch is similar to the effect of a permanent perception hearing loss. The striking correspondence between the fine structure of binaural pitch matching curves for AM signals, as existing for unpaired ears without fatigue, compared with curves calculated from pure tone results usually does not exist any more. The rough shifts in both curves, however, are similar. The results indicate that, in the sequence of processes that occur between stimulation at the outer ear with complex sounds and sensation, the spectral components of the stimulus are still present as such at the place in the system where the cause of fatigue and perception deafness is localized i.e. in the organ of Corti. The areas in the organ of Corti that correspond with the separate spectral components, evidently, do all play a role in the process of perception, and may be other places, corresponding with combination tones, as well. The described experiments do not enable us, to separate and manipulate with the separate components. Further experiments will be carried out with synthesized harmonic three and two component signals, such, that...
amplitudes and phases of the components can be varied mutually independently with monotic as well as dichotic stimulation.

References:
Van den Brink, G. (1969): Experiments on binaural diplacusis and tone perception,
Van den Brink, G. (1971): Two experiments on pitch perception: diplacusis of harmonic AM signals and pitch of inharmonic AM signals,
1. Introduction

As long as masking is considered phenomenologically as the threshold increment of a sound due to the presence of another (masking) sound (part 2 of the A.S.A. 1960 definition), the investigator of masking is not likely to gain much insight in the function of the hearing organ. For that purpose it is necessary to trace and to locate the mechanisms underlying the masking phenomenon. From literature data it appears that at different levels in the hearing organ contributions must arise to psychoacoustically measurable masking. The existence of dichotic masking implies that a central component is involved. The fact that (non-'remote') monotic masking is more prominent, on the other hand, suggests that peripheral contributions are not negligible. The picture is made still more complex by the finding that in non-simultaneous (forward) masking several components can be distinguished having different time constants (cf, e.g., Botsford 1971).

In this paper we will restrict ourselves to the peripheral aspects of non-simultaneous masking that are related to the peripheral auditory frequency analysis. Two arguments for this restriction are: (1) our knowledge of the function of peripheral auditory processing might be practicable, especially regarding peripheral frequency analysis, less is known, however, about central processing, and (2) we have the feeling that the peripheral aspect is often underestimated, especially so in non-simultaneous masking.

In simultaneous masking the similarity between pure-tone
(or narrow-band) masking curve and tuning curve (as determined for several mammals) strongly suggest the cochlear frequency analysis as a common underlying mechanism. (In this paper we will use the term *cochlear frequency analysis* avoiding the question whether this is mainly brought about by basilar membrane motion, or that some mechanical or neural "sharpening" within the cochlea plays a significant part). From the two curves mentioned above at any rate the tuning curve provides a measure of the peripheral auditory frequency selectivity. The high selectivity observed must be of relevance to non-simultaneous masking, as it will produce a not negligible stretching in time, especially at short signals. The stretching causes responses to stimuli, originally separated in time, to overlap. Therefore is it desirable to consider critically the statement that backward masking is attributable to time dependent properties of the auditory nervous system (e.g., Jeffress, 1970).

2. Experiments on backward masking

In the literature several experiments have been described in which backward masking was determined for a number of different acoustical stimuli. The masking effect appears to depend on stimulus parameters such as spectral and time-composition, and masker intensity. We will restrict ourselves to experiments in which the duration $T_p$ of the masked sound (probe $P$) is relatively short and in which further the spectrum of the masker $M$ encloses that of $P$.

As regards the spectral composition of $M$ and $P$ three main categories can be distinguished:

A. $P$ narrow-band and $M$ narrow-band ($f_p$ and $f_M$)
B. $P$ narrow-band and $M$ broad-band ($f_p$)
C. $P$ broad-band and $M$ broad-band.

Thus categorized, a number of references to the literature are given in Table I.

For a quantitative comparison of the data we would start from the following manner of the data presentation. In category A we take as distance $-\Delta t$, the interval between $P$-onset and $M$-onset. In group B we use the interval between the middle of
P and M-onset. Under that condition measurements with different probe duration (T_p) are well comparable, and in B4 maximal masking was obtained at Δt=0 (Fig. 1). For the dependent variable, the amount of masking, most authors give the threshold increment of P. Others, however, give the threshold related to some reference level. In most cases practically all backward masking occurs for -Δt < 10 ms. Therefore, for the threshold L_p to be presented in the diagrams we choose the threshold at -Δt = 10 ms as a reference. In Fig. 2 to Fig. 5 a number of literature data are compared with each other. The data of C6 (subj. SJM, 70 dB SPL) fit the drawn line of Fig. 3 closely. For further details concerning the various measuring procedures the reader is referred to the original articles.

**Fig. 1.** Definition of Δt in different stimuli.

**Fig. 2.** Backward and forward masking for periodic short tonal signals (A1) compared with masking of a periodic pulse on a periodic tone (B4). Repetition frequency 50 Hz, f_p = 1 kHz.

**Fig. 3.** Backward masking in case of a long noise masker and a short probe (tonal in B1, B2, and B5, and a click in C5).
**Table I**

1. Repetition frequency of the stimulus was 50 Hz.
2. Monotic and dichotic listening conditions examined.
3. Duration 8 cycles of probe frequency.
4. The threshold of the part-tone of probe frequency was determined. Repetition frequencies varied from 25 up to 400 Hz. (Table I continued on page 5)
3. Discussion

In the discussion on our experiment B4 (Duifhuis 1971, 1972), of which some results are depicted in Fig. 2 and Fig. 4 it was concluded that the observed threshold increment is due to cochlear interaction of P and M. The threshold increment for $\Delta t < 0$, however, can also be considered to be backward masking and as such be compared with other literature data. A great similarity is immediately seen, which also holds quantitatively if the masker level is taken into account.

The major part of backward masking occurs within approx. 10 ms (in Fig. 3 perhaps 20 ms). The maximum backward masking at $\Delta t = 0$, depends on the intensity of M, but apparently also on the composition of P. In C1 up to C4 (Fig. 5) the composition of M equals that of P, so that the masking at $\Delta t = 0$ can be derived directly from the DL in loudness.

(Table I continued)

5. $\tau$ is the decay time of the click.

6. The stimuli were presented in a continuous low-level masking noise background (2dB spectrum level).
Generally, the masking is a few decibels below the sensation level of M. If P and M are not equally composed, the masking level at $\Delta t = 0$ is generally somewhat lower.

The observed quantitative similarity leads to the following postulate:

Postulate 1: Backward masking has a dominant short term component which is brought about by temporal overlap of cochlear (filter) responses.

In further support of this postulate the following arguments are put forward. Measurements of Elliott (1962b) showed that with the increase of $f_p$ (0.5, 1, and 4 kHz, respectively) backward masking extended over a shorter interval. This is in agreement with our own results (e.g., Fig. 4) and is interpreted as being in agreement with the assumption that the relative bandwidth along the cochlear partition is in first approximation constant. This implies that rise times and decay times of the cochlear filters decrease with increasing frequency, so that with increasing frequency the overlap of responses diminishes. The trend signalized here is also found with the comparison of B1 and B2 with B5 in Fig 3. The masking as function of $\Delta t$ is smoother at $f_p = 0.5$ kHz that at $f_p = 1$ kHz (B5, and B1, B2, respectively). Recently Patterson (1971) confirmed this finding for a single value of $\Delta t$.

(Further experiments on this topic are under study, some results of which are likely to be presented at the symposium). As the response time of the (cochlear) filter is approx. inversely proportional to the bandwidth, which is approx. proportional to the tuning frequency, it appears to be more convenient to express $\Delta t$ in the dimensionless quantity $\Delta k$, so that $\Delta k = \Delta t f_p$. (cf. Duifhuis, 1971). The two curves of Fig. 4 then almost coincide. A further argument is provided by the effect of phase on the masking level as was established in Duifhuis (1971, 1972 Sec. 4.3). A waveform interaction predicts such an influence of phase. This could not be verified in most of the other (other than B4) experiments, because there the phases of P and M are not related in a well-defined way.

The data from category C, where P as well as M are broadband, are more difficult to describe quantitatively with the
proposed postulate. The neat time definition in the acoustic stimulus is done away with by the complexity of the time patterns arising over the whole cochlear partition. Therefore, in fact these stimuli are not very suitable for the investigation of the mechanisms underlying non-simultaneous masking. In category A and B it can be made plausible that the probe P because of its restricted bandwidth scans the response of the cochlear partition at the place which is maximally sensitive to f_p. With broad-band stimuli, however, interaction occurs over the entire cochlear partition. If apically two responses already overlap almost completely, they can still be separated at the base. Therefore, the high-frequency components determine to a considerable extent the listener's capability of separating P and M in time. From Fig. 5 it can be seen that in C1 up to C4 backward masking comes almost completely about within 5 ms. On the basis of the suggestion that high-frequency components are relevant, the shortening of the interval over which backward masking extends if compared to the results in category B, was to be expected. This shortening applies to a less extent to C5 and C6. The different character of M in experiments C1 up to C4 compared with C5 and C6 causes the results to be not simply comparable. It is recommended to verify whether this dissimilarity results in such a perceptive difference that different threshold criteria are used.

For a further verification of postulate 1 we therefore expect more from measurements with a coherent masker (e.g., pulse) than from measurements with a noise masker. In order to check the dependence on frequency one can, besides utilizing a tone burst for P, either imbed signals of the type C1 to C4 in noise (band-stop filtered) or send them through a band-pass filter. The use of a periodic masker can provide an advantage with regard to the elimination of adaptation effects.

Postulate 1 implies another postulate, since the cochlear overlap of originally separated time patterns works two ways. Therefore, forward masking can also be attributed, at least partly, to the cochlear filter mechanism.
Postulate 2: In forward masking two components are to be distinguished. The first is concerned with the effect of cochlear interaction, which extends over approx. 20 ms and which describes the major amount of short-term forward masking. The other component can be described with an exponentially decreasing threshold having a time constant of the order of magnitude of 75 ms.

The argumentation regarding the cochlear interaction mentioned in the postulate, is analogous to that for backward masking. As for the second component of forward masking, the following is remarked. Generally, forward masking, like T.T.S., is presented as a threshold increment which as a function of the logarithm of time decreases linearly. This description encounters objections for the limit values of $\Delta t$. Botsford (1971) showed for T.T.S. that a description of such a trend with two exponentially decreasing factors also covers the data reasonably well. Such a mechanism is only then more plausible than a description with a logarithmic $\Delta t$, if it is assumed that the auditory system treats amplitudes logarithmically. Fig. 6 shows that the threshold increment (in dB on a log scale) for intervals greater than approx. 20 ms can be reasonably well described with a time constant $\tau = 75$ ms ($\exp(-t/\tau)$). The presented data are from Zwislocki et al. (1959), Stein (1960), and Plomp (1964), Gruber and Boerger (1971), and Wilson and Carhart (1971). The deviation from the exponential trend occurs in the region where cochlear interaction is to be expected ($\Delta t > 20$ ms). It should be noticed that these deviations are somewhat compressed on the logarithmic dB-scale.

In conclusion we would remark that apparently a significant amount of backward masking, as well as a portion of forward masking can be attributed to filter properties of the peripheral ear. The proposed postulates are based on a relatively high estimate of the peripheral frequency resolving power, which is supported by other experiments, but which deserves further confirmation. We will not exclude the possibility of a comparable masking effect at a higher neural level, but we believe that in short-term backward masking this is of secondary importance.
Fig. 6. Forward masking. The lines represent a time constant of 75 ms. Significant deviations occur at $\Delta t < 20$ ms. For details concerning the different experimental procedures the reader is referred to the original papers. The indicated parameter is the masker level.

A further verification of the proposed postulates, deserves attention. For this, new data on the cochlear frequency resolving power will be of the highest importance.

4. References.


Ronken, D.A. (1970): Monaural Detection of a Phase Difference between Clicks,
Stein, H.J. (1960): Das Absinken der Mithörschwelle nach dem Abschalten von weissem Rauschen,
    Acustica 10, 116-119.
Wilson, R.H., and Carhart, R. (1971): Forward and Backward Masking: Interactions and Additivity,
The first successful recording from individual cochlear nerve fibres was by Tasaki (1954) in the anaesthetized guinea pig. His classical frequency response curve, for a fibre subserving the characteristic frequency of 7 kHz, resembled sufficiently the broadly tuned "resonance curve" of the basilar membrane for the conclusion to be drawn that the neural frequency response function merely reflected the mechanical response of the basilar membrane. On the other hand, data subsequently obtained from the cat cochlear nerve, notably by Kiang and colleagues (1965, 1967), but confirmed by others (e.g. Simmons and Linehan 1968, de Boer 1969; Evans et al 1970, 1971), indicate the cat neural frequency response to be substantially narrower.

The present experiments were designed to investigate possible species differences between cat and guinea pig, and to make a detailed comparison between the frequency response of single cochlear fibres in the guinea pig and the available mechanical data in the same species from the measurements of von Békésy (1944) and Johnstone and colleagues (e.g. 1970).

Using ultrafine micropipettes, recordings were made from several hundred single cochlear nerve fibres in the pentobarbitone anaesthetized guinea pig. The fibres were positively identified on grounds of latency, spike polarity, and histology. Stimuli were delivered in a closed system and stimulus levels monitored at the tympanic membrane. Frequency threshold curves (FTCs) were determined in the classical way using visual and aural criteria of 'threshold' of response to a 100 msec tone (with 5 msec rise and fall times) occurring 4/sec, and by a frequency sweep method. They were corrected to threshold dB SPL at the tympanic membrane in the closed bulla condition.

The minimum thresholds of the fibres approached within 10-20 dB of the behavioural threshold reported in the literature (e.g. Heffner et al 1971). Exceptions to this were a few high frequency fibres and fibres from preparations where there was evidence of malfunction of the cochlea either from abnormally low perfusion or local damage. Their thresholds were over 70 dB SPL. Data from these fibres are shown with open symbols in Figs. 3 ff. With these fibres excepted, the range of thresholds at a given frequency in any one animal was less than 20 dB.

The great majority of frequency threshold curves obtained resembled those for the cat (Fig. 1; cf. Fig. 1b of Kiang et al, 1967). Thus, these
curves are sharply tuned and asymmetrical (on a logarithmic frequency scale) for fibres above 2 kHz, and progressively less sharp and less asymmetrical for fibres of lower characteristic frequencies (CF). The lower frequency fibres also show a curious inflexion on the high frequency cut-offs. All these curves are substantially sharper than the corresponding frequency response functions derived for the basilar membrane by von Békésy (1944) and Johnstone and colleagues (1970). About 20% of the fibres, however, had anomalous frequency response properties (Fig. 2). Nearly all of these were the high threshold fibres obtained from abnormal cochleas as described above. The FTCs of these units resembled the mechanical frequency response functions.

![Frequency threshold curves of 8 cochlear nerve fibres from 6 guinea pigs, in dB SPL at the tympanic membrane. Lower dotted curves: analogous measurements of vibration amplitude of the guinea pig basilar membrane by von Békésy (<1 kHz) and Johnstone et al (curve at 18 kHz). The mechanical curves have been corrected to relate to sound pressure at the tympanic membrane in the closed bulla condition and are positioned arbitrarily on the intensity scale.](Attachment://image.png)
The slopes of the low and high-frequency cut-offs of the portion of the FTCs within 5 and 25 dB of threshold are shown in Figs. 3 and 4 respectively. Fibres denoted by open symbols are from the abnormally high threshold population. The cochlear nerve low frequency cut-offs were steeper than the corresponding mechanical measures by factors of between 2-12 for fibres with CF below 2 kHz, and 6-36 for fibres above 2 kHz. Corresponding ratios for the high frequency cut-offs were: 0.5-3 for fibres below 2 kHz, and 1-7 for fibres with CF above 2 kHz. The slopes of the high frequency cut-offs represent the minimum values, for they increase with stimulus level, to over 1000 dB/octave in some cases.

The relative sharpness of each guinea pig FTC was measured as the "310dB" value, namely the ratio of characteristic frequency to the bandwidth at 10 dB above threshold (Fig. 5). The neural bandwidths were thus narrower than the corresponding basilar membrane values by factors of between 3-10 and 1.3-5, for fibres with CF above and below 2 kHz respectively.

Fig. 2. Frequency threshold curves of 5 abnormally high threshold cochlear fibres from 4 guinea pigs.
Fig. 3. Slopes of the low frequency cut-offs of the frequency threshold curve versus CF of fibres from 12 guinea pigs. (Slopes measured over region: 5-25 dB above minimum threshold). Open symbols in this and subsequent figures: fibres with abnormally high threshold. Dotted line through star symbols: analogous measurements from basilar membrane frequency response data of von Békésy (1 kHz) and Johnstone et al (16 kHz).

Fig. 4. Slopes of high frequency cut-offs of frequency threshold curves. For circumstances and symbols, see Fig. 3.
These results are quantitatively in agreement with similar measures for cat cochlear nerve fibres (Evans & Wilson, 1971), except for a greater spread of data and the high threshold anomalous fibres found in the guinea pig.

This, then, brings the guinea pig in line with the consistent body of data on the frequency response of the cat (e.g. Kiang et al 1965; 1967; Matsuki et al 1958; Simmons and Linehan 1966; de Boer 1969; Evans & Wilson 1971). The neural measures are consistently narrower than the corresponding values for the basilar membrane response of the guinea pig. At intermediate frequencies (3-10 kHz), this difference between the neural and mechanical 'tuning' approaches an order of magnitude. Unfortunately, no actual measurements of the guinea pig basilar membrane motion exist at regions corresponding to these intermediate frequencies, and the above comparisons have been made by interpolation between the low frequency data of von Békésy and the high frequency data of Johnstone and colleagues.

Fig. 5. "Q10dB" of frequency threshold curves versus CF of fibres from 12 guinea pigs. Symbols as in Fig. 3.
The question therefore arises whether the mechanical data are in error or whether the interpolation is unjustified. This question has received added force from recent measurements by Rhode (1971) at a point in the squirrel monkey cochlea subserving frequencies of about 7 kHz. While Rhode published slope values in substantial agreement with the findings of Johnstone et al (1970), he found significant non-linearities. These are such that if a constant amplitude criterion is used for constructing a frequency response function from Rhode's data, then this function has cut-off slopes and bandwidth approaching those measured in cochlear nerves of comparable CF. However, the fact that this derivation depends upon a non-linearity specifically not found by von Bekesy or by Johnstone and colleagues, or by Wilson and Johnstone (this symposium), makes it difficult to interpret.

Another possible explanation for the apparent discrepancy between the neural and mechanical frequency response functions was suggested by Huxley (1969). He questioned whether the surgical opening of the cochlea necessary to carry out the direct measurements of the basilar membrane action could destroy a high degree of resonance in the mechanical elements. This suggestion specifically looked at in a second series of experiments (Evans 1970), i.e., recordings were made from single cochlear fibres in the guinea pig, as above, but after the scala tympani had been surgically opened over nearly half the first (basal) turn. Fig. 6 shows frequency threshold curves obtained from a number of fibres in one guinea pig after making an exposure of the basilar membrane and draining the perilymph from the scala tympani, as indicated. The curves do not differ systematically from those obtained from intact cochleas. They are still substantially shorter than the mechanical measurements of Johnstone and colleagues (lower dotted curve).

Experiments on single cochlear fibres in the cat indicate that the apparent difference in sharpness of frequency response between the mechanical and neural elements of the cochlea is not due to non-linear sharpening mechanisms (de Boer, 1969; Evans, Rosenberg and Wilson 1970; Evans and Wilson 1971), or to lateral inhibition (Evans, Rosenberg and Wilson, 1971; Wilson and Evans, 1971). The cochlea appears to act as a simple linear filter.

The simplest conclusions to be drawn from these considerations are either (a) that the measurements of the basilar membrane responses are still grossly in error; (b) that the notion of the basilar membrane does not represent the effective mechanical input to the hair cell transduction system, or that (c) the broadly tuned linear basilar membrane filter is followed by a second more sharply tuned filter. The present experiments and those reported by Wilson and Johnstone (this symposium) support the last suggestion. A consistent finding in our measurements in the cat cochlear nerve (e.g. Evans and Wilson 1971) and in the present study in the guinea pig, is the wide range of 'tuning' properties at any given characteristic frequency. This together with the finding in the guinea pig of broadly tuned fibres from pathological cochleas, offers circumstantial evidence for the existence of a second filter whose sharp tuning is physiologically vulnerable.
Fig. 6. Frequency threshold curves of 6 cochlear fibres obtained after opening scala tympani of first turn and draining perilymph to expose basilar membrane over the region indicated. Lower dotted curve as in Fig. 3.

REFERENCES


Evans, E. F. (1970). Narrow 'tuning' of the responses of cochlear nerve fibres emanating from the exposed basilar membrane. J. Physiol. 203, 75-76P.


1. Introduction
Results of threshold measurements suggest a selectivity of the ear which is an order of magnitude greater than that of "resonance curves" of the basilar membrane displacement. From experiments of Scholl (1962) and Elliott (1967) it has been concluded that the ear's selectivity might increase with time after the beginning of the stimulation. If this hypothesis is correct, the masking contour produced by a short narrow band masker should be much broader than the masking contour of a long masker. In order to test this prediction, the following simultaneous and forward masking experiments were performed.

2. Simultaneous Masking
The masking contour of a critical band wide noise (masker) centered at 1,65 kHz was measured with tonal signals by the method of tracking. Five subjects with normal hearing and psychoacoustical experience participated. The presentation was monaurally through a free-field correction network over an earphone (BEYSER DT 48).
In Fig. 1 the dashed line indicates the masking contour for a continuous masker (SPL 70 dB) and 500 ms signal duration, i.e. steady state condition. The dash-dotted curve represents the corresponding displacement of the basilar membrane; the great difference in selectivity is clearly visible.
Fig. 1 Simultaneous masking of a critical band wide noise (SPL 70 dB) centered at 12.5 Bark (hatched area). SPLS of signal is given as function of its critical band rate z and frequency f_s, respectively. For continuous masker, the masking contour corresponds to the dashed curve (signal duration T_s = 500 ms) and the filled triangles (T_s = 10 ms), respectively. For 10 ms masker duration (T_s = 10 ms) the open circles represent the central values with interquartile ranges. The dash-dotted line indicates the selectivity of the basilar membrane displacement.

When the signal duration T_s was reduced from 500 ms to 10 ms (masker still continuous), the resulting masking contour is given by the filled triangles. Comparison with the dashed line shows that the whole masking contour seems to be shifted.
upward by about 10 dB. The symbols in Fig. 1 represent the central value of at least 100 threshold determinations, the bars indicate the interquartile ranges.

For a gated masker with 200 ms duration, 500 ms silent interval and 10 ms signal duration, the results are quite similar to those represented by the filled triangles. Signal and masker starting simultaneously, were switched on and off Gaussian shaped and fed through third octave band filters to avoid spectral effects.

The open circles in Fig. 1 show the masking contour for a duration as short as 10 ms for signal and masker. We find almost no difference between open and filled symbols, i.e. for short (10 ms) as well as for continuous maskers, the same form of the masking contour can be observed. Thus, our simultaneous masking experiment indicates that no increase of the selectivity during stimulation by the masker takes place.

Green (1963) had observed that the threshold at the slopes was up to 40 dB higher for a gated masker than for a continuous masker. In our experiment, the difference between gated and continuous threshold (filled and open symbols in Fig. 1) was only few dB. This striking discrepancy has to be explained.

As mentioned above, in our measurements the masker was switched on and off with a Gaussian shaped signal. In Green's experiment, the masker switching signal was shaped rectangularly. The consequences of this methodical difference shall be discussed by means of Fig. 2.

The upper part of Fig. 2 shows the selectivity of the ear as measured in threshold experiments (thin lines) in comparison with the spectra of different shaped 10 ms tone impulses at 2 kHz (thick lines). The thick solid line represents the spectral pattern of a third octave band filter; the dashed, dotted and dash-dotted lines give the spectral patterns of rectangularly, trapezoidally and Gaussian shaped 10 ms impulses, respectively. The corresponding time-functions are shown in the lower part of Fig. 2.
Fig. 2 Masking (thin solid lines) of tonal signals by narrow band maskers with different center frequency $f_c$. The masked threshold $SPL_{TH}$ is shown as a function of frequency $f$ and critical band rate $z$, respectively. THQ indicates threshold in quiet. For comparison spectral patterns (thick lines) are shown produced by a third octave band filter (solid) and by single 2 kHz tone impulses with rectangular (dashed), trapezoidal (dotted) and Gaussian (dash-dotted) shape (see lower part of figure).

It can easily be seen in Fig. 2 that spectral effects play an important role in temporal masking experiments. Care has to be taken that the short time spectra of masker and signal, respectively, do not exceed the thin lines in Fig. 2, which represent the ear's selectivity in a steady state condition. In another paper (Zwicker and Fastl, 1972) it will be demonstrated in detail that for the differences between Green's and our results mainly the spectral effect is responsible.

Thus, we tested our apprehension that there might be no increase of the ear's selectivity by some forward masking experiments.
5. Forward Masking

Two experienced observers measured the threshold of 5 ms tonal signals masked by a sinusoidal masker with a frequency of 2 kHz and durations of 200, 30, 10 and 5 ms, respectively. The SPL of the masking tone was 80 dB. Again, signal and masker were switched on and off Gaussian shaped and fed through third octave band filters. The time pattern of the stimulus presentation is given in the upper part of Fig. 5.

![Figure 5](image)

**Fig. 5** Forward masking contours of 5 ms signals as function of masker duration $T_M$. SPL$_s$ of signal is given on the ordinate, signal frequency $f_s$ on the abscissa. The masker is represented by a 2 kHz tone with 80 dB SPL, the time-delay $t_d$ between masker- and signal-offset is parameter. The signal occurs only after each second masker impulse.

If a sharpening of the masking contour during stimulation takes place, it should become broader with decreasing masker duration, at least for short delay times.
The lower part of Fig. 3 shows somewhat simplified masking patterns, corresponding to the masking conditions mentioned above. The delay-time $t_d$ between masker offset and signal offset is the parameter (at $t_d = 0$ ms signal and masker end simultaneously).

For forward masking, Fig. 3 indicates that the masking pattern remains nearly constant for all masker durations. Perhaps the slopes are even steeper for short masker durations. Thus, the result of this experiment suggests also that there is no increase of the selectivity of the masking contour during stimulation.

In a comparable experiment, Elliott (1967) came to somewhat different results, which have been interpreted as an increase of the ear's selectivity with time. Since Elliott used trapezoidally shaped signals and maskers, we can conclude from Fig. 2 (dotted line) that spectral effects were involved, especially at the low frequency slope of the masker.

4. Discussion

The experiments on simultaneous and forward masking at short masker durations suggest that no time-dependent sharpening mechanism is involved in the great selectivity of the auditory system. Measurements of single fibre responses as well as experiments with the cubic difference tone support this conclusion. In order to test this conclusion further, experiments on backward masking are in progress.

References

Elliott, L.L. (1967): Development of Auditory Narrow-Band Frequency Contours,

Green, D.W. (1969): Masking with Continuous and Pulsed Sinusoids,

PITCH AND POWER SPECTRA OF SHORT TONE PULSES

Leo M. Grobben

E.N.T. Dept., State University Utrecht Netherlands.

A. THE PITCH OF SHORT TONES

Discrimination of pitch of short tone pulses has been studied by many workers (1-10). Depending upon methods of experimentation different values have been found for the frequency difference limen (D.L.) All the authors, however, demonstrated that the D.L. even under different definitions increases with decreasing duration. Gabor (5) developed a theoretical method of signal representation in which both time and frequency dependence are considered; he used experimental work of others (12, 13) to "prove" his concept in a not very convincing way. Many comments (6,9,10,11) on this approach stated that some aspects of Gabor's work were not quite clear and justified. One of the most important criticisms is the remark, that Gabor's theory is only valid for a linear system, which does not apply for the hearing organ (3,14).

Recently Ronken (10) investigated again the time (Δt) bandwidth (Δf) constraint by comparing pitch discrimination of short tone pulses, with different envelopes but with the same carrier frequencies (1000 Hz) and the same bandwidth. At first sight he found very different values of the Δf . Δt product, which, however, got much more alike when he used a more suitable Δt definition based on a signal to noise ratio concept.

Experiments

With a method earlier described (15) we determined pitch discrimination of tone pulses as a function of time duration for both
square and gaussian envelope of the pulse. Two normal hearing subjects, (D.V. and L.G.) and one severely perceptive hard of hearing, (A.B.), all of them trained in this kind of psychoacoustic measurements, had to match the pitch of a tone pulse $p$ with variable carrier frequency with the pitch of a reference tone $r$. In the block diagram (figure 1) the scheme of the apparatus used is represented. The normal hearing subjects listened with condensor telephone receivers and the hard of hearing subject with headphones because of lack of output of the former. The sound level was 40 phones. Both $p$ and $r$ were listened to in binaural presentation.

Figure 2 shows some of the results: the standard deviation $\delta f$ of 25 matchings as a function of the number $n$ of periods of the pulse. In the case of the square envelope (start and stop phase always being 0) the number of cycles is unmistakable. In the case of the gaussian envelope this number is derived from the tone definition as described in figure 6. The most interesting part of the graphs in figure 2 is their midportion where $\delta f$ decreases with increasing $n$. Plotted in a double logarithmic diagram (same scale on both axes) the curves show slopes of about $-45^\circ$ around this midportion. For small $n$ $\delta f$ seems to be rather constant; pitch discrimination there is almost impossible as one hears a click with very few pitch clues. For large values of $n$ (longer tone) the curves are getting flat again: a constant $\delta f$ for longer duration. In figure 3 the product $\delta f \cdot \delta t$ for the three subjects is represented as a function of frequency of the reference tone for both square and gaussian envelope.

In figure 5 our $\delta f$ values for pulses of 300 msec are plotted in the graph Rakowsky (16) showed, comparing different investigations of pitch discrimination of tones.

As one aspect of training and spread of our measurements, the slopes of the $\delta f - n$ curves for two subjects are presented in figure 4, in chronologic order.

**Results**

From figure 3 one can conclude that in the case of normal hearing subjects even with the rather "pessimistic" definition of duration of the gaussian pulse, the $\delta f \cdot \delta t$ product is on the average lower than
in the case of the square pulse. The hard of hearing subject does not give this result clearly. Especially in the high frequency range, where her hearing loss is the largest, discrimination of the gaussian pulse frequency is worse. Figure 7 shows her tone audiogram.

B. PITCH DISCRIMINATION OF SHORT TONES IN NOISE

Comparison between the $\delta f \cdot \delta t$ products in cases of square and gaussian envelopes suggested that the presence of side lobes in the Fourier frequency power spectrum could be responsible for the difference just mentioned. Masking of those side lobes would perhaps change the frequency discrimination. Rajcan (17) recently showed an interesting correlation between Fourier power spectrum and tone pulse pitch in listening to tone burst with different start and stop phases.

Experiments

In the case of $n=4$ the Fourier power spectrum of a square-enveloped tone pulse has rather large components even far outside the carrier frequency. As is illustrated in figure 8 side lobes of the frequency spectrum are masked by band rejected noise, the cut-off frequencies being carefully chosen. The experimental procedure however was exactly the same as described above. In order to be able to make a justified comparison between $\delta f$'s in the masked and in the unmasked situation, matchings of the two kinds are executed alternatively. After 6 or 7 matchings a pause of some 10 minutes was introduced whereas a whole series of 25 adjustments with and 25 without noise took about one hour.

Results

In figure 9 quotients of $\delta f$ masked and $\delta f$ unmasked for three normal hearing test subjects are plotted as a function of $f_r$ for different envelopes and noise configurations. From this figure one can conclude that in general:

1. $\delta f(\text{sq.env})$ decreases (pitch discrimination improves) if noise, that masks the most important side lobes of the Fourier frequency power spectrum is added ( and ).

2. This effect probably would be stronger if noise, usually a disturbing factor in all observation, did not deteriorate the pitch discri-
mination. This can be seen from the effect on $f$ of noise with a wide rejection gap. (cut-off frequencies at 3rd or 4th side lobe maximum.)

Pitch of gaussian pulses can be better discriminated without than with disturbing noise. The absence of side bands renders their masking unnecessary; hence discrimination shall be reduced by any type of noise.

REFERENCES.

Kay, I and Silverman, R.A. (1957) Inform. and contr. 1, 64. (9).
Liang Chih-an en Chistovitch (1960) Sovjet Physics Acoustics 6, 196. (8).
Majernik (1964) 4e Con.int. de cybernétique, Namur, 556. (3).
PITCH AND POWER SPECTRA OF SHORT TONE PULSES

Figure 1
Block diagram of the apparatus

Figure 2
\( \delta f \) for different numbers of periods (3 subjects) LG and DV normal, AB hard of hearing.
Figure 3
\( \delta f \cdot \delta t \) as a function of \( f_r \) for two envelopes
(3 subjects)

Figure 4
Slopes in chronologic order
(2 subjects)
Figure 5
D.L. for pitch as determined by several authors (after Rakowsky)
The dots mark the results of Grobben.

Figure 6
Definition of a tone pulse with gaussian envelope.

Figure 7
Audiogram of subject A.B.

Figure 8
Noise bands in relation to power spectrum.
Figure 9  
\[ \frac{\delta f_{\text{masked}}}{\delta f_{\text{unmasked}}} \]
as a function of \( f_r \) (n=4) for different envelopes and noise configurations.

- \( \square \) square pulse with intensive noise bands
- \( \blacksquare \) square pulse with moderate noise bands
- \( \bigcirc \) square pulse with noise only with components far from the carrier frequency
- \( \bigtriangleup \) gaussian pulse with moderate noise bands

Acknowledgements

I wish to thank my co-workers F. van Rhijn and D. Vogelpoel for their help in experimentation and constructive thinking and professor J.J. Groen for his discussions and assistance in preparing this manuscript.
PSYCHOPHYSICAL EXPERIMENTS ON GRATING ACUITY

T. Houtgast

Institute for Perception TNO, Soesterberg, The Netherlands

Summary

The "internal" modulation depth of auditory grating stimuli is probed with a test tone. Two psychophysical methods are applied: (1) a traditional masking paradigm and (2) a method based on the pulsation threshold phenomenon. This phenomenon can be described as follows: when a test tone and a masker are alternated continuously, the test-tone level can be adjusted to what can be called a "pulsation threshold" (for levels above this value the test-tone bursts are perceived as a pulsating tone, whereas for lower levels they are perceived as a continuous tone). As an auditory grating is made progressively narrower the "internal" modulation depth of the grating tends to decrease because of the limited frequency resolution of the ear. The amount of decrease as a function of grating width gives an estimate of the ear's frequency selectivity. The frequency selectivities revealed by our two methods are substantially different. These differences are discussed in terms of a second stage of frequency selectivity (sharpening) following a first stage of poor (mechanical) frequency resolution.

1. Introduction

For studies of the ear's frequency selectivity an attractive stimulus is a random noise with a sinusoidally rippled spectrum. Thus the intensity as a function of (linear) frequency is simply sinusoidal in form. The parameters of interest are the peak-to-trough depth (called modulation depth) and the peak-to-peak frequency distance. In general, the limitations of the ears frequency resolution are revealed by a decrease in the "internal" modulation depth as the peak-to-peak distance is decreased.
A highly interesting series of experiments with such stimuli was performed recently by Wilson and Evans (1971). Psychophysical data on the limits of grating resolution were obtained, together with electrophysiological measurements on cochlear nerve fibres in the cat, revealing the decrease in the "internal" modulation depth as a function of the decrease in peak-to-peak distance of the grating. These data are roughly consistent and predictable if the frequency response area of single nerve fibres are considered as reflecting the smearing functions effective upon the spectral grating. The high frequency resolution revealed by these data, compared with the relatively poor frequency selectivity as observed in the mechanical part of the inner ear, suggests that after a first stage of poor (mechanical) resolution, there may exist a second stage of sharpening.

As a contribution to the investigation of the ear's frequency selectivity a series of psychophysical experiments was performed with grating stimuli. The "internal" modulation depth was probed with a test tone, as a function of the peak-to-peak distance of the grating. Two methods were used which, in previous experiments, consistently produced different estimates of the ear's frequency selectivity; this is of interest with respect to the suggested two stages of frequency resolution.

2. Experiments

**Stimuli.** Fig. 1 indicates how the grating was produced and placed in the appropriate orientations with respect to the probe frequency $f_0$. The relative intensity of the delayed noise was always fixed at 0.36, resulting in a modulation depth of 12 dB. In a series of measurements the probe frequency $f_0$ was fixed. The delay ($\tau$) and the phase ($\phi$) of the added noise determines the orientation of the grating around the probe frequency. The peak-to-peak distance of the grating is determined by the parameter $n$, indicating the number of peaks between 0 Hz and the probe frequency. For each value of $n$, four grating orientations were investigated, as indicated in Fig. 1 (example for $n = 2$).
Methods. The test tone was switched on and off continuously, 125 msec on, 125 msec off, etc. (with smooth 20 msec rise and decay curves). The subject controlled the level of these tone bursts by means of a knob. The grating stimulus, the masker, was presented in two different ways, resulting in two different methods. (1) Continuous masker: The masker is presented continuously and the subject is instructed to adjust the level of the tone bursts to that value at which the tone bursts are just perceptible. This method will be referred to as "masking". (2) Pulsating masker: The masker is presented only in the 125 msec gaps between the tone bursts. In this case, for high probe levels a pulsating tone is perceived whereas, for a region of lower levels, the series of tone bursts gives rise to the
perception of a continuous tone. The subject is instructed to adjust the level of the tone bursts to the highest value for which no pulsation of the tone is perceived. This method is referred to as "pulsation". (It was suggested before, Houtgast (1971), that this pulsation method might be related directly to the neural representation of the masker).

The usual problems associated with adjustment procedures are probably not important here since the data of interest always are the differences between the adjustments for the grating orientations cos top and cos valley and between sin pos and sin neg (see Fig. 1). The stimuli were presented monaurally be earphone. The level at which the grated noise was presented was such that the level of the tone bursts at masked threshold (method 1) was on the average about 45 dB SL.

Results. Data were obtained from four subjects, with \( f_0 = 1000 \) Hz and \( n \) ranging from 0 to 12. The average threshold differences between the conditions cos top - cos valley and between sin pos - sin neg are presented in Fig. 2, (top panel). The bar at the base-line for \( n = 0 \) indicates the reliability of the data points (+ 2 st. error). The subsequent processing of these data is directed to the question: what linear weighting function around the probe frequency (thus what filter characteristic) can account for the blurring of the grating, as reflected by the cos (top-valley) curve, and the phase shift of the grating, as reflected by the sin (pos - neg) curve? In other words: what filter around \( f_0 \) gives output differences for grating orientations cos (top-valley) and for sin (pos-neg), as a function of \( n \), which correspond to the measured data? For this purpose we need to define a relation between the level differences at the filter output and the measured threshold differences. For \( n = 0 \) the "grating" is flat and the conditions cos top and cos valley actually refer to two flat spectra, 12 dB apart. Thus, any linear filter will show a level difference at the output of 12 dB for these conditions. The "masking" data show about the same difference. The "pulsation" data for \( n = 0 \) show a considerable smaller difference. (This might be related to observations of Møller (1970) on cochlear nucleus units showing that the slope of the function firing rate versus stimulus level is steeper for tone stimulation than for noise stimulation). The most reasonable relation between the measured level differences and the level difference at the output of a linear filter is a proportional one, the proportionality being determined by the data for \( n = 0 \). With this assumption, the data completely determine the shape of the filter. When the threshold differences cos (top-
Fig. 2. The uppermost panels present the average data of four subjects obtained with a 1000 Hz probe, in both masking and pulsation conditions. The middle panels present the intensity-weighting functions which account for the data; their equivalent-square bandwidth $\Delta f$ is indicated. The lower panels present these weighting functions as ordinary filter characteristics; the $-3$ dB bandwidth is indicated. The dotted line indicates the accuracy limit, given the accuracy of the data points as indicated in the top graph by the vertical bar at $n = 0$.

valley) are referred to as $C_n$ (dB) and the differences $\sin$ (pos-neg) as $S_n$ (dB), the filter, described as an intensity-weighting function $W(f)$, is obtained from

$$W(f) = 0.5 + 1.13 \sum_{n=1}^{12} \left[ \frac{1.2 C_n}{C_0} \cos 2\pi \frac{f}{f_0} + \frac{10}{12} \frac{1.2 S_n}{S_0} \sin 2\pi \frac{f}{f_0} \right]$$
This function, estimated from both the "masking" and "pulsating" data, are presented in Fig. 2, (middle panel). The selectivity of these filters may be expressed by the bandwidth $\Delta f$ of a rectangular filter covering the same area and with the same top value $W(f_0)$. This bandwidth can be obtained from

$$\Delta f = f_0 \left[ 1 + 2.26 \sum_{n=1}^{12} \frac{1.2 C_n/C_0}{10} \right]^{-1}$$

The intensity-weighting function can be converted into a more traditional attenuation curve in dB by defining $W(f_0)$ as 0 dB. (Of course, the negative-going part of $W(f)$ cannot be expressed in this way.) These curves are presented in Fig. 2, bottom. The -3 dB bandwidth is indicated. The dotted line indicates the accuracy limit, given the accuracy of the data points presented in the top graph.

In conclusion, the grating data can be used to estimate a linear weighting function, or filter characteristic (accepting the proportionality between the filter output level and the threshold level based upon the data for $n = 0$). The selectivity of $W(f)$, as expressed by $\Delta f$, is determined by the cos(top-valley) data; the sin(pos-neg) data are required to specify the shape (asymmetry) of $W(f)$.

For a number of other values of probe frequency $f_0$ (roughly 250, 500, 2000 and 4000 Hz) the cos(top-valley) data were obtained from two subjects. The frequency selectivity revealed by these data, expressed in the equivalent bandwidth $\Delta f$, is presented in Fig. 3.
3. Speculative discussion

The results obtained with the two different methods reveal quite different degrees of frequency selectivity. It is tempting to relate this to the two stages of frequency resolution as mentioned in the introduction: a poor (mechanical) resolution (accounting for the stimulation pattern along the basilar membrane), followed by some process of sharpening (the sum of the two processes accounting for the neural representation). Within this scheme it appears that the effect of the second stage (sharpening) is not (fully) reflected in the "masking" data. It is not unreasonable to assume that detectability is determined for the most part only by the amount of mechanical interaction between signal and masker. On the other hand, the pulsation threshold, by hypothesis implying continuity in nervous activity (Houtgast, 1971) may indeed correctly reflect the overall result of the two processes.

Unfortunately, this second process cannot be studied separately since any change in the stimulus also involves the first stage. Its nature can only be revealed by comparing the result of the first stage (stimulation pattern along the basilar membrane) and the overall results of the two stages (neural representation). The characteristics of the first one may be obtained from direct observations of membrane movements, Mossbauer measurements and psychophysical measurements which are assumed to reflect mechanical interactions (e.g., beats of mistuned consonants, roughness and, perhaps, masked-threshold data). The characteristics after the second stage may be obtained from electrophysiological observations and from psychophysical measurements assumed to reflect the neural representation (e.g., the limits of grating perception, timbre differences and, perhaps, pulsation threshold data).

The general differences between the stimulation pattern and the neural representation reveals the following properties of that intriguing second-stage.

(1) It should cause a considerable increase in the frequency selectivity for frequencies higher than the characteristic frequency of a given single fibre.

(2) In combination with the first stage it should roughly behave linearly (that is, the single-fibre activity caused by a broad-band signal is predictable, to some extent, as the sum of the activities caused by the individual spectral components).
(3) Is it possible that, despite (2), the process has some non-linear properties which may account for two-tone suppression, combination tones and the different slopes of the firing rate versus stimulation level curves for tone and noise stimuli?

(4) Can the process be such that data based on the masked threshold of a test tone do not (completely) reveal its action?

If such a process could be formalized it would certainly throw light on a variety of different phenomena.

References


THE PRE-RESPONSE STIMULUS ENSEMBLE OF NEURONS IN THE COCHLEAR NUCLEUS

Peter I.M. Johannesma

Lab. of Biophysics and Medical Physics, Univ. Nijmegen, Nijmegen, The Netherlands.

INTRODUCTION

Presentation of an auditory stimulus
\[ x(t), \quad 0 \leq t \leq t_0 \]
may elicit a sequence of action potentials
\[ z(t) = \sum_{n=1}^{N} \delta(t-t_n), \quad n = 1, N \] (1)
from a neuron in the auditory system.

Periodically repeated presentation of the stimulus
\[ x(t) : x(t+kt_0) = x(t), \quad k = 1, K \] (2)
allows the formation of the ensemble of elicited responses.

Definition: The Post Stimulus Response Ensemble (PSRE) of stimulus
\[ x(t), \quad 0 \leq t < t_0, \]
by the repetitive presentation of stimulus \( x(t) \):
\[ z(t+kt_0) : 0 \leq t < t_0, \quad k = 1, K \] (3)
If no systematic changes dependent on \( k \) are visible in this ensemble, e.g. no habituation, then the PSRE is homogeneous. In this case the average value of the PSRE, i.e. the Post Stimulus Time Histogram (PSTH) or event-density \( n(t) \)
\[ n(t) = \frac{1}{K} \sum_{k=1}^{K} z(t+kt_0), \quad 0 \leq t < t_0 \] (4)
is the function normally used for the representation of the dynamic aspects of the single cell response.

The relation between stimulus \( x(t) \) and pulse sequence \( z(t) \) is always essentially nonlinear; the relation between stimulus \( x(t) \) and event-density \( n(t) \) may be nonlinear, but is usually not essentially nonlinear. For the study of the stochastic dynamics, i.e. the dynamic relation between stimulus \( x(t) \) and averaged or expected response \( n(t) \), the computation of the crosscorrelations forms a general and useful
approach (Wiener, 1958).

The first-order crosscorrelation

\[ R(\tau) = \frac{1}{T_0} \int_{-T_0}^{T_0} dt \cdot x(t-\tau) \cdot n(t) \]  

represents the linear relation between \( x(t) \) and \( n(t) \); the harmonic part of the (average) response.

The second-order crosscorrelation

\[ R(\sigma,\tau) = \frac{1}{T_0} \int_{-T_0}^{T_0} dt \cdot x(t-\sigma) \cdot x(t-\tau) \cdot n(t) \]  

represents the dependence of the average response \( n(t) \) on the quadratic properties of the stimulus \( x(t) \); e.g. on the second harmonic for a sinusoidal stimulus.

Higher-order crosscorrelations represent more complex and usually less important aspects of the stimulus-response relations.

This approach to the analysis of single-cell responses on repetitive stimuli, will be generalised in this paper for non-repetitive stimuli where no PSTH can be constructed and the event-density \( n(t) \) cannot be measured. The exploitation of the fact that the response is a pulse sequence allows a more general analysis including a relation with the concepts and methods used in pattern recognition.

THEORETICAL DESCRIPTION OF STIMULUS-RESPONSE RELATIONS FOR SENSORY NEURONS.

The correlation of stimulus \( x(t) \) and PSTH \( n(t) \)

\[ R(\tau) = \frac{1}{T_0} \int_{-T_0}^{T_0} dt \cdot x(t-\tau) \cdot n(t) \]

\[ = \frac{1}{T_0} \int_{-T_0}^{T_0} dt \cdot x(t-\tau) \cdot \frac{K}{K} \sum_{k=1}^{K} z(t+kt_0), \text{ because of Eq. (4)} \]

\[ = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{T_0} \int_{-T_0}^{T_0} dt \cdot x(t-\tau) \cdot z(t+kt_0) \]

\[ = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{T_0} \int_{-T_0}^{T_0} dt \cdot x(t+kt_0-\tau) \cdot z(t+kt_0), \text{ because of Eq. (2)} \]

\[ = \frac{1}{T_0} \int_{-T_0}^{T_0} dt \cdot x(t-\tau) \cdot z(t), \quad T_0 = Kt_0 \]  

(7)
This equation states that the crosscorrelation between stimulus $x(t)$ and PSTH or event-density $n(t)$ is, always and exactly, equal to the crosscorrelation between stimulus $x(t)$ and pulse sequence $z(t)$. The same can easily be derived for the second-order crosscorrelation

$$R(\sigma, \tau) = \frac{1}{T_0} \int_{0}^{T_0} dt \ x(t-\sigma) \ x(t-\tau) \ n(t)$$

$$= \frac{1}{T_0} \int_{0}^{T_0} dt \ x(t-\sigma) \ x(t-\tau) \ z(t)$$

(8)

and for the crosscorrelation of arbitrary order.

These mathematical results imply that the information with respect to the response of a cell on the repetitive stimulus $x(t)$ contained in the PSTH $n(t)$ is also present in the crosscorrelations between stimulus $x(t)$ and pulse-sequence $z(t)$.

Eq. (7) and Eq. (8) will from here on form the basic definitions of $R(\tau)$ and $R(\sigma, \tau)$ which apply both for repetitive and non-repetitive stimuli.

The fact that the response $z(t)$ is a pulse-sequence, leads to the following equations

$$R(\tau) = \frac{1}{T_0} \int_{0}^{T_0} dt \ x(t-\tau) \ z(t)$$

$$= \frac{1}{T_0} \int_{0}^{T_0} dt \ x(t-\tau) \ \sum_{n=1}^{N} \delta(t-t_n), \text{ because of Eq. (1)}$$

$$= \frac{1}{T_0} \sum_{n=1}^{N} \int_{0}^{T_0} dt \ x(t-\tau) \ \delta(t-t_n)$$

$$= \frac{1}{T_0} \sum_{n=1}^{N} x(t_n - \tau)$$

(9)

Eq. (9) states that the crosscorrelation of stimulus $x(t)$ and pulse-sequence $z(t)$ equals the average value of the stimulus preceding an action potential (de Boer and Kuyper, 1968).

In the same way, it can be shown that

$$R(\sigma, \tau) = \frac{1}{T_0} \int_{0}^{T_0} dt \ x(t-\sigma) \ x(t-\tau) \ z(t)$$

$$= \frac{1}{T_0} \sum_{n=1}^{N} x(t_n - \sigma) \ x(t_n - \tau)$$

(10)

Analogous results again follow for the higher-order crosscorrelations.

The meaning of these equations can be made more clear by definition of the pre-response stimulus ensemble.
**Definition:** The Pre-Response Stimulus Ensemble (PRSE) is the ensemble of stimuli preceding an action potential

\[ x_n(t) = x(t-n\tau), \quad n = 1, N \]  

The next step is to recognise the fact that the \( m \)th-order cross-correlation functions of stimulus \( x(t) \) and pulse-sequence \( z(t) \) defined as

\[ R_m = R(\tau_1, \tau_2, \ldots, \tau_m) = \int dt \ x(t-\tau_1) \cdot x(t-\tau_2) \cdots x(t-\tau_m) \ z(t) \]  

is identical with the \( m \)th-order moment-function of the PRSE, defined as

\[ R_m = R(\tau_1, \tau_2, \ldots, \tau_m) = \sum_{n=1}^{N} x_n(\tau_1) \cdot x_n(\tau_2) \cdots x_n(\tau_m) \]

Several important conclusions may now be drawn.

**Conclusion 1:**

The information on stimulus-response relations contained in the PSTH \( n(t) \) is also present in the crosscorrelation-functions of stimulus \( x(t) \) and pulse sequence \( z(t) \) and in the moment-functions of the Pre-Response Stimulus Ensemble.

**Conclusion 2:**

For the definition of the Pre-Response Stimulus Ensemble and computation of the moment-functions \( R_m \), there is no need for the stimulus \( x(t) \) to be repetitive and/or simple. In fact complex, non-repetitive stimuli may be more adequate in this approach because of the potential richness of PRSE (e.g. Gaussian white noise, Poisson-distributed clicks, randomly in amplitude and frequency-modulated tones).

**Conclusion 3:**

Since the stimulus does not have to be repetitive the correlation method does not need controllability of the stimulus; observability may be sufficient. If the stimulus is not controllable formation and analysis of the PRSE seems the most adequate approach.

**Conclusion 4:**

The characteristics of the PRSE may be investigated and represented through the moment-functions \( R_m \), by means of other characteristics or directly through the probability density distribution.

In order to be able to extract quantitative information from the PRSE and to establish the relation with pattern recognition we introduce a geometrical representation of the Pre-Response Stimulus Ensemble. Given a complex, repetitive or not, stimulus \( x(t) \) with highest-frequency components \( \leq W \), the assumption is made that apriori physiological knowledge implies that the response \( z(t) \) does not depend on the part of the stimulus \( x(s) \) with \( s < t-T \). (For the Cochlear Nucleus \( T \) might be \( \approx 20 \) m sec.) As a consequence
PRE-RESPONSE STIMULUS ENSEMBLE

Johannesma

Each element of the PRSE

\[ x_n(\tau) \equiv x(t_n - \tau); \quad n = 1, N; \quad 0 \leq \tau \leq T \]

can be represented by 2WT sample-values at

\[ 0 \leq \tau = m \frac{W}{2W} \leq T; \quad m = 1, M; \quad M = 2WT \]

This implies that the function \( x_n(\tau) \) may be replaced by the 2WT-dimensional vector \( \hat{x}_n \). Defining now a 2WT-dimensional signal space, then each 2WT-dimensional signal vector defines a point in this space and the PRSE forms a cloud, possibly even a cluster, in this space. The characteristics of this cloud (e.g. location, size, form) will now become the point of interest. However, these characteristics can only be evaluated through a comparison with some standard form.

**Definition:** The (original or complete) Stimulus Ensemble (SE) is the ensemble of stimuli contained in the complete stimulus of bandwidth \( W \) and duration \( T_0 \)

\[ \hat{x}_p = x(t_p - T) \]

where \( 0 \leq t_p = p \frac{W}{2W} \leq T_0; \quad p = 1, P; \quad P = 2WT, \quad T_0 >> T \)

\[ 0 \leq \tau = m \frac{W}{2W} \leq T; \quad m = 1, M; \quad M = 2WT \]

This implies that stimuli are taken over the total duration \( T_0 \) and each again sampled at \( \frac{1}{2W} \) over a time \( T \).

The relation between stimulus \( x(t) \) and response \( z(t) \) can now be represented in five Probability Density Functions (PDF).

1. \( f(\hat{x}) \) is the P.D.F. of the stimulus ensemble;
   this function describes the probability density of occurrence of stimulus \( \hat{x} \) and may or may not be under experimental control.

2. \( f(z) \) represents the probability of occurrence of an action potential per unit of time \((1/2W)\) and equals the average frequency \((N/T_0)\) for the given stimulus ensemble.

3. \( f(\hat{x}, z) \) is the combined PDF of stimulus and response ensemble;
   this function represents the probability density of occurrence of a stimulus \( \hat{x} \) and an action potential at the end of \( \hat{x} \).

4. \( f(\hat{x}|z) \) is the conditional PDF of stimulus with respect to response;
   this function gives the probability density of occurrence of stimulus \( \hat{x} \)
   when an action potential is known to occur.

5. \( f(z|\hat{x}) \) is the conditional PDF of response with respect to stimulus;
   this function gives the probability of occurrence of an action potential directly following the presentation of stimulus \( \hat{x} \), is identical with the event-density \( n(t) \) and follows directly from the PSTH.

These five probability density functions are, however, not independent.
The Bayes' relation implies
\[ f(\mathbf{x}) \cdot f(z|x) = f(x,z) = f(z) \cdot f(x|z) \] (14)
Since stimulus \( x(t) \) and response \( z(t) \) are observable, the PDF of the
stimulus ensemble \( f(\mathbf{x}) \) and of the response ensemble \( f(z) \) can be measured.
This implies that, because of Eq. (14) the following two approaches yield
the same information.

**Forward approach**
Determination of \( f(z|x) \) from the average value of the Post-Stimulus
Response Ensemble (PSTH). This approach fits well into the framework
of system theory. It represents response prediction from the experimenters
point of view.

**Backward approach**
Determination of \( f(x|z) \) from the distribution of the Pre-Response
Stimulus Ensemble. This type of approach, which is strongly related to
the theory of pattern recognition is focused upon stimulus estimation
from the animal's point of view.

With the simple relation between these two approaches:
\[ \frac{f(z|x)}{f(x)} = \frac{f(x,z)}{f(z) \cdot f(z)} = \frac{f(x|z)}{f(x)} \] (15)
In words: the normalised response density is equal to the quotient of
aposteriori and apriori stimulus density.

The forward approach directly studies \( f(z|x) \), i.e. the PSTH for a limited
set of stimuli \( \mathbf{x} \). These stimuli are presented repetitively; as a
consequence full stimulus control is needed.

The backward approach evaluates \( f(x|z) \) or \( f(x,z) \), usually but not
necessarily for complex stimuli, and compares these with \( f(x) \).

The determination of the functions \( R = R(\tau_1, \tau_2, \ldots, \tau_m) \) can be
formulated both in the forward approach as the crosscorrelations between
stimulus \( x(t) \) and PSTH \( n(t) \) (Eq. (5) and Eq. (6)) and in the backward
approach as the crosscorrelations between pulse-sequence \( z(t) \) and
stimulus \( x(t) \) (Eq. (7) and Eq. (8)) as well as the moments of the Pre-
Response Stimulus Ensemble (Eq. (9) and Eq. (10)).

Since the Pre-Response Stimulus Ensemble is a selection out of the
original Stimulus Ensemble it is possible to define the complement of the
PRSE with respect to the SE. The collection of stimuli contained in the
SE but not in the PRSE forms the Complementary PRSE. A comparative
evaluation of PRSE and CPRSE is a well-known problem in the realm of
pattern recognition.
EXPERIMENTAL APPLICATION ON EXTRACELLULAR SINGLE CELL RECORDINGS IN THE COCHLEAR NUCLEUS OF THE ANAESTHETISED CAT.

First-order analysis

Gaussian white noise (random or pseudorandom) with a frequency content \( W \) large with respect to the characteristic frequency of the cell was used as auditory stimulus.

Computation of first-order crosscorrelation

\[
R(\tau) = \int dt \, x(t-\tau) \, z(t) = \sum_{n} x(t_n-\tau) = \sum_{n} x_n(\tau)
\]
gave three types of results.

a. \( R(\tau) \) equals zero. This is the case for cells with a \( CF \geq 3kC \); but, differing from results for primary auditory fibers, this may also occur for cells with a much lower \( CF \).

b. Complex \( R(\tau) \): either consisting of two well discernible types of oscillations or with a complex envelope in time and frequency.

c. Simple \( R(\tau) \): describable as an amplitude modulation of a sine wave of approximately the C.F. of the cell.

Five parameters were extracted from \( R(\tau) \):

- average time \( \tau_0 \)
- time duration \( \sigma_\tau \)
- phase \( \phi \)
- average frequency \( \omega_0 \)
- spectral width \( \sigma_\omega \)

The uncertainty product of time duration and spectral width was for the simple \( R(\tau) \) only slightly above the theoretical limit of 0.5

\[
0.5 < \sigma_\tau \cdot \sigma_\omega < 0.7
\]

(16)

This implies that the average pre-spike stimulus is (nearly) as narrow in the frequency-domain as is compatible with its duration in the time-domain.

The waveform could be well approximated through

\[
R(\tau) = c \left( \frac{\tau-a}{\beta} \right)^{\gamma-1} \cdot e^{-\frac{\tau-a}{\beta}} \sin(\omega_0 \tau + \phi)
\]

(17)

where the parameters \( a, \beta \) and \( \gamma \) follow from \( \tau_0, \sigma_\tau \) and \( \sigma_\omega \).

The form of \( R(\tau) \) is not in an obvious way related with the type of neuron characterised by the PSTH on CF tone-bursts (Pfeiffer, 1966). The quality of \( R(\tau) \) seems clearly related with the amount of phase-lock in tone-burst response and with cell-type (compare Lavine, 1971). \( R(\tau) \) forms the average value of the PRSE. If \( R(\tau) \) differs significantly from zero, then the following analysis of the dispersion of this ensemble appears attractive. For each stimulus \( x_n \), the energy and similarity with the average pre-response stimulus \( R(\tau) \) is computed:
Energy: \( e_n = \int_0^T \! d\tau \{ x_n(\tau) \}^2 = \vec{x}_n \cdot \vec{x}_n \)  

(18)

Similarity: \( \rho_n = \frac{\int_0^T \! d\tau \, x_n(\tau) \, R(\tau)}{\left[ \int_0^T \! d\tau \{ x_n(\tau) \}^2 \right]^{1/2} \cdot \left[ \int_0^T \! d\tau \{ R(\tau) \}^2 \right]^{1/2}} = \frac{\vec{x}_n \cdot \vec{R}}{||\vec{x}_n|| \cdot ||\vec{R}||} \)  

(19)

The square root of the energy \( \sqrt{e_n} = ||\vec{x}_n|| \) represents the length of \( \vec{x}_n \), while \( \rho_n \) equals the cosine of the angle between arbitrary pre-response stimulus \( \vec{x}_n \) and average pre-response stimulus \( \vec{R} \).

Distribution of similarity \( \rho \) and energy \( e \) for the Stimulus Ensemble \( f(\rho,e) \) and for the Pre-Response Stimulus Ensemble \( f(\rho,e;z) \) result in the probability density function of spike generation as function of similarity and energy of a signal \( x(t) \)

\[ f(z | \rho,e) = \frac{f(\rho,e;z)}{f(\rho,e)} \]  

(20)

This function supplies information concerning selectivity and threshold of the neuron.

Second-order analysis

For this analysis, which can be applied both on low- and on high-frequency cells, the stimuli were Gaussian white noise and/or randomly in amplitude and frequency modulated tones. Subject of investigation is the second-order crosscorrelation of stimulus \( x(t) \) and pulse sequence \( z(t) \)

\[ R(\sigma,\tau) = \int dt \, x(t-\sigma) \, x(t-\tau) \, z(t) = \vec{R} \cdot \vec{x}_n(\sigma) \cdot \vec{x}_n(\tau) = \vec{R} \cdot \vec{r}_n(\sigma,\tau) \]

which equals the ensemble-averaged second moment or the autocorrelation of the PRSE. In an analogous way as the first-order moment \( R(\tau) \) may be represented as a 2WT-dimensional vector \( \vec{R} \), the second-order moment \( R(\sigma,\tau) \) can be regarded as a 2WT \( \times \) 2WT-dimensional matrix \( R \). The eigen-vectors of this matrix form an orthogonal set spanning the signal-space; the eigen-vectors with the largest eigen-values form the main axes for the representation of the PRSE. Though this approach does well fit in with the theory of pattern recognition (Karhunen - Loève expansion) it does not appear well matched to the general characteristics of auditory signal transformation. A better suited description may be found by considering the signal both as function of time and as function of frequency.
Definition: The ambiguity-function of the signal \( x_n(\tau) \), \( 0 \leq \tau \leq T \), is given through the expression (Rihaczek, 1968)

\[
\epsilon_n(\omega, \tau) = x(\tau) e^{-i\omega \tau} x^*_n(\omega)
\]

(21)

where \( x^*_n(\omega) = \int_0^T d\sigma e^{i\omega \sigma} x_n(\sigma) \)

Some symbol manipulations leads to

\[
\epsilon_n(\omega, \tau) = \int d\sigma e^{-i\omega \sigma} r_n(\tau, \tau - \sigma)
\]

(22)

\[
E(\omega, \tau) = \sum_n \epsilon_n(\omega, \tau) = \frac{1}{N} \int d\sigma e^{-i\omega \sigma} R_n(\tau, \tau - \sigma) = \int d\sigma e^{-i\omega \sigma} R(\tau, \tau - \sigma)
\]

(23)

This implies that the (average) ambiguity-function is the Fourier-transform of the (average) autocorrelation.

The ambiguity-function has some interesting and relevant properties for auditory signal processing. The integral over all frequencies gives the energy density as function of time

\[
\int_{-W}^{W} d\omega \epsilon_n(\omega, \tau) = e_n(\tau), \quad \int_{-W}^{W} d\omega E(\omega, \tau) = E(\tau)
\]

(24)

integration over time gives energy density as function or frequency

\[
\int_0^T d\tau \epsilon_n(\omega, \tau) = e_n(\omega), \quad \int_0^T d\tau E(\omega, \tau) = E(\omega)
\]

(25)

and integration over both frequency and time gives the total energy of the signal

\[
\int_{-W}^{W} d\omega \int_0^T d\tau \epsilon_n(\omega, \tau) = e_n, \quad \int_{-W}^{W} d\omega \int_0^T d\tau E(\omega, \tau) = E
\]

(26)

The ambiguity-function \( \epsilon_n(\omega, \tau) \) may be considered as the complex energy of \( x_n(\tau) \) as function of both frequency and time; \( E(\omega, \tau) \) is the average energy differentiated with respect to frequency and time of the PRSE. The deviations of the ambiguity-functions of the individual pre-response stimuli \( \epsilon_n(\omega, \tau) \) from the average ambiguity-function of the PRSE \( E(\omega, \tau) \) can be evaluated in an analogous way as the deviations of the pre-response stimuli from the average pre-response stimulus.

The (average) uncertainty product \( \sigma_\omega \cdot \sigma_\tau \), i.e. the combined frequency-time resolution of this cell for complex stimuli can be computed from the (average) ambiguity function.

There does not necessarily exist a general apriori relation between the spectral density of the first-order crosscorrelation
PRE-RESPONSE STIMULUS ENSEMBLE

\[ I(\omega) = |\tilde{y}(\omega)|^2 = \int dt \ e^{-i\omega t} R(t) |^2 \]  

(27)

and the spectral density derived from the ambiguity function of the PRSE

\[ E(\omega) = \int dt \ E(\omega, \tau) \]  

(25)

The proportion of these two densities

\[ 0 \leq C(\omega) = \frac{I(\omega)}{E(\omega)} \leq 1 \]  

(28)

appears to be the obvious definition of a synchronisation-coefficient, at least for complex stimuli.

A related approach realisable with analog equipment in real time and resulting in the real energy as function of frequency and time can be implemented with a set of band-pass filters and quadratic rectifiers.

If \( h_k(\tau) \) is the impulse response of the band-pass filter with center frequency at \( \omega = \omega_k \) then the filtered signal is

\[ x_k(t) = \int dt h_k(\tau) x(t-\tau) \]  

(29)

and the energy density of frequency components around \( \omega = \omega_k \) at time \( t \) is

\[ e(\omega_k, t) = \langle x_n(t) \rangle^2 \]  

(30)

The crosscorrelation of the energy density and the sequence of action potentials

\[ E(\omega_k, \tau) = \int dt e(\omega_k, t-\tau) z(t) \]

\[ = \sum_n e(\omega_k, t_n-\tau) = \sum_n e(\omega_k, \tau) \]  

(31)

equals the sum of the pre-spike energy-densities.

The relation with the second-order crosscorrelation or the average autocorrelation of the PRSE is

\[ E(\omega_k, \tau) = \int d\lambda \ h_k(\lambda) \int d\mu \ h_k(\mu) \ R(\tau+\lambda, \tau+\mu) \]  

(32)

and with the average ambiguity-function of the PRSE

\[ E(\omega_k, \tau) = \int d\omega \int d\sigma \ H_k^*(\omega, \sigma) E(\omega, \tau+\sigma) \]  

(33)

where

\[ H_k(\omega, \sigma) = h_k(\sigma) e^{-i\omega \sigma} H_k^*(\omega) \]  

(34)

is the ambiguity-function of the filter impulse response.

The real, time and frequency dependent, energy density \( E(\omega_k, \tau) \) is an integration over frequency and time with a weighting function determined by the characteristics of the band-pass filter. This function may be considered as the sonogram of this neuron averaged over the PRSE.

An inherent problem with this hardware analysis is that frequency and time resolution of the sonogram depend on the characteristics of the filters.

Knowledge of the probability density function of the energy \( e(\omega, \tau) \) for the original Stimulus Ensemble
PRE-RESPONSE STIMULUS ENSEMBLE

\[ f(e(\omega, \tau) = f(e(\omega)) \]

and for the Pre-Response Stimulus Ensemble
\[ f(e(\omega, \tau), z) \]

allows the computation of the related forward distributions. 

**Dynamic Response Area (DRA)** defined as the probability density of occurrence of an action potential a time \( \tau \) after presentation of a signal with energy \( e \) at frequency \( \omega \):
\[
\begin{align*}
    f(z|e(\omega, \tau)) &= \frac{f(e(\omega, \tau), z)}{f(e(\omega, \tau))} \\
    &= \frac{\int_{\tau}^{T} d\tau e(\omega, \tau), z}{f(e(\omega, \tau))}
\end{align*}
\]

(35)

**Response Area (RA)** defined as the probability density of occurrence of an action potential induced by the presentation of a signal with energy \( e \) at frequency \( \omega \)
\[
\begin{align*}
    f(z|e(\omega)) &= \frac{f(e(\omega), z)}{f(e(\omega))} \\
    &= \frac{\int_{0}^{T} d\tau e(\omega, \tau)}{f(e(\omega))}
\end{align*}
\]

(36)

**Dynamic Response (DR)** defined as the probability density of occurrence of an action potential a time \( \tau \) after the presentation of a signal with energy \( e \)
\[
\begin{align*}
    f(z|e(\tau)) &= \frac{f(e(\tau), z)}{f(e(\tau))} \\
    &= \frac{\int_{-W}^{W} d\omega e(\omega, \tau), z}{f(e(\tau))}
\end{align*}
\]

(37)

For the original Stimulus Ensemble any wide-band stimulus may, in principle, be chosen. The central question, which can only be answered experimentally, is to what extent the normalised distributions \( f(z|x) \), \( f(z|\rho, e) \) and \( f(z|e(\omega, \tau)) \) which characterise the cell, are independent of the stimulus ensemble. For a model study see Johannesma, 1971.

In the backward approach random and pseudo-random Gaussian white noise and randomly or pseudo-randomly in amplitude and/or frequency modulated tones are used for determination of DRA \( f(z|e(\omega, \tau)) \), RA \( f(z|e(\omega)) \) and DR \( f(z|e(\tau)) \) (compare Van Gisbergen e.a., 1971).

In the forward approach tone-bursts are used for determination of the frequency dependent dynamic characteristics (DRA), tones of gradually increasing intensity for the response area (RA) and amplitude modulated noise for investigation of wide-band dynamic response (DR).

Preliminary experimental data and computational results will be presented to illustrate applicability and effectiveness of this approach.
References

Boer, E. de and Kuyper, P. (1968): Triggered correlation,

Gisbergen, J.A.M. van, Grashuis, J.L., Johannesma, P.I.M., and
with stochastic stimuli,

Johannesma, P.I.M. (1971): Dynamical aspects of the transmission of
stochastic neural signals,
Proc. first European Biophysics Congress, Wiener Med. Akademie, V,
329-333.

Lavine, R.A. (1971): Phase-locking in response of single neurons in
cochlear nuclear complex of the cat to low-frequency tonal stimuli,
J. Neurophysiol. 34, 467-483.

Pfeiffer, R.R. (1966): Classification of response patterns of spike
discharges for units in the cochlear nucleus: tone burst stimulation,
Exp. Brain Res. 1, 220-235.

Rihaczek, A.W. (1968): Signal energy distribution in time and frequency,
IEEE Tr. on Information Theory, IT 14-3, 360-374.

Stratonovich, R.L.: Topics in the theory of random noise;

Wiener, N.: Nonlinear problems in random theory;
I Introduction

When looking at the great complexity of the apparatus which transforms sound into activity of primary neurons, one is led to think that also the relation between the wave-form of the sound stimulus on one side, and the course of the actual adequate stimulus of a sensory cell on the other, will be very complex.

Since also the exact location of primary spike generation is not known it is difficult to think of a simple model of the transformation of sound into nerve spikes.

A most simple model is justified however when one assumes spike generation at the post-synaptic element of the (inner) haircell.

In its most elementary form we can think the whole model to consist of a linear filter followed by a trigger mechanism (figure 1).

We will show how this very simple model, the modified Weiss model, is compatible with the experimental facts.

If the model, in principle, is valid, one would expect that a filtered version of the sound stimulus should be the signal that modulates the receptor potential of the haircell.

It will be shown that indeed this filtered version of the stimulating waveform has very characteristic properties at the spike moments.
II. Analysis of the modified Weiss model.

One possible way of analysing systems like the system in fig. 1a, which transform continuous signals into trains of pulses, is the method of reverse correlation. The applied stimulus is white noise; the method is essentially crosscorrelation of the input waveform with the output spike signal. The result is an estimate of the linear part of the system.

In the case of the Weiss model, the result will be a linear combination of the impulse response \( h(t) \) of the filter, and its time derivative \( h'(t) \), the relative weight of the two components being determined by the threshold value \( b \). In formula:

\[
(1) \quad h^*(t) = b \cdot h(t) + \sqrt{\frac{b}{\pi}} h'(t) \quad (\text{de Boer & Kuyper, 1967})
\]

If the threshold value \( b \) were known, the differential equation (1) could be solved and exact reconstruction of the signal \( y(t) \) would be possible.

The second step in the analysis, is to treat the obtained \( h^*(t) \) as the impulse response of a linear filter \( C^* \), and to filter the original signal \( x(t) \) with this filter \( C^* \). See fig.1b.
fig.1b. The function $h^*(t)$ obtained by cross correlation of the white-noise input signal with the output spike signal is a first estimate of the filter impulse-response $h(t)$. The signal $y^*(t)$ leads to a direct estimate of $b$ and thus to $h(t)$.

How well the thus generated signal $y^*(t)$ resembles the true signal $y(t)$, depends on the value of $b$. It is clear from formula (2) that when the $b$-value is high, $y^*(t)$ is a good approximation of $y(t)$ but when $b$ is close to zero $y^*(t)$ will be close to the time derivative of $y(t)$.

$$y^*(t) = b \cdot y(t) + \sqrt{\frac{b}{\pi}} \cdot y'(t)$$

This signal $y^*(t)$ has very interesting properties at the original spike moments, i.e. the moments when $y(t)$ crossed level $b$ upwards. As a matter of fact it can be shown, mathematically, that $y^*(t)$ will be at a maximum (or very near to a maximum) on every one of those instants, irrespective of $b$.

Thus on a spike-moment the signal properties of $y^*(t)$ do not have a specific feature which reveals the $b$-value (it will always coincide with a maximum); however the distribution of the values of $y^*(t)$ on the spike moments directly discloses the missing constant.
III Simulation of the model

The above mentioned steps i.e. determination of $h^*(t)$, generation of $y^*(t)$ and finally computation of the amplitude histogram were performed in an electronic model, simulating figure 1a. This was done to check that the theory was rightly evaluated, and to determine the influence of the inaccuracies involved in the actual electronic treatment of the steps, because these procedures were also applied to the experimental data.

The result of the simulation with a low $b$-value is given in figure 2.

![Figure 2](image)

The upper one is an amplitude-histogram of the signal values of $y^*(t)$ on spike-moments. The lower one shows the distribution of the $y^*(t)$ values on arbitrary moments.
IV Comparison with experimental results.

In experiments with cats the unitary spikes picked up with a microelectrode from the eighth nerve were recorded, together with the stimulating waveform (white noise), with an Ampex FM tape recorder. In many cases it was possible to determine the crosscorrelate $h^*(\tau)$ and to generate the filtered version $y^*(t)$ of the actual stimulus $x(t)$. It has been reported before (de Boer and de Jongh, 1971) that the spike-moments indeed have a tendency to coincide with maxima of this signal $y^*(t)$. Of course one cannot expect this correspondence to be as exact as in the case of the simulation, firstly because of the inherent inaccuracy involved in the determination of $h^*(\tau)$ and secondly because of jitter to be expected in the spike generation.

In figure 3 we see the amplitude histogram of the signal values of $y^*(t)$ at spike moments (unit 4, cat 11; 900 Hz). If the model applies, we must conclude that the threshold in this case is zero or near zero. However, a comparison with the result of the electronic simulation with zero threshold shows a discrepancy: in the cat data the higher $y^*(t)$ values are clearly favoured against lower ones. In the language of zero-crossings this means: a zero-crossing of $y(t)$ with a steep slope is more likely to initiate a spike than a zero-crossing which is less steep.

We will comment on this later.

fig.3
Figure 4 displays a short piece of the $y^*(t)$ signal, obtained from cat data; the spike moments are marked with a vertical bar. When looking at it, the question arises: why does every top bear a spike? Why do spikes coincide with particular tops and not with others? Stated otherwise: what method is there in the determination of which zero-crossings initiate spikes?

\[ \text{fig.4, Cat 11, Unit 4, cf 900Hz} \]

V Repetitive wave-form approach.

We thought we could get a clue that might lead to an answer to this question by studying the response of a single unit to a complex wave-form that was repeated many times. Pieces of a thus obtained histogram are shown in figure 5.

The result of the one case that is worked out up till now is indeed revealing in more than one respect:

1. Nearly every zero-crossing may give rise to a spike.

2. Zero-crossings with steep slopes, in general yield greater probability of firing than zero-crossings which are less steep.

3. No fixed relation seems to exist between the steepness of the slope and the firing probability.

4. A long time constant of a few milliseconds seems to play a role in the firing probability:
this one can best be seen when the wave-form remains nearly constant for a few periods while the corresponding peaks in the histogram decrease (points of interest are marked with an X in figure 5). This decay is also apparent after a period of high mean activity.

VI Discussion

From a comparison between outer- and inner haircell innervation and physiological properties we feel inclined to believe that cochlear sharpening is localised at the Basilar membrane/haircell transition. Inner haircells are more sharply tuned and make contact with many afferent fibres, whereas outer haircells have a many to one relationship with afferent fibres. Also: asynchronous activity in say 8 fibres (the afferents from one inner haircell) generated by the same signal (the y(t) signal) could enable a good representation of signals with frequencies up to several KC.

Since the spikes occur only in the positive fase of the \( y^*(t) \) signal, the receptor potential must be able to follow a fast phenomenon like this (the \( y^*(t) \) signal in question had its central frequency around 900 Hz) accurately, which implies a very short passive-membrane time constant (<500 μsec).

It is clear that time constants of this order of magnitude will favor heavily steeper slopes against fainter ones. This could explain why large amplitudes of the \( y^*(t) \) signal coincide with a great firing-probability.

The long time constant which is apparent from the shown histogram may be caused by a temporary failure of transmitter stores to fulfill the high demand.
SPIKES IN THE 8th NERVE AND THE SOUND STIMULUS. de JONGH 8

fig.5. Piece of the $y^*(t)$ signal obtained from repetitive $x(t)$ signal together with histogram of the spike activity.

References.


Proceedings 7th I.C.A. Budapest.
PERCEPTION OF OCTAVE COMPLEXES

P.J.J. Lamoré

Dept. of Biological and Medical Physics, Medical Faculty Rotterdam, The Netherlands

Introduction

It is commonly known that two frequency components of M and N Hz respectively, when presented simultaneously, can give a beat sensation if the frequency ratio is slightly different from m : n (m and n being small integers, m < n).

So far three explanations of the origin of the beats have been proposed, as extensively reviewed by Plomp (1967):

1) The two frequency components give rise to a combination tone of some order which interferes with the lower tone or with some other combination tone. For instance, beats produced by two frequency components of p and 3p + b Hz respectively (b few Herz) would be generated by interference of the third order combination tone of the frequency (3p + b) - 2p = p + b and the lower primary tone.

2) Each of the two primary components gives rise to aural harmonics, produced by distortion in the ear. So, with frequency components 2p and 3p + b, the beats would be generated by interference of the third harmonic of the lower component (2p Hz) and the second harmonic of the higher one (3p + b Hz) which yields a beat rate of 2b.

3) The beats are related to the periodic variations of the waveform of the superimposed sinusoids.

In the article just mentioned Plomp's results favour the last theory with strong arguments.

Although the two described frequency components are separated more than one critical bandwidth in most cases, yet some interaction takes place in the ear manifesting itself as beats.
To study that interaction in more detail, we confined ourselves to the study of stimuli with two frequency components wherein the frequency of the higher component is a multiple of the frequency of the lower one. Under these conditions the distance between maximally stimulated areas on the basilar membrane is as great as possible, as is the distance between the areas where the two components are supposed to interact. This facilitates the experimental procedures. For instance the region of the membrane between the maximally stimulated areas can be masked without completely masking one of the partials at the same time. Moreover, in this way we can avoid conditions where combination tones play a role (Plomp, 1965).

Two frequency components with a frequency ratio slightly different from 1 : n can be described as two exactly tuned frequency components with one component shifting continuously in phase. Experiments with phase-locked frequency components enable us therefore to consider each of those phase relations separately. Listening to two frequency component stimuli with a frequency ratio 1 : n, either phase-locked or slightly mis-tuned we perceived alterations of loudness and/or timbre and/or pitch under certain conditions related to the phase. The effects depend upon the intensities of both components and affect only the lower component, only the higher or both components. Of the three possibilities just described, we started the investigation by studying the most simple one, that is when only perception of the higher component is affected.

The aim of the investigation is to study how the interaction on the basilar membrane takes place and how the stimulus is processed, using two frequency component stimuli as described.

**Methods, apparatus and procedure**

Experiments were carried out with stimuli consisting of two phase-locked frequency components with a frequency ratio 1 : 2. The pitch and the masking level of the higher component were measured as a function of the phase of the lower component. The pitch was measured by matching with a pure tone. The white noise masking level was determined with a two alternative forced choice procedure. By varying the intensity of the noise,
two points of the psychometric function were determined and the estimate of the 75% correct responses level was chosen as the masking level.

Fig. 1. Temporal stimulus configuration for the masking level experiments.

In the measurements of the masking level the stimuli were presented as illustrated in Fig. 1. The stimuli were always preceded by a warning signal. The duration of the C-puls (higher frequency component) varied from 200 msec. at 3200 Hz to 500 msec. at 400 Hz.

Fig. 2. Schematic block-diagram of the apparatus.
A schematic block-diagram of the apparatus is presented in Fig. 2. Two phase-lock generators were triggered with the output of an oscillator, direct and after dividing the frequency by two respectively. The distortion of the components of the electrical signal over the earphone was less than -55 dB. The observer was seated in a sound-isolated booth. The stimuli were presented monaurally.

For the used combinations of the intensities of the frequency components the pitch of the higher component was measured as a function of the phase. After that the masking level was determined. The frequency of the warning was adjusted to a value so that its pitch correspond with the pitch as measured earlier. The measurements were carried out as a function of the phase of the lower frequency component. The phase of the higher component was always set at a value, resulting in an electrical waveform over the earphone of the form:
\[
\sin 2\pi f t + \sin(2\pi 2f t + 90°),
\]
given a phase of 70° of the lower component. This resulted in easily surveyable plottings.

Results

The measurements described in the preceding paragraph were carried out for the frequency combinations 200 and 400 Hz, 400 and 800 Hz, 760 and 1520 Hz, 1600 and 3200 Hz. The results are shown in Figs. 3 - 6. For each frequency combination the measurements were started by adjusting the intensities of the both components in such a way that the effect of the phase upon the pitch and the loudness of the higher component was optimally audible. Then the pitch and the masking level of the higher component were measured as a function of the phase of the lower component. After that the same measurements were carried out with either the intensity of the lower component or the intensity of the higher component increased or decreased by 5 dB. For the frequency combination 760 and 1520 Hz measurements were carried out over the whole range of intensity combinations where any effect on pitch and masking level was detectable. For the other frequency combinations 4 or 5 measurements were done. Each block in the Figs. 3 - 6 gives the masking level and the pitch of the higher component as a function of the phase of the lower one at a specific intensity combination of both
Fig. 3. Pitch and masking level of the higher frequency component as a function of the phase of the lower frequency component for different intensities of each of the components.

Frequency combination 760/1520 Hz.
Fig. 4. Pitch and masking level of the higher frequency component as a function of the phase of the lower frequency component for different intensities of each of the components. Frequency combination 400/800 Hz.
Fig. 5. Pitch and masking level of the higher frequency component as a function of the phase of the lower frequency component for different intensities of each of the components. Frequency combination 200/400 Hz.
Fig. 6. Masking level of the higher frequency component as a function of the phase of the lower frequency component for different intensities of each of the components. Frequency combination 1600/3200 Hz.

Fig. 7. Waveforms of a signal consisting of two phase-locked frequency components with equal amplitudes as a function of the phase of the lower component.
components. The value of the pitch of the higher component is plotted as the ratio of the measured pitch and the pitch of the higher component on itself.
The value of the phase at which the masking level is minimal, turns out to correspond to the phase value at which the pitch drops from a maximum value to a value equal or somewhat lower than the pitch of the higher component on itself, thus without the lower component.
The pitch shifts as a function of the phase as shown in the right parts of the blocks in Figs. 3 - 6 represent the phenomenon which is known as the sweep-tone effect.
The accuracy of the masking level measurements is 1 dB for the highest levels and 5 dB for the lowest levels. In the same way the accuracy of the pitch measurements varies from less than 5 Hz to 10 or 20 Hz. So far all measurements were carried out by only one observer. In a preliminary testing three more observers heared a shift of the pitch and a change of the loudness of the higher component as a function of the phase. To all observers the higher component gave the impression of a pure tone of which the frequency was varied.

Discussion
It is still an open question whether the time structure of a stimulus at some place along the basilar membrane or the excitation pattern of the stimulus along the membrane is the determining factor in the processing of pitch by the auditory system. To meet the aim of this investigation it is a necessity to consider both possibilities. Because this investigation is still in an early stage, several experiments, although already planned, have not been carried out yet. Therefore a conclusion about the results obtained so far is preliminary and moreover applies only to the link with the time-structure of the stimulus. Let us, in a preliminary reasoning, look how far we can relate the waveform of the stimulus to its pitch and its masking level as a function of the phase. Our starting point is the idea that the time-structure of the stimulus at some place along the basilar membrane determines the pitch of an acoustical signal (a.o. Schouten et al., 1962; Ritsma, 1967;
Bilsen and Ritsma, 1969). Analogously to the electrophysiological experiments of Brugge et al., (1969), we assume that interaction phenomena only occur at a place of the basilar membrane where the amplitudes of both frequency components are about equal (within a 20 dB range). Fig. 7 illustrates the different waveforms of a signal consisting of two phase-locked frequency components of a frequency ratio 1 : 2 with equal amplitudes. Evidently, the waveform of the signal depends strongly upon the phase relation of the components.

The question arises what waveform is "seen" by the haircells at the area concerned. The acoustical waveform of the stimulus from the earphone passes the outer and the middle ear and reaches the area of interaction in the cochlea as a travelling wave. If all those steps would have either a flat phase characteristic or would yield a travelling time in the transmission of the wave, the waveform of the stimulus would not change during its course. However, the contrary is the case (Zwislocki, 1957; Möller, 1963; Johnstone et al., 1970; Rhode, 1970). This implies a change of the waveform.

Furthermore we assume for reasons of simplicity that mainly one of the possible waveforms as illustrated in Fig. 7, is actually present at the area of interaction. Then the question arises which one of the waveforms of Fig. 7 is being detected for each phase condition in the Figs. 3 - 6. In all blocks of the Figs. 3 - 6 where a pitch shift occurs as a function of phase, the pitch increases with increasing phase. For some value of the phase a maximum is reached. At this value the pitch drops to a value equal or sometimes somewhat lower than the pitch of the higher component sounded by itself. The phase region where the pitch drops is the only phase condition where - sometimes by extrapolation, sometimes really measured - ambiguity of pitch could be expected. Looking at the waveforms in Fig. 7 and considering the peak intervals and the peak heights, only Fig. 7d might provide the possibility of complete ambiguity. For that reason we might conclude that in the case the pitch drops, the waveform at the area of interaction on the basilar membrane is as it is in Fig. 7d. We assume this waveform to be the actual
waveform under the circumstances that the pitch drop occurs. For lower values of the phase down to zero the pitch is always higher than the pitch of the higher component alone. In the same way for higher values of the phase the pitch is always equal or somewhat lower than the one without interaction. Looking at Fig. 7 this means that the pitch is determined in some way by taking the interval between a lower top and the next higher top. Furthermore, it is a striking coincidence that the masking level for the higher component is minimal for a phase value at which the pitch drop takes place, i.e. when the waveform is like it is in Fig. 7d. When the difference in height between succeeding peaks is maximal the masking level is maximal. Although it is tempting to suppose a link between the difference in height and the masking level, more experimental results are needed to confirm this. We are fully aware of the fact that the results and the reasoning just given represent only a rough first approach. The first thing to do now is to measure where exactly the interaction on the basilar membrane takes place. Masking procedures are considered for this purpose. Measurements as described were also carried out by Terhardt (1971) in a different approach. His results are comparable with ours.

References
Acustica 22, 63-73.
Møller, A.R. (1963): Transfer function of the middle ear,
Plomp, R. (1965): Detectability threshold for combination tones,
Plomp, R. (1967): Beats of mistuned consonances,
Rhode, W.S. (1970): Observations of the vibration of the basilar membrane in squirrel monkeys using the Mösbauer technique,
Ritsma, R.J. (1967): Frequencies dominant in the perception of pitch of complex sounds,
Schouten, J.F., Ritsma, R.J., and Cardozo, B.L. (1962): Pitch of the residue,
    Acustica 25, 53-61.
Zwislocki, J. (1957): Some impedance measurements on normal and pathological ears,
The intensity of natural sounds usually varies with time at various rates. Many sounds have a more or less rapidly changing spectral composition. The well-known static description of the response areas of single units in the ascending auditory pathways with regard to frequency and intensity reflects only the static response and does not contain information about how the responses are related to the dynamic parameters of the sound. Only a few studies have reported on the influence on firing in primary Eighth Nerve fibers and cochlear nucleus units in response to sounds where the parameters varied in time.

In the present study the responses of single units in the rat cochlear nucleus have been recorded in responses to sinusoidally amplitude modulated pure tones and broadband noise and to sinusoidally frequency modulated tones. The resulting modulation of the neural discharge frequency has been quantitatively determined from cycle histograms, locked to the modulation wave form.

Generally it has been found that the gain functions, which relate the frequency modulation of the recorded discharges to the modulation of stimuli, resemble those of a lowpass filter. In some situations, such gain functions have been shown to have a more or less pronounced peak near the cut-off frequency representing a tuning to a particular modulation frequency (Ratliff et al. 1967, Ratliff et al. 1969).
Figure 1 shows an oscillogram of the utilized amplitude modulated sound together with a histogram of the neural discharges of a unit in the cochlear nucleus of a rat in response to an amplitude modulated tone. The smooth curve superimposed on the histogram is a single sine wave matched to fit the histogram according to the least mean square error criterion. The figures in the upper corner of this diagram give the relative amplitude and the phase angle of the matched sine wave. The waveform of
the amplitude modulated tone is seen in the lower part of Fig. 1. The
degree of modulation is the ratio between A and B. It is usually expressed
in per cent.

Figure 2 shows the relative amplitude (in dB and phase angle in
degrees) of the sine wave which fits the histogram in the range of modulat-
ion frequency from 1 to 1000. The cycle histograms of the activity are seen
for different modulation frequencies. In this case the stimulus was an
amplitude modulated tone at the unit's CF (30.0 kHz) and the modulation was
20%. Its intensity was 60 dB SPL (the threshold of the unit was 45 dB SPL).

A pronounced enhancement of the modulation wave in the neural discharge
pattern is seen in the responses of this unit at or around a modulation
frequency of 300 Hz. At this frequency the modulation of 20%, i.e. of about
+1.9 dB, gives rise to nearly 40% modulation of the density of the record-
ed neural spike train.

Other units, however, have a less pronounced peak in their gain func-
tions.
Figure 3 shows a sample of gain functions obtained from responses to amplitude modulated tones at CF. The ratio (in dB) between the modulation of the neural discharge frequency and the modulation of the stimulus tones together with the phase angles between modulation of the response and the stimulus is shown. A modulation was either 10 or 20% and the sound intensity was about 20 dB above the unit's threshold at CF.

In many units the gain function obtained in response to amplitude modulated tones was similar over a large range of sound intensities while in other units the gain functions changed in a systematic way when the sound intensity was raised from just above threshold. In such units, low modulation frequencies were attenuated more at high sound intensities than at intensities just above threshold. In a narrow frequency region around the frequency of the peak in the gain function the degree of modulation was reproduced almost unchanged or only slightly attenuated for a wide range of sound intensity. In most units the phase angles of the gain functions were influenced very little by the change in sound intensity.

The modulation of sinusoidally amplitude modulated broadband noise was also preserved in the recorded discharge pattern of cochlear nucleus units. The amount of modulation of the discharge frequency was, however, in many units slightly less for noise as compared to tones.

The distortion of the shape of the cycle histograms was found to be rather small even in cases where the sound stimulation resulted in an almost 100% modulation of the neural discharge frequency.

Frequency modulated tones gave in these neurons a response pattern which was very similar to the amplitude modulated tones provided that the tone had a frequency which was higher or lower than the CF of the unit. Cycle histograms of the activity evoked by tones the frequency of which was equal to the CF of the unit often showed a distorted shape with the high content of second harmonic. The sensitivity to frequency modulation was higher for tones above CF than below. In the frequency range of 15 kHz a frequency modulation of the size of ± 50 Hz often gave a response pattern which was nearly 100% modulated when the carrier frequency was slightly higher than the CF of the unit.

It is thus clear that changes in amplitude or frequency are enhanced in the response pattern of those units within certain ranges of rate of change, while it is suppressed in other rates. The units can thus be regarded to function like a resonant filter which is tuned to a certain modulation frequency. At that frequency, very small changes in amplitude
or frequency of a pure tone or of the amplitude of a broadband noise are clearly reproduced in the temporal pattern of the neural discharges. It seems rather plausible that this enhancement of amplitude or frequency change in a certain range of modulation frequencies occurs in the neural excitation process in the cochlear nucleus. The fact that sinusoidal modulation is reproduced with very little distortion suggests that even nonsinusoidal amplitude or frequency changes may be reproduced with high fidelity in the temporal pattern of the neural discharges. This reproduction of changes in amplitude or frequency occurs over a large range of sound intensities extending to levels where the main firing rate has reached a steady state value.

Ratliff, F., Knight, B.W., Toyoda, Jun-ichi, and Hartline, H.K. (1967): Enhancement of flicker by lateral inhibition,
Science 158, 392-393.
Ratliff, F., Knight, B.W., and Graham, N. (1969): On tuning and amplification by lateral inhibition,
INTRODUCTION.

Two basic types of model have been proposed to account for the way in which we perceive complex tones. In one class, the "pattern recognition" models, the pitch of a complex is assumed to be derived from certain primary sensations, such as the pitches of certain partials contained in the complex. Examples of this type of model have been presented by Thurlow (1963), Whitfield (1970) and Walliser (1968, 1969). These models all require that at least one partial should be analysable from the complex, and in support of this idea Terhardt (1970) has presented data comparing a) the frequency difference between two tones necessary for the perception of two separated pitches b) the "existence region" of the tonal residue (Ritsma, 1962).

Terhardt finds that the curve "pitch separation" is almost identical with the lower boundary of the existence region, and concludes that "partials lying in a certain frequency region contribute to the perception of periodicity pitch only if those partials generate separate pitches". This conclusion is open to question however, since it is based on a comparison of the results of different subjects, and since it may not be appropriate to use the results for a two-tone complex as a basis for comparison.

A second class of models, the temporal models, assume that the pitch of a complex tone is derived from the time-pattern of vibration at some point on the basilar membrane. These models require an interaction of the individual components, so that the periodicity of the input waveform is preserved. Thus on this type of model there will not necessarily be a relationship between the audibility of the partials in a complex and the pitch of the complex as a whole. Clearly then, one way of distinguishing between these two types of models is to see if a residue pitch can be heard when none of the individual components are separately audible.

Experiment 1. The audibility of partials in a multi-tone complex.

Plomp (1964) found that for multi-tone complexes containing the first twelve harmonics, the partials can only be distinguished if their frequency separation exceeds the critical band width. For two simultaneous tones the pitches can be distinguished for separations less than the
critical band, at least for frequencies up to 5 KHz. A temporal theory of pitch perception can account for this discrepancy as follows. Consider the excitation pattern along the basilar membrane for the different types of tonal complex (fig.1).

Fig. 1. Idealised patterns of excitation on the basilar membrane in response to the two types of tonal complex. The arrows indicate points on the basilar membrane whose pattern of vibration corresponds uniquely to one of the component tones.

For any region on the basilar membrane where the excitation patterns of the individual components overlap, these components will interfere, and the pattern of firing will correspond to the interference pattern (Rose et al., 1969; Brugge et al., 1969). However for a two-tone complex there will be certain regions on the basilar membrane (see arrows) whose pattern of vibration is determined solely by one or the other component tones. Thus the patterns of neural firings derived from these regions will signal the separate periodicities of the component tones. For a multi-component complex this will only occur for the lowest and highest components in a complex, or for components which are widely spaced in frequency.

This line of reasoning can explain Plomp's results, and also indicates that the lowest partial in a multi-component complex should be particularly easily discriminable. This would make Walliser's model more plausible, since the model assumes that the pitch of a complex tone is determined as a subjective sub-harmonic of the pitch of the lowest partial. A further prediction may be made: a band of noise in the frequency region just below the lowest partial in a complex will disturb the pattern of vibration in that region, so that the partial will now no longer have a unique temporal representation on the basilar membrane. Thus that partial should now be no more easily audible than a partial within a multi-tone complex. Experiment 1 was designed to test these predictions.

Method. A method similar to that used by Plomp (1964) was tried, but was found to have certain drawbacks. Given two comparison tones the subjects sometimes seemed able to make their judgments by choosing the tone which was least "dissimilar" to the complex, while at the same time they seemed unable to hear the partial in question as a separate entity. For this reason a method involving only a yes-no decision was used. The subject operated a two-way switch to hear either the multi-tone complex or a single tone. This tone was either identical to the lowest partial in the complex, or was of frequency about ¼ J.N.D's higher than this (so that the comparison tone always fell within the frequency range of the complex).
The probability of the comparison tone being present in the complex was 0.5.

Stimuli were recorded on tape. For a given frequency of the lowest partial, multi-component signals were recorded at a number of different values of the spacing of the components. It was found that subjects maintained a fairly constant "false alarm" rate, of about 10%, so that presumably they were able to maintain a stable criterion. From the trials on which the correct answer was "yes", the frequency separation at which the partial could be separately heard 75% of the time was found by interpolation. Stimuli were recorded in random order, with the provision that no stimuli with the same lowest component were presented on successive trials. Each stimulus lasted 10 seconds and was followed by 5 seconds silence in which the subject had to write down his decision. Each component was at approximately 40 db S.L.

Lowest partials with frequencies 250, 500, 1000, 2000, 4000, 5000, 6000 and 8000 Hz were used. The components in each complex were approximately equally spaced, but only the separation in frequency between the two lowest partials was set exactly. Thus the complex as a whole was non-harmonic, and did not produce any (residue) pitch.

Results.

Condition 1. No noise.

A pilot experiment showed that the audibility of the lowest partial was relatively little affected by the number of other components in the complex. A complete set of data was only obtained for complexes containing four components. The mean results of four subjects are shown plotted in fig.2.

![Fig. 2. Mean separation of the partials necessary for 75% correct identification of the pitch of the lowest partial. Also shown are the results of Plomp (1964) for a two-tone complex.](image-url)
The results of Plomp (1964) for two-tone complexes are shown in the same figure. The results are quite similar, and the results of individual subjects fall around those of Plomp (mean of two observers). These results indicate that the lowest partial in a complex is about as discriminable as a partial in a two-tone complex. However, in this experiment the subjects appeared unable to "hear out" the lowest partial when that partial lay above 5 KHz (unless the other partials were so far separated as to be almost inaudible). Plomp did not report such an effect, and presumably the discrepancy is due to the method differences.

**Condition 2. With Noise.**

For this part of the experiment a 1/3 octave band of noise, with centre frequency 1/3 octave below that of the lowest partial, was added to the complex. The level of the noise was set up as follows. The filter was set so that its centre frequency coincided with that of the lowest partial, and the noise level was adjusted until it was just sufficient to mask the lowest partial presented alone. The centre frequency of the noise was then re-set to 1/3 octave lower. The edge of the noise band was thus separated in frequency from the lowest partial by 1/6 octave.

Results of the same four subjects are shown in fig.3,

> Fig. 3. As for fig.2, but with a 1/3 octave band of noise in the frequency region below the lowest partial. Also shown are the results of Plomp (1964) for a partial within a multi-tone complex.

which also shows the results of Plomp for a partial within a multi-tone complex. Again the results are similar, except that the subjects in this experiment were unable to "hear out" the lowest partial when it lay above 5 KHz. Fig. 4 compares the results for the noise and no-noise conditions. As predicted by the temporal model, the addition of the noise increases the frequency separation necessary for the correct identification of the pitch of the lowest partial, and the amount of difference is similar to that found by Plomp for a two-tone complex and a multi-tone complex.
Experiment 2. Range over which a residue-like pitch may be heard.

The "pattern recognition" models would predict that a complex tone will only have a well-defined pitch when at least one partial in the complex is separately audible. The experiment of the previous section indicates that the limiting case will be reached when the lowest partial is no longer separately perceptible. These models generate two further predictions:

1) Variations in ability amongst individual subjects, in hearing out partials in a complex tone, should be reflected in corresponding variations in the existence region of the tonal residue for those subjects.

2) The effects of noise on the perceptibility of individual partials should be reflected in corresponding changes in the existence region of the tonal residue.

The experiment in this section was designed to test these predictions.

There are a number of difficulties associated with the determination of the "existence region". One is the possibility that combination tones in the frequency region below the frequency of the lowest partial physically presented may influence the percept. To check on this possibility three conditions were used: 1) Stimuli at 40 db S.L. 2) Stimuli at 15 db S.L. 3) Stimuli at 15 db S.L., with a 1/3 octave band of noise in the frequency region below the lowest partial. A second difficulty is that the transition between tonal and atonal is not very well defined, so that subjects have difficulty in making consistent judgments. To try and overcome this a stimulus having certain time-varying properties was used. The stimulus was similar to a sine-wave modulated 100%, except that the carrier frequency alternated between successive sections of the envelope, as in fig. 5.
AUDIBILITY OF PARTIALS IN A COMPLEX TONE.

Fig. 5. Diagrammatic representation of the "cyclic pitch" stimulus. The "carrier" alternates between $f_c$ and $f_c + df$. The envelope repetition rate is $g$.

The two carrier frequencies were chosen to differ by about 2 Hz. Under these conditions a pitch is heard, which varies up and down in a cyclic manner about two times a second. This pitch cannot be derived from the envelope repetition rate $g$, since this is constant, and so the pitch cannot be confused with a "roughness" sensation. Further, the time-varying qualities of this stimulus serve to make the pitch very distinct, so that the subject now has a much clearer decision to make - is a varying pitch audible or not? The stimulus has the disadvantage that it is rather more complicated than the A.M. stimuli used by Ritsma (1962).

Method.

Stimuli were band-pass filtered, using a filter with attenuation rate of 24 dB/octave, with the centre frequency set to equal the "carrier" frequency ($f_c$). This removed any lower partials, and reduced the level of the component corresponding to the envelope rate ($g$) to -60db. Photographs of oscilloscope tracings showed that the main effect of the filtering was to reduce the slight irregularities in the waveform at points between adjacent envelope sections; there were no measurable changes of the position of peaks in the fine structure within the envelope.

Subjects were instructed to listen for a varying pitch. To discourage them from listening to individual partials a comparison tone was provided. This consisted of a pulse-train of rate $g$ Hz, which had been filtered so as to have a timbre similar to that of the "cyclic pitch" stimulus. Subjects were told that the pitch they were to listen for would not be greatly different from that of the comparison tone. They were allowed to switch backwards and forwards between the two stimuli as often as they wished. The task was to adjust a knob (which controlled $g$) until the varying pitch was only just audible. Each match was repeated four times, going alternately up and down, and it was found that subjects were able to perform the task very consistently. In the "noise" condition a 1/3 octave band of noise with centre frequency 1/3 octave below $f$ was added at an intensity just sufficient to mask a tone at 15 dB S.L. of frequency within the passband of the filter.

Results.

The results were similar to those of Ritsma (1962) except that the cyclic pitch could be heard for $g$ up to 1400 Hz (c.f. 800 Hz found by Ritsma), and that, at low S.L.'s the pitch could only be heard for ratios
Fig. 6. The range of values of the envelope repetition rate, \( g \), over which a cyclic pitch is heard, plotted as a function of "carrier" frequency, \( f_c \). Also shown are the results of Ritsma (1962), for a.m. tones.

Fig. 7. Comparison of the separation of the partials necessary for a) 75% correct identification of the pitch of the lowest partial and b) audibility of the cyclic pitch.
AUDIBILITY OF PARTIALS IN A COMPLEX TONE. MOORE

of f/g up to about 12 (c.f. 25 found by Ritsma). (see fig.6). Fig. 6 compares the results of the noise and no-noise conditions at 15 db S.L. with the results of the first experiment at about 40 db S.L. Note that increasing the S.L. tended to increase the range over which the cyclic pitch was audible, so that at 40 db S.L. the dotted curves (showing the existence region of the cyclic pitch) would have been shifted further down. We may note two points:

1) In the two noise conditions the cyclic pitch can be heard when the separation of the partials is less (by about 50% on average) than the separation necessary for the pitches of any of the partials to be separately audible.

2) The addition of the noise has a greater effect on the "separation of the partials necessary for separate audibility of the lowest partial" than on the "existence region of the cyclic pitch", especially at low frequencies. These differences are significant.

Both of these findings are contrary to the predictions from the pattern recognition models. The results of individual subjects also fail to confirm the predictions. The models would predict a correlation between the individual scores in the two experiments e.g. if a subject required a larger-than-average separation of the partials to "hear out" the lowest partial, then that subject should also require a larger than average separation of the partials in order to hear the cyclic pitch. An estimate of the ranks of the subjects was obtained by summing the ranks in individual conditions. These ranks are shown below:

<table>
<thead>
<tr>
<th>subject</th>
<th>C.M.</th>
<th>P.L.</th>
<th>D.J.</th>
<th>B.M.</th>
</tr>
</thead>
<tbody>
<tr>
<td>lowest partial</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>cyclic pitch</td>
<td>1</td>
<td>2</td>
<td>3½</td>
<td>3½</td>
</tr>
</tbody>
</table>

There is no correlation between the two sets of ranks (Spearman's rank correlation coefficient, r = -0.1). Note that the subjects with the highest and lowest ranks in the "lowest partial" experiment (D.J. and B.M.) obtained equal ranks in the "cyclic pitch" experiment.

We may conclude that these results do not support the predictions from the pattern recognition models. This does not mean that such models are ruled out; they may well be appropriate at low harmonic numbers. However it is clear that these models do not provide an explanation for the limits of the existence region of the tonal residue. It is also clear that such models cannot provide a complete explanation of the way in which we perceive complex tones.
References.

Brugge, J. et al. (1969); Time structure of discharges in single auditory nerve fibres of the squirrel monkey in response to complex periodic sounds. 

Plomp, R. (1964); The ear as a frequency analyzer. 

Ritsma, R.J. (1962); Existence region of the tonal residue I, 

Rose, J.E. et al. (1969); Neural correlates of combination tones, 

Terhardt, E. (1970); Frequency analysis and periodicity detection in the sensations of roughness and periodicity pitch. 

Thurlow, W.R. (1963); Perception of low auditory pitch: a multi-cue mediation theory, 
Psychol. Rev., 70, 461-470.

Walliser, K. (1968); Zusammenwirken von hullkurvenperioden und tonheit bei der bildung der periodentonhohe. 
Doctoral Dissertation, Technische Hochschule, Munchen.

Walliser, K. (1960); Uber ein funktionsschema fur die bildung der periodentonhohe aus dem schallreiz, 
Kybernetik, 6, 65-72.

DIRECT COMPARISON OF ABSOLUTE AND RELATIVE PITCH

Andrzej Rakowski

The Academy of Music, Warszawa, Poland

1. Introduction. The term "absolute pitch" refers to the faculty of recognizing and defining the pitch of a musical tone without the use of a reference tone. This faculty is not attributed to the whole population but to less than 1% of it. People not possessing absolute pitch can accurately recognize a tone as higher, lower or identical with a reference tone only shortly after this reference tone is given.

The specific nature of absolute pitch may be best revealed in experiments concerning tonal memory. Such an experiment was conducted by Bachem /1954/ and successfully supported his theory /Bachem 1937/ of "chroma" identification by the possessors of absolute pitch. Harris /1952/, Wickelgren /1969/ and Massaro /1970/ reported experiments in which observers compared tones of different pitch over increasing intervals of time. In all these experiments the observers' task was essentially the same as in AX method of threshold measurements. The error of comparisons, increasing with time interval between A and X signals, was used to indicate the decay of pitch memory trace.

2. Method. The present experiment was so originally designed as to use in it the method of adjustments in which standard and variable signals were separated by various time intervals. For practical reasons the method actually used had to be essentially varied in comparison with the classical method of adjustments or "method of average error". The standard signal S was exposed only once, for a relati-
vely long time /10 sec./ then, after time interval at in which no signal was heard, the variable signal V appeared at the lower or /in equal number of cases/ upper edge of a frequency scale. The subject’s task was to adjust the frequency of a variable tone so as to make this tone equal in pitch with the standard tone S.

Owing to observers’ great experience in pitch comparisons and to the appropriate design of a tone generator, this task was accomplished in a relatively short time Δt /average 3 sec/ which was practically invariant with varying time delay at. The variable signal V of first rapidly then slowly changing frequency, acted as retroactive interference signal with the pitch memory trace of signal S, thus increasing the variability measure of adjustments. However, due to the fact of these conditions being equal for all comparisons, the interfering effect of signal V could be taken as uniformly masking factor not influencing substantially the display of results for various delays at.

Six audiologically normal music students took part in the experiment as subjects. Three of them had genuine, infallable absolute pitch according to Bachem’s /1937/ criterion, and were chosen with the aid of especially designed test method. All subjects passed an intense training in pitch discrimination tasks.

Testing was conducted individually in a sound isolated booth, in sessions not exceeding 2 hours with three intervals. Subjects listened to tones binaurally through the earphones with frequency response equalised for free field conditions. The loudness level for all signals was 40 phones, and frequencies used were 110, 400, 1000, 3000 and 7000 Hz. Time delay Δt was 0, 5, 10, 25 seconds, 1, 5, 30 minutes and 24 hours. The programme was carried out by the experimenter but the clock dial was visible to the subject and he was additionally warned by a signal light two seconds before the end of a silence interval. Rehearsing standard pitch S in silence interval was encouraged except for the longest delay time 24 hours in which subjects were staying outside the laboratory.
Results and discussion

The result of each setting was read out from the digital frequency meter with the accuracy up to 0.1 Hz. Each subject made 20 settings at any given set of measuring conditions and standard deviations were taken as comparative measure of pitch memory decay. The display of these computed standard deviations is shown in Fig. 1 in two separate graphs. On the left the results of 3 subjects not possessing the absolute pitch /NAP/ are presented, and on the right the results of 3 subjects with absolute pitch /AP/. Points represent mean values and vertical lines show the dispersion of results.

The inspection of graphs in Fig 1 shows, that the performances of two groups of subjects differ markedly for longer time delays. At the delays ranging from 10 sec to 5 min the possibility of accurate adjustments based on the exact sensation of pitch stored in memory seems to expire. Facing this fact subjects were forced to develop another criterion in their matching tasks. The NAP listeners tried to keep the standard in memory as long as possible and then having only a slight trace of its sensation tried to make the V tone
as similar to S as possible, or merely guessed. The very important factor here was the memory for a frequency region in which the given tone was located; the accuracy of this matching however was of the order of magnitude smaller than at shorter delays.

The method adopted by the possessors of AP was obviously different. When direct contact was lost with the pitch S stored in memory, they tried to recall it with the aid of their own "imprinted" pitch standards, namely 12 pitches of a chromatic musical scale. E.g. they remembered that 1000 Hz sounds a little lower than C₆ and tried to remember only this small deviation. This technique was particularly useful with S tone belonging to the normal musical scale. Such was the case with S = 110 Hz /A₂/. This tone, as belonging to a group of "imprinted" standards was equally familiar to AP subjects after any time delay.

In spite of very clear instructions given to AP subjects, they were not able in the course of experiment to imprint in the memory additional pitch standards not dependent on the chromatic scale. E.g. they did not remember a 1000 Hz tone "as such". In this aspect they performed similarly to NAP subjects. If such a fact could be confirmed in a long lasting experiment, it might throw a new light on the theory of absolute pitch as a faculty acquired through "imprinting" in early childhood a limited number of standards in individual's memory.

4. References.
Harris, J.D. /1952/: The decline of pitch discrimination with time; J.Exptl.Psychol. 43, 96.
INTENSITY DISCRIMINATION OF NOISE BANDS AS A FUNCTION OF BANDWIDTH AND DURATION

M. Rodenburg

Dept. of Biological and Medical Physics/ Dept. of Otolaryngology, Medical Faculty Rotterdam, The Netherlands

Frequency resolution and temporal integration are two important properties of the auditory system. In this study we investigated whether these processes play a role in intensity discrimination of noise bands.

1) Determination of the modulation threshold of noise bands as a function of the bandwidth.

A band of noise can be regarded as a sinusoidal signal with irregularly varying amplitude and phase. The rate of the amplitude fluctuations is inversely proportional to the bandwidth and is independent of the center frequency of the noise band. For small bandwidths these fluctuations may influence the intensity discrimination. For large bandwidths the fluctuations are rapid and smoothed out by the auditory system. In that case the difference limen for noise is 0.5 - 1 dB. When the bandwidth is small the fluctuations are slow and deteriorate intensity discrimination. Zwicker (1956) has found that the difference limen increases when the bandwidth decreases below 1000 Hz. Besides the difference limen appeared to be independent of the center frequency. Bos and De Boer (1966) have confirmed that the difference limen for noise bands depends almost exclusively on the bandwidth and not on the center frequency. Maiwald (1967) investigated the effect of the bandwidth, too. He used noise bands with a lower cut-off frequency of 6.4 KHz and varied the upper cut-off frequency. The difference limen decreased with increasing bandwidth up to 3.4 KHz. For band-
widths beyond 3.4 KHz the difference limen was constant. He concluded that this was due to the fact that the critical band limits the bandwidth beyond this value. In that case we would expect that the relation between the difference limen and the bandwidth will be a decreasing function up to a bandwidth corresponding with the critical band. Beyond that bandwidth the difference limen will be independent of the bandwidth. This transition should occur at a smaller bandwidth for noise bands with a low center frequency, since the critical bandwidth decreases with frequency. Therefore we investigated the effect of the bandwidth at different center frequencies. A modulation method was used in this experiment. The modulation threshold was determined for noise bands of several bandwidths at the center frequencies 500 Hz, 2000 Hz and 8000 Hz at a modulation frequency of 10 Hz. A forced-choice method was used. The percentage correct answers was determined as function of the modulation depth. The modulation depth at the 75% point was taken as the modulation threshold. Two subjects participated in this experiment. The experimental results are given in Fig. 1.

Fig. 1.
Modulation threshold as a function of the bandwidth of the noise at three center frequencies.
The results of the two observers are similar. The modulation threshold decreases with increasing bandwidth up to a certain value and is constant beyond it. This value is 200 Hz for a center frequency of 500 Hz, 600 Hz for a center frequency of 2000 Hz and 2500 Hz for a center frequency of 8000 Hz.

2) Determination of the difference limen of noise pulses as a function of the duration.

From threshold measurements with tone pulses as a function of their duration it is known that the activity of the auditory system includes integration. Most authors assume that the integration time is about 200 msec. and independent of the frequency (Zwicker and Feldtkeller 1967, Zwislocki 1960, 1965). Plomp and Bouman (1959), however, found that the integration time varied with frequency. At 250 Hz the integration time appeared to be 375 msec. and at 8000 Hz only 150 msec. It is obvious that the energy of noise pulses fluctuates. The standard deviation $\sigma_T$ of the energy distribution decreases with increasing duration. Rice (1954) derived an equation which gives the ratio of the standard deviation and the mean of the energy distribution for relatively narrow bands of noise:

$$\frac{\sigma_T}{\bar{E}} \sim \frac{1}{\sqrt{T(f_b-f_a)}} , \text{if } (T(f_b-f_a) \gg 1).$$

In this formula $\bar{E}$ is the mean energy of the noise pulse with duration $T$, $\sigma_T$ the standard deviation, $f_b$ and $f_a$ the upper and lower cut-off frequencies. According to this formula the ratio $\sigma_T/\bar{E}$ decreases with increasing $T$. The D.L. will also decrease with increasing $T$. When the duration of the noise pulses is longer than the integration time of the ear, it is the integration time, rather than the duration of the noise pulses, which determines the threshold. The D.L. of noise pulses as a function of the duration should thus decrease with increasing $T$ only up to the integration time. For longer durations the D.L. should be constant. This provides another possibility to estimate the integration time of the auditory system. The D.L. for intensity of noise pulses has been determined as a function
of the duration. The duration was varied from 25 msec. up to 1 sec. White noise or bandpass filtered noise (bandwidth 300 Hz) at the center frequencies of 500 Hz, 2000 Hz and 8000 Hz were used. Two noise pulses of different intensities were presented with an interval of 0.5 - 1 sec. between the two pulses. The sequence of the noise pulses was determined by a forced-choice generator. A psychometric curve was determined and the 75% point taken as the difference limen. The experiments were carried out at 30 dB sensation level of the pulse with the lower intensity. Two subjects participated in this experiment. Fig. 2 shows the experimental results. For white noise with a duration longer than 100 msec. the two subjects show difference limens of respectively 0.5 and 0.9 dB. For shorter durations the difference limen increases. For a center frequency of 8000 Hz our estimate of the integration time (inflection points in the curves) is the same as for white noise.

Fig. 2.
Difference limen of noise pulses as a function of the duration.
Discussion
In the first experiment we saw that the modulation threshold decreased with increasing bandwidth up to a critical value. These values are about twice as large as the critical bands for these center frequencies. The rate of increase with increasing center frequency is comparable with that of the critical band. This makes it likely that the critical band mechanism plays a role in the detection of amplitude modulation of wide band noise as was proposed by Maiwald.

The modulation threshold of white noise as a function of the modulation frequency shows a low-pass characteristic with a cut-off frequency of about 60 Hz (Rodenburg, 1972). Attenuation of the fluctuations beyond this frequency causes an increase in the modulation threshold with increasing bandwidth. For an ideal rectangular filter the expected number of maxima of the envelope per second is \(0.641(f_b - f_a)\), \(f_b\) and \(f_a\) being the upper and lower cut-off frequencies, respectively (Rice 1954). This means that the expected number of maxima is the same as the cut-off frequency of the attenuation characteristic (60 Hz), when the bandwidth is about 90 Hz. The curves of Fig. 1, indeed, show a breakpoint at this bandwidth.

The results of the experiments with the noise pulses of different durations show that the integration time is a function of the center frequencies. Our data show that the integration time is about a factor 2 longer for low frequencies than for high frequencies, which is in agreement with the results of Plomp and Bouman. For white noise the integration time is the same as for high frequencies. Evidently the system uses the smallest integration time available given the frequency content of the stimulus.

The difference limen for noise bands of 300 Hz bandwidth and a duration between 200 msec. and 1 sec. is larger than for white noise. This is due to the reduction of the bandwidth. The difference limen increases with increasing center frequency for durations longer than the integration time. This increase is mainly due to the decrease of the integration time, although it cannot be explained completely by this effect. This is illustrated by the fact, that the difference limen at 8000 Hz and
100 msec. is still larger than at 500 Hz and 100 msec., although in this case the durations are the same. *)

References

*) Possible influences of the critical band mechanism would result in still a greater discrepancy.
Identical sounds that start within about a twentieth of a second of each other at two different sources are normally heard as a single sound. The fused sound appears to originate at the source that leads in time. The leading sound's predominance is well known as the precedence effect (Wallach, Newman, & Rosenzweig, 1949) or Haas effect (Haas, 1951). However, the two sounds need not be identical in order to fuse into a single image. The frequency of a tone presented through an earphone to one ear may differ from the frequency to the other ear by 2-15% and still the listener reports a single pitch (Thurlow & Bernstein, 1957). Even when heard as two pitches at much larger frequency differences, the resulting chord may have only a single spatial location. Hence, pitch fusion and position fusion are independent, suggesting separate neural mechanisms for pitch and localization. The present experiments investigated the limits of position fusion for tones of different frequencies presented through loudspeakers in a free field and through earphones.

PROCEDURE

Free-field measurements were made in a 4 m by 4 m anechoic room. Two double-cone loudspeakers were placed 3 m from the listener's head, one 45° to the left and the other 45° to the right of the median plane. Tones were presented separately through third-octave filters via each speaker for 6 or 500 msec approximately every 2 seconds until the subject had marked the position or positions of the sound or sounds and their temporal order on a prepared scoring sheet. The onset time difference, ΔT, between the two speakers was varied from 0 to 500 msec; in half the trials the left speaker, which always had the lower frequency, came on first. The
frequency separation, $\Delta F$, between the tones was varied from 0 to 4200 Hz. Tones were presented at approximately 45, 65, and 85 phons as determined for each of the four subjects by loudness matches.

Earphone measurements were made with TDH-39 earphones. Tones lasted 700 msec, had a 1-msec rise-fall time, and were repeated after a silent interval of 600 msec. They were set equally loud for each listener with the lower frequency always at 70 dB SPL. The $\Delta T$ was varied from 0 to 2 msec, and $\Delta F$ from 0 to 4530 Hz with the geometric mean constant at 2000 Hz. On a prepared sheet, the subject marked the position of the sound image or images in his head. Six subjects, including the author, made two judgments of each combination of $\Delta F$ and $\Delta T$.

RESULTS

1. **Position fusion.** Figure 1 shows how the percentage of

![Figure 1. Fusion as a function of $\Delta F$. (In loudspeakers, tone duration was 500 msec and level was 65 phons.)](image)
FREQUENCY SELECTIVITY AND SOUND LOCALIZATION

SCHARF

fused judgments, in which the sound image had a single position, decreased as the frequency separation between the two tones increased. Fusion under earphone and free-field listening varies in much the same way with $\Delta F$. Fusion functions for $\Delta T = 2$ msec are similar to the one shown, but at longer $\Delta Ts$ the functions decrease more rapidly under both listening conditions.

The interactive effect of $\Delta T$ and $\Delta F$ on fusion is shown in Figure 2 for free-field listening at 45 and 65 phons. The closer the two tones in frequency, the longer the onset time disparity at which they usually fuse. Fusion occurs more readily at 45 than at 65 phons, and more readily at 65 than at 85 phons (not shown). Measurements with 6-msec tones gave similar results except that fusion occurred more often at small $\Delta Ts$. Data for $\Delta F = 0$ Hz were like those for 160 Hz. At the narrow $\Delta Fs$ fusion breaks down between 20 and 60 msec, values similar to those reported for speech and music (Haas, 1951; Wallach, et al., 1949).

2. Localization. Figure 3 shows that the precedence effect holds when two separated loudspeakers are energized by the same pure tones. Data at 85 phons were similar to those at 65 phons. Generally, localization was better (i.e. was toward the leading speaker) at the higher levels than at 45 phons, and also for 6-msec tones than for the 500-msec tones shown in Figure 3. Although the sound seemed to come from the same side as the leading loudspeaker, the listener heard the 500-Hz and 2000-Hz tones in back 40% of the time.

As Figure 4 shows, even when the left and right tones differed as much as 1900 Hz listeners continued to locate the fused sound on the side of the leading speaker, but only when the left, low-frequency speaker led. When the right, high-frequency speaker led, the sound was often heard in the center or toward the left. This asymmetry did not apply to the unfused sounds. Their separated images were almost always localized at the loudspeakers. Their temporal order was correctly reported at a $\Delta T$ of 60 but not 20 msec.

3. Lateralization. The lateralization (localization of a
Figure 2. Number of fusion judgments as a function of onset time difference at four frequency separations and two levels.
Figure 7. Localization as a function of onset time difference at three frequencies and two levels. (Frequencies were the same in both loudspeakers. Δf was 0 Hz.)

RIGHT LEADS LEFT SPEAKER (ΔT) IN MSEC

PERCENTAGE JUDGED LEFT (O) OR RIGHT (●)

5000 HZ

2000 HZ

500 HZ

45 PHONS

65 PHONS

100 80 60 40 20 0

-60 -20 20 60

-60 -20 20 60

-60 -20 20 60

-60 -20 20 60
sound image inside the head) of fused tones presented through earphones is shown in Figure 5. Each point is based on 6 to 12 judgments by three to six subjects. Between ΔFs of 0 and 300 Hz, which is the critical band at 2000 Hz, lateralization depends on ΔT in much the same way. At 600 Hz, where five subjects still fused, the sound image remained more centered and the low-frequency side was heavily favored. The means at 1200 Hz are for only three subjects and may not be representative.

Figure 4. Localization as a function of onset time difference at four ΔFs. (Level was 65 phons, duration 500 msec.)
Figure 5. Lateralization of a dichotic tone presented thru earphones as a function of onset time difference at six $\Delta F$s.
DISCUSSION

In position fusion and localization we see little evidence of the auditory system's fine frequency selectivity. This finding is not inconsistent with the van Bergeijk-Békésy model of localization (van Bergeijk, 1962). Localization is presumably determined by the difference between the number of relevant neural units on the left and right sides of the auditory nervous system that respond to a given sound. For example, a sound reaching the left ear first would activate more right-side units than left-side units. It is reasonable to assume that a relevant unit would continue to be sensitive to interaural onset time differences even when the tones to the two ears differ in frequency, provided both tones are within the unit's response area or tuning curve. Since these binaural units have typical tuning curves and respond to a wide range of frequencies, small interaural frequency differences ought not disturb localization. Similarly, fusion would hold so long as some minimum number or proportion of the involved units are affected by inputs from both ears.

Consideration of the typical tuning curve's shape suggests two specific psychophysical predictions. First, fusion should occur at wider $\Delta F$s as intensity increases because a given unit responds to a wider range of frequencies at higher intensities. Although Perrott (1970) did measure somewhat less pitch fusion near threshold than at higher levels, the present free-field results plus earlier earphone measurements show the opposite tendency—fusion gets poorer as SPL increases.

A second, more successful prediction is that the localization of a dichotic tone pair would be biased toward the low-frequency side. The low-frequency member of the pair ought to affect more units than the high-frequency member since the typical tuning curve is steep toward high frequencies and shallow toward low frequencies. The lateralization data do show a consistent bias toward the low-frequency side at a $\Delta F$ of 600 Hz and a smaller bias at 300 and 220 Hz. Perhaps, the asymmetry in free-field localization (Figure 4) also reflects a low-frequency bias.

(The writer greatly appreciates the hospitality of Prof. J. Jauhiainen and Dr. E. Lampio at the Acoustics Lab. of the Helsinki University of Technology as well as the technical advice of J. Toivanen and the assistance of M. Pulakka. Research supported by a grant from N.I.H.)


AN INTEGRABLE MODEL FOR THE BASILAR MEMBRANE

M. R. Schroeder

Drittes Physikalisches Institut, Universität Göttingen, Germany

ABSTRACT
Assuming a logarithmic relationship between place and characteristic frequency, proportionality between the latter and phase velocity and a loss-factor independent of place, the wave equation for the basilar membrane displacement can be integrated analytically. The main theoretical results are as follows:

1. The low-frequency phase-slope equals approximately $-K\omega_r(x)$ where $K$ is a dimensionless constant and $\omega_r(x)$ is the characteristic frequency at place $x$.

2. The asymptotic high-frequency amplitude slope equals $-6(K+1)$ dB/octave.

3. Just above resonance the phase reaches a minimum value of about $-K$ Radians.

The Rhode (1971) low-frequency phase-slope data yields $K = 27 \pm 3$. The asymptotic high-frequency amplitude slope should therefore be $(168 \pm 18)$ dB/octave.

Surprisingly, the great frequency selectivity of the cochlea as reflected in the large numerical value of the high-frequency amplitude-slope is not determined by hydro-mechanical losses, which primarily influence behavior near resonance, but the constant $K$ whose physical meaning is approximately "$2\pi$ times the number of wavelengths" on the membrane for low frequencies. With the above value for $K$, the number of wavelengths becomes about 5 - considerably more
than found by von Békésy.

Nevertheless, it is interesting to note that the phase-slopes and high-frequency amplitude-slopes of von Békésy (1960, pp. 446-463) are not inconsistent in the light of the present theoretical calculations: If the membrane becomes stiffer at the high amplitudes at which von Békésy made his measurement, both phase- and amplitude-slopes would be decreased.

INTRODUCTION

Figure 1 shows a simplified electrical model of a section of the basilar membrane [cf. Zwislocki (1950) and Peterson and Bogert (1950)]. The total membrane is considered a series connection of infinitely many, infinitely short sections with parameters $L_1$, $R$, $L$, $C$ depending on the place $x$. At the basal end ($x = 0$) the volume velocity (represented in the model by the current through $L_1$) is prescribed by the motion of the stapes; at the apical end the "boundary condition" is given by the vanishing pressure difference (represented by the voltage across the terminals), i.e., a short-circuit in the electrical analog.

The most apparent omission in the model shown in Fig. 1 is a component representing the restoring force due to elastic coupling between adjacent segments. However, this force can be represented in the model by including in the shunt inductance $L$ a term proportional to the membrane surface tension $\sigma$ and inversely proportional to the phase velocity squared. Since, in this paper, the shunt impedance is not determined from the membrane mass density and tension but rather from the resonance frequency at a place $x$ (cf. Eq. 3), the effect of the membrane tension is automatically included as far as behavior at resonance is concerned.

However, because the basilar membrane is a dispersive transmission medium, i.e., the phase velocity $c$ depends on frequency, the inductance $L$ is not a constant, but decreases with frequency (cf. Eq. 27). One result of this dispersion
is a broadening of the frequency response at a given place in addition to that due to losses.

**THE TRANSFER FUNCTION**

The (natural) logarithm, \( g \), of the voltage ratio between output and input of a section (when terminated in its characteristic impedance\(^1\)) is given by

\[
g = -2 \arcsinh \left( \frac{1}{2} \left[ \frac{j \omega L_1}{(R+j \omega L+1/j \omega C)} \right]^2 \right),
\]

For \( \Delta x \to 0 \) one may therefore write

\[
g = -\left( \frac{j \omega L_1}{(R+j \omega L+1/j \omega C)} \right)^2.
\]

Introducing the "resonance frequency" of the shunt arm

\[
\omega_r(x) = (L \cdot C)^{-\frac{1}{2}}
\]

and its loss-factor (reciprocal "Q")

\[
\delta(x) = \omega_r \cdot R \cdot C,
\]

and the phase velocity for low frequencies

\[
c(x) = (L_1 \cdot C)^{-\frac{1}{2}} \cdot \Delta x,
\]

Eq. 2 may conveniently be rewritten as

\[
g(\omega, x) = -j(\omega/c) \cdot \left[ 1 - (\omega/\omega_r)^2 + j \delta \omega/\omega_r \right]^{-\frac{1}{2}} \cdot \Delta x,
\]

where \( c \) and \( \omega_r \), and therefore \( g \), are functions of \( x \).

The logarithmic voltage ratio \( G(\omega, x) \) between a place \( x \) and the basal input (\( x = 0 \)) is obtained by integration of Eq. 6:

\[
G(\omega, x) = \int_0^x g(\omega, \xi) d\xi/\Delta x
\]

or with the integration variable \( z = \omega/\omega_r(\xi) \):

\[
G(\omega, x) = -j \omega \int_{z_2}^{z_1} c^{-1} \cdot \left[ 1 - z^2 + j \delta \cdot z \right]^{-\frac{1}{2}} (dz/d\xi) d\xi,
\]

where \( z_1 = \omega/\omega_m \) and \( z_2 = \omega/\omega_r(x) \), and \( \omega_m \) is the resonance...
frequency at the basal end: $\omega_m = \omega_r(0)$.

THREE ASSUMPTIONS

We will now make the following assumptions:

1. A logarithmic relation between place $\zeta$ and the characteristic ("resonance") frequency $\omega_r(\zeta)$:

$$\zeta = d \cdot \ln[\omega_m/\omega_r(\zeta)]$$

where $d$ is the distance along the membrane for which the characteristic frequency changes by a factor $e = 2.718...$

2. Proportionality between (low-frequency) phase velocity $c(\zeta)$ at a place $\zeta$ and its characteristic frequency $\omega_r(\zeta)$:

$$c(\zeta) = \frac{\omega_r(\zeta)}{k}$$

where $k$ is a constant with the dimension cm$^{-1}$.

The proportionality, Eq. 10, implies that the ratio of $L$ ("membrane" mass) and $L_1$ (fluid mass) is independent of place.

3. Independence from place of the loss-factor $\delta$:

$$\delta(\zeta) = \delta.$$  

At least two of these conditions (Eqs. 9 and 11) appear to be reasonably well met above some species-dependent frequency (about 1 kHz in man).

With Eqs. 9-11, analytic integration of Eq. 8 becomes possible. The formal result is

$$G(\omega, x) = -j \cdot K[\arcsin(z_2 - j \delta) - \arcsin(z_1 - j \delta)]$$

where

$$K = k \cdot d, \quad z_1 = \frac{\omega}{\omega_m} \quad \text{and} \quad z_2 = \frac{\omega}{\omega_r} = \left(\frac{\omega}{\omega_m}\right) \cdot \exp(x/d).$$

A factor $(1 + \delta^2)^{-1/2}$ in the arguments of the arccos-functions has been omitted because it can be absorbed into $\omega_m$ whose value would be changed by less than 3 % for $\delta < 0.5$. 
The physical significance of the dimensionless constant $K$ introduced here is that $K/2\pi$ is approximately the number of wavelengths on the basilar membrane for radian frequencies small compared to $\omega_m$ (see Section on "Phase above Resonance").

[The three conditions given in Eqs. 9 to 11 are sufficient but not necessary for the integrability of Eq. 8. As the least restrictive condition leading to equivalent results, it suffices that the loss-factor be constant and that $L_1$ be proportional to $L'[\delta \ln(LC)/\delta x]^2$. Thus, there is a possibility of analytical treatment of cases with other than logarithmic place-frequency relationships.]

**PHASE AT LOW FREQUENCIES**

Splitting $G(\omega, x)$ into its amplitude and phase parts:

$$G(\omega, x) = \ln[A(\omega, x)] + j\phi(\omega, x),$$

Eq. 12 yields for $\omega \ll \omega_r$ and $\delta \ll 2$

$$\phi(\omega, x) = -K(\omega/\omega_r)(1-\omega_r/\omega_m),$$

i.e., a linear dependence of phase on frequency.

Assuming $2\omega_m = 2\pi \cdot 50$ kHz, Rhode's (1971) low-frequency phase data yields for the important dimensionless constant $K$:

$$K = 27 \pm 3$$

for squirrel monkey. Similar values may obtain for related species (including man).

**AMPLITUDE ABOVE RESONANCE**

For $\omega > \omega_r (z_2 > 1)$, Eq. 12 yields

$$A(\omega, x) = \{\omega/\omega_r + [(\omega/\omega_r)^2-1]^{1/2}\}^{-K},$$

where the identity

$$j \arcsin(u) = j\frac{\pi}{2} + \ln[u+(u^2-1)^{1/2}]$$

has been used and the loss-factor $\delta$ neglected,
For $\omega \gg \omega_r$, Eq. 16 may be further simplified and rewritten to bring both frequency and place dependence into evidence:

$$A(\omega, x) = \left(\frac{2\omega}{\omega_m}\right)^{-K} e^{-kx}.$$  \hspace{1cm} (18)

Thus, the asymptotic slope of amplitude vs. frequency equals $-6 \cdot K$ dB/octave.

Including $+6$ dB/octave from the stapes displacement-to-pressure transformation and $-12$ dB/octave for the transformation from the latter to the basilar membrane displacement, the theoretical slope becomes $-6(K+1)$ dB/octave.

With the experimental value of $K = 27 \pm 3$, from the low-frequency phase measurements, one obtains for the overall asymptotic slope

$$-(168 \pm 18)$$

a value somewhat above the range quoted by Rhode (-50 to -140 dB/octave) and Johnstone et al. (1970) (-95 dB/octave for guinea pig at 18 kHz). For frequencies nearer resonance and small loss-factors, the attenuation rate is even larger than the asymptotic values.

The significance of the theoretical relation between low-frequency phase-slope and high-frequency amplitude-slope lies in the possibility of determining the latter from the former for comparison with neural tuning curves. Considerations of additional sharpening mechanisms should thereby be facilitated with important implications for a refined understanding of hearing.

Equation 18 also shows that the amplitude, sufficiently above resonance, decays exponentially with place with a decay length $k^{-1} = d/K$. With $d = 0.6$ cm (from Rhode's 2-position experiment), $k^{-1} = 0.02$ cm.

PHASE ABOVE RESONANCE

Expansion of Eq. 12 just above resonance ($z > 1.1$) gives for the phase
AN INTEGRABLE MODEL FOR THE BASILAR MEMBRANE  SCHROEDER

\[ \phi(\omega, x) = -K \left( \frac{1}{2} \pi - \frac{1}{2} \delta \cdot \left( \frac{\omega}{\omega_r} \right)^2 \right)^{\frac{1}{2}} \cdot \arcsin \left( \frac{\omega}{\omega_m} \right) \right) . \]  

(19)

This phase angle reaches a minimum about 20% above resonance:

\[ \phi_{\text{min}} \approx -1.15 \cdot K . \]  

(20)

It is interesting to note that the minimum phase angle is practically independent of place for \( \omega_r < \omega_m \). (Of course, near the stapes the phase approaches zero.)

The significance of \(-\phi_{\text{min}}\) is that it equals \(2\pi\) times the number of wavelengths on the basilar membrane. With Eq. 20 and \( K = 27 \), the number of wavelengths becomes about 5 — in marked contrast to earlier conceptions which usually envisaged considerably smaller numbers of wavelengths.

For comparison with experimental data, the phase angle of the stapes-displacement to fluid-pressure transformation (roughly \( \pm \pi \)) and from the latter to the basilar membrane displacement (close to \(-\pi\)) must be added. Thus

\[ \phi_{\text{minB}} = -1.15 \cdot K - 1.57 . \]  

(21)

With \( K = 27 + 3 \), one obtains,

\[ \phi_{\text{minB}} = -32 \pm 4 \text{ Radians}, \]  

(22)

which agrees reasonably well with Rhode's data (-22 to -28 Radians).

AMPLITUDE RATIO AT RESONANCE

Expansion of Eq. 12 at resonance \( z_1 = 1 \) for \( \delta << 1 \) and \( \omega_r < \omega_m \) yields (up to linear terms in \( \delta \)):

\[ \ln[A(\omega_r, x)] = -\frac{1}{2} K \left( (2\delta)^2 - \delta \right) . \]  

(23)

At frequencies considerably below resonance \( z_2 << 1 \), one has

\[ \ln[A(\omega, x)] = \frac{1}{4} K \cdot \delta \cdot \omega^2 \left( \omega_r^{-2} - \omega_m^{-2} \right) \]  

(24)

For comparison with experiment the transformations from the stapes displacement to the fluid pressure and the latter to the basilar membrane displacement must be included. Thus, the overall ratio of the membrane displacement amplitude at resonance to the amplitude at some lower frequency \( \omega \) becomes
\[ \ln[A(\omega_r, x)/A(\omega, x)] = \ln(\omega_r/\omega \cdot \delta) \]
\[ -1/4 K [2 \cdot (2\delta)^2 - 26 + \delta \cdot \omega^2 (\omega_r^{-2} - \omega_m^{-2})]. \quad (25) \]

For frequencies 10 times below resonance, Rhode's data show a level decrease of roughly 30 dB. With these values and \( K = 27 \), Eq. (25) yields
\[ \delta = 0.02, \quad (26) \]
a value for the loss-factor which seems unexpectedly small.

GROUP DELAY TO THE CHARACTERISTIC PLACE

The group delay is obtained by differentiating \(-G(\omega, x)\) with respect to \( \omega \) and taking the imaginary part. With Eq. 12
\[ \frac{\delta G}{\delta \omega} = -j K \cdot \omega_r^{-1} \left[ (1 - (z_2 - j \delta)^2)^{-1/2} - 1 \right] \cdot \omega_r/\omega_m \]
\[ \left[ (1 - (z_1 - j \delta)^2)^{-1/2} \right]. \quad (27) \]
Thus, at resonance \( (z_2 = 1) \) and for \( z_1 = \exp(-x/d) \ll 1 \):
\[ \tau_{gr} = K \cdot \omega_r^{-1} \left[ (2\delta)^{-1/2} - \omega_r/\omega_m \right]. \quad (28) \]

To this must be added the (usually negligible) group delay between fluid-pressure and basilar-membrane displacement \((2/\omega_r \cdot \delta)\).

Comparing the group delay for the resonance frequency with that at low frequencies (from Eq. 13):
\[ \tau_{go} = K \cdot \omega_r^{-1} (1 - \omega_r/\omega_m), \quad (29) \]
we see that the former is larger by a factor that depends strongly on \( \delta \).

With \( \tau_{gr}/\tau_{go} = 2 \) (from Rhode's data) and for \( \omega_r/\omega_m = 1/7 \), one obtains
\[ \delta \approx 0.1, \]
a value considerably larger than the loss factor obtained from the amplitude at resonance. It is improbable that this discrepancy is due to experimental error. More likely, it reflects an inappropriateness of one (or several) of the assumptions made - or even a deficiency of the model itself.
IMPULSE RESPONSE

In order to determine the impulse response of the basilar membrane, one can proceed as follows. The pressure difference between the scalae at place x as a function of time, \( p(t,x) \), for a pressure difference at the stapes \((x=0)\) given by a Dirac delta-function is obtained by evaluating the inverse Fourier transform of \( \exp[G(\omega,x)] \), with \( G(\omega,x) \) from Eq. 12.

For radian frequencies an octave or more below \( \omega_m \), the second arcsin-term in Eq. 12 can be replaced by its argument. Within this approximation, it represents a negative delay of \( K/\omega_m \), or roughly 0.1 msec, and a gain of \( K\cdot\delta/2 \), or about 6 dB.

The first arcsin-term has a well-known Fourier transform:

\[
p(t,x) = K \cdot t^{-1} \cdot \exp(-\delta \cdot \omega_r t/2) \cdot J_K(\omega_r t),
\]

where \( J_K \) is a Bessel function of the first kind.

The impulse response given by Eq. 30 has an interesting scaling property: a multiplication of \( \omega_r \) by a factor \( M \) is equivalent to a time compression by \( M \) and an amplitude expansion by the same factor. The total energy contained is likewise multiplied by \( M \), a fact which also follows from consideration of the transfer function in the frequency domain. Thus, for \( \omega_r < \omega_m/2 \), \( p(t,x)/\omega_r \) plays the role of a "universal" impulse response for the model under consideration. (In computing \( p(t,x) \), use can be made of the fact that a Bessel function with a half-integer index can be represented by elementary functions.)

For comparison with experimental data, the impulse response between stapes displacement and basilar membrane displacement \( m(t,x) \) is of interest. It is obtained from Eq. 30 by differentiating \( p(t,x) \) with respect to time, multiplying with the input impedance of the cochlea as seen from the stapes (essentially a real constant for radian frequencies an octave or more below \( \omega_m \)) and convolving with the impulse response of the shunt arm at place \( x \). Thus,

\[
m(t,x) = \text{const.} \frac{\partial}{\partial t} p(t,x) \ast [\exp(-\delta \cdot \omega_r t) \cdot \sin \omega_r t]/\omega_r,
\]
where * stands for convolution in the time domain.

CONCLUSION

A simplified model of the basilar membrane has been made specific by assuming that resonance frequency and (low-frequency) phase velocity are (similar) exponential functions of place and that the loss factor associated with the membrane motion is independent of place. The model so specified can be integrated analytically yielding relatively simple, explicit expressions for many physical observables such as phase delay, amplitude slopes, group velocities, impulse responses etc. Several of Rhode's and Johnstone's experimental Mössbauer results can be matched with the assumption of but one critical parameter: K. (In fact, even von Békésy's measurement do not contradict the model, albeit for a different value of K.) Thus, the same value of $K$ ($\approx 27$) can predict the low-frequency phase-slope, the high-frequency amplitude-slope and the correct value for the minimum phase reached above resonance.

FOOTNOTES

1 This assumption is justified if the parameters of the model vary sufficiently slowly as a function of place. More specifically, the logarithmic derivatives with respect to $x$ of $j\omega L_1$ and $(R+j\omega L+1/j\omega C)^{-1}$ must be small compared to the geometric mean of these two quantities. This condition of quasi-homogeneity is, in fact, fulfilled even near resonance. Thus, there should be no significant standing waves on the basilar membrane. Indeed, standing waves have never been seen, except possibly beyond the characteristic place at levels considerably below the resonance level (cf. Rhode (1971), Fig. 8). Did Nature know this when she designed the cochlea?

2 Exact values for $\omega_m$ may not be available. W.C. Stebbins et al. (1966) give $2\pi \times 45$ kHz for Macaca. Rhode (private communication) mentions that squirrel monkey can hear up to $2\pi \times 60$ kHz.
REFERENCES


Fig. 1: Equivalent electrical circuit for one section of the basilar membrane. The voltage across the capacitor is proportional to the membrane displacement.
ANALYSIS OF FLUID-ELASTIC INTERACTION IN COCHLEA

C. R. Steele

Presently, Eidgenössische Technische Hochschule, Zürich

ABSTRACT

A model of the cochlea is analysed by the "phase integral" asymptotic method, with which an a priori assumption of "long" or "short" waves is not necessary. The possibility for independent rotation of the arches of Corti, bending of the remaining portion of the basilar membrane, and deflection of the tip of the bony shelf is admitted. Results offer an explanation for the differences in the observations by Békésy (1960) and the "place" from damage evidence and the phase from recent work by Rhode (1971), using the Mössbauer effect. For frequencies over 1 KHz, it appears that no "long waves" occur in the cochlea. The model gives excellent agreement with Békésy's data for "click" propagation along the cochlea.

COCHLEAR MODEL

The model in Fig.1 is distinguished by the taper angle of the scalae walls, which permits a solution for the 3-dimensional fluid motion as a Fourier series in $\theta$. For the fluid pressure on the basilar membrane,

$$ p(r,\theta,o) = B_0(r) + B_1(r) \cos N\theta + ... $$

$$ + B_{1/2}(r) \sin N\theta/2 + ... $$

(1)

where $B_0$ gives the average, $B_1$ the first deviation due to the flexing of the basilar membrane (Mode I, Fig.2), and $B_{1/2}$ the asymmetric effects due to the rotation of the arches of Corti (Mode II, Fig.2) and the displacement of the bony shelf tip (Mode III, Fig.2). Because of the narrow taper, $N = 350$, which provides a convenient large parameter for the asymptotic expansion of the solution.

The stiffnesses are (Mode I) $K$, plate bending; (Mode II) $k$, against uniform rotation of arches, and $GJ$, against relative rotation between arches; (Mode III) $\mathcal{K}$, against uniform displacement, and $EI$, against relative displacement between arches. The simplifying assumptions are that the edges are hinged, and that Modes II and III are elastically uncoupled. Values are chosen so that the model has exactly the behavior under the static point load that Békésy measured, and fits Békésy's pressure-volume displacement values, except near the helicotrema, where the model is about 3 times more compliant. It appears that the use of a more realistic, but more complicated, clamped edge condition would resolve this difference.

It is found that, even with the addition of the actual tectorial membrane and the soft cells, the mass of the organ of Corti is negligible in comparison with the fluid loading, in agreement with Békésy's observation. Only the damping due to fluid viscosity is considered.
PLATE STRIP IN INFINITE FLUID

A simple solution can be obtained for the plate strip (Mode I) in a fluid, if the wall constraints are removed. For an inviscid, incompressible fluid and a massless plate, there occurs a point of transition $r_t$, given by

$$K(N/r_t)^5 = 2\rho\omega^2$$

(2)

at which a traveling wave develops, moving toward the wide end of the plate, due to a prescribed uniform, far-field pressure in the fluid, fluctuating with the frequency $\omega/2\pi$. Thus, using only $K$ from Békésy's point load tests, Eq.1 gives the lower curve in Fig.3. If the arches of Corti are assumed to be perfectly flexible, so the total width of the basilar membrane $3b/2$ is used, then $N$ in Eq.2 is increased by 50%, and the upper curve in Fig.3 is obtained. The actual localization is between these two extremes. This indicates that 1) the basilar membrane is indeed, in the terminology of mechanics, to a good approximation not a "membrane" but a plate with negligible tension, 2) the development of traveling waves is an intrinsic property of the tapered plate in a fluid - independent of size of fluid chambers, mass of the plate, or damping. This is in agreement with Békésy's experimental conclusions.

SOLUTION

The solution for the model of Fig.1 is obtained, for a pure tone excitation at the stapes, in the form

$$\eta(r) = \psi(r)[1+O(N^{-1})] \exp\left[i(\omega t - \rho(r) - \delta(r))\right]$$

(3)

Once the phase function $\rho = \int \mu dr$ is determined from the "eikonel" equation, the inviscid amplitude function $\psi$ and the damping due to viscosity $\delta$ are easily determined. An approximation to the eikonel equation is

$$\frac{\mu b}{\rho \omega^2} \tanh \mu l = \frac{b^2}{K + EI \mu^4} + \frac{9}{64} \frac{b^4}{k + GJ \mu^2} + \frac{8}{\pi^2} \frac{b}{K((\frac{\pi}{b})^2 + \mu^2)^2}$$

(4)
For long wavelengths; i.e. \( \mu \to 0 \), the right-hand-side gives the volume compliance under a static pressure difference in the scalae. For short wavelengths, the stiffnesses in the \( r \)-direction \( (E, G, J, K) \) become significant. On the left-hand-side is the fluid effect. The height \( l \) of the model scala, which is chosen to be 4 mm so that the area is about that of the cochlea, appears only in the hyperbolic function. Since

\[
\tanh \mu l \sim \begin{cases} 
\mu l & \text{for } \mu l < 1 \\
1 & \text{for } \mu l > 1
\end{cases}
\]

the assumption of "longwave" theory, that the pressure on the basilar membrane is the average of the pressure over the scala cross-section, is satisfied when \( \mu l < 1 \).

The 2-mode model (i.e. with \( \lambda \to \infty \)), shows the transition to occur at about the "Transition Pt.-Flex.Arches" curve in Fig.3. The significant activity is thus "shortwave". Of particular interest is that the maximum arch motion in the model is near Békésy's curve for the maximum Reissner's (rather than the basilar) membrane motion, while the model maximum basilar membrane displacement is near the damage curve, particularly at 1 KHz where the end effect of the helicotrema and Mode III are not so significant. Fig.4 shows the phase at the points of maximum arch and membrane motion with experimental values from Rhode (1971) and Békésy (1960). It seems clear that the components of the organ of Corti can have significantly different motion.

Mode III is of dominant importance near the stapes, as shown by the impedance in Fig.5. The peculiar behavior of Békésy's experimental points near 1 KHz is now seen to be the transition from the low frequency, longwave asymptote to the high frequency, shortwave asymptote. It appears therefore that, for frequencies over 1 KHz, no "longwaves" occur in the cochlea.
Eq. 3 can be easily extended to transient excitation analysis, by a Laplace transformation over frequency and then a "saddle-point" estimation. The result for the front and tail of a "click" propagation is shown in Fig. 6.

Finally it might be mentioned that all results are reasonably insensitive to variations in the model parameters. Figs. 5 and 6 are independent of the fluid viscosity. Unfortunately, the computations have not yet been performed for the 3-Mode model curves in Figs. 3, 4, and 6.

References:
Fig. 5
Impedance of Cochlea at Stapes

Fig. 6
Click Propagation along Cochlea
Introduction
During the past decades a lot of pitch phenomena have been reported which cannot be explained within the scope of the "classical" place theory. Therefore, different mechanisms of pitch extraction like time interval measurement and autocorrelation processes have been proposed. But these possible explanations suffer from the disadvantage that crucial parts of the hypothesized mechanism have been attached to the stage of nervous processing where there are almost no observable phenomena supporting or disproving any assumption. Thus, the general result of our scientific progress is a great improvement in the knowledge of discrete pitch phenomena accompanied with an almost complete confusion what the basic principles of pitch perception are concerned.

In this paper, the phenomena observed within the area of pitch perception of pure and complex tones shall be considered. It will be shown that, at least within this area, there must not be confusion, because the pitch phenomena which play a role in this area are compatible with a generalized place principle. The contents of this paper is an excerpt from a more detailed study which is in publication (Terhardt, 1972a, b).

1. Spectral Pitch and Virtual Pitch
It has not always been regarded with full consequence in the past that a complex tone generates two basically different kinds of pitch, although this was clearly stated already by Schouten (1938, 1940). Schouten observed that one and the same complex tone may generate two subjective pitch attributes which differ in quality but which have the same height: First, there is the perception of the sinusoidal fundamental component (if this component is present in the spectrum) and, secondly, there is the "residue", having a sharp timbre and the same pitch as the fundamental.
It follows from this observation that we have clearly to distinguish between both kinds of subjective attributes. We call the "classical" kind of pitch which follows Ohm's law, spectral pitch; the second kind of pitch which may be perceived without the presence of a directly corresponding spectral clue shall be called virtual pitch. It would be highly premature to conclude from the existence of the "residue" that the "classical" place theory must be basically wrong. Further experimental results like those considered in the following sections have to decide whether and how both kinds of pitch have something to do with each other.

2. Sound Analysis

In Fig. 1 some important data on the limits of the ear's frequency analyzing power are shown (slightly simplified and averaged). \( \Delta f \), plotted on the ordinate, is the minimal frequency distance of a part tone of a sound from its neighbours which is necessary for the perception of a spectral pitch corresponding to the tone's frequency \( f_T \) (abscissa). In the special case of a harmonic sound (complex tone), \( \Delta f \) is identical with the fundamental frequency \( f_b \).

Two different situations are distinguished:

If there are part tones of lower as well as of higher frequency than \( f_T \) present, curve b is valid (after data of Plomp, 1964). If the part tone with the frequency \( f_T \) has only higher neighbours (i.e., if the part tone is the lowest component of the sound), curve c has to be considered as the limit of perceptibility of a corres-

---

**Fig. 1**

a: Critical bandwidth \( \Delta f \) as a function of frequency level \( f_T \) (see text).

b: Frequency distance \( \Delta f \) of a part tone with frequency \( f_T \) from adjacent tones which is necessary for perception of a corresponding spectral pitch (see text).

c: Frequency distance \( \Delta f \) of two pure tones which is necessary for perception of a pitch corresponding to the lower tone with frequency \( f_T \) (see text).

d: Boundary of existence region of the "residue" as a function of the frequency \( f_T \) of the lowest component (see text).

n = number of harmonic of a complex tone with fundamental frequency \( f_b \).
PITCH PERCEPTION OF COMPLEX TONES

Ponding spectral pitch (after data of Thurlow and Bernstein, 1957; Plomp, 1964; Terhardt, 1968a).

For comparison purposes the size of the critical bands after Zwicker and Feldtkeller (1967) (curve a), the boundary of the existence region of the "residue" (curve d, after Ritsma, 1962, 1963; Walliser, 1968), and the relation between $f_T$ and $f_B$ for certain harmonic numbers $n$ are plotted in Fig.1.

With respect to these data it is interesting to discuss two points. The first point is the possible principle underlying the observed frequency analyzing abilities of the ear. Here, we must conclude that the data at least do not disprove a possible place principle: It was stated already by Plomp (1964) that the limit represented by curve b coincides rather well with the critical bandwidth; this coincidence is relevant with respect to the question of place principle because the critical bandwidth becomes widely accepted to be a measure of the inner ear's frequency selectivity (Zwicker, 1970). The high "selectivity" which seems to be represented by curve c may have to do with two different effects. On one hand, it is known that an "edge tone" is perceived if a sound spectrum has a sharp limit. This should hold also for the perception of a spectral pitch corresponding to the lowest partial of a high-pass filtered complex tone. On the other hand, the curve c originally was found in the case where only two pure tones with the frequency distance $\Delta f$ were presented. In this case there occur beats with the beat frequency $\Delta f$. If we consider the consequences which these beats have with respect to the time structure of the resulting excitation pattern along the cochlear partition we find that there must occur and vanish two separate excitation maxima periodically with the beat frequency $\Delta f$. The nervous system which analyzes the spatial excitation pattern has also a certain time resolution and, therefore, should be able to find out these two maxima which periodically occur, if the beat frequency is not too high. We recognize in Fig.1 that the two curves c and b differ significantly only in the beat frequency range $\Delta f \lessapprox 300$ Hz. A time resolving power of the nervous system as represented by an upper frequency boundary for the resolution of beats of roughly 300 Hz, on the other hand, is in good agreement with other psychoacoustic observations on perception of "auditory flutter" and roughness (Miller and Taylor, 1948; Terhardt, 1968a, b). Thus, the limits of the sound analyzing abilities of the ear as represented in Fig.1 may be understood simply on the basis of psychoacoustic results.
The second interesting point, regarding Fig. 1, is the relation between frequency resolution of partials and virtual pitch. Plomp (1967) and Ritsma (1967) found that certain partials of a complex tone are dominant with respect to the perception of virtual pitch. Surprisingly, these dominant partials are partials with lower harmonic numbers out of the range of about n = 3 ... 6 which are rather good separated by the ear (see Fig. 1, curve b). Furthermore, at a foregoing Symposium in 1969 the author pointed out that the perception of separate spectral pitches seems to be not only "by chance" connected with the perception of virtual pitch but that the pitch separation seems to be an essential condition for the existence of virtual pitch (Terhardt, 1970). This was concluded from a comparison of the existence region of the "residue" (curve d in Fig. 1) with the minimal frequency separation of the lowest partial from its higher neighbour (curve c). Thus, we again have to state that certain partials of a complex tone are dominant with respect to virtual pitch perception not in spite of their resolution by the ear but, rather, because of the resolution. This conclusion seems to become accepted by other authors, too (Smoorenburg, 1970; Houtsma and Goldstein, 1972).

3. Scales of Spectral Pitch

After the work of Stevens et al. (1937), Békésy (1960), Zwicker and Feldtkeller (1967) and others, it is not necessary to explain the close relations which were found to exist between

1. the "mel-scale", resulting from magnitude estimation of spectral pitch;
2. the "jnd-scale", resulting from the integration of jnd's of spectral pitch;
3. the scale of distance of excitation maximum from the helicotrema as a function of tone frequency;
4. the critical band scale, resulting from experiments on masking and loudness.

We merely have to mention these close relations, which may be represented by one and the same function underlying these scales, the "Tonheit" - function (Zwicker and Feldtkeller, 1967). These close relations strongly suggest that the spectral pitch is closely related to the situation of the excitation pattern along the cochlear partition.

4. Basic Properties of Virtual Pitch

The virtual pitch of complex tones has been extensively investigated (see, e.g., Schouten, 1938, 1940; de Boer, 1956; Flanagan and Guttman, 1960a, b; Schouten et al., 1962; Ritsma, 1962, 1963, 1967; Plomp, 1967; Walliser, 1969a).
We may summarize the main properties of the virtual pitch of complex tones as follows.

(1) The "virtual character".
As stated already, the presence of a part tone which corresponds to the perceived pitch is not essential. Furthermore, the "subjective intensity" of the pitch sensation depends strongly on the number and frequencies of the presented part tones. Recently, Smoorenburg (1970) and Houtsma and Goldstein (1972) investigated this kind of pitch under extreme conditions (i.e., with presentation of only two harmonics) under which one hardly can speak of a "pitch sensation" (what the pitch corresponding to the fundamental is concerned) but perhaps rather should speak of a "pitch imagination". These investigations point out impressively the true character of virtual pitch:

In contrast to the spectral pitch, the virtual pitch of complex tones can be perceived with all degrees of "subjective intensity". This "subjective intensity" may range without discontinuities from "almost pure imagination" (when only two harmonics are presented) towards "strong pitch sensation" (when a full complex tone spectrum is presented).

(2) The ambiguity.
In the perception of spectral pitch, there have almost no ambiguities ever been reported. Indeed, for subjects with normal ears it seems to be difficult to err perceptually in that way that one takes a pure tone for its octave or its fifth. On the contrary, in the perception of virtual pitch such errors occur often.

(3) The relation to dominant spectral pitches.
For "residue tones" (i.e., high-pass filtered complex tones) Walliser (1968, 1969a, b) has shown that the perceived virtual pitch always can be determined as a subharmonic of the lowest partial present in the complex tone spectrum. As was already pointed out by Terhardt (1970), we may summarize the behaviour of virtual pitch in general, i.e., virtual pitch of complex tones with lower harmonics as well as without, by generalizing the rule of Walliser: Virtual pitch is always a subharmonic of a dominant partial; and dominant are such partials which are resolvable.

The most important conclusions which we may draw from this rough summary of the basic virtual pitch phenomena are: Firstly, the distinction between spectral pitch and virtual pitch which we introduced in section 1 is justified and confirmed; from (1) and (2) it follows that one cannot assume that spectral pitch and virtual pitch might be established by one and the same mechanism. Secondly, in spite of this distinction, there seems to exist a
relation between the virtual pitch of a complex tone and the spectral pitches of its partials. This will become even more evident in the following section.

5. Pitch Shifts and Octave Enlargement

Last, but not least, the following phenomena have to be considered.

(1) Pitch shifts of pure tones with constant frequency (i.e., shifts of spectral pitch) due to changes of SPL.

After the investigations of Stevens (1935), Snow (1936), Cohen (1961) and Walliser (1969c) it is evident that low tones (below about 1 kHz) become lower and high tones (above about 3 kHz) become higher when the SPL is increased.

(2) Pitch shifts of complex tones with constant fundamental frequency (i.e., shifts of virtual pitch) due to changes of SPL.

For a complex tone with a complete spectrum of partials this effect is of negligible quantity (Fletcher, 1934; Lewis and Cowan, 1936). This lack of pitch shift is in good agreement with the results mentioned above: The virtual pitch of a complex tone was found to be subharmonic to the pitches of dominant partials; these dominant partials are situated in the frequency region of about 500...1500 Hz (3rd to 6th harmonic), and in this frequency region the spectral pitches are practically not influenced by the SPL.

When a complex tone without lower harmonics ("residue tone") is presented, the virtual pitch depends on the SPL in the same way as the spectral pitch of the lowest harmonic which is present (Walliser, 1969a). Thus, the principle after which the virtual pitch is subharmonic to the pitch of dominant partials holds in this case, too.

(3) Pitch shifts of pure tones due to partial masking.

Recently, this phenomenon was investigated by Terhardt and Fastl (1971).

Fig. 2 shows the pitch shifts in terms of the equivalent relative frequency shift $\nu$ (ordinate) as a function of the frequency $f_T$ of the tone (abscissa), for three different spectral configurations.

If the spectral energy distribution of the masking sound is constant around the tone frequency, the pitch arises for tones above about 1 kHz (curve a). If the energy of the masking sound (this may be a low-pass noise or a second tone) is situated just below the tone, there is a pronounced positive shift for all tone frequencies (curve b). In the reversed case (masker energy above tone frequency), the pitch shift behaves as is demonstrated by curve c: Low tones become lower, high tones become higher.

These pitch shifts may be explained by assuming the existence of two different influences: One influence is the mere increase of sound energy in
Fig. 2 Pitch shift $v$ of a pure tone with frequency $f_1$, due to partial masking. $v$ is the relative frequency difference between the unmasked and the masked tone for equal pitch. a: Masking by white noise. b: Masking by low-pass noise just below $f_1$ or by a second adjacent lower tone. c: Masking by high-pass noise just above $f_1$ or by a second adjacent higher tone. d: Pitch shift corresponding to a constant place shift of $-4$ mels (see text).

the spectral neighbourhood of the tone; this influence causes a rise of the pitch of higher tones (curve a). The second, additional influence would be suggested by a place theory: Due to partial masking, the maximum of the excitation pattern of the tone shifts a little, away from the adjacent masker pattern. Curve d represents this hypothetical effect, assuming a frequency-independent place shift equivalent to 4 mels. It can be seen that, indeed, curve c may be interpreted as the result of the common influences of both effects: with good approximation is $c$ the sum of $a$ and $d$.

Although these pitch shifts do not directly prove any principle of spectral pitch perception, the situation is not as contradictory as former data might have made believe us. On the basis of the data of Fig. 2, we have to concede a place principle a good chance.

(4) Pitch shifts of complex tones due to partial masking.

As was found by Walliser (1969a), the shift of the virtual pitch of a "residue tone" has the same size and direction as the shift of the lowest (i.e., dominant) partial present in the spectrum. Thus, also in this case, the virtual pitch remains subharmonic to the pitch of the dominant partial.

(5) Interaural pitch differences (binaural diplacusis) of spectral pitch. This effect, as far as normal ears are concerned, is well known due to the investigations of, e.g., Jeffress (1944) and van den Brink (1970a, b). As was discussed already in the literature, this effect may be explained in the easiest way by a place theory.

(6) Interaural pitch differences of virtual pitch. This phenomenon was investigated recently by van den Brink (1970b) with "residue tones". The result was that, again, virtual pitch remains subhar-
monic to the pitch of dominant partials, i.e., the interaural virtual pitch differences are the same as the interaural spectral pitch differences of the present harmonics.

(7) Pitch difference between a pure tone and a complex tone of the same fundamental frequency ("virtual pitch shift"). The virtual pitch of a complex tone (with or without lower harmonics) is a little lower than the spectral pitch of a pure tone which has the same frequency as the fundamental (Walliser, 1968, 1969a; Terhardt, 1971a). This effect supports the conclusion that virtual pitch and spectral pitch must be two basically different subjective attributes which are not established by one and the same mechanism.

(8) Octave enlargement for successive pure tones. This phenomenon is well known since Ward's (1954) investigations. Further measurements of the effect have been performed by Walliser (1969d), Terhardt (1971a), and Lindqvist and Sundberg (1971). Fig. 3 shows the relative frequency deviation \( \Omega \) of the higher tone from the "physically true" octave frequency, as a function of the "true" octave frequency \( 2f_{T1} \), i.e., twice the frequency of the lower tone. The curves a,b,c show that, in the whole frequency range of interest, two successive pure tones must have a frequency ratio of a little more than 2 in order to spread out a subjectively correct octave interval.

This phenomenon, which looks rather strange at first sight, fits into the remainder of our knowledge on pitch perception of complex tones rather well by means of the following hypothesis: "Our sense of musical intervals, and thereby of the octave interval, is acquired; we acquire the knowledge of harmonic relations in earliest life, when we learn to recognize and to
understand speech sounds" (see Terhardt, 1969/70, 1971b). In order to understand how this hypothesis explains the octave enlargement we have to consider the pitch intervals which correspond to the harmonic frequency intervals of the partials of voiced speech sounds, i.e., of complex tones. Since, as was mentioned above, the spectral pitch of a pure tone may be shifted by partial masking, we must expect that also the spectral pitch of a certain harmonic of a complex tone may be influenced by the remainder of the spectrum. Such pitch shifts have been determined by subjective measurements (Terhardt, 1971b). The result was that the spectral pitch of the fundamental of a complex tone is shifted downward, and those of the 2nd and higher harmonics are shifted upward. Thus, the pitch intervals between the 1st and 2nd harmonics and also between the 2nd and 4th harmonics are enlarged in comparison with the correspondent pitch intervals of successive pure tones. This kind of "octave enlargement" which is caused directly by the pitch shifts due to mutual masking of harmonics is demonstrated by the curves d and e in Fig.3. The agreement of the tendencies of both kinds of octave enlargement (curves a, b, c and d, e, respectively) supports the hypothesis that our knowledge of the harmonic pitch intervals is acquired.

(9) Octave enlargement for successive complex tones.

When the fundamental frequencies of two successive complex tones are adjusted in such a way that the perceived pitch jump sounds like a correct octave interval, the ratio of both fundamental frequencies has a value which is a little greater than two, as is the case with pure tones (Terhardt, 1971a; Lindqvist and Sundberg, 1971). Thus, for spectral pitch as well as for virtual pitch, the pitch interval which is judged to be a correct octave interval is one and the same.

6. General Conclusions

After this rough consideration of the pitch phenomena which are relevant in the area of pitch perception of complex tones, we may summarize the following general conclusions.

Spectral pitch and virtual pitch are basically different percepts. Spectral pitch seems to be a "primary sensation" which is closely related to the situation of the cochlear excitation. Virtual pitch seems to be derived from the same clues as the spectral pitches of the dominant partials (or from the spectral pitches as such) by a process of high complexity. An essential means involved in this process is the knowledge of harmonic pitch intervals which is stored in the hearing mechanism. This knowledge seems to be acquired during the learning process which enables us in earliest life to recog-
nize speech sounds.
A place theory, based on psychoacoustic results and generalized according to this concept, has been worked out and accounts for the phenomena considered in this paper (Terhardt, 1972b).

References
Fletcher, H. (1934): Loudness, pitch and the timbre of musical tones and their relations to the intensity, the frequency and the overtone structure, J. Acoust. Soc. Amer. 6, 59-69.


THE RELATION OF THE MEDIAL GENICULATE BODY TO THE TONOTOPIC ORGANIZATION OF THE AUDITORY PATHWAY

I. C. Whitfield

Neurocommunications Research Unit,
University of Birmingham, England

The lower levels of the auditory system appear to be strictly tonotopic in the sense that the great majority of units have well-defined characteristic frequencies, and that there is an orderly spatial arrangement of units with respect to their characteristic frequencies. This situation obtains at least as far up as the inferior colliculus (Rose, Greenwood, Goldberg & Hind, 1963). However, at the primary auditory cortex not only is the degree of tonotopic organisation very weak (Evans, Ross & Whitfield, 1965; Rigby, Ross & Whitfield, 1968) but many units respond over so wide a band as to have virtually no characteristic frequency (Evans & Whitfield, 1964). The status of the medial geniculate body in this respect has been in considerable doubt. Early experiments of Ades Mettler & Culler (1939) suggested a spiral arrangement of frequency, while the first micro-electrode recordings (Galambos, 1952) failed to demonstrate any tonotopicity at all. However, more recent anatomical studies of the colliculo-geniculate projection (Goldberg & Moore, 1967) have indicated that a tonotopic arrangement might be expected to exist in the principal division of the medial geniculate with the high/low frequency organisation along a medio-lateral axis. Morest's (1965) anatomical studies of the organisation of the medial geniculate body itself are not in conflict with such a proposal. More recently still Aitkin & Webster (1971) in anaesthetized and temporal lobectomised cats have reported finding a strict tonotopic organisation in the principal division with low
frequencies placed laterally and high frequencies medially. In contrast to these findings Whitfield & Purser reported at the Symposium on Basic Thalamic Structure and Function (New York, 1971) an anterior/posterior trend of characteristic frequency with high frequencies placed anteriorly, but with a scatter about the mean, very similar to that seen in the cortex (Fig. 1).

![Figure 1.](image)

Regression of frequency on position for W-shaped units in the ventral subnucleus of a single animal. The main graph shows the points, the regression line and the 90% limits (- - - -) for the prediction of a further observation for the "upper inhibitory" characteristic frequencies. Curves a and b indicate the corresponding data for the excitatory and lower inhibitory characteristic frequencies respectively. The observed points are not shown in the latter two cases. The penetrations are along the anterior posterior axis, zero on the abscissa scale corresponding approximately to the F5.5 plane. (From Whitfield & Purser, 1971.)

The animals from which these results were obtained were intact unanaesthetized cats. The penetrations in these experiments were made along a line parallel to the antero-posterior axis of the animal, which does not exactly follow the "axis" of the medial geniculate itself; in particular the most anterior part of the nucleus was not sampled. Similarly the vertical penetrations of the Horsley-Clarke co-ordinates
used do not follow the vertical curvature of the laminae. Nevertheless the results obtained are not easily explicable in terms of failure to sample precisely along some true tonotopic axis, and do bear a remarkable resemblance to the pattern found in the cortex.

Rose & Woolsey (1949) showed by degeneration experiments a connexional relationship between the antero-posterior axis of the nucleus and the corresponding axis of area AI of the cortex, such that the anterior part of the nucleus was connected with the anterior part of AI and the more posterior parts of the nucleus with progressively more posterior regions of AI. The arrangement of the thalamo-cortical connections is thus consistent with the findings of similarity between the distribution of units in AI and in the geniculate nucleus.

We thus find that anatomical studies of colliculo-geniculate connections (Rose & Woolsey, 1958; Morest, 1965; Goldberg & Moore, 1967) all suggest a medio-lateral tonotopic organisation in the principal division of the nucleus with high frequencies medially (Fig. 2).

Figure 2.
The projection of the inferior colliculus on the medial geniculate nucleus, showing the maintenance of a tonotopic arrangement. (After Goldberg & Moore, 1967.)
This is precisely what is found electrophysiologically under anaesthesia and in the absence of temporal cortex. However, in the unanaesthetized intact animal this picture seems to be modified, and the nature of the modification strongly suggests that it is due to cortical influence. Morest's (1965) anatomical studies show the presence of cortico-thalamic fibres whose terminals course in a medio-lateral plane and hence presumably provide an input whose organisation is perpendicular to the ascending input from the colliculus (Fig. 3).

Some further physiological evidence of a dual input to units of the medial geniculate is provided by the behaviour of units with W-shaped response curves (Purser & Whitfield, 1972a). These units have fairly wide band inhibitory response curves with a central excitatory peak (Fig. 4). The bandwidth of the excitatory section is commonly similar to that of units at the cochlear nucleus level, while the wide band inhibitory section is comparable with the bandwidths of inhibitory units found in the cortex. We have on occasion observed the disappearance
over a period of time of the central excitatory region while the inhibitory limbs remain unchanged (Fig. 4). It seems possible, therefore, that these complex responses are composites built up from two separate inputs, and lend support to the idea of a cortical and a sub-cortical influence on geniculate units.

![Figure 4.](image)

Response of a 'W-shaped' unit in the medial geniculate body of the cat to different frequencies. The excitation originally present was replaced by inhibition in the course of about 15 mins. (From Purser & Whitfield, 1972a.)

It is now fairly well established that while the auditory cortex is inessential for discrimination between single steady tones, its role is an extremely important one in those auditory functions which depend on the temporal relations of sound patterns for their discrimination (see Kelly & Whitfield, 1971; Swarbrick & Whitfield, 1972). Such a difference in behaviour, although not sharp, would nevertheless presumably involve a major reorganisation of connectivity away from a tonotopic arrangement and towards more complex integrations. It is suggested that the medial geniculate body may be an important site of this reorganisation.
REFERENCES

Ades, H. W., Mettler, F. A. and Culler E. A. (1939): Effect of lesions in the medial geniculate bodies upon hearing in the cat,
Am. J. Physiol. 125, 15-23.

Aitkin, L. M. and Webster, W. R. (1971): Tonotopic organization in the medial geniculate body of the cat,
Brain Research, 26, 402-5.

Evans, E. F., Ross, H. F. and Whitfield, I. C. (1965): The spatial distribution of unit characteristic frequency in the primary auditory cortex of the cat,
J. Physiol. 179, 238-47.

Evans, E. F. and Whitfield, I. C. (1964): Classification of unit responses in the auditory cortex of the unanaesthetized and unrestrained cat,
J. Physiol. 171, 476-93.

Galambos, R. (1952): Microelectrode studies on medial geniculate body of cat, III. Response to pure tones,
J. Neurophysiol. 15, 381-400.

Goldberg, J. M. and Moore, R. Y. (1967): Ascending projections of the lateral lemniscus in the cat and monkey,

Kelly, J. B. and Whitfield, I. C. (1971): Effects of auditory cortical lesions on discriminations of rising and falling frequency-modulated tones,
J. Neurophysiol. 34, 802-816.

Morest, D. K. (1965): The laminar structure of the medial geniculate body of the cat,
J. Anat. Lond. 98, 611-630.

Purser, D. and Whitfield, I. C. (1972a): W-shaped excitatory/inhibitory response curves from the medial geniculate nucleus of the unanaesthetized intact cat,
J. Physiol. 221, 17-18P.

Purser, D. and Whitfield, I. C. (1972b): Thalamo-cortical connections and tonotopicity in the cat medial geniculate body,
J. Physiol. in press.
Rigby, D.C., Ross, H.F. and Whitfield, I.C. (1968): Frequency organisation in the second auditory area (AII) of the cat,
   J. Physiol. 194, 67-68P.
Rose, J.E. and Woolsey, C.N. (1949): Relationships of thalamic connections to cellular structure and evocable electrical activity in the auditory region of cat,
Rose, J.E. and Woolsey, C.N. (1958): The relations of thalamic connections and functional organisation of the thalamic auditory system of the cat,
Swarbrick, L. and Whitfield, I.C. (1972): Auditory cortical units selectively responsive to stimulus shape,
   J. Physiol. in press.
   In Symposium on Basic Thalamic Structure and Function, to be published in Brain Behavior & Evolution.
In this paper we will describe a model of pitch-perception based on what might be called "auditory pattern-recognition." In order to clarify what we mean by "auditory pattern-recognition", we propose the following analogy between a classic visual pattern-recognition problem and pitch-perception. The letter "A", as typed on this typewriter, has something in common (its "A-ness") with the same letter printed by hand, in a newspaper, etc. In every case it is clearly recognized as the letter "A" despite the fact that the physical features of the various letters "A" (e.g., size, type-style, orientation, etc.) are different. It is obviously not a simple matter to describe the exact process by which the nervous system performs the recognition. However, in very general terms we might describe it as a sequency of transformations of what we will call "patterns of nervous activity". The reasoning is as follows: when one looks at a letter (or anything else), the physical stimulus, in this case an image, in 'transformed' by the sense organ into a "pattern of peripheral nervous activity". At this peripheral level, each different physical stimulus is assumed to produce a different "pattern". Next, this peripheral activity pattern is transformed into another pattern of activity. This second transformation is assumed to work in such a way that after the transformation the patterns produced by all the letters "A" will be similar (and all the letters "B", etc.). The second transformation thus serves to extract from the input stimulus (actually the peripheral activity pattern produced by the stimulus) those features which are relevant for "letter-perception". Now consider the pitch problem. The musical note "middle-C" is always "middle-C" (that means it has the same pitch) regardless of whether it is produced by a clarinet, a piano, a tuning fork, etc. This can also be called a pattern-recognition phenomenon. To describe the 'recognition' process in the same
terms as our visual analogy, we can say it consists of a sequence of transformations. When a note (or other sound) is presented, the acoustic waveform is 'transformed' by the sense organ into a "peripheral activity pattern". At this level, each different acoustic stimulus, each different note, has a somewhat different representation. Now, in order to extract pitch, the nervous system transforms the peripheral activity pattern into another pattern, such that all stimuli with the same pitch are then similarly represented. In other words, the transformed peripheral activity patterns of stimuli with the same pitch will look alike; the 'important' features of the transformed patterns will be the same. This general approach to pitch-perception is what we will formalize later in our "pattern-transformation model of pitch".

A characteristic feature of all pattern-recognition schemes is information loss. In our visual example information about the fine details (e.g., size, orientation, type-face) of the various letters is lost. Only information relevant to letter-recognition is preserved. In the auditory example information about which instrument produced the note is lost, and only information relevant to pitch is preserved. Of course, this does not imply that if two stimuli are classified as belonging to the same 'pattern' (letter, or note, in our examples) they are indistinguishable. Obviously, various letters "A" look different, and various "middle-C"s sound different, but for the purposes of the particular recognition task at hand this may be ignored. Thus, building a pattern-recognition model of either the letter-perception or the pitch-perception process amounts primarily to defining just what information is "lost" and what is "preserved".

In our "pattern-transformation model" of pitch we have assumed that what is "lost" is information about the temporal fine-structure of the stimulus; details of the acoustic waveform are assumed to be irrelevant (for pitch), and only the power-spectrum of the waveform is "preserved". This point of view contrasts sharply with many theories which propose that pitch is derived from an analysis of the stimulus fine-structure (e.g., Schouten, 1940; Ritsma, 1962). Thus, in order to justify our position we will present a short summary of the recent research on this problem.

Models in which pitch is derived from some operation on the temporal fine-structure of the stimulus are "phase-sensitive". That is, since the relative phases of the spectral components of a waveform determine its temporal fine-structure, changes in these phase relations would be expected to affect the pitch of the waveform. There have been several experiments directed specifically at this issue. Some of the results appear to favor a phase-sensitive pitch model, but many do not.
One experiment which showed a phase effect was reported by Licklider (1957). Licklider produced a 16-component harmonic stimulus (16 successive upper harmonics of some missing fundamental) with a device which allowed independent control of the amplitudes and phases of the individual components. He observed (no data are given) that when the phases were adjusted to produce a waveform with pronounced peaks and valleys a very strong pitch was heard. When the phases were set haphazardly, to produce a waveform with a rather flat envelope, the pitch seemed to disappear. Unfortunately Licklider's description of this experiment is brief, and no actual data (e.g. pitch matches) are included, so it is difficult to know how heavily to weight this evidence on the phase question. This applies also to the experiment of Mathes and Miller (1947), who reported an effect similar to Licklider's, obtained with 3-component stimuli. More recently, what appears to be a clear phase-effect has been reported by Ritsma and Engel (1964). In this experiment, listeners matched the pitch of four 3-component harmonic stimuli. The phase of the center component had been shifted 90° with respect to the phases of the other components (QFM).

Ritsma and Engel report that for two of the stimuli, very few matches were made to the fundamental frequency, but were distributed around two other frequencies, one above and one below the fundamental. This result contrasts with data from other experiments in which similar stimuli had been used, except that all three components had been added in cosine phase (AM) (Schouten, Ritsma, and Cardozo, 1962, for example). In these latter experiments the pitch of the comparable stimuli always corresponded to the fundamental frequency. Ritsma and Engle explained their results with a fine-structure model in which pitch was assumed to be determined by the time interval between major peaks in the stimulus waveform (pitch = 1/time).

There are other data which support a phase-insensitive model of pitch. For example, Patterson (1972), in an experiment quite similar to Licklider's, found absolutely no influence of phase relations on pitch. Patterson measured the pitch of several 6- and 12-component stimuli in which the components had been added either in cosine phase or in random phase. There were no differences between the pitch-matches to the two types of stimuli, despite the fact that the temporal fine-structures were grossly different. In another important study, Bilsen (1968) measured the pitch of a stimulus which consisted of a random noise plus its delayed repetition. This waveform has a random, non-periodic fine-structure, but has a power-spectrum with equally spaced spectral peaks. By shifting the positions of the spectral peaks Bilsen was able to show that as far as pitch was concerned, the
noise stimulus behaved much like a stimulus made up of discrete spectral components. We have recently performed two additional pitch-matching experiments on the phase question (Wightman, 1972). The first experiment was an attempt to replicate the phase effect reported by Ritsma and Engel. Our test stimuli were the same 3-component harmonic complexes as were used in the Ritsma and Engel experiment; the phase of the center component was shifted $90^\circ$ (QFM). In addition, we tested stimuli made up of the same components, but, with the components added in cosine phase (AM). Ritsma and Engel did not report data from this latter condition. In the second experiment we repeated some of the conditions studied by Patterson. Four 12-component stimuli (lowest component around 1400 Hz, spacing 200 Hz) were presented, in three different phase configurations (one cosine phase condition, and two different random-phase conditions). In both of these experiments the principal aim was to compare the pitches of waveforms having the same power-spectra but different phase-spectra (thus different fine-structure). The results of both experiments suggest that stimulus fine-structure is irrelevant for pitch. In the first experiment, not only did we fail to obtain the phase effect reported by Ritsma and Engel, but, more importantly, the pitch matches were virtually identical in the QFM and AM conditions. The fine-structures of these two types of waveforms are quite different. The data from the second experiment corroborate Patterson's finding that with multi-component (wide-band) stimuli as well, pitch is unaffected by changes in the phase relations among the components. In our replication of Patterson's experiment, the pitch matches to each complex were the same regardless of the phase configuration.

In summary, taking all the available data into account, including that from our own two experiments, we feel the weight of evidence clearly favors a phase-insensitive model of pitch.

A pitch extractor that operates on the power spectrum of a stimulus, rather than on the temporal fine-structure, is phase-insensitive; all waveforms which have the same power spectrum (regardless of the temporal features of the waveforms) are predicted to have the same pitch. An autocorrelator is one example of such a mechanism (autocorrelation is simply the Fourier transform of the power spectrum). Licklider (1951) was perhaps the first to propose a model of pitch (duplex theory) based on autocorrelation. However, even though autocorrelation is formally phase-insensitive, Licklider argued that the temporal aspects of stimulation could not be completely ignored. In his model, autocorrelation was computed by analyzing periodicities in nerve firings (this represents a computation in time-domain, formally equivalent to Fourier transforming the power spectrum).
It might be expected that the temporal features of the stimulus would influence the temporal patterns of nerve-firings. Thus Licklider's model cannot fairly be called a phase-insensitive model. The pitch extractor that we will outline here is quite closely related to Licklider's, except that in our model the temporal features of stimulation are ignored, for the most part. We compute autocorrelation by Fourier transforming a "pattern". The "pattern" is assumed to represent the power-spectrum of the stimulus.

An outline of our "pattern-transformation" model is given in Figure 1. There are three operational stages. At Stage I the physical (acoustic) stimulus is transformed by the ear into what we will call a "peripheral activity pattern" or PAP. This first transformation is assumed to resemble a spectral analysis of the incoming stimulus. Thus place in the PAP corresponds to frequency, and amount of activity at each place correlates with power. It is only the amount of activity at each place that is important, not the temporal distribution of the activity. The spectral analysis at this stage is obviously only a coarse analysis, due to the limited resolving power of the analyser (cochlea). Also, it is certainly contaminated by non-linearities; we will ignore these for the moment. The output of the first stage of our model, then, is assumed to be a pattern of neural activity which roughly represents the power spectrum of the stimulus. Stage II of the model is a Fourier transformer. This is thought to be nothing more (nor less) than a network of neural inter-connections which takes one pattern of activity (output of Stage I) and transforms it (à la Fourier) into another pattern. The only mathematical operations required of this network are addition and multiplication by a constant; both are reasonable from a physiological point of view. Now, because the output of Stage I is roughly a power-spectrum, and the Stage II transformation is Fourier, the output of Stage II roughly represents an autocorrelation function. Actually, it is another "pattern of activity" in which place corresponds to the time ("t") dimension of autocorrelation, and the amount of activity at each place corresponds to the value of the autocorrelation, R(t).

Stage III of the model is the pitch-detector. This device operates on the transformed PAP (output of Stage II) and outputs information about the pitch of the input stimulus. It is not really necessary to define exactly how this stage of the model works, since we cannot ever measure pitch absolutely. We can measure only how the pitch of one stimulus compares with the pitch of another stimulus. The basic assumption of our approach is that after the Stage II transformation, stimuli which have the same pitch are similarly represented. This means that the 'important' features of their
respective transformed PAP's are the same. We feel it is not unreasonable to assume that the 'important' features are simply the positions in the patterns where activity is greatest. Thus, for the moment, we will assume that Stage III is simply a "peak-picker"; it locates the positions in the transformed PAP where activity is maximal. Two stimuli will be predicted to have the same pitch, then, if the maxima in their respective transformed PAP's occur in the same places. With a single stimulus, we will assume its pitch (or pitches, since there is often more than one) is predicted by the position of the highest peak in the transformed pattern. Peaks which are located near the place in the pattern corresponding to t=0 in the autocorrelation function must be ignored. All stimuli, even those which evoke no pitch percept, would produce maxima at this place. It is simply a property of the Fourier transform of a power spectrum. Thus no information about pitch could possibly come from that region.

The predictions of the model are almost completely determined by the assumptions we make about the operation of Stage I, the transformation from acoustic stimulus to PAP. The Fourier transform (Stage II) is parameter-free, and for the moment we are making very simple assumptions about how Stage III operates. Since Stage I is thought to represent virtually the entire peripheral auditory system, describing how it might work is at best a very complicated matter. In order to get a first look at the possibilities of our pitch model, we made some simplifying assumptions. First, we assumed that the action of the middle-ear can be represented by a simple low-pass filter (1000 Hz cut-off frequency; attenuation of higher frequencies at the rate of about 12 dB/oct.). Second, we assumed
that the spectral analyzing properties of the inner-ear can be modeled with
a set of bandpass filters. The transfer characteristic of each of these filters
is assumed to be triangular in shape (on log-log coordinates), and the slopes
of the two sides of the triangle (in dB/oct) are assumed to be independent
of the center frequency of the filter involved. In addition, we assumed that
the high-frequency side the slope is greater (about a factor of 2) than
on the low-frequency side. Thus, our simplified Stage I transformer has only
two free parameters, the two slopes of the filters.

Given our simple assumptions it is easy to compute the expected PAP
for any stimulus. The PAP for a pure sinusoidal stimulus for example, is
simply a triangle (on log-log coordinates), extending more to high frequen-
cies than to low frequencies. For any other stimulus, we simply decompose
the stimulus into its component sinusoids, compute the contributions of
each to the total PAP, and (assuming linearity for the moment) add them up.

From preliminary experiments with the model it has become clear that
there are few data from pitch-matching experiments which are seriously
at variance with the predictions of the model. A wide variety of complex-
tone stimuli was tested; even with our admittedly oversimplified represen-
tation of the Stage I transformation, the pitches predicted by the model
(given by the positions of the peaks in the transformed PAP) were in every
case close to the pitches obtained from listeners' matches. Moreover,
changes in the Stage I parameter values over a wide range had little ef-
fet on the predictions. This was encouraging, and convinced us that the ap-
proach was worth pursuing. However, it is our current feeling that compa-
rison of the predicted pitches with pitch-matching data is probably not a
powerful test of the validity of our approach. Luckily, the model predicts
more than just the 'value' of the pitch; it gives numerical estimates of
the strengths of the various pitches as well (the height of a particular
peak is assumed to represent the 'strength' of the pitch). These pitch
strength predictions constitute one of the novel features of our model.
Thus it was of considerable interest to explore the influence of various
stimulus and model parameters on predicted pitch strength. A few general
examples of our first tests are shown in Figure 2. In this figure we have
presented graphic representations of the output of Stage II (Fourier
transformer) of the model. Each panel represents a pattern of activity:
place is given on the abcissa of each plot, amount of activity on the
ordinate. Pitch strength can be estimated from the sizes of the various
peaks in the patterns. Recall that the maxima near the t=0 place (left)
must be ignored. The positions of the various peaks give the predicted
pitch values.
In the first two panels, we show how the assumed precision of spectral resolution (Stage I) affects the predicted pitch strengths. Spectral resolution here is defined as the 3-dB bandwidth (in percent of center frequency) of the pattern produced at Stage I by a sinusoid. This bandwidth is, of course, determined by the two filter-slopes; for these computations we assumed that the ratio of the two slopes was 2. The input stimulus was a 9-component harmonic complex centered at 1400 Hz (200 Hz fundamental). It is clear that the assumption of high-resolution (right panel, 5% bandwidth) predicts a stronger, less ambiguous pitch than does a low-resolution (left, 30% bandwidth) assumption. In the second two panels we consider the relation between predicted pitch strength and the number of components making up the complex-tone stimulus. Bandwidth was fixed at 10% for this example; the lowest component frequency in the two stimuli was 800 Hz, and the separation
between the components was 200 Hz. The 12-component stimulus is predicted to have a much stronger pitch (right) than the 3-component stimulus. This prediction is consistent with informal reports (no data are available) about these stimuli. Typically, the observation is that the larger the number of components (particularly if the lowest component is at a reasonably low frequency) the stronger the pitch. In the third pair of plots we show the effect of the center frequency of the stimulus. In this example we fixed bandwidth (10%) the number of components (5), and the spacing (200 Hz). Clearly the stimulus centered at 1000 Hz is predicted to have a better pitch than the one centered at 2000 Hz.

A few detailed examples can be given of our estimates of pitch strength. Assume, for the purposes of this discussion, that pitch strength is given simply by the absolute height of the relevant peak in the transformed pattern. Now consider the function relating the strength of the pitch of harmonic complexes and the component frequencies. Figure 3 shows several such functions, computed for different numbers of stimulus components (spacing between components was constant at 200 Hz). It is clear that in all cases pitch strength is predicted to decrease as the component frequencies are raised. In addition, it is always the case that increasing the number of components in the complex increases predicted pitch strength. But the increase is considerably greater if the component frequencies are low. For example, if all the components are above about 2000 Hz, there is little difference between the strength of the pitch of a 3-component stimulus and that of a 12 component stimulus. However, if the frequency of the lowest component is 600 Hz, there is a very large difference.

In Fig. 4 are shown additional functions relating predicted pitch strength and component frequency region. In this case the parameter of interest is the frequency difference, or spacing, between the stimulus components (for these examples all stimuli have 6 components). In general, it appears that if the frequency of the lowest component is fixed, pitch strength increases as the components are moved farther apart. Moreover, if we assume that the stimuli have no pitch if the strength drops below a certain limit, we can compute a kind of "existence-region" for the pitch of these stimuli. Assume the cut-off strength value is .05. Then, we would predict that the stimulus with a 125 Hz spacing (fundamental) would lose its pitch if all the components were above about 2000 Hz, the pitch of the stimulus with 200 Hz spacing would disappear at 3200 Hz, and the pitch of the stimulus with the 500 Hz spacing at 5000 Hz (off the figure). These values for the upper limits of pitch perception for these stimuli agree
Pitch as auditory Pattern-Recognition

**Fig. 3**

**NUMBER OF COMPONENTS**

- **STRENGTH**
- **FREQUENCY or LOWEST COMPONENT (kHz)**

- (12)
- (6)
- (3)

**Fig. 4**

**SPACING**

- **STRENGTH**
- **FREQUENCY or LOWEST COMPONENT (kHz)**

- (500)
- (200)
- (125)
rather well with Ritsma's measurements of the "existence-region" (1962). Unfortunately, Ritsma's data also suggest that with component spacings of greater than 500 Hz the upper limit decreases, and our model, in its simplest form, does not make this prediction.

The pitch strength estimates provided by our pattern-transformation model are in general agreement either with available data (such as Ritsma's data on the "existence region") or with informal reports. Unfortunately, there have been few, if any, experimental attempts actually to measure pitch "strength". There is no doubt about the validity of the concept. Any listener will admit, for example, that a 200 Hz pulse-train has a very strong pitch, and a 3-component harmonic complex centered at 2000 Hz has a very weak pitch, even though he may claim that the pitches of the two stimuli are the same. The difficulty is in quantifying pitch strength. Thus, as a check on the validity of our approach it seems highly desirable to develop objective experimental measures of pitch strength. Experiments along these lines are currently in progress.

REFERENCES
Schouten, J.F. (1940) "Five articles on the perception of sound (1938-1940)", Institute for Perception, Eindhoven.
Previous measures of basilar membrane (BM) motion have either been direct optical observation of displacement using stroboscopic illumination (Békésy, 1960) or indirect measurement of velocity using the Mössbauer technique (Johnstone & Boyle, 1967; Johnstone, Taylor & Boyle, 1970; Rhode, 1971). Both methods have inherent disadvantages.

The present interim report is of a third method using a capacitive probe. Although it requires a dry BM at the point of measurement and has a somewhat greater area of pick-up than a Mössbauer source, it has the advantages of being more sensitive and inherently more linear. It is used to measure the amplitude and phase response of the BM and incus, to measure linearity over a wide range, and to measure the amount of the combination tone $2f_1 - f_2$.

The capacitive probe is a miniaturised version of Békésy's probe (1960, pp. 54-57), with an active tip diameter of 0.15 mm closely surrounded by a guard ring of 0.2 mm diameter. Changes of capacitance produced by the moving object modulate the amplitude of a radio frequency current passing through the tip. This is then amplified and demodulated. The probe is attached to a moving armature system which is used both for fine adjustments of the probe-membrane distance and also for calibration of the system by vibration of the probe through a known distance. For a solid vibrating object this method gives an accurate calibration. For the BM, however, with a radial amplitude profile comparable with the tip diameter it is calculated that 5 dB should be added to the measured values. The variations along the BM turn out to be small over the diameter of the tip.

Guinea pigs of 200-500 gm were anaesthetised with Nembutal and paralysed with Flaxedil and artificially respired. Sound was introduced closed field from a Bruel & Kjaer 4134 microphone driver through hollow ear bars and monitored close to the tympanic membrane. The posterior lateral aspect of the bulla was opened widely and remained open throughout the experiment.

After a satisfactory N1 click threshold was measured on the cochlea bone below the point to be opened, the basal turn
was picked and pared away until sufficient access for visual placement of the capacitance probe was obtained. The scala tympani was then drained with tissue and the probe introduced so that its tip was normal to the BM and about 20 μ distant. The distance was continuously monitored by reference to the DC component of the demodulated signal, and maintained constant by adjustment of the electromagnetic drive. Once the limit of adjustment (20 μ) had been reached due to the build up of fluid it was necessary to withdraw the probe and drain, otherwise the response became contaminated by fluid ripples at low frequencies.

The frequency response of the BM was measured in two ways. Firstly, automatically from 20 Hz - 20 kHz by filtering the output of the probe amplifier with a Bruel & Kjaer 2020 Slave Filter (3.16 Hz bandwidth) driven by a 1024 Sine Random Generator and recorded on a 2305 Level Recorder. Secondly, point by point, up to 50 kHz using a Brookdeal 401 Lock-In Amplifier (PSA) phase-locked to a Dymar 741 signal generator. The in-phase and quadrature components were measured (successively) to give the amplitude and phase of the vibration. The integrating time constant was set according to the signal to noise ratio, within the range 0.3-30 sec, giving bandwidths down to 0.008 Hz. For phase-locked measurements of the 2f₁-f₂ component of vibration the PSA was synchronised to a signal derived from the primaries f₁ and f₂ via a diode mixer and the slave filter set at 2f₁-f₂.

When measurements on the BM had been completed the position was noted and photographed and the probe repositioned to measure the vibration at the end of the long process of the incus. The cochlea was then further dissected to obtain measurements at other positions on the BM and finally to reveal its basal extremity.

The interim results are selected from data based on 16 guinea pigs, and 31 BM placements. Opening and draining the first turn caused the click threshold for N1 response to rise by less than 3 dB provided that no other damage was done. From then onwards, however, the N1 threshold would rise progressively. In some cases measurements were made while the cochlea was in excellent physiological condition and again later after it had deteriorated. In no case was there any suggestion that the mechanical properties depended upon the physiological state of the cochlea. Some results recorded from the footplate of the incus and also directly from the stapes agreed in amplitude with those from the long process of the incus. As the latter was more accessible this site was used routinely.

Fig. 1a shows some typical raw data for BM and incus response obtained by the two methods. The agreement between automatic plotting and the PSA measurements is generally better than that shown in Fig. 1a for the incus. This animal was chosen for illustration because it has the widest range of BM placements: it was, however, dead for the later part of the experiment.

The incus response is flat up to 1 kHz indicating the constant volume displacement characteristic of the closed sound
system; the 1-3 kHz level is rather variable from animal to animal followed by one or two sharp dips between 3 and 8 kHz not reflected in the BM response; the 10-20 kHz region is again regular and smooth followed by a rather high but variable (from animal to animal) cut-off up to 40-50 kHz.

Fig. 1a. Automatic plots and PSA measurements: from the probe tube microphone at the tympanic membrane; from the capacitance probe on the BM and on the long process of the incus. At constant voltage to the sound generator.

In some cases the very high rate of cut-off leads to an unconvincing function for the BM/incus ratio. It appears possible that the sharp dips and the higher rates of cut-off may result from changes in the direction of incus/stapes motion. The component of motion measured by the capacitive probe is $58^\circ \pm 5^\circ$ from the stapes axis which is substantially the same component as that measured by Johnstone et al. (1967).
Fig. 1b is based on the same data as Fig. 1a, plus those from two other points on the BM, reduced to constant incus amplitude and constant SPL. The curves have been smoothed for the reasons outlined above neglecting the sharp dips in incus response, and the dips in sound pressure due to reflections within the ear cavity above 20 kHz.

![Graph showing BM/incus ratio and amplitude of vibration at 100 dB SPL for three points on the BM.](image)

**Fig. 1b.** BM/incus ratio and amplitude of vibration at 100 dB SPL for three points on the BM. The sharp dips in the incus and BM responses and the interference dips in the sound calibration above 20 kHz (Fig. 1a) have been smoothed out. Corrections have been applied for incus angle, probe tip diameter, and the frequency response of the probe tube microphone, capacitive probe etc.

It can be seen from Fig. 1b that the steep high frequency slope does not continue indefinitely but reaches a plateau about 35 dB below the peak. This feature, common to all our
CAPACITIVE MEASURES OF B.M. VIBRATION

Otherwise the results are substantially in agreement with those of Johnstone. For constant incus amplitude the initial low frequency slope is 12 dB/octave, with a mean of 9 dB/octave from 0.1 to 20 kHz, and a maximum high frequency slope before the plateau of 100 dB/octave. For constant SPL the mean low frequency slope is only 2 dB/octave with a high frequency slope of 130 dB/octave. The shape of the low frequency characteristic agrees rather well with the differential CM measurements of Dallos (1969a) in the first turn, encouraging the view that the differential CM measured at low SPLs might reflect BM displacement.

![Fig. 2. Phase of BM response at three positions and the incus relative to the driver voltage.](image)

The poor filter characteristics of points on the BM compared with corresponding single cochlear nerve fibres recorded under comparable conditions (Evans, 1970) suggests the need for a second filtering mechanism beyond the BM. This requirement is made stronger by the presence of this plateau in a region where the neural response becomes steeper.

Fig. 2 shows examples of the phase response of the BM and incus relative to the driver voltage. The BM leads the incus by 0.25 cycles at low frequencies but eventually lags behind at high frequencies. In the 2-8 kHz region the incus response again shows anomalous behaviour with an additional phase lag not reflected in the BM response. Just above the BM peak frequency the phase curve steepens and then runs parallel with
the incus curve forming a phase plateau. The depth of this phase plateau below the incus curve increases with distance along the BM. For basal positions the phase shift can be less than 0.25 cycles. Such a low value is unlikely to be useful in a phase sensitive sharpening mechanism. The overall amplitude and phase characteristics at constant SPL are similar to those of an "m" derived low-pass filter.

The nature of the amplitude and phase plateaux rather suggests that the whole cochlea might be moving at these frequencies. This possibility was tested in a later experiment with E.F. Evans. The amplitude of motion of the cochlear shell was found to be more than 40 dB below the required level. Comparable measurements of spiral lamina motion have not yet been made although earlier findings showed the spiral lamina motion to be 20-30 dB below BM motion, in agreement with the findings of Johnstone et al. and Rhode.

Fig. 3 shows an impulse response computed (with the assistance of G.F. Pick) from the amplitude and phase data. The number of peaks present indicates that it might be possible to explain multiple-peaked click PSTs observed at lower frequencies in cochlear nerve fibres without recourse to a second "ringing" filter.

![Impulse response of the BM computed from the phase and amplitude data.](image)

Fig. 3. Impulse response of the BM computed from the phase and amplitude data.

Linearity of BM vibration level was measured over a wide range in a number of animals. Fig. 4 shows measurements at one BM placement for a number of different frequencies.
Within the physiologically reasonable range up to 110 dB SPL the BM was invariably linear to within about 1 dB contrary to the findings of Rhode. Above 110 dB SPL, however, significant deviations from linearity occurred. These were generally limited to frequencies above the peak response and were sometimes a "stretching" type as illustrated or, at other frequencies or in other cochleas, a "saturating" distortion. It is therefore not possible to explain away the differences between the shapes of the mechanical and neural filter characteristics on the grounds that the mechanical response becomes sharper at lower SPLs. Nor can the difference be attributed to some form of, say electro-mechanical, feedback to the BM dependent upon good physiological condition as this has no effect on the mechanical parameters.

Fig. 4. Measurement of linearity at five frequencies, shown relative to the BM response curve in the inset. The lines represent linearity.

Fig. 5 shows a graph of peak frequency of the BM response (or highest frequency before the response starts to drop rapidly) plotted as a function of distance from the extreme basal end of the BM. These data appear to fit in with the differential CM points derived from Dallos' data rather better than Békésy's direct measurements.
Fig. 5. Frequency of peak BM response as a function of distance from the basal end. Also shown are the points of Johnstone et al. measured from the stapes (to which 1 mm has been added), the data of Békésy, and cochlear microphonic data from Dallos (1971).

<table>
<thead>
<tr>
<th>G.P. Peak of BM response (kHz)</th>
<th>2f₁-f₂ (kHz)</th>
<th>f₁ (kHz)</th>
<th>f₂ (kHz)</th>
<th>SPL of f₁ (dB)</th>
<th>SPL of f₂ (dB)</th>
<th>Average level of 2f₁-f₂ relative to primaries (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>149L 20</td>
<td>16</td>
<td>18</td>
<td>20</td>
<td>101</td>
<td>101</td>
<td>-55</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>18</td>
<td>20</td>
<td>111</td>
<td>111</td>
<td>-37</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>22</td>
<td>24</td>
<td>100</td>
<td>99</td>
<td>&lt;60</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>22</td>
<td>24</td>
<td>110</td>
<td>109</td>
<td>-54</td>
</tr>
<tr>
<td>153L 21</td>
<td>18</td>
<td>20</td>
<td>22</td>
<td>89</td>
<td>90</td>
<td>&lt;51</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>20</td>
<td>22</td>
<td>99</td>
<td>100</td>
<td>-52</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>22</td>
<td>24</td>
<td>90</td>
<td>94</td>
<td>-55</td>
</tr>
<tr>
<td>156R 23</td>
<td>20</td>
<td>22</td>
<td>24</td>
<td>103</td>
<td>102</td>
<td>&lt;50</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>22</td>
<td>24</td>
<td>113</td>
<td>112</td>
<td>-41</td>
</tr>
</tbody>
</table>

*Considered as an equivalent sound input.

The table gives the results of measurements at 2f₁-f₂ for
f₂/f₁ ratios near 1.1. Although the levels of vibration at 2f₁ - f₂ were measured absolutely, they were converted to equivalent levels of sound input and expressed relative to the average level of the primaries. Instrumental distortion including the sound generator and capacitive probe was less than -60 dB.

At 110 dB SPL there is clearly a significant amount of distortion in the ear. As this becomes much less at lower levels, it is consistent with an over-loading distortion. At 90 and 100 dB SPL the level of 2f₁ - f₂ reached 55 dB below the primaries in the better cases. This would be about 40 dB below the level required to account for the psychophysical and neurophysiological results (Goldstein, 1967; Goldstein & Kiang, 1968).

Goldstein (1967, 1970) and Smoorenburg (1971) each saw the need to invoke a filter followed by a nonlinearity and a further filter to account for the observed properties of the combination tone 2f₁ - f₂. As 2f₁ - f₂ is not present on the BM the non-linearity cannot be hydrodynamic (Tonndorf, 1970), but may be in the coupling between BM and excitation of the hair cells. Against this Dallos (1969b) has shown that 2f₁ - f₂ in cochlear microphonics does not vary appropriately with f₂/f₁ and sound level. If in fact cochlear microphonics reflect the receptor potentials, the non-linearity must be central to the receptor potentials. The second filter cannot be closely coupled to the BM and like the non-linearity it presumably comes after the mechanical transduction process but must precede the generation of nerve impulses. Perhaps the simplest assumption might be that the non-linearity and second filter are electrical in nature taking input from the local CM current or potential.

References
Dallos, P. (1969a). J. Acoust. Soc. Amer. 45, 62, Fig. 1, 1001, Fig. 3.
Dallos, P. (1971). J. Acoust. Soc. Amer. 49, 1141, Fig. 1-2, 1820, Fig. 1.
The stimulation of the sensory cells in the organ of Corti possibly depends on the tectorial membrane. The size as well as the attachment of this membrane may affect the process of stimulation. Since the tectorial membrane in the domestic pig is not only very wide but also very thick in relation to what we know about the same membrane in humans, the tectorial membranes of several other species were investigated. Most of them like cat, guinea pig, and mouse have similarly large tectorial membranes, whereas rabbit and especially squirrel monkey showed membranes which seem to be more similar to the human tectorial membrane.

The inner ear, the cochlear duct and especially the tectorial membrane of squirrel monkey were investigated using the same methods as described in an article on inner ears of domestic pigs (ZWICKER 1971). The geometry of scala tympani, scala vestibuli, cochlear duct, basilar membrane and organ of Corti was found to be in very good agreement with the data reported by IGARASHI (1964). On the other hand some differences could be observed for the tectorial membrane with regard to its size and attachment. But these differences are relatively small in comparison with the differences between the tectorial membrane of the domestic pig (D) and that of the squirrel monkey (S) as shown in the schematic drawings of Fig. 1. In the basal winding (DI, SI in Fig. 1) the difference is still small but becomes larger in the second winding (D II, S II) and is remarkable in the apical winding (III) near the helicotrema.
There the tectorial membrane of the domestic pig (D III) does not only reach but even overlaps the organ of Corti. The membrane is thick and seems to be fastened quite well at the limbus. A special net of fibres (Randfasernetz) at the outer rim of the tectorial membrane in radial direction could not be observed. The membrane did not seem to be attached to the Hensen cells or any other part of the organ of Corti.

The tectorial membrane of the squirrel monkey is much thinner and not as wide. It just reaches in radial direction up to the most outward row of the outer sensory cells and does not end as abruptly as in domestic pig but tapers off into single fibres which seem to be attached in the two apical windings to the Deiter - cells between and behind the outer sensory cells. This connection appears to be designed in such a way that through holes and tunnels the liquid in the sulcus spiralis is connected with the liquid in the scala media.

Although most of the single parts of the complicated structure of the cochlear position reach a strong similarity in monkey and pig, the tectorial membranes look quite different even though the membranes may be used for the same purpose in both animals. These observations give rise to the supposition that the ductus cochlearis is constructed in such a way that liquid can flow between the reticular membrane and the tectorial membrane. NEUBERT (1950) has assumed an alternating flow at this place with the frequency of the stimulation. He proposed this alternating current to be responsible for the stimulation of the sensory cells. In order to test this possibility the inner ears of domestic pigs were stimulated with a vibrator pin at the stapes. Only very small alternating currents could barely be observed but relatively strong directed currents were seen. In order to see this in the binocular microscope the amplitude of vibration has to be very high, it corresponded to about 20 dB above the threshold of pain.

Whether this observed current is a secondary effect or a primary effect causing the stimulation of the hair-cells and transferring the wide area of basilar membrane vibrations into a much smaller area of directed current is not clear as yet. Further investigations of inner ears of animals as well
as measurements on enlarged hydromechanical models are necessary. Nevertheless, it is of interest to realize that very young kittens which are deaf if under 10 days of age have a connection between tectorial membrane and organ of Corti (LINDEMANN, et al 1971) which does not allow streaming of liquid. This connection opens at just that time after birth, then they begin to hear.

References.
Igarashi, M. (1964): The Inner Ear Anatomy of the Squirrel Monkey,
U.S. Naval School, Pensacola, Florida, Monograph No 8
Lindeman, H. H. et al (1971): The Sensory Hairs and the Tectorial Membrane in the Development of the Cat's Organ of Corti,
Acta Otolaryng 72, 229-242.
Neubert, K. (1950): Die Basilarmembran des Menschen und ihr Verankerungssystem,
Z. Anat. 114, 539-588.
Zwicker, E. (1971): Die Abmessungen des Innenohrs des Haus­s­weines,
Acustica 25, 232-239.
Evidence from aural combination tones and musical tones against classical temporal periodicity theory.

Julius L. Goldstein
Research Laboratory of Electronics, M.I.T., Cambridge, Massachusetts 02139

ABSTRACT

Classical temporal periodicity theory has attributed the distinct, though nominally related perceptual phenomena of aural combination tones and musical pitch of complex tones to aural detection of temporal periodicities created by linear interference among Fourier components in the sound. The growing body of psychophysical data on combination tones is compatible instead with a nonlinear mechanism that generates stimulus-like intermodulation products in the inner ear. Action potentials from the cochlear nerve in anesthetized cats give evidence of stimulus-like intermodulation products with properties similar to those measured psychophysically. Although the phase properties of these action potentials are not fully understood, it is already evident that the phase properties predicted by classical periodicity theory do not accord with the psychophysical data. New experiments on musical pitch of complex tones have demonstrated that musical perception is little affected whether the stimulus spectrum is presented normally (monotically) to one ear or split (dichotically) between subjects' two ears. This central
formation of musical pitch contrasts with the absence of aural combination tones with the dichotic stimuli. [Work supported by National Institutes of Health (Grant 5 PO1 GM14940-05)].

INTRODUCTION

Classical periodicity theory has attributed the distinct perceptual phenomena of aural combination tones and pitch of complex tones to aural detection of temporal periodicities created by linear interference among Fourier components in the sound. Recent work has indicated instead that aural combination tones originate as stimulus-like intermodulation products as a consequence of essential nonlinearities in the cochlear frequency analyzer, and that musical pitch of complex tones originates centrally in consequence of measurement across the frequency dimension of an internally represented short-term spectrum.

1. Aural Combination Tones

1.1. Amplitude properties from psychophysics

Auditory perception of complex tones reveals the apparent presence of a stimulus modification whereby a peripheral nonlinear mechanism adds intermodulation products to the stimulus (Helmholtz, 1863). The properties of this phenomenon differs greatly from predictions by simple nonlinear mechanisms (Hermann, 1891; Zwicker 1955) and so the older periodicity theory (Young, 1800) for aural combination tone has persisted as an alternative (Meyer, 1956). Interest in these phenomena was awakened by recent psychophysical demonstrations that significant combination tones occur at low sound levels, that they correspond to a stimulus modification in which the amplitudes...
of the intermodulation products grow nearly in proportion to stimulus amplitude, and that they depend upon frequency parameters in a manner strongly suggesting cochlear origin (Goldstein, 1967). These findings mark the initiation of a reversal of the dominant view of the past century that aural combination tones, independently of mechanism, are negligible at low sounds levels (see review by Plomp, 1965).

All known properties of aural combination tones match those of stimulus tones. The earliest reported are i) a tonal sensation that matches that of a simple tone of frequency $f_{ct} = mf_1 + nf_2$, where $m$ and $n$ are integers, $f_1$ and $f_2$ but not $f_{ct}$ are stimulus tones, ii) a loudness that fluctuates in the presence of a probe tone of low intensity and of frequency near $f_{ct'}$, iii) masking accomplished with least power by choosing the center frequency of a narrow band of noise at $f_{ct}$, iv) elimination (or cancellation) by a null adjustment in amplitude and phase of a stimulus tone of frequency $f_{ct}$ which is added to the primary stimulus [see Fig. 1], and v) agreement between the cancellation amplitude and the amplitude of an "equilibration" stimulus tone which in the presence of the lower primary ($f_1$) matches the loudness of the combination tone [see Fig. 2].

Additional psychophysical studies have confirmed known and added new information (Zwicker, 1968; Helle, 1969; Schroeder, 1969; Greenwood, 1971; Hall, 1972; Smoorenburg, 1971). Greenwood, (1971) investigated masking patterns produced by complex tones and complex bands of noise and found that narrow bands of noise in the stimulus produce aural distortion phenomena very similar to those of stimulus tones. Masking patterns revealed increments and decrements in level that were accounted for by
aural modification of the masker or maskee with stimulus-like intermodulation products. Hall (1972b) demonstrated that combination tones of different order but of equal frequency \((f_2 - f_1 \text{ and } 2f_1 - f_2, \text{ for } f_2/f_1 = 3/2)\) summed to yield a cancellation amplitude that depended upon stimulus phase in precisely the vectorial form predicted by a linear interaction of simple harmonic oscillations, i.e., in accord with stimulus-like intermodulation products. Smoorenburg (1971) using an equilibration technique, showed that post-stimulus masking at the combination frequency \(2f_1 - f_2\) produced by a two-tone masker \((f_1 < f_2)\) matches that of a stimulus consisting of the cancellation tone plus the lower primary \(f_1\).

1.2. Amplitude properties from auditory-nerve physiology

The crucial question for auditory theory is whether there is physiological validity in the compelling psychophysical conception that aural combination tones exist as stimulus-like cochlear intermodulation products [Goldstein and Kiang, 1968; Goldstein, 1970]. Possibly aural combination tones are mediated by neural periodicities that reflect aural detection of temporal periodicities created by linear interference among Fourier components in the stimulus (Meyer, 1956; Rose, et al., 1969). Goldstein and Kiang (1968) found that action potentials from single fibers in the auditory nerve of anesthetized cats give evidence of stimulus-like intermodulation products with amplitude properties similar to those measured psychophysically in awake humans.

As would be expected from stimulus-like intermodulation, auditory-nerve responses were found that were selectively tuned as well as phase locked to combination tones. Figure 3 shows
two cases of selective responses to combination tones. The stimulus comprises the simple tones of frequencies $f_1$ and $f_2$ and the combination frequency $2f_1 - f_2$ approximates the characteristic frequency of the nerve fiber. In Fig. 3a the stimulus is periodic, consisting of third and fourth harmonics of the reference fundamental. The PST histogram period is one cycle of the reference fundamental. No significant response above spontaneous is elicited by either stimulus tone presented alone. With both tones simultaneous a dramatic response to the combination tone $2f_1 - f_2$ is obtained. The phase-locked time pattern corresponds to $2f_1 - f_2$, which in this case is the second harmonic of the reference fundamental. For stimulus frequencies above about 5 kHz, no phase locking to the fine structure is measurable for simple-tone stimuli, but time locking to the envelope of tone bursts can be measured. In Fig. 3b a dramatic response is obtained only when both tones are simultaneous.

When a nerve fiber is responsive to one or both stimulus tones, more complex response patterns are observed. Figure 4 illustrates an example in which the fiber appears to be responding to a superposition of lower primary plus the aural combination tone $2f_1 - f_2$ (top row). Cancellation of the aural combination tone (second row) simplifies the response pattern to one approximating neural response to a simple tone of frequency $f_1$. A salient property of the cancellation measurements is that the cancellation tone alone (third row) generates a phase-locked response that is in antiphase with the phase locking at the combination frequency arising from the primary stimulus (top row). These physiological experiments demonstrated that cancellation levels (with $f_2 > f_1 > f_{cf}$) have similar
dependence upon primary stimulus level and frequencies as found psychophysically (see Fig. 5).

Using intracochlear differential electrodes across the scalae Dallos (1969) searched cochlear microphonic potentials for distortion components that mirror the psychophysical and physiological amplitude data, but he found no correspondence. It is not clear whether the cochlear microphonic expected from a stimulus-like combination tone could be measured in the presence of the potentials generated by the stimuli used, because no measurements were reported with stimuli that simulate the anticipated aural distortion.

1.3. Phase properties of $2f_1 - f_2$

Although the stimulus-like aural combination tones measured psychophysically and physiologically have similar peculiar amplitude properties, their phase properties differ greatly. At low and moderate sound levels, the psychophysical phase decreases typically at 6-10 degrees per decibel increase in stimulus level (Fig. 6), while the physiological phase is independent of stimulus amplitude (see Goldstein and Kiang, 1968, Figs. 2, 3, & 7; Goldstein, 1970, Fig. 9). In addition the psychophysical data show a systematic dependence upon stimulus frequencies (Fig. 7); however, no corresponding physiological data are available.

It is instructive to examine the phase properties predicted by classical periodicity theory. A simplified model for auditory nerve phase locking is given in Fig. 8a. Phase locking to combination frequencies is generated here only by the non-linearity of the transducer. The shape of the expected PST (or period) histogram for an arbitrary combination frequency
can be described as a function of the memoryless rectifier and its signal (Goldstein, 1972). The particular form of the rectifier given, exponentiation, provides a good description of salient temporal aspects of responses to simple tones and wide band noise (Colburn, 1970; Siebert, 1970). Examples of expected histograms given by this model are shown in Fig. 8b; the four histograms are normalized for equal area. Because of symmetries (and absence of memory following the rectifier in this simplified model) the expected histograms to the four third-order combination tones are identical.

Consider the properties of a cancellation tone that is added to the primary stimulus (see Fig. 8a) and chosen to eliminate phase locking at the combination frequency \(2f_1 - f_2\) (i.e. null the fundamental Fourier coefficient of the expected histogram). The cancellation phase referred to the stimulus, \(\phi_{ct}\), and the cancellation amplitude referred to the rectifier input are

\[
\phi_{ct} = \pi + (2\alpha - \beta - \gamma), \quad c \sim 2I_2(a)I_1(b)/I_0(a)I_0(b), \quad \text{where} \quad I_m(z) \text{ is the } m\text{th order modified Bessel function of the first kind.}
\]

The psychophysical cancellation data in Fig. 7 are not fit by the above theory. The theory indicates that \(\phi_{ct}\) approaches \(\pi\) as \(f_2/f_1\) approaches one, while the data imply \(\phi_{ct}\) should approach zero. An expansion nonlinearity is responsible for the theoretical phase, while the psychophysical data imply a compression nonlinearity. Moreover, the phase characteristics responsible for the the psychophysical data possess far too much curvature (Goldstein, 1967, sect. D) to be identified with the phase characteristics of single auditory nerve fibers (Goldstein, Baer and Kiang, 1971). By coincidence only, over the range of
transducer inputs \((a = b)\) that are descriptive of the phase locking to primary tones measurable in auditory nerve fibers, the cancellation level shares some characteristics of the data (see table I).

### Table I

| \(20 \log_{10} a/c\) | 27.8 | 20 | 13.6 | 12 | 16.3 |

Further analysis of the transducer model for neural phase locking provides a valuable theoretical basis for distinguishing between stimulus-like and transducer-like combination tone phenomena in auditory nerve data (Goldstein, 1970, 1972). In response to a two-tone signal this model with an arbitrary memoryless rectifier predicts phase locking to combination tones having the following pertinent properties.

i) The expected period histogram for a combination tone is either in phase or out of phase with respect to the period histograms for the input frequencies. Thus with \(a\) and \(b\) as the phases of the input tones, the expected period histogram for a combination frequency \(pf_1 + qf_2\) will have its maximum at the phase of either \(pa + qb\) or \(\pi + pa + qb\). It is useful to think of the former in phase case as arising from an expansion nonlinearity; while the out of phase case results from a compression nonlinearity.

ii) The depth of phase locking at combination frequencies satisfies a well defined rank ordering. A powerful ranking relation can be obtained by defining a measure of phase locking, \(S\) for synchrony, as the \((\text{Maximum height of the expected period histogram} - \text{Height of the expected period histogram a half cycle beyond the maximum})/\text{(Average height of the}\)
expected period histogram). The ensuing ranking relations are
that synchrony to a combination frequency \( pf_1 + qf_2 \) cannot exceed either, the synchrony to \( f_1 \) divided by the magnitude of \( p \), or the synchrony to \( f_2 \) divided by the magnitude of \( q \).

On the basis of this theory additional physiological data (e.g., Goldstein and Kiang, 1968, Fig. 2) can be given more certain interpretation as requiring a transducer input containing energy at the combination frequency. Moreover, on the basis of this theory we have discovered that auditory nerve responses to narrow band two-tone stimuli, with equal amplitudes and frequencies \( f_1 < f_2, 2f_1 - f_2 = \omega_f \), usually exhibit stimulus-like phase locking to \( 2f_1 - f_2 \) and phase locking to \( f_2 \) that is transducer-like. For these cases the transducer model requires inputs of \( f_1 \) and \( 2f_1 - f_2 \), while expansion nonlinearity in the rectifier generates phase locking to \( f_2 \).

1.4. Conclusion on aural combination tones

Evidence is accumulating that psychophysical and physiological combination tone phenomena arise from a nonlinear mechanism that generates stimulus-like intermodulation products in the inner ear. The amplitudes of the combination tones referred to the stimulus are similar for the psychophysical and physiological data, but the phases are different. This discrepancy in phase could be caused by the presence of a level dependent modification of the peripheral nonlinearity in the alert human subjects for the psychophysical experiments, and its absence in the anesthetized cats in the physiological experiments. While the discovery of such unknown mechanism would be of great importance, it would not modify our conception of aural combination tones as peripheral intermodulation products. We must however, also recognize the possibility that
the phase enigma could be pointing to a gap in our conceptual understanding of central processing of cochlear signals (Goldstein, 1971).

2. **Pitch of Complex Tones**

2.1. Basic properties

Classical periodicity theory for musical sounds containing no fundamental energy appears to have a simpler denouement than aural combination tones. According to the modern and most influential formulation, the residue theory (Schouten 1940a), musical notes are signalled directly by temporal periods in the cochlear output that arise from beating among Fourier components of the stimulus which the ear's filters fail to resolve.

The single outstanding empirical property (Hermann, 1912; Schouten, 1940b; deBoer, 1956; Schouten, Ritsma and Cardozo, 1962) is that the note communicated - i.e., the musical pitch reported - for inharmonic complex tones does not correspond to a difference frequency. The musical pitch of a complex tone changes nearly in proportion with the uniform frequency perturbation of its component frequencies away from an harmonic complex. The proportionality factor may be interpreted - according to either spectral or temporal theories of pitch processing - as being inversely equal to the harmonic position of the "effective carrier" of the internal signal (de Boer, 1956; Schouten, Ritsma and Cardozo, 1962). It was first discovered that when the complex tone is bandpass, the effective carrier is approximated by the center frequency of the band (first effect of Schouten, et al., 1962); however a systematic saturation of the effective carrier occurs
Aural Combination Tones and Musical Pitch

GOLDSTEIN

11

(second effect of Schouten, et al., 1962) as the bandpass region is centered about higher harmonics of a complex tone (de Boer, 1956; Schouten, et al., 1962).

Neither spectral nor temporal theories could explain the saturation effect for effective carriers that are lower in frequency than any Fourier component of the complex tone stimulus. The discovery that aural combination tones manifest a nonlinear stimulus modification that occurs at low as well as high stimulus levels supplied a rational basis for attributing the saturation effect to pitch mediation by stimuli augmented with combination tones (Goldstein and Kiang, 1968; Ritsma, 1970; Smoorenburg, 1970); however neither spectral nor temporal theory was clearly favored.

Attention was turned toward spectral theories when it was discovered that the effective carrier for wide band signals is approximately quadruple the reference fundamental (Ritsma, 1967) and that the effective carriers for bandpass complex tones never exceed about ten times the reference fundamental (Ritsma, 1970; Smoorenburg, 1970; Goldstein, 1967b). In the region of the fourth harmonic, other psycho-physical data on phase effects (Mathes and Miller, 1947; Goldstein, 1967c) and on masking (Zwicker and Feldtkeller, 1967) suggest that little temporal interaction among stimulus partials occurs in the auditory periphery at low stimulus levels. Secondly, in the region to the tenth harmonic, which is effective in communicating musical pitch, data on behavioral resolution of partials and combination tones for "continuous" complex tone stimuli suggest (Plomp, 1964; Goldstein, 1967) that the auditory system has the capability of physically resolving all Fourier components in the stimulus or its
Aural Combination Tones and Musical Pitch

peripheral modification.

2.2 New experiments

The first conclusive demonstration against the necessity for cochlear periodicity detection in the mediation of pitch of complex tones was the discovery that musical pitch is also evoked by dichotically presented two-tone complexes in which only one Fourier component is in each ear (Houtsma and Goldstein, 1970, 1971, 1972). Moreover, the sufficiency of cochlear periodicity detection (of temporal intervals corresponding to pitch) is rendered highly unlikely by the similarity of musical pitch for normal and dichotic modes of presentation at low stimulus levels (e.g., 20 dB SL). At moderate and high stimulus levels aural combination tones are known to play a significant role in the normal presentation mode.

The ability of subjects to perceive musical pitch of complex tones was investigated with the musical interval recognition paradigm shown in Fig. 9. Several reasons underlie the design of this paradigm. The use of sequences of sounds reflects first, the fact that recognition of pitch intervals rather than absolute pitch is the more natural task in music for which a natural language exists, and second, our experience that isolated complex tones are more apt to be perceived analytically as a mixture of partial tones than as sequences of complex tones. Finally, the randomization of the harmonic numbers used in presenting any one complex tone effectively prevented subjects from reporting intervals on the basis of only a single Fourier component in each sound.

Subjects's ability to identify the eight intervals given in Fig. 9 was scored in an eight alternative forced choice experiment as a function of harmonic number n and of reference
Experimental results for harmonic two-tone sounds at 20 dB SL, given in Fig. 10, (Houtsma and Goldstein, 1972) demonstrate the similarity of the monotic (both components of each sound in one ear) and dichotic (one component in each ear) modes of presentation. Similar experiments with harmonic complex tones of up to six components and with inharmonic complex tones confirm the behavior found with the simpler sounds as well as basic properties discovered by earlier investigators of musical pitch.

2.3. Conclusion on musical pitch

The new experiments demonstrate that the musical pitch of complex tones (sans fundamental energy) originates in a mechanism that reads across a central projection of the cochlear frequency map. No single place in the spectrum can mediate the musical pitch of dichotically presented complex tones as would be required by classical place or periodicity theories. Although clear constraints on new theories exist (Houtsma and Goldstein, 1971, 1972), it appears premature to decide whether to opt for a generalized place theory that operates on the short-term spectral envelope or for a generalized periodicity theory that utilizes temporal information from a broad range of places (cf. Whitfield, 1970).

3. Conclusion on Classical Periodicity Detection

Arguments that evidence for "periodicity detection" from musical pitch suggests by parsimony a similar temporal mechanism for aural combination tones have been dissipated. Aural combination tones have the properties of stimulus-like intermodulation products and these occur only when two or more simple tones are present in the same ear. Musical pitch with
essentially identical properties is evoked when a two-tone complex is presented either dichotically or monotonically. Moreover, pitch is effectively communicated only when the stimulus tones are sufficiently disparate in relative frequency. Thus the distinct phenomena of aural combination tones and musical pitch have very different points of origin in the auditory system. Aural combination tones arise from peripheral intermodulation and are stimulus-like. Musical pitch originates from hierarchical processing across the characteristic frequency dimension of central projections of the cochleae, while peripheral interaction among Fourier components of a complex tone is irrelevant.
Aural Combination Tones and Musical Pitch

GOLDSTEIN

FIG. 1: Relative level of cancellation tone required to nullify perception of aural combination tone. Properties of \( f(n) \) are non-Helmholtzian. (From Goldstein 1967, 1970; see also Hall, 1972).

\[
f(n) = f_1 - n(f_2 - f_1)
\]

FIG. 2

Relative level of equilibration stimulus-tone which in presence of the lower of the two primaries matches the loudness of a combination tone. Loudness balance and cancellation (Fig. 1) data are similar. (Loudness balance procedure as in Goldstein (1967) except that here all stimuli were monaurally presented via earphones.)
FIG. 3: (a, Left) PST histograms of responses to single-tone (top two histograms) and two-tone (bottom histogram) stimuli. (Unit 451-10, CF=2.69 kHz.) The stimulus frequencies $f_1$ and $f_2$ are harmonically related. Each histogram is synchronized with one cycle of the stimulus period. The rate of spontaneous activity for this fiber was low (5 spikes per second), so that the effect of the combination tone is particularly striking.

(b, right) PST histograms of responses to single-tone (middle histograms) and two-tone (bottom histogram) stimuli. (Unit 466-19, CF=7.88 kHz.) The stimuli were tone bursts and the histograms are synchronized with the onset times of the bursts. The presentation time and duration of the tone bursts are shown by the horizontal lines at the bottom of the figure. The tone bursts were presented at a rate of 10 per second. (From Goldstein and Kiang (1968), Figs. 5 and 10).
Aural Combination Tones and Musical Pitch

UNIT K504-19 CF = 1.58 kHz

PST HISTOGRAM SYNCHRONIZED TO:

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>2f_1 (2.10 kHz)</th>
<th>f_2 (2.72 kHz)</th>
<th>2f_1 - f_2 (1.48 kHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>f_1, -54 dB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f_2, -54 dB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f_1, -54 dB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f_2, -54 dB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f_{CT}, -74 dB</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

FIG. 4: Temporal patterns of auditory nerve fiber response to a two-tone signal, manifesting an effective stimulus comprising the aural combination tone (2f_1 - f_2) plus the lower primary (f_1).

RECTIFIER MODEL (TOP)

\[ x(t) = 1.2 \cos \omega_1 t + 1.2 \cos \left\{ (2\omega_1 - \omega_2) t - 65^\circ \right\} \]

\[ y(t) = \exp(x) \]

FIG. 5: Relative level of cancellation tone required to nullify auditory-nerve phase locking at the combination frequency 2f_1 - f_2. (From Goldstein, 1970).
FIG. 6: Phase of the psychophysical cancellation tone for the aural combination tone $2f_1 - f_2$ generated with various two-tone stimuli. The phase is referred to the primaries: 

$$\phi = \alpha - (2\beta - \gamma),$$

where $\alpha$, $\beta$ and $\gamma$ are the phases at $2f_1$, $f_2$, $f_1$ and $f_2$ respectively. The primary stimulus tones are of equal level. The levels given are nominal, being based upon earphone voltage and coupler-calibration curves. The cancellation adjustments were null, being critical in both directions for both phase and level.

FIG. 7: Cancellation phase as in Fig. 6 except here the primary frequencies are varied systematically. The lower primary tone is the fixed parameter for each curve (From Goldstein, 1967).
Aural Combination Tones and Musical Pitch

FIG. 8: (a, top) Simplified model for auditory-nerve phase locking. The nonlinearity of the rectifier within each channel is the only source of phase locking to frequencies not represented in the stimulus spectrum.

(b, bottom) Expected period (or PST) histograms for various reference frequencies—the two primaries and the four third-order combination frequencies. The stimulus to the exponential rectifier is \( a/I_2(a) \{ \cos(\omega_1 t + \alpha) + \cos(\omega_2 t + \beta) \} \). The Bessel function \( I_0(a) \) normalizes the area of the expected histogram. The synchrony measure \( S \) quantifies the depth of phase locking (see text).
FIG. 9: Experimental paradigm for musical interval identification experiments: (a) musical intervals to be identified, (b) time envelope of the stimulus for each musical interval, (c) the three possible two-tone stimuli for each of the two notes; for each note a random choice is made among the three possible stimuli. (From Houtsma and Goldstein, 1972).

FIG. 10: Performance contours (percent correct) for musical interval identification experiments with pairs of randomly chosen successive upper harmonics (From Houtsma and Goldstein, 1972).
REFERENCES


Dallos, P. (1969): The combination tone $2f_1-f_2$ in microphonic potentials,

Goldstein, J.L. (1967): Auditory nonlinearity,

(1967b): unpublished data,
Laboratory of Psychophysics, Harvard.

(1967c): Auditory spectral filtering and monaural phase perception,


(1971): Progress report on aural combination tones,

(1972): Neural phase locking to combination tones predicted by simple transducer models,

and Kiang, N.Y.S. (1968): Neural correlates of the aural combination tone $2f_1-f_2$,

(National Educational Consultants, Baltimore, Md.)

Greenwood, D.D. (1971): Aural combination tones and auditory masking,

(1971b): Comparison of masking by combination bands and bands of noise of external origin,

Hall, J.L. (1972): Auditory distortion products $f_2-f_1$ and $2f_1-f_2$,
Aural Combination Tones and Musical Pitch

Goldstein

(1972b): Monaural phase effect: cancellation and reinforcement of distortion products \( f_2 - f_1 \) and \( 2f_1 - f_2 \), J. Acoust. Soc. Am. in press.


List of Participants

F.A. Bilsen, Lab. for Technical Physics, Technical University Delft, Delft, Holland
E. de Boer, Physical Lab. of the E.N.T. Dept., University of Amsterdam, Amsterdam, Holland
G. van den Brink, Dept. of Biological and Medical Physics, Medical Faculty Rotterdam, Rotterdam, Holland
B.L. Cardozo, Institute for Perception Research, Eindhoven, Holland
H. Duifhuis, Institute for Perception Research, Eindhoven, Holland
E.F. Evans, Medical Research Group, Dept. of Communication, University of Keele, Keele-Staffs., England
H. Fastl, Institut für Elektrotechnik, Technische Hochschule München, München, W. Germany
J.A.M. van Gisbergen, Lab. for Medical Physics, University of Nijmegen, Nijmegen, Holland
J.L. Goldstein, Dept. of Electrical Engineering, M.I.T. Research Laboratory of Electronics, Cambridge-Mass., U.S.A.
C.A.A.J. Greebe, Institute for Perception Research, Eindhoven, Holland
L.M. Grobben, E.N.T. Dept., State University, Utrecht, Holland
T. Houtgast, Institute for Perception T.N.O., Soesterberg, Holland
P.I.M. Johannesma, Lab. for Medical Physics, University of Nijmegen, Nijmegen, Holland
J.R. Johnstone, Dept. of Communication, University of Keele, Keele-Staffs., England
Th.A. de Jong, Institute for Perception Research, Eindhoven, Holland
H.R. de Jongh, Physical Lab. of the E.N.T. Dept., University of Amsterdam, Amsterdam, Holland
H. Kronberg, III Physikalisches Institut, Universität Göttingen, Göttingen, W. Germany
P. Kuyper, Physical Lab. of the E.N.T. Dept., University of Amsterdam, Amsterdam, Holland
P.J.J. Lamoré, Dept. of Biological and Medical Physics, Medical Faculty, Rotterdam, Holland
V. Mellert, III Physikalisches Institut, Universität Göttingen, Göttingen, W. Germany
A.R. Møller, Dept. of Physiology II, Karolinska Institute, Stockholm, Sweden
B.C.J. Moore, Dept. of Psychology, University of Reading, Reading, England

J.J.M. Neelen, Institute for Perception Research, Eindhoven, Holland

L.P.A.S. van Noorden, Institute for Perception Research, Eindhoven, Holland

R. Plomp, Institute for Perception T.N.O., Soesterberg, Holland

L. Pols, Institute for Perception T.N.O., Soesterberg, Holland

A. Rakowski, Panstwowa Szkoła Muzyczna, Warszawa, Poland

R.J. Ritsma, Audiological Institute, E.N.T. Dept., University of Groningen, Groningen, Holland

M. Rodenburg, Dept. of Biological and Medical Physics, Medical Faculty Rotterdam, Rotterdam, Holland

B. Scharf, Dept. of Psychology, North Eastern University, Boston-Mass., U.S.A.

J.F. Schouten, Institute for Perception Research, Eindhoven, Holland

M. Schroeder, III Physikalisches Institut, Universität Göttingen, Göttingen, W. Germany

C.R. Steele, Stanford University, Stanford-Calif., U.S.A.

E. Terhardt, Institut für Elektroakustik, Technische Hochschule München, München, W. Germany

L.L.M. Vogten, Institute for Perception Research, Eindhoven, Holland

R. Weber, III Physikalisches Institut, Universität Göttingen, Göttingen, W. Germany

F.L. Wightman, Dept. of Psychology, University of California, San Diego, La Jolla-Calif., U.S.A.

J.P. Wilson, Dept. of Communication, University of Keele, Keele-Staffs., England

E. Zwicker, Institut für Elektrotechnik, Technische Hochschule München, München, W. Germany