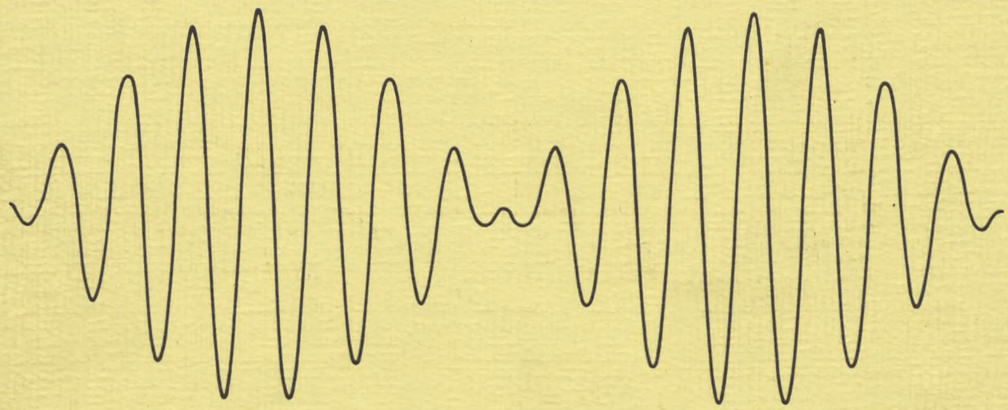
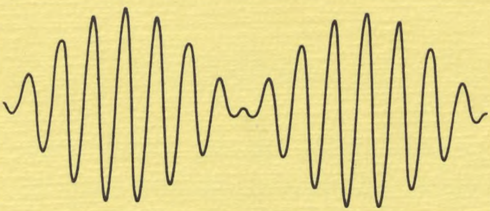
ASPECTS OF TWO-TONE PERCEPTION



G. F. Smoorenburg

ASPECTS OF TWO-TONE PERCEPTION



G. F. Smoorenburg

ASPECTS OF TWO-TONE PERCEPTION

PROEFSCHRIFT

ter verkrijging van de graad van doctor
in de wiskunde en natuurwetenschappen
aan de Rijksuniversiteit te Utrecht,
op gezag van de Rector Magnificus
prof. dr. Sj. Groenman,
volgens besluit van het college van dekanen
in het openbaar te verdedigen op
maandag 13 september 1971
des namiddags te 4.00 uur (precies)

door

GUIDO FRANCISCUS SMOORENBURG

geboren op 2 augustus 1943 te Haarlem

Promotor: prof. dr. M. A. Bouman

Natura non nisi parendo vincitur

De natuur wordt slechts overwonnen
door haar te gehoorzamen.

(Francis Bacon, 1620)

INTRODUCTION

The title of this dissertation 'aspects of two-tone perception' is ambiguous. For what is a tone? Generally, any auditory sensation having pitch is called a tone. But in this case the word 'tone' is used in the restricted sense of the sound produced by a sinusoidal vibration. Such a tone is an element of the auditory sensation. Its pitch is unequivocally related to the frequency of the vibration. A mixture of sinusoidal vibrations of various frequencies is perceived as a number of tones with pitches corresponding to the stimulus frequencies. It is a fundamental property of the ear that the frequency components of a stimulus are represented perceptually as tones. The ear acts as a spectral analyzer.

The close relation between a frequency component and a tone has become so self-evident that the word 'tone' may stand for both the frequency component of the stimulus and the percept. In the title of this dissertation 'tone' stands for frequency component; the perception of stimuli with two frequency components was studied. The distinction between the two uses of the word 'tone' is unimportant in so far the *frequency component - tone* relation holds strictly. However, the subject of this dissertation is that this is not generally true. The two-frequency stimulus f_1, f_2 may evoke more than two tones in the ear. For example, a difference tone corresponding to the frequency component $f_2 - f_1$ may be perceived. Such a tone is conceivably the product of a nonlinearity of the ear. Frequency analysis by the ear succeeding the nonlinearity would reveal the distortion in the perception of additional tones. These additional tones are called combination tones.

The difference tone (f_2-f_1) is audible only at high stimulus levels. It can be described as the product of an overloading type of nonlinearity. The combination tones which are studied in this dissertation correspond to frequency components $f_1-k(f_2-f_1)$, where k is a small integer. These combination tones are audible even at very low stimulus levels which suggests that they are the products of a more essential nonlinearity. The reason for the study of these combination tones was the working hypothesis that such a study might help to understand the nonlinear detection of the ear. Two-frequency stimuli were applied in this research since they have the smallest number of variables.

This dissertation consists of three papers. The abstracts of these papers are given below.

ABSTRACTS

PITCH PERCEPTION OF TWO-FREQUENCY STIMULI

Experiments on the pitch of complex tones produced by two frequency components are reported. An exploratory experiment revealed that subjects perceive the pitches of individual part-tones or the stimulus as a whole with a pitch corresponding to about the fundamental frequency. The latter pitch was investigated more thoroughly. In case of adjacent harmonics, the pitch corresponded to the (absent) fundamental frequency. A shift of the frequencies away from such a harmonic situation while maintaining a constant frequency difference resulted in a pitch shift. For higher harmonic numbers, the pitch shift was larger than could be met by current theories. The large pitch shift was explained by taking into account an auditory nonlinearity which generates combination tones of the type $f_1-k(f_2-f_1)$. Sound-pressure level dependence of the pitch shift could be explained in the same manner. When the combination tones were masked, the large pitch shift dimin-

ished. With regard to the pitch mechanism, the results suggested that the detection of the low pitch of the complex tone requires spectral information.
(This paper has been published in The Journal of the Acoustical Society of America, Vol. 48, 924-942, October 1970.)

AUDIBILITY REGION OF COMBINATION TONES

This paper describes research on combination tones of the type $f_1-k(f_2-f_1)$. Such tones are audible during stimulation by the two frequency components f_1 and f_2, $f_1<f_2$. The results of the experiments suggest that these combination tones are audible only in a restricted frequency region below f_1. The lower limit of this 'audibility region', for the combination tone $2f_1-f_2$ (k=1), was determined primarily by the level of the lower stimulus component L_1 and additionally by L_2. The width of the audibility region appeared to be highly subject dependent. The lowest stimulus level for which $2f_1-f_2$ was audible was 15-20 dB SL if both stimulus components had equal levels. If the lower component was set at 40 dB, then the higher one could be reduced to 4 dB before $2f_1-f_2$ reached threshold. Higher order combination tones could be perceived up to k= 5 or 6. At 40 dB SL the lower limit of the audibility region was approximately the same regardless of the order (i.e. independent of k). The width of the audibility region could not be explained completely by the amount of interaction of the mechanical stimulation patterns corresponding to f_1 and f_2. In addition, it is demonstrated that these combination tones may play a misleading part in masking experiments. The detection threshold of a simple tone masked by another simple tone may be determined by the detection of a combination tone of the type $f_1-k(f_2-f_1)$, even at a low level of the masker.
(This paper has been submitted for publication in The Journal of the Acoustical Society of America.)

COMBINATION TONES AND THEIR ORIGIN

Combination tones corresponding to $f_1-k(f_2-f_1)$, with k a small positive integer, are often audible during stimulation by the two frequency components f_1 and f_2 ($f_1<f_2$). In this paper the Cubic Difference Tone $2f_1-f_2$ (k=1) is studied mainly. The research reported is directed at the type of nonlinearity generating the CDT and at the site of CDT generation. The first experiment shows that the generation of the CDT is affected by a dip (a threshold elevation in a narrow frequency region). The CDT was perceived only when the level of f_2 exceeded the elevated threshold. In the second experiment the cancellation method is used. The results suggest that the high-frequency slope of the pattern of stimulation upon which the nonlinearity operates is comparable with the slope revealed in masking experiments. In the last experiments the cancellation method is reconsidered. Estimates of the CDT level found by various other measuring methods in which the probe tone was presented nonsimultaneously with the stimulus f_1,f_2 were all lower than the cancellation level. The results are discussed in terms of a nonlinearity intimately coupled with a frequency selectivity. A simple compressing type of nonlinearity is proposed which accounts rather well for the CDT data. The lower estimates of the CDT level found for nonsimultaneous probe tones are attributed to suppression effects introduced by the nonlinearity.
(*This is a preliminary version of a paper to be submitted for publication in The Journal of the Acoustical Society of America.*)

ACKNOWLEDGMENTS

I am much indebted to prof. dr. M. A. Bouman and to dr. R. Plomp who supervised this research. Further I thank ir. T. Houtgast for the stimulating discussions we had together. The experiments could be realized thanks to the cooperation of the subjects T. Houtgast, A. M. Mimpen, W. Klein, J. Verschuure, and F.L. Wightman. Drs. W. Larkin,

F. Ph. van Eyl, and F. L. Wightman were of invaluable help in improving the readability of the manuscripts.

This research was fully supported by the Netherlands Organization for the Advancement of Pure Research (ZWO).

SAMENVATTING

Een belangrijke eigenschap van het gehoororgaan is frekwentieanalyse. Bij aandachtig luisteren naar een toon voortgebracht door een muziekinstrument is het meestal mogelijk een aantal deeltonen waar te nemen. Deze deeltonen korresponderen met frekwentiekomponenten van de door het muziekinstrument voortgebrachte trilling. Soms blijken er echter deeltonen waarneembaar te zijn die geen overeenkomstige frekwentiekomponenten in de stimulus bezitten. Zo kan een stimulus bestaande uit de twee frekwentiekomponenten f_1 en f_2 ($f_1<f_2$), indien voldoende luid, een deeltoon voortbrengen die korrespondeert met de frekwentiekomponent f_2-f_1. Deze deeltoon wordt de verschiltoon genoemd. Het ontstaan van de verschiltoon wordt toegeschreven aan vervorming van het signaal door overbelasting van het gehoororgaan. Men neemt aan dat deze vervorming in het gehoororgaan voorafgaat aan de frekwentieanalyse zodat de door de vervorming geintroduceerde frekwentiekomponenten worden geanalyseerd en vervolgens hoorbaar worden als deeltonen. Deeltonen die slechts hoorbaar zijn bij een gelijktijdige aanbieding van twee of meer frekwentiekomponenten worden kombinatietonen genoemd.

In dit proefschrift worden kombinatietonen bestudeerd die korresponderen met de frekwentiekomponenten $f_1-k(f_2-f_1)$, waarbij k een klein positief geheel getal is. Deze kombinatietonen blijken reeds te worden waargenomen bij de zeer lage geluiddrukniveaus van 15 à 20 dB boven de drempel. Gezien deze lage niveaus kan in dit geval niet gesproken worden van vervorming als gevolg van overbelasting. Aangenomen wordt dan ook dat deze kombinatietonen de produkten zijn van een meer essentiële niet-lineariteit van het gehoororgaan. De experimenten hadden tot doel deze niet-li-

neariteit langs psychofysische weg te bestuderen. De resultaten worden gepresenteerd in de vorm van drie artikelen. In het eerste artikel wordt de toonhoogteperceptie bestudeerd van een signaal dat uit slechts twee opeenvolgende harmonischen bestaat. Voor hogere harmonischen blijkt de toonhoogteperceptie gebaseerd te zijn op de kombinatietonen van het type $f_1-k(f_2-f_1)$. Het tweede artikel laat zien dat deze kombinatietonen slechts voorkomen in een beperkt frekwentiegebied onder f_1. De niet-lineariteit blijkt gekoppeld te zijn aan de frekwentieanalyse. Het laatste artikel gaat dieper in op de plaats waar deze kombinatietonen worden gegenereerd en op de aard van de niet-lineariteit.

STELLINGEN

1 De bestudering van kombinatietonen is een effektieve methode om langs psychofysische weg de niet-lineaire signaalverwerking in het gehoororgaan te analyseren.

2 De in dit proefschrift gemaakte veronderstelling dat kombinatietonen het gevolg kunnen zijn van een komprimerende mechanische niet-lineariteit impliceert dat de amplitude van de trilling der cochleaire scheidingswand op drempelniveau aanmerkelijk groter kan zijn dan de huidige onbevredigende schattingen van ongeveer 10^{-12}m.

3 De Boer's fase regel is slechts van kracht indien de in het gehoororgaan opgewekte kombinatietonen voldoen aan de bij de fase regel behorende stimulus voorwaarde. Voor de kombinatietonen van het type $f_1-k(f_2-f_1)$ is dit het geval; voor de verschiltoon f_2-f_1 niet.

E. de Boer, "A note on Phase Distortion and Hearing", Acustica *11*, 182-184, (1961).

4 Niet-lineariteiten worden meestal mathematisch beschreven met Taylor reeksontwikkelingen. In geval van essentiële niet-lineariteiten biedt een beschrijving met Tsjebysjev polynomen vaak een gunstig alternatief.

5 Een specifikatie van de vervorming van een elektronische versterker voor slechts het grootste vermogen kan misleidend zijn.

6 Een korrelatie tussen het moment van optreden van aktiepotentialen in de gehoorzenuw en de fase van een van de stimulus afgeleide frekwentiekomponent toont niet noodzakelijk aan dat deze frekwentiekomponent ook in het gehoororgaan gerepresenteerd is. Intervalhistogrammen van de aktiepotentiaaltreinen bieden daaromtrent adekwatere informatie.

7 De omhullende van het tijdpatroon van een stimulus is geen eenduidige maat voor de waargenomen "ruwheid".
E. Terhardt, "Frequency Analysis and Periodicity Detection in the Sensations of Roughness and Periodicity Pitch", in *"Frequency Analysis and Periodicity Detection in Hearing"*, R. Plomp and G.F. Smoorenburg Eds., A.W. Sijthoff, Leiden (1970).

8 Met stimuli bestaande uit twee frekwentiekomponenten kan worden aangetoond dat de positieve sompotentialen (+SP) zoals gedefinieerd door Kupperman de aktiviteit van de basilaire membraan niet volkomen representeren.
R. Kupperman, "The SP in Connection with the Movements of the Basilar Membrane", in *"Frequency Analysis and Periodicity Detection in Hearing"*, R. Plomp and G.F. Smoorenburg Eds., A.W. Sijthoff, Leiden (1970).

9 Iedere onderzoeker doet er goed aan op de hoogte te zijn van de door Bacon beschreven 'idola' (vooroordelen) die het verwerven van zuivere kennis van de natuur in de weg kunnen staan.
Francis Bacon, *Novum Organum I* (1620).
see Benjamin Farrington, *Francis Bacon, Philosopher of Industrial Science*, Collier Books, New York (1949).

10 Het vak dat in dit proefschrift wordt beoefend zou men met "audionomie" kunnen betitelen.

CURRICULUM VITAE

De schrijver doorliep het Coornhert Lyceum te Haarlem en voltooide de middelbare schoolopleiding aan de Rijks hbs te Amersfoort. In 1961 werd de studie in de wiskunde en natuurwetenschappen begonnen aan de Rijks Universiteit te Utrecht. Het doktoraalexamen experimentele natuurkunde met groot bijvak toegepaste wiskunde werd in 1967 afgelegd. Daarna was de schrijver wetenschappelijk medewerker bij de afdeling medische en fysiologische fysica van de Rijks Universiteit te Utrecht op een aanstelling van de Stichting voor Zuiver Wetenschappelijk Onderzoek ZWO. Sinds 1 januari 1971 is de schrijver in dienst van het Instituut voor Zintuigfysiologie RVO-TNO te Soesterberg.

ASPECTS OF TWO-TONE PERCEPTION

PROEFSCHRIFT

ter verkrijging van de graad van doctor
in de wiskunde en natuurwetenschappen
aan de Rijksuniversiteit te Utrecht,
op gezag van de Rector Magnificus
prof. dr. Sj. Groenman,
volgens besluit van het college van dekanen
in het openbaar te verdedigen op
maandag 13 september 1971
des namiddags te 4.00 uur (precies)

door

GUIDO FRANCISCUS SMOORENBURG

geboren op 2 augustus 1943 te Haarlem

Promotor: prof. dr. M. A. Bouman

Natura non nisi parendo vincitur

De natuur wordt slechts overwonnen
door haar te gehoorzamen.

(Francis Bacon, 1620)

INTRODUCTION

The title of this dissertation 'aspects of two-tone perception' is ambiguous. For what is a tone? Generally, any auditory sensation having pitch is called a tone. But in this case the word 'tone' is used in the restricted sense of the sound produced by a sinusoidal vibration. Such a tone is an element of the auditory sensation. Its pitch is unequivocally related to the frequency of the vibration. A mixture of sinusoidal vibrations of various frequencies is perceived as a number of tones with pitches corresponding to the stimulus frequencies. It is a fundamental property of the ear that the frequency components of a stimulus are represented perceptually as tones. The ear acts as a spectral analyzer.

The close relation between a frequency component and a tone has become so self-evident that the word 'tone' may stand for both the frequency component of the stimulus and the percept. In the title of this dissertation 'tone' stands for frequency component; the perception of stimuli with two frequency components was studied. The distinction between the two uses of the word 'tone' is unimportant in so far the *frequency component - tone* relation holds strictly. However, the subject of this dissertation is that this is not generally true. The two-frequency stimulus f_1,f_2 may evoke more than two tones in the ear. For example, a difference tone corresponding to the frequency component f_2-f_1 may be perceived. Such a tone is conceivably the product of a nonlinearity of the ear. Frequency analysis by the ear succeeding the nonlinearity would reveal the distortion in the perception of additional tones. These additional tones are called combination tones.

The difference tone (f_2-f_1) is audible only at high stimulus levels. It can be described as the product of an overloading type of nonlinearity. The combination tones which are studied in this dissertation correspond to frequency components $f_1-k(f_2-f_1)$, where k is a small integer. These combination tones are audible even at very low stimulus levels which suggests that they are the products of a more essential nonlinearity. The reason for the study of these combination tones was the working hypothesis that such a study might help to understand the nonlinear detection of the ear. Two-frequency stimuli were applied in this research since they have the smallest number of variables.

This dissertation consists of three papers. The abstracts of these papers are given below.

ABSTRACTS

PITCH PERCEPTION OF TWO-FREQUENCY STIMULI

Experiments on the pitch of complex tones produced by two frequency components are reported. An exploratory experiment revealed that subjects perceive the pitches of individual part-tones or the stimulus as a whole with a pitch corresponding to about the fundamental frequency. The latter pitch was investigated more thoroughly. In case of adjacent harmonics, the pitch corresponded to the (absent) fundamental frequency. A shift of the frequencies away from such a harmonic situation while maintaining a constant frequency difference resulted in a pitch shift. For higher harmonic numbers, the pitch shift was larger than could be met by current theories. The large pitch shift was explained by taking into account an auditory nonlinearity which generates combination tones of the type $f_1-k(f_2-f_1)$. Sound-pressure level dependence of the pitch shift could be explained in the same manner. When the combination tones were masked, the large pitch shift diminished. With regard to the pitch mechanism, the results suggested that the detection of the low pitch of the complex tone requires spectral information.

(*This paper has been published in The Journal of the Acoustical Society of America, Vol. 48, 924-942, October 1970.*)

AUDIBILITY REGION OF COMBINATION TONES

This paper describes research on combination tones of the type $f_1-k(f_2-f_1)$. Such tones are audible during stimulation by the two frequency components f_1 and f_2, $f_1<f_2$. The results of the experiments suggest that these combination tones are audible only in a restricted frequency region below f_1. The lower limit of this 'audibility region', for the combination tone $2f_1-f_2$ (k=1), was determined primarily by the level of the lower stimulus component L_1 and additionally by L_2. The width of the audibility region appeared to be highly subject dependent. The lowest stimulus level for which $2f_1-f_2$ was audible was 15-20 dB SL if both stimulus com-

ponents had equal levels. If the lower component was set at 40 dB, then the higher one could be reduced to 4 dB before $2f_1-f_2$ reached threshold. Higher order combination tones could be perceived up to k= 5 or 6. At 40 dB SL the lower limit of the audibility region was approximately the same regardless of the order (i.e. independent of k). The width of the audibility region could not be explained completely by the amount of interaction of the mechanical stimulation patterns corresponding to f_1 and f_2. In addition, it is demonstrated that these combination tones may play a misleading part in masking experiments. The detection threshold of a simple tone masked by another simple tone may be determined by the detection of a combination tone of the type $f_1-k(f_2-f_1)$, even at a low level of the masker. (*This paper has been submitted for publication in The Journal of the Acoustical Society of America.*)

COMBINATION TONES AND THEIR ORIGIN

Combination tones corresponding to $f_1-k(f_2-f_1)$, with k a small positive integer, are often audible during stimulation by the two frequency components f_1 and f_2 ($f_1<f_2$). In this paper the Cubic Difference Tone $2f_1-f_2$ (k=1) is studied mainly. The research reported is directed at the type of nonlinearity generating the CDT and at the site of CDT generation. The first experiment shows that the generation of the CDT is affected by a dip (a threshold elevation in a narrow frequency region). The CDT was perceived only when the level of f_2 exceeded the elevated threshold. In the second experiment the cancellation method is used. The results suggest that the high-frequency slope of the pattern of stimulation upon which the nonlinearity operates is comparable with the slope revealed in masking experiments. In the last experiments the cancellation method is reconsidered. Estimates of the CDT level found by various other measuring methods in which the probe tone was presented nonsimultaneously with the stimulus f_1,f_2 were all lower than the cancellation level. The results are discussed in terms of a nonlinearity intimately coupled with a frequency selectivity. A simple compressing type of nonlinearity is proposed which accounts rather well for the CDT data. The lower estimates of the CDT level found for nonsimultaneous probe tones are attributed to suppression effects introduced by the nonlinearity.
(*This is a preliminary version of a paper to be submitted for publication in The Journal of the Acoustical Society of America.*)

ACKNOWLEDGMENTS

I am much indebted to prof. dr. M. A. Bouman and to dr. R. Plomp who supervised this research. Further I thank ir. T. Houtgast for the stimulating discussions we had together. The experiments could be realized thanks to the cooperation of the subjects T. Houtgast, A. M. Mimpen, W. Klein, J. Verschuure, and F. L. Wightman. Drs. W. Larkin, F. Ph. van Eyl, and F. L. Wightman were of invaluable help in improving the readability of the manuscripts.

This research was fully supported by the Netherlands Organization for the Advancement of Pure Research (ZWO).

CONTENTS

PITCH PERCEPTION OF TWO - FREQUENCY STIMULI

INTRODUCTION

The human ear is able to analyze a stimulus consisting of two frequency components. If the frequency difference is not too small, a person can hear two part-tones with pitches corresponding to the individual stimulus frequencies. On the other hand, a set of harmonics gives rise to a single-pitch sensation, notwithstanding the analyzing faculty of the human ear. This single or over-all pitch, which corresponds closely to the fundamental frequency, will be designated as "the pitch of the complex tone". The present experiments were an attempt to study the properties of this pitch produced by a stimulus consisting of two frequency components.

Although data on this question are few, the experiments were thought to be fruitful for three reasons.

(1) Results of experiments on the spectral region dominant in the perception of pitch of complex tones show that the most effective spectral region covers only two low harmonics (Ritsma, 1967a).

(2) The time pattern of two summated frequency components is rather similar to the time pattern of a sinusoidally amplitude-modulated (SAM) signal, consisting of three frequency components, which gives rise to a clear pitch of the complex tone.

(3) The number of frequency components itself does not seem important. The marked difference for the signals in (2) —two versus three frequency components— is not recovered perceptually because of the introduction of combination tones such as $2f_1-f_2$.

This paper has been published by The Journal of the Acoustical Society of America, Vol. 48, 924-942, October 1970.

1. TERMINOLOGY

The applied terminology is, as much as possible, in line with the American Standard Acoustical Terminology (American Standard Association, 1960). According to this standard, however, some terms have both a physical and a perceptual meaning. An attempt to a clearer distinction will be made here:

tone	an auditory sensation having pitch,
simple tone	a tone generated by a sinusoidal vibration and characterized by its singleness of pitch,
part-tone	a tone similar to a simple tone,
complex tone	a tone that is not a part-tone,
fundamental tone	a part-tone, where the corresponding simple tone is generated by the fundamental frequency component,
combination tone	a part-tone not present in the sensations produced by single constituent frequency components of a stimulus,
partial	a frequency component,
harmonic	a partial whose frequency is an integral multiple of the fundamental frequency.

The word tone is used only in a perceptual sense because its origin is such and because a more specific word for the perceptual aspect is lacking. As for the physical aspect, one can dispense with the word "tone" because purely physical terms suffice in describing a stimulus. The concept "partial" is also dualistic. Hence, distinction is made between partial and part-tone for the physical and perceptual aspects, respectively. A part-tone is similar to a simple tone not only in regard to pitch but also, for instance, with respect to timbre. Thus, a set of high harmonics which is not analyzed into part-tones gives rise to a sensation that is not called a part-tone itself, although its pitch may match the pitch of a simple tone. The sensation of such a set of harmonics is an example of a complex tone. A tone is called complex if it is not similar to a simple tone. This perceptual description is desirable —although "complex" may originally refer to the physical properties— in a framework where "tone" is used in a perceptual sense only. A certain stimulus consisting of several harmonics may be perceived in different ways; a part-tone may be noticed if the attention is drawn to it but it may vanish in a complex tone as well. These strict definitions are adopted in this paper.

2. PREVIOUS RESEARCH

The pitch of a complex tone produced by a complete set of harmonics, up to the 10th or 20th, generally has been considered to be the pitch of the fundamental tone. This idea stems already from Ohm (1843, 1844) and from von Helmholtz (1863). Even then Seebeck (1841, 1843, 1844a, 1844b) objected that a periodic sound wave with a very faint fundamental partial has a pronounced low pitch. The important role ascribed to the fundamental tone becomes even more doubtful when one recalls that the human ear is relatively insensitive to the low frequency of a fundamental partial. Some investigators (e.g., Schaefer and Abraham, 1904; Fletcher, 1924) tried to overcome the problem of "the case of the missing fundamental" by assuming that the fundamental tone in case of an absent fundamental partial is reintroduced in the ear by a nonlinear distortion. Schouten (1938), however, disproved this assumption. He used a special complex stimulus which did not introduce a fundamental tone as was demonstrated by absence of beats when a simple tone, with a frequency slightly different from the fundamental frequency, was added. It appeared that the pitch of the complex tone was not affected by the absence of the fundamental tone. This leads to the conclusion that pitch of a complex tone is not necessarily determined only by the pitch of the fundamental tone. A detailed historical review was given by Plomp (1967).

As a consequence, the question of the relative contributions of the partials, the fundamental one included, to the pitch of an arbitrary complex tone arises. These contributions may, of course, depend on the spectral envelope. The question was investigated by Plomp (1967) with a method based upon conflicting pitch information. The harmonics of a stimulus were separated into two groups, after which the fundamental frequencies of each group were drawn apart to some extent. The frequencies of the partials of both groups remained the same integral multiples of the group's fundamental frequency. Such a signal offers the opportunity to conclude from the pitch sensation of the total stimulus which group determines the pitch. An insight into the relative contributions of the partials was achieved by separating the set of harmonics at different frequencies. Plomp used a flat spectrum and a spectrum with a -6 dB/oct slope which more closely resembles the slopes found for speech vowels and for sounds produced by musical instruments. The pitch of the signals with both slopes appeared to be determined by "the third and higher harmonics for fundamental frequencies up to about 700 Hz and by the fourth and higher harmonics for fundamental frequencies up to about 350 Hz." Ritsma (1967a) used this method of conflicting pitch information in experiments that led to the concept of dominant frequency regions in pitch perception of complex tones. He found that "for fundamental frequencies in the range of 100 to 400 Hz, and for sensation levels up to at least 50 dB above threshold of the entire signal, the frequency band consisting of the third, fourth, and fifth harmonics tends to dominate the pitch sensation as long as its amplitude exceeds a minimum absolute level of about 10 dB above threshold." Moreover, he found that the dominant frequency region most effective in determining the pitch covers only two partials. An earlier finding by Schouten and 't Hart (1965) agrees with these results. They reported that the primary pitch of the strike note of bells is determined by the second and third partial. A secondary pitch, which sometimes occurs, is determined by the fourth, fifth, and sixth partial. This conclusion could be drawn because partials of bells are not harmonically related, which gives a natural situation of conflicting pitch information.

However, the complex stimuli can be descibed by the frequency spectrum as well as by the time pattern, and there is evidence that the time structure is involved in pitch detection. Schouten (1940a, 1940b) reconciled the limited frequency-analyzing power of the human ear with pitch perception based upon periodicity detection. He stated:"The lower harmonics can be perceived individually and have almost the same pitch as when sounded separately. The higher harmonics, however, cannot be perceived separately, but are perceived collectively as one component the residue with a pitch determined by the periodicity of the collective waveform, which is equal to that of the fundamental tone".

The periodicity concept seems to be supported by a distinct class of experiments in which sets of inharmonic partials are used. Such experiments were initiated by Schouten (1940c) and continued by de Boer (1956), who applied amplitude-modulated stimuli with seven and with five frequency components. De Boer got the impression that five is the lowest number of frequency components from which a stable and distinct residue could be obtained. The inharmonic signals were produced by shifting the carrier frequency away from a harmonic situation in which the carrier frequency is an integral (of four or more) multiple of the modulating frequency. Then, a clear pitch shift was perceived while the modulating frequency (or the envelope frequency) remained the same. On the basis of time intervals between pronounced deflections of the time pattern, de Boer was able to predict this pitch shift to a first approximation. Yet, for spectra of widely separated components he also thought of a complementary mechanism like "the evaluation of an appropriate common divisor to the presented frequencies". The small experimental deviations from the first approximation were explained by assuming a nonsymmetrical weighting of the frequency components involved: the lower frequency components should play a more important part. This tendency is understandable by now in terms of the dominant frequency region in pitch perception of complex tones. Later experiments with SAM signals showed that three frequency components are already sufficient to achieve a tonal residue (Ritsma, 1962). In this case, the effect of pitch shift for inharmonic signals was established by Schouten, Ritsma, and Cardozo (1962). They described their results with a model operating in the time domain.

An influence of phase upon pitch was demonstrated by Ritsma and Engel (1964). Without affecting the power spectrum, they changed the time structure from a SAM signal into a so-called quasi-frequency-modulated (QFM) signal by shifting the phase of the carrier frequency component over 90°. A QFM signal appeared to give rise to several pitches, all corresponding to time intervals between pronounced deflections that can be distinguished in the particular time structure.

The perception of pitch of complex tones has been discussed in terms of both frequency spectrum and time structure. Their relative contributions to pitch detection is still an open question. The problem was investigated by Flanagan and Guttman (1960a, 1960b), Guttman and Flanagan (1964), and Rosenberg (1965). All applied pulse trains of several polarity patterns, which in some experiments were filtered or partly masked by noise. Different polarity patterns can give different frequency spectra for the same pulse repetition frequency. For these signals, three modes of pitch perception could generally be distinguished. With increasing frequency of pulse repetition, the pitch appeared to be related, successively, to the pulse repetition frequency itself for pulse rates up to 100 Hz, to the fundamental frequency for pulse rates from 200 to 500 Hz, and to the lowest partial for pulse rates above 1000 Hz. The first transition region was found to be particularly interesting. Consistent pitch matchings made in that region did not correspond to either stimulus parameter of the three pitch modes. Starting with the assumption that pitch is related to time intervals between unilateral deflections of the basilar membrane, an explanation was given by taking into account the resonance properties of this membrane as described by Flanagan (1962). Responses of the membrane to pulses of opposite polarities are interspaced. Deflections of both responses in one direction differ in time by half a period of the resonance frequency. Thus, in case of bipolar pulse trains, the time intervals between unilateral deflections at successive points along the basilar membrane with their respective resonance frequencies will supply continuously varying pitch information. This explanation implies that a situation of conflicting pitch information arises if a wide-band stimulus is applied. The same reasoning was used by Ritsma (1967b) to check the concept of dominant frequency regions in the perception of pitch of complex tones. Pulse trains of alternating polarity and with pulse repetition frequencies from 100 to 250 Hz appeared to give rise to pitch sensations corresponding to the time intervals between the pulses, plus or minus half an oscillation of a frequency, which was accepted as the central frequency of the dominant region. According to the findings of this experiment, the dominant frequency region appeared to be centered around the sixth to fifth harmonic for pitches corresponding to 100-250 Hz, respectively.

Ritsma discussed the results of his experiment in terms of time structure detection. Yet, Plomp (1964, 1968) has shown that the human ear easily analyzes these dominant partials. If pitch is determined by the time structure, the problem arises of why harmonics, easily separable by the ear, dominate in pitch perception. A time-structure detection requires the cooperation (e.g., addition) of at least two frequency components in order to furnish the pitch information. One may defend the opposite statement that this very region is the dominant one because pitch is determined by frequency-pattern recognition requiring a good frequency resolution. The results of many experiments can be explained in terms of both ideas. Walliser (1968, 1969c) proposed a combined functional model in which the pitch of a set of high harmonics is determined roughly by the frequency of the envelope of the time structure and then determined precisely by that subharmonic of the pitch corresponding to the lowest partial in the stimulus, which has a pitch corresponding as closely as possible to the envelope frequency. As more critical experiments on the relative contributions of frequency-pattern recognition and periodicity detection to the pitch of complex tones are lacking, the results of the investigations reported below will be discussed in the light of both concepts.

3. DISPARITY AMONG SUBJECTS

The time-consuming pitch matchings reported in Secs. 4-7 were carried out by two observers, GS and TH. Their results showed a great deal of similarity. We noticed, however, that a two-frequency stimulus consisting of, for example, the eighth and ninth harmonic was not perceived similarly by every listener. Whereas some observers, GS and TH included, perceived the stimulus as a whole with a pitch corresponding to about the fundamental frequency, other listeners were able to hear only the pitches of individual part-tones. A closer look at this phenomenon of interpersonal differences was thought to be worthwhile, the more so as differences of opinion about the pitch produced by signals consisting of a few harmonics may be due to such a disparity in pitch perception among subjects.

3.1. Experiment

The way in which two-frequency stimuli are perceived with respect to pitch was studied with 42 subjects and a well-suited pair of signals. Each signal was produced by two sine-wave generators without mutual synchronization. One signal was composed of the frequency components f_1=1800 Hz and f_2=2000 Hz, the other signal of the components f_1=1750 Hz and f_2=2000 Hz. If the two signals are presented successively, one might expect that a pitch drop corresponding to the frequency drop of the lower frequency component would be heard. On the other hand, the pitch may rise because the first signal consists of harmonics with a fundamental frequency of 200 Hz and the second signal of harmonics with a fundamental frequency of 250 Hz.

Combination tones may play a part in these pitch judgments. According to Plomp (1965), a difference tone f_2-f_1 is not likely to be detected by any observer at sensation levels of the stimulus components up to 30 dB. Because a higher sensation level (SL) of 40 dB was used in the present experiment, an octave band of noise with a center frequency of 250 Hz was added to the stimuli. The noise band masked simple tones of 200 Hz and of 250 Hz at 30 dB (SL), which assured that no subject would hear a difference tone. All other distortion components that might be introduced by the ear at 40 dB SL are the combination tones of the type f_1-$k(f_2-f_1)$, $f_1<f_2$ (Plomp, 1965; Goldstein, 1967). This means that the pitches of all audible combination tones (the difference tone being masked sufficiently) will change in the same direction as the pitch of the part-tone corresponding to the lower frequency component. Taking into account the presence of combination tones, there is still the following unequivocal relation: if the pitch of the signal f_1,f_2=1750,2000 Hz is judged to be higher than the pitch of the signal 1800,2000 Hz, then the judgment must have been based upon the complex tone as a whole; if the pitch is judged to be lower, then the judgment must have been based upon one or more pitches of individual part-tones. However, the statement may be wrong if a person judging on the basis of individual part-tones does not follow the pitch change of a part-tone with a certain k (k=0 included), but rather compares pitches of part-tones with different k. For example, the pitch interval of a major third corresponding to the fundamental frequencies 250 and 200 Hz may also be heard if the pitches of the part-tone corresponding to 1750 Hz (k=0) and of the combination tone corresponding to 1400 Hz (k=2) are compared. This unlikely possibility was eliminated by interviewing the subjects.

Twenty-five times in one run the two signals were presented successively in a random order. The presentation times were 160 msec each with a pause of 160 msec. This situation resembles somewhat the dynamics of music. To avoid clicks, transient times of 10 msec were applied. A new pair of signals was presented every 4 sec. The octave band of noise was presented continuously. Distortion products of the apparatus itself were all well below the threshold of hearing. The experiment was completely automized. The stimuli were presented binaurally through headphones in a sound-insulated room.

The subject was told that he would hear two tone bursts with different pitch. He had to make a forced-choice judgment of whether the pitch of the second burst was lower or higher than the pitch of the first. No information feedback was given. The 42 subjects were all coworkers of the Institute. Only a few listeners were acquainted with the background of the experiment; only seven were working in psychoacoustics. The experiment was repeated one month later.

3.2. Results and Discussion

Nearly all pitch judgments in any one run appear to be based upon either the part-tones or the complex tone perceived as a whole. Moreover, not a single subject who had a stable criterion in the first session reversed this criterion at the second session. This result is brought out in more relief by mentioning that between the two sessions a different experiment with similar stimuli was performed in which all subjects participated. The added data from both sessions are given in Table I.

No one reported an ambiguous pitch. Even the scores N=5-9, 10-15, and 16-20 do not seem to be caused by criterion reversals during one run. Half of the subjects (four) with these scores were not at all able to hear the direction of the pitch jump of simple tones of 1750 and 1800 Hz solely. The other subjects making these scores merely had difficulties in judging a pitch jump of a part-tone in the presence of other part-tones.

Thus, it appears that one of the two criteria dominates for each subject; there is not just rivalry between the two possible ways of judging the pitch. The ratio of subjects judging on the basis of one or the other criterion is the same within the subset of subjects with musical interest. There also appears to be no correlation between either criterion and the group of psychoacousticians; out of a total of seven subjects, four were judging on the basis of individual part-tones and there were three who heard the pitch jump of the complex tone.

In order to be sure that the unequivocal relation between direction of perceived pitch jump and criterion holds generally, it still should be verified that there were no comparisons of pitches of part-tones with different k. Every subject with scores N=21-25 reported an interval of a major third. Some subjects were able to name the interval; others were asked to reproduce the perceived pitch jump by singing (cuckoo). Generally, the subjects making these scores judged the task to be easy. The pitch jump was their immediate impression: they generally did not report spontaneously that they perceived a compound stimulus. The subjects who scored N=0-4 reported a pitch interval that was usually estimated to be a minor second, or they reproduced vocally a slight pitch jump. The task was judged differently from being easy to being difficult. Most subjects reported that they could hear more than one tone; they never reported that they could also hear a major-third relation between pitches of some part-tones, not even upon request. Thus, subjects with good analyzing capabilities did not compare pitches of individual part-tones corresponding to frequencies equal to the same multiple of the two fundamental frequencies. In view of this, it is improbable that subjects with scores N=21-25 and who never reported the perception of part-tones did compare the pitches of these particular part-tones. We may infer that the perception of a pitch jump corresponding to the fundamental frequencies was based upon pitches of the complex tones perceived as a whole and that the judgments in opposite direction were based upon the pitches of individual part-tones or perhaps just upon timbre.

Table I. Disparity among 42 subjects for pitch perception of two-frequency stimuli, added over two sessions. The subjects were presented with two successive stimuli in a random order. One stimulus consisted of the frequencies 1800,2000 Hz; the other of the frequencies 1750,2000 Hz. The table gives the number of subjects who perceived out of a total of 25 stimulus pairs a specified number of times N a pitch corresponding to the fundamental frequency, which means that N times "the pitch of the complex tone" was perceived. In the other cases, pitch jumps according to individual part-tones were perceived.

Score of "pitch of the complex tone" judgments N out of 25					
N	0-4	5-9	10-15	16-20	21-25
Number of subjects	35	7	8	2	32

This experiment was not intended to obtain quantitative data on disparity among subjects in pitch perception. It merely should demonstrate that there are important individual differences. The choice of parameters may influence the results; perhaps both in the sense of a parameter-dependent criterion and in the sense of having or having not simultaneously two criteria at one's disposal. For example, presentation time, sound-pressure level (SPL) of each partial, and frequency difference may play a part.

These findings show some resemblance to the results obtained by Schodder and David (1960). They investigated the influence of a fixed-frequency component on the frequency discriminability of another component as a function of frequency difference. As we have done in this section, they chose the higher frequency to be fixed. Owing to this choice, they could find a similar disparity among subjects for frequency differences within the critical bandwidth. The just-noticeable frequency difference of the lower component appeared to be related to individual frequency components in some cases and to the "envelope frequency" in others. There was no effect for frequency differences that are large with respect to the critical bandwidth.

4. PITCH PERCEPTION OF INHARMONIC STIMULI

It is apparent that for a two-frequency stimulus the primary pitch impression of about half the subjects was based upon the complex tone perceived as a whole and not on some individual part-tones. Inharmonic signals, as mentioned in Sec. 2, are well suited for further investigations into the pitch of these complex tones perceived as a whole. In the present experiment, the two frequencies are varied while the frequency difference is kept constant. All situations in which the frequencies of the two components are no integral multiples of the frequency difference are usually designated as inharmonic although such an inharmonic signal may be composed, for instance, of the seventh and ninth harmonic. This designation is due to the properties of pitch of complex tones which will be established for the two-frequency stimuli.

4.1. Method

The pitch of a complex tone should be measured primarily by matching with a simple tone of a certain SPL. However, this appears to be a very difficult task for the kind of stimuli used and it requires much training. To avoid the difficult interpretation of learning effects in such training —such as learning to recognize pitch intervals between part-tones of the test signal and the simple tone— the principle of matching with a simple tone was abandoned except for some reference measurements. A similarity of pitch is easier to recognize if the timbres of the two complex tones do not differ too much. For this reason, it was decided to match with a harmonic signal in the frequency region of the test signal.

Results of preliminary pitch matchings with harmonic 90%-SAM signals and with harmonic two-frequency stimuli were similar. The harmonic two-frequency signals, which produce a less pronounced but still distinct pitch for both observers, were adopted as a matching signal. But, if we accept such a harmonic signal as a pitch reference, then we have to know the pitch of these signals in relation to the pitch of simple tones. One can take the view that a harmonic stimulus and a simple sinusoidal stimulus give rise to the same pitch if the frequency of the sinusoid is equal to the fundamental frequency of the harmonic signal. But as long as this idea is not proved for all situations and, moreover, as long as discrepancies are reported (Walliser, 1968, 1969a,1969c), it is not justifiable simply to start with this idea. Therefore, the actual matching signals are given and it has been verified whether the different harmonic stimuli with equal fundamental frequencies did produce the same pitch.

The comparison of pitches of complex tones with about the same timbre has the disadvantage that one cannot exclude the possibility that pitches of individual part-tones are matched. Of course, this may be done inadvertently. Sometimes, after the pitches of the complex tones had been matched, an observer still noticed a pitch jump of a part-tone. Pitches of part-tones can be the changing element that remains and, consequently, attracts the attention. We are confronted here with an essential difficulty in the measuring procedure; a finished matching is not always totally satisfying, because the pitches of individual part-tones usually do not coincide. Nevertheless, if one changes the fundamental frequency of the matching signal, the pitch difference of the complex tones is heard at once. One might try to eliminate the problem of interfering pitches of part-tones by using a matching signal in a frequency region adjacent to the test signal in order to avoid coinciding frequency components. But then one also has to think of the presence of combination tones. Exclusion of any two coinciding part-tones necessitates a large timbre difference between matching and test signal which makes the pitch matchings more difficult. In view of these problems, it was finally decided to use matching signals with different harmonic numbers. The numbers varied from those of the test signal to two or three higher or lower than the test signal. The use of different matching signals offered the opportunity to check whether perhaps the pitches of individual part-tones were matched. If they were matched, the results would depend on the constituent harmonics of the matching signal.

The constant frequency difference used was 200 Hz, a value giving a distinct pitch over a wide range of stimulus frequencies. Moreover, this value was chosen to facilitate a comparison of the results with those of de Boer (1956) and Schouten et al. (1962). The SL of the individual components was 40 dB, a comfortable level and still low enough to avoid a significant influence of the difference tone (with a constant pitch).

4.2. Apparatus and Procedure

The test signal was generated by two free-running sine-wave generators (Hewlett-Packard 200 CD). By means of a special trimming procedure, the harmonic distortion was reduced to about -60 dB. After independent attenuation sufficient to avoid mutual synchronization, the two sinusoids were added with a passive impedance-matched mixer (Fig. 1). Next, the signal was led through the switch S into the headphone (Beyer DT 48 S) with a series resistance to match the low impedance

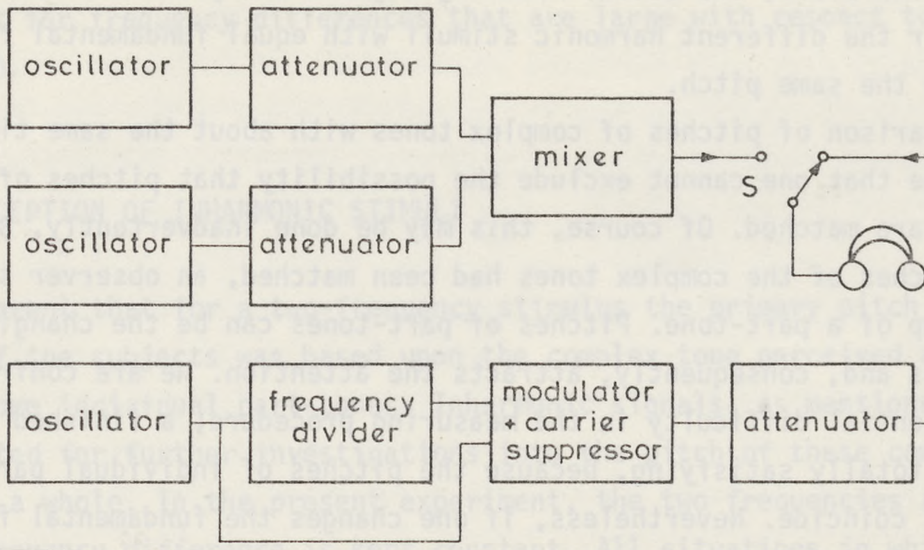


Fig. 1. Block diagram of the apparatus.

of the headphones, 5Ω.

The harmonic matching signal was generated with a modulating procedure. The modulating frequency was derived from the carrier frequency by means of a specially designed frequency divider (Fig. 2) producing a sinusoid. The carrier and the modulating sinusoid were fed into an amplitude modulator producing an SAM signal which has three frequency components. The two-frequency stimulus was then obtained by cancelling the carrier-frequency component. This was realized by adding to the SAM signal the carrier-frequency component in opposite phase. The frequencies of the two remaining components are given by the carrier frequency plus and minus the modulating frequency. This means that the divisor had to be an odd integer in order to obtain a matching signal harmonic in the sense of being composed of two frequencies that are integral multiples of their frequency difference. For example, if the divisor equals 2n+1 and if the carrier frequency is $(n+\frac{1}{2}).g$, then the modulating frequency is given by $\frac{1}{2}.g$ and the two frequency components are $n.g$ and $(n+1).g$, g being the fundamental frequency. Owing to this divider, the frequencies of the matching signal could be varied over a large range (one to two decades) without requiring adjustments. The distortion components of the final signal were below -36 dB (worst case) with respect to the two remaining frequency components. The phase relation between these components was chosen in such a way that the top of an oscillation of the fine structure did coincide with the maximum of the envelope (cosine addition). This was an arbitrary choice; the pitch produced by a two-frequency stimulus is independent of the phase relation.

The observer was seated in a sound-insulated room. The signals were presented monaurally to avoid problems such as diplacusis. The test signal was compared with the matching signal by means of a three-position switch whose middle position was silent. Each series of pitch matchings was started with a harmonic situation in which the frequencies of the test signal are successive integral multiples n and n+1 of the difference frequency, 200 Hz. This was called a central harmonic situation (n,n+1). The subject had to adjust the carrier frequency of the matching signal for equal pitches. Then the frequencies of the test signal were shifted upwards by 50 Hz and again an adjustment had to be made. This was repeated up to the next harmonic situation (n+1,n+2). Finally, a series was completed by lowering the frequencies in steps of 50 Hz from the initial central harmonic situation down to situation (n-1,n). Thus, a series consisted of nine pitch matchings. Each session with three series lasted about an hour. Two subjects participated.

4.3. Results

Both subjects had the impression that the way in which the pitch was perceived

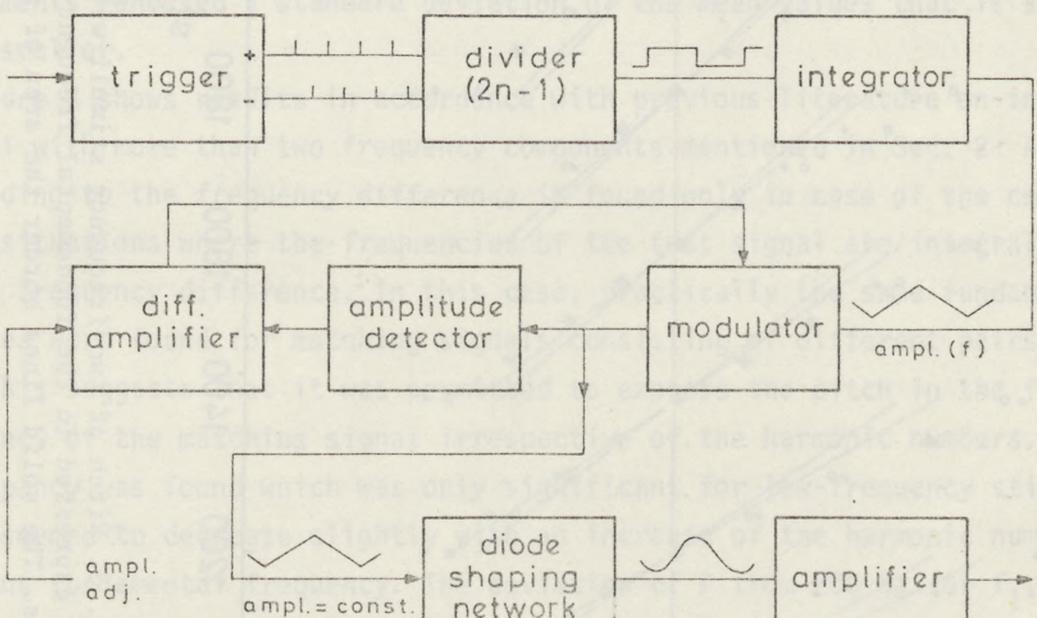


Fig. 2. Block diagram of the frequency divider. A division by 2n-1 is obtained directly with a scaler of n by counting n pulses, triggering a flip-flop, and using the state of this flip-flop to switch between pulse trains triggered by positive and by negative zero crossings of the input sinusoid. Next a triangular wave with a frequency-dependent amplitude is obtained by integration. A combination of a modulator, a special amplitude detector with minimal ripple at a reasonable integration time, and a differential amplifier functions as an amplitude equalizer. Then the triangle is shaped into a sinusoid by means of a diode shaping network (function generator). In this way, filters are avoided so that a sinusoid with a constant amplitude and a fixed phase with respect to the input is produced over a wide frequency range without requiring adjustments. Distortion components are below -45 dB.

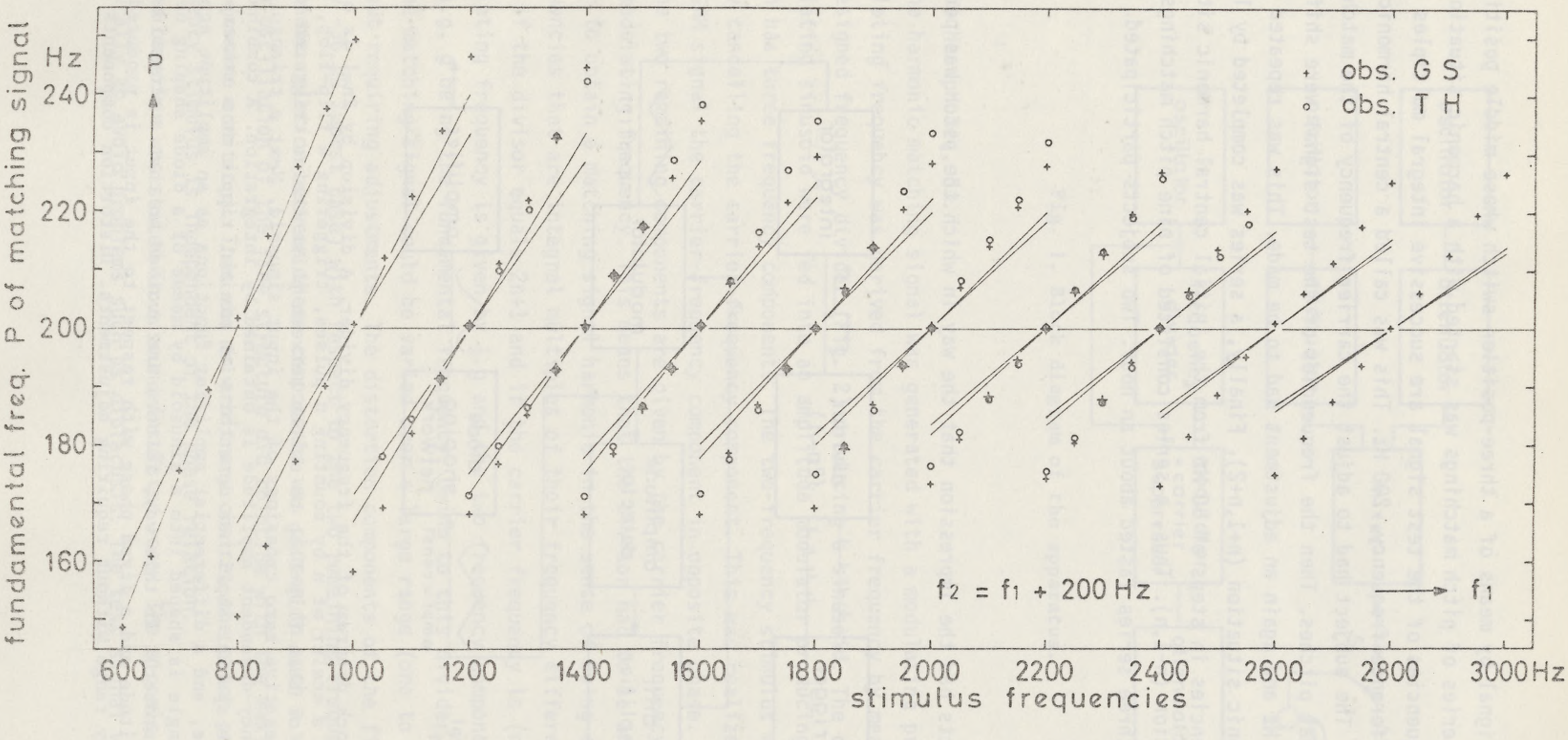


Fig. 3. Pitch of two-frequency stimuli with a constant frequency difference of 200 Hz expressed in the fundamental frequency P of two-frequency harmonic matching signals. The solid lines border the area of predicted pitch.

depended on the harmonic numbers of the stimulus. With higher harmonic numbers, a rattlelike sensation gradually became a perceptually distinct element. The pitch of the complex tone seemed to be associated with this rattle. In addition to pitch it was also possible to attribute to this rattle a timbre which was lower than would be expected on the basis of the stimulus frequencies.

The results of the pitch measurements of the two observers are shown in Fig. 3. The pitch is expressed in the (absent) fundamental frequency P of the various harmonic matching signals. Each data point is the average of nine adjustments performed on different days. Three matching signals with different pairs of successive harmonics were used, each three times. For test signals with central harmonic situations consisting of the frequencies up to the pair f_1,f_2=1600,1800 Hz, the harmonic numbers of the matching signal were the same or one or two higher than the harmonic numbers of the test signal. For test signals of higher frequencies, the harmonic numbers were the same or one or two lower. The standard deviation of each average pitch value was too small to plot in the graph; the standard deviation of each adjustment amounted from typically 0.75 Hz in a central harmonic situation to 2.5 Hz in situations with maximal pitch shift. Nine independent adjustments rendered a standard deviation of the mean values that is a factor three smaller.

Figure 3 shows results in accordance with previous literature on inharmonic stimuli with more than two frequency components mentioned in Sec. 2. A pitch corresponding to the frequency difference is found only in case of the central harmonic situations where the frequencies of the test signal are integral multiples of the frequency difference. In this case, practically the same fundamental frequencies were found for matching signals consisting of different pairs of harmonics. This suggests that it was permitted to express the pitch in the fundamental frequency of the matching signal irrespective of the harmonic numbers. A slight discrepancy was found which was only significant for low-frequency stimuli. The pitch seemed to decrease slightly with an increase of the harmonic numbers and a constant fundamental frequency. The deviation of P from 200 Hz for f_1,f_2=800,1000 Hz (Fig. 3) illustrates this effect. Larger effects might have been found if a larger difference between the harmonic numbers of the matching and the test signal had been used.

A frequency shift Δf away from a central harmonic situation is followed by a pitch shift Δp. The magnitude of the pitch shift depends on the frequencies of the test signal and will be compared with the solid lines in Sec. 4.4. The pitch shift follows the frequency shift over quite a range resulting in an ambiguity of pitch: a test signal with certain frequencies has more than one pitch. Notice that the pitch in, for example, the situation $f_1=(n-\frac{1}{2}).200$ Hz and $f_2=(n+\frac{1}{2}).200$ Hz does not correspond to the periodicity 100 Hz but to some values around 200 Hz. This illustrates why a stimulus whose frequencies are harmonically related but not integral multiples of the frequency difference is designated as inharmonic.

Figure 3 shows also that the frequency range over which adjustments were made is smaller for observer TH than for observer GS. The low-frequency side is limited by the prominence of the part-tones. Observer TH had more difficulties in matching the pitches without being diverted by the individual part-tones than observer GS. The high-frequency side is not limited by problems arising from the measuring technique but by the tonality of the stimulus. Above a certain frequency, the stimulus loses its tonal character abruptly and the standard deviation of the adjustments increases progressively.

To complete the measurements, each observer also made adjustments with a simple tone as a matching signal. During this measurement, the frequencies were again increased in steps of 50 Hz, starting from a central harmonic situation up to the next harmonic situation, and then the reverse was done for the low-frequency side of the central harmonic situation. Some series of observer GS were replaced by a random sequence of frequency pairs. The results are given in Fig. 4. To enable an easy comparison, a solid line is drawn that represents the best linear fit to the matchings with two-frequency harmonic signals as given in Fig. 3. Accidental adjustments belonging to other central harmonic situations are also given. The adjustments are rather difficult, as mentioned before. Observer TH found the inharmonic situations less clear, especially at the low-frequency side. At that side, his adjustments differ significantly from the results of Fig. 3. Perhaps this is linked up with a less-pronounced lower pitch in ambiguous pitch situations. The data represent first results. There was no prior training in order to avoid learning effects.

The pitch in case of the central harmonic situation in Fig. 4 corresponds to frequencies lower than 200 Hz. Thus, the pitch of the harmonic signals does not only seem to decrease slightly with increasing harmonic numbers as mentioned before, but the pitch of these harmonic signals appears also to be lower than the pitch of a simple tone of the fundamental frequency. The magnitude of this effect depends on the observer. The results from observer GS suggest that it is also present in inharmonic situations. This slight but consistent effect has been reported extensively by Walliser (1968, 1969a, 1969c). It will be disregarded in the following, where the slope $\Delta P/\Delta f$ is important.

4.4. Discussion

The pitch of complex tones can be predicted on the basis of frequency pattern or of time structure. If it is assumed that pitch is determined by recognition of

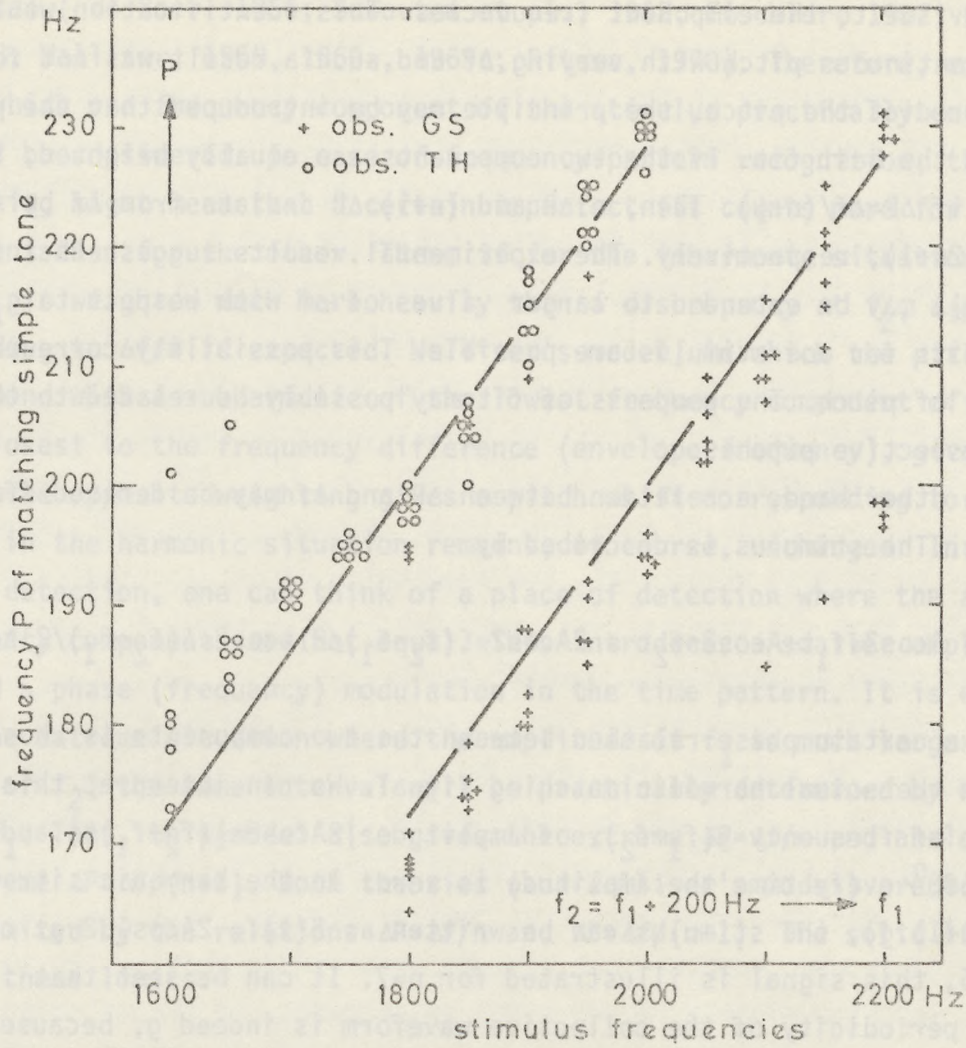


Fig. 4. Pitch of two-frequency stimuli with a constant frequency difference of 200 Hz expressed in the frequency P of a matching simple tone. For observers TH and GS, respectively, the solid lines represent the best linear fit to corresponding data from Fig. 3 obtained with a two-frequency harmonic matching signal.

a pattern arising from the projection of the frequency spectrum upon the auditory system, then the information of at least two frequency components should be present. The pitch of a two-frequency stimulus may then be determined by the projections of f_1 and f_2. They are to be compared with reference patterns, each pattern corresponding to n.P, present in the auditory system and connected with the sensation of a pitch "P". (The limited frequency-analyzing power of the auditory system may impede a pattern recognition at high harmonics; this question is discussed in Sec. 8.) In all harmonic situations, where f_1=n.g and f_2=(n+1).g, a pitch corresponding to the fundamental frequency g is predicted. After an increase of the stimulus frequencies with Δf, small with respect to g, the pattern does not fit any reference pattern if the projections of the test-signal components must coincide with successive components of a reference pattern. The restriction "successive" is reasonable because otherwise P may be identified with the greatest common divisor to the component frequencies. This identification would result in a very discontinuous pitch with varying Δf and such a result was not found. To be able to predict the pitch, the principle may be introduced that the pitch corresponds to the best fit. If the two components are equally weighted, the best fit is found if $\Delta P=\Delta f/(n+\frac{1}{2})$. Then, $n.\Delta P$ and $(n+1).\Delta P$ deviate from Δf by $-\Delta f/(2n+1)$ and $+\Delta f/(2n+1)$, respectively. The experimental results suggest that the relation $\Delta P=\Delta f/(n+\frac{1}{2})$ may be extended to larger values of Δf with respect to g. Consequently, several fits for one stimulus are possible. This possibility corresponds with the ambiguity of pitch. The goodness of fit may possibly be related to the prominence of the respective pitches.

On the other hand, a relation between ΔP and Δf may be derived from the time structure. The stimulus is described by

$$S(t) = A\cos 2\pi f_1 t + A\cos 2\pi f_2 t = 2A\cos 2\pi\{(f_2-f_1)/2\}t\cos 2\pi\{(f_2+f_1)/2\}t.$$

As such, a certain phase relation between the two components is chosen in agreement with the actual harmonic matching signal. We can interpret this signal as a sine wave of frequency $\frac{1}{2}(f_1+f_2)$, of amplitude $|2A\cos 2\pi\frac{1}{2}(f_2-f_1)t|$ and with a phase jumping 180^o every time the amplitude is zero. In the harmonic situation ($f_1=n.g$ and $f_2=(n+1).g$), the stimulus can be written as $S(t) = 2A\cos\frac{1}{2}.2\pi gt\cos(n+\frac{1}{2}).2\pi gt$. In Fig. 5, this signal is illustrated for n=7. It can be seen that in the harmonic case the periodicity of the collective waveform is indeed g, because in the time interval τ the fine structure shows $7\frac{1}{2}$ oscillations and the phase has once jumped 180^o at zero amplitude. This periodicity is lost if both frequencies are increased by Δf ($\Delta f<g$) away from the harmonic situation. Then the pitch can be predicted with the aid of a simple model proposed by de Boer (1956), which, of course, also holds for harmonic stimuli. The pitch is presumed to be related to the time interval between, for example, pronounced peaks of the waveform. Different time intervals (in the harmonic situation of Fig. 5: τ, τ', and τ'') may be obtained by taking various peaks. This agrees with the ambiguity of pitch. The pitch corresponding to the time interval τ is determined by $(n+\frac{1}{2})$ oscillations of a frequency $(n+\frac{1}{2}).g$ in the fine structure. After a frequency shift Δf, the time interval is determined by $(n+\frac{1}{2})$ oscillations of a frequency $(n+\frac{1}{2}).g + \Delta f$. Thus, a pitch corresponding to $\Delta f/(n+\frac{1}{2})$ may be expected. This model has the advantage (in contrast to frequency-pattern recognition) that no approximation is involved in the prediction of the pitch of inharmonic stimuli.

Both theories, one based upon a frequency-pattern recognition and the other upon a detection in the time domain, predict a pitch shift corresponding to $\Delta f/(n+\frac{1}{2})$. However, the results of comparable experiments with more frequency components in the stimulus suggest that the weighting of the frequency components is asymmetrical (de Boer, 1956; Schouten et al., 1962; Fischler, 1967; Fischler and Cern, 1968; Walliser, 1968, 1969a, 1969c; Ritsma, 1970). Therefore, extreme situations in which one frequency component of the stimulus practically determines the pitch will be considered. In case of frequency-pattern recognition, the asymmetrical weighting might mean that a certain imperfect fit counts more for one frequency component than for the other. Then, if for the inharmonic situation a discrepancy at f_1 is weighted much more heavily than a discrepancy at f_2, a pitch shift corresponding to $\Delta f/n$ is expected. Walliser's model, in which the pitch is assumed to correspond with a subharmonic of the lowest frequency component of the stimulus that is closest to the frequency difference (envelope frequency), gives the same relation. The opposite weighting gives a pitch shift corresponding to $\Delta f/(n+1)$. The pitch in the harmonic situation remains, of course, unchanged. In case of time structure detection, one can think of a place of detection where the amplitudes of the frequency components are not equal. This introduces a smaller amplitude modulation and a phase (frequency) modulation in the time pattern. It is easy to see that in the extreme situation where the amplitude of f_1 is much larger than the amplitude of f_2, the time interval will be practically determined by n oscillations of f_1. Thus, this model also gives the extreme $\Delta P=\Delta f/n$ and analogously the other extreme $\Delta P=\Delta f/(n+1)$. Both theories (and Walliser's model) predict a pitch in an area limited by the relations $\Delta P=\Delta f/n$ and $\Delta P=\Delta f/(n+1)$. The solid lines in Fig.3 represent these limits.

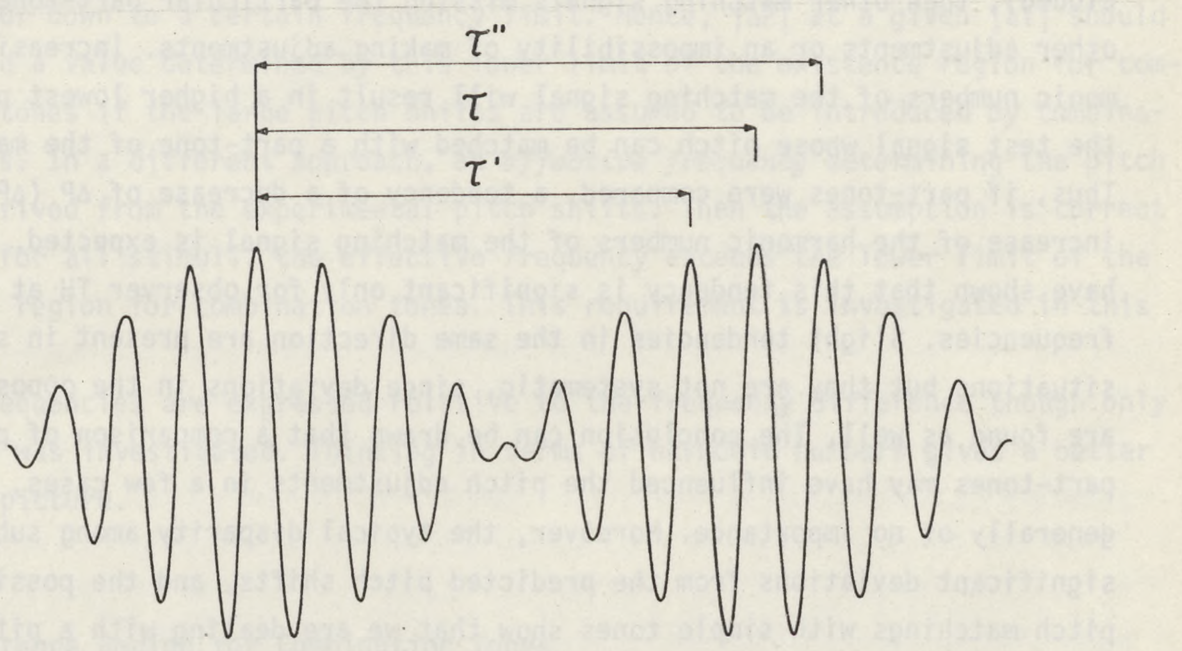


Fig. 5. Waveform of a two-frequency stimulus consisting of the seventh and eighth harmonics with equal amplitudes. Between different peaks there are indicated several time intervals τ, τ', and τ'' which can be related to pitches produced by this signal.

Owing to the small standard deviation of the average pitch settings, it can be concluded that most pitches deviate significantly from the predicted values in spite of the generality of the theoretical speculations which cover all theories mentioned previously. These significant deviations are not restricted to two-frequency stimuli. The results of Schouten et al. (1962) show significant deviations from the predictions of Fischler, which can be seen more clearly in the representation of these data by Ritsma (1970). Moreover, a calculation similar to Fischler's, but with an extreme weighting, does not result in a right prediction, because for higher harmonics and for most observers the pitch shift found is significantly larger than Δf divided by the lowest harmonic number of the stimulus components. Also, the results of Walliser (1968; cf. Fig. 32A) obtained with inharmonic stimuli consisting of more than three frequency components show this deviation from the general prediction, the prediction of his particular model included. We can infer that all theories yet presented in this paper fail to predict correctly the pitch of inharmonic complex tones.

As the preceding speculations start from the assumption that the pitches of complex tones were matched, it should be checked that no pitches of some part-tones were compared. This may be examined by considering the individual adjustments. However, a statistically significant conclusion requires a standard deviation of the individual settings even smaller than was obtained. The possibility that part-tones were compared can also be examined by tracing the influence of different matching signals on the pitch adjustments. If the matching would be based on a comparison of the pitches of certain part-tones (combination tones included), then other matching signals missing the particular part-tone would give other adjustments or an impossibility of making adjustments. Increasing the harmonic numbers of the matching signal will result in a higher lowest part-tone of the test signal whose pitch can be matched with a part-tone of the matching signal. Thus, if part-tones were compared, a tendency of a decrease of ΔP ($\Delta P=\Delta f/n$) with an increase of the harmonic numbers of the matching signal is expected. Minute tests have shown that this tendency is significant only for observer TH at low stimulus frequencies. Slight tendencies in the same direction are present in some other situations but they are not systematic, since deviations in the opposite direction are found as well. The conclusion can be drawn that a comparison of pitches of part-tones may have influenced the pitch adjustments in a few cases, but it is generally of no importance. Moreover, the typical disparity among subjects, the significant deviations from the predicted pitch shifts, and the possibility of pitch matchings with simple tones show that we are dealing with a pitch characterized by distinct properties.

5. PITCH IN RELATION TO COMBINATION TONES

5.1. The Possible Relation

The deviations of the pitch shifts in Fig. 3 (and Fig. 4) from the prediction $|\Delta f|/(n+1)<|\Delta P|<|\Delta f|/n$ are all in the same direction; $|\Delta P|$ tends to be larger than expected. This is especially true for higher harmonic numbers. The results in the cited literature show the same trend.

The experimentally determined pitch shift is comparable with predicted pitch shifts for lower harmonic numbers. This raises the question as to whether perhaps combination tones play a part in pitch detection. Recent publications (Zwicker, 1955, 1968; Plomp, 1965; Goldstein, 1967; Helle, 1969/1970) show that a nonlinear mechanism introduces a set of combination tones that can be described by $f_1-k(f_2-f_1)$ if the stimulus consists of the frequency components f_1 and f_2, with $f_1<f_2$ and f_2/f_1 not too large and k a small integer. These combination tones correspond to lower stimulus frequencies in both harmonic and inharmonic situations; a simultaneous increase of $f_1=n.g$ and $f_2=(n+1).g$ by Δf gives combination tones that can be described by $(n-k).g$ also increased by Δf. Thus, the large pitch shift might be explained by taking into account combination tones. The import of this idea is investigated in Secs. 6 and 7.

A characteristic property of these combination tones is that, at moderate SPL, they are found in a restricted existence region below f_1. (Extensive measurements on this subject will be reported soon.) In each stimulus condition, combination tones occur down to a certain frequency limit. Hence, $|\Delta P|$ at a given $|\Delta f|$ should not exceed a value determined by this lower limit of the existence region for combination tones if the large pitch shifts are assumed to be introduced by combination tones. In a different approach, an *effective frequency* determining the pitch may be derived from the experimental pitch shifts. Then the assumption is correct only if, for all stimuli, the effective frequency exceeds the lower limit of the existence region for combination tones. This requirement is investigated in this section.

The frequencies are expressed relative to the frequency difference though only one value was investigated. Thinking in terms of harmonic numbers gives a better over-all picture.

5.2. Existence Region for Combination Tones

The combination tones are found in the frequency region $f_\lambda<f_1-k(f_2-f_1)<f_1$, where the lower limit f_λ of this existence region depends mainly on the SPL of f_1

and additionally on the SPL of f_2 and on k. The limit f_λ was determined for the SPL and for all f_1=n.200 Hz used in the pitch-matching experiment of Sec. 3. We measured f_λ by gradually decreasing f_2, starting from $f_2=2.f_1$, and observing the appearance of successive combination tones described by successive k. This method turned out to be very reliable: the appearance of a new combination tone (crossing the lower limit of the existence region) is noticed easily and its recognition is facilitated because the pitch of the combination tone changes in a direction opposite to the pitch of the part-tone corresponding to f_2. The frequencies f_2' at which the successive combination tones appeared were noted. In addition, the pitch of a simple tone was matched to the pitch of each combination tone, raised above its threshold at f_2' by introducing a small decrease of f_2, in order to check the assignation of the correct value of k to the respective combination tones. Then $f_\lambda=f_1-k(f_2'-f_1)$ was calculated for successive k.

In this way, f_λ was revealed as a function of k and the associated values of $f_2'-f_1$. The value of f_λ appropriate to the pitch-matching experiment was obtained by interpolating f_λ for the situation f_2-f_1=200 Hz. When f_2 was decreased, the appearance of successive combination tones could be observed down to a certain f_2-f_1 before all part-tones merged into one beating sound. For the highest stimulus frequencies giving rise to a low pitch of the complex tone (f_1=13.200-15.200 Hz for observer GS and f_1=11.200-13.200 Hz for observer TH), successive combination tones could be observed down to about f_2-f_1=200 Hz by observer GS and down to about f_2-f_1=180 Hz by observer TH. Then the highest values of k were 6 and 4-5 for observer GS and TH, respectively. Smaller frequency differences were found for lower stimulus frequencies. Thus, $f_\lambda(f_2-f_1=200\text{ Hz})$ could be obtained without extrapolations for all stimulus frequencies. The lower limit λ relative to the frequency difference f_2-f_1=200 Hz, $\lambda=f_\lambda/(f_2-f_1)$, has been plotted in Fig. 6 as a function of f_1, f_2. The standard deviation has not been given simultaneously because it is too small (one- or two-tenths of a scale unit).

This representation of the lower limit of the existence region for combination tones was preferred above a representation of the lowest-audible combination tones. In the latter case, the lower limit consists of segments of straight lines described by $f_1-k(f_2-f_1)$, with k as high as possible but still $f_1-k(f_2-f_1)>f_\lambda$. These line segments parallel f_1 and f_2 in Fig. 6 and leap by 200 Hz at stimulus frequencies where such a line intersects λ. Apart from the fact that, using λ, these lines can be constructed in Fig. 6, the latter representation was rejected because it was considered to be too limited. For example, the time structure in channels with characteristic frequencies near the lower limit of the existence region for combination tones may have a component corresponding to a subthreshold combination tone. If pitch is related to the time structure, then this component may affect the pitch and the representation of the lower limit as given by λ in Fig. 6 will be more adequate.

This determination of the lower limit of the existence region for combination tones does not necessarily imply that all combination tones corresponding to frequencies within the existence region are actually present. The results of Goldstein (1967) suggest that all these combination tones are really present with monotonically increasing loudness from the lower limit up to a frequency (about 10% below f_1) whose corresponding combination tone can still be distinguished from the part-tone corresponding to f_1.

5.3. Effective Frequency for Pitch Detection

According to the pitch theories based upon frequency-pattern recognition as well as time-structure detection, $\Delta f/\Delta P$, the reciprocal of the slope of the ob-

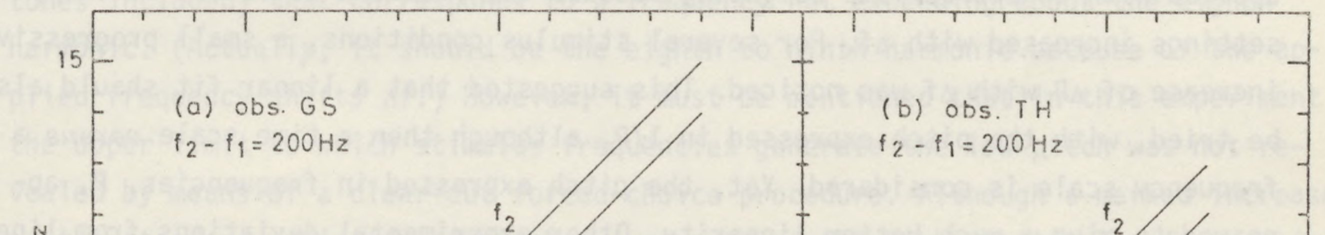


Fig. 6. Effective frequency η for pitch detection and lower limit λ of the existence region for combination tones as a function of stimulus frequencies for stimulus components of both 40 dB SL. Frequencies are expressed relative to the constant frequency difference $f_2-f_1=200$ Hz; the stimulus frequencies f_1 and f_2 are indicated by straight lines. If pitch originates from primary tones and combination tones, then $\lambda<\eta<f_2$ should hold. If for observer TH two values of η are given, the upper curve holds for P<200 Hz and the lower for P>200 Hz.

lique lines in Fig. 3 equals the number of the harmonic most effective in perception of the pitch of the complex tone. Also, $\Delta f/\Delta P$ may equal a noninteger resulting from an appropriate weighting of several harmonics. Thus η, the effective frequency relative to the frequency difference, may be defined by $\eta=\Delta f/\Delta P$. The theories predict η as a function of Δf to be a constant determined by the weighting of the stimulus frequencies and characterized by the central harmonic situation: $n<\eta<(n+1)$. Although the predicted constants of proportionality are mostly incorrect, the linear relationship (η=const) may actually still hold. On the other hand, there may be a one-to-one relationship between η and the stimulus frequencies independent of the initial central harmonic situation. If combination tones are involved in pitch detection, another possibility would be that their levels varying with the frequency shifts Δf around a central harmonic situation affect η. An analysis of the data for individual Δf values showed that as a first-order approximation $\eta(\Delta f)$=const. Thus, the effective frequency could be calculated from a best linear fit to experimental $\Delta P(\Delta f)$. A best fit in the sense of least squares was applied with an appropriate weighting because the standard deviation of the pitch settings increased with Δf. For several stimulus conditions, a small progressive increase of ΔP with Δf was noticed. This suggested that a linear fit should also be tried, with the pitch expressed in $1/P$, although then a time scale *versus* a frequency scale is considered. Yet, the pitch expressed in frequencies, P, appeared to give a much better linearity. Other experimental deviations from linearity that can be distinguished in Fig. 3 are not systematic and were therefore regarded as being insignificant.

The calculated effective frequencies were also plotted in Fig. 6. The standard deviation was again too small to represent in the figure. The standard deviation was determined by 81 adjustments (nine series of nine frequency settings) and amounted typically from 0.05 for low stimulus frequencies up to 0.10 for the higher frequencies. Reproduction of the results could be examined by comparing these values of η with values revealed more than half a year later, in an experiment presented in Sec. 6. Differences of 0.21, 0.20, and 0.33 were found for three different situations, respectively.

For the situations with low harmonics of observer TH, one straight line does not fit the data satisfactorily (see Fig. 3). Two different values of η were found for each frequency pair; a higher one for pitches with $P<200$ Hz and a lower one for $P>200$ Hz, both plotted in Fig. 6(b). In the situations $(n,n+1)=(6,7)$ and $(n,n+1)=(7,8)$, the individual settings suggested that adjustments were made erroneously between individual part-tones of the test signal and the matching signal for $P<200$ Hz. This may be caused by a less pronounced lower pitch of the complex tone eluding observation in the situations of ambiguous pitch. The bifurcation in case of $(n,n+1)=(8,9)$ may be due to a dependence on Δf of a combination tone near the lower limit of the existence region (if combination tones are involved in pitch detection).

5.4. Effective Frequency in Relation to Existence Region

In agreement with the requirement that combination tones play a part in pitch detection, the effective frequency η (*re* 200 Hz) appears to lie above the lower limit λ of the existence region for combination tones. Figure 6 shows also that the deviation of η from the stimulus frequencies f_1, f_2 (*re* 200 Hz) increases with increasing stimulus frequencies. Moreover, for both observers, it seems that η is limited by a value of about 8, resulting in coincidence of η and λ. The coincidence occurs at the highest stimulus frequencies that still gave rise to the low pitch of the complex tone. This suggests that generation of this low pitch by two-frequency stimuli requires at least one stimulus-introduced part-tone (combination tones included) that corresponds to a frequency not exceeding about the eighth harmonic. (Actually, it should be the eighth to ninth harmonic because of the applied frequency shifts Δf.) However, it must be mentioned that in this experiment the upper limit to which stimulus frequencies generate the low pitch was not revealed by means of a clear-cut forced-choice procedure. Although a marked increase of the standard deviation of pitch adjustments at the highest stimulus frequencies indicated that the upper limit was reached, the limit was actually determined by the observers judging reliable pitch matchings to be impossible for higher frequencies.

5.5. Disparities among Subjects for Combination Tones and Pitch

In Sec. 3, a disparity among subjects was found in pitch perception of two-frequency stimuli. The disparity appeared to be consistent and this suggested we investigate whether there is a correlation with the existence region for combination tones. Therefore, the lower limit of the existence region (for $2f_1-f_2$) was determined three times for each of the 42 subjects, at the SPL used in the pitch experiment and for f_1=1800 Hz. The results showed quite a large variation among subjects; the lower limit varied from 1150 to 1550 Hz. This variability could not be explained by a different performance of "good" and "bad" listeners. Although the frequency distribution of the results looked unimodal, there appeared to be a significant correlation between the lower limit of the existence region and the two groups of subjects. Both the nonparametric Wilcoxon test and an application of student's t distribution showed with a significance of more than 99.5% that the lower limit of the existence region for combination tones was lower for subjects judging pitch on the basis of the complex tone perceived as a whole than for subjects judging pitch on the basis of individual part-tones. Yet, the difference of the mean lower limits, which are, respectively, 1267 and 1367 Hz for both groups, seems to be too small to account directly for the large disparity in pitch perception.

6. PITCH AS A FUNCTION OF SPL

6.1. Introduction

The results of the preceding section support the idea that combination tones are involved in pitch detection. The relation between the effective frequency determining the pitch and the lower limit of the existence region for combination tones invites an additional investigation into the pitch shift of two-frequency stimuli as a function of SPL. The lower limit of the existence region rises (the existence region diminishes) with a decrease of SPL of f_1. Owing to this dependence on SPL, a nice experimental situation can be created: An equal variation of both levels does not change the time pattern while the existence region changes and, *vice versa*, a limited change of the amplitude of f_2 does not affect the existence region while the time pattern changes, especially with respect to the amplitude variations.

6.2. Experiment

The apparatus configuration of Fig. 2 was used again. In case of equal levels, the sensation levels of the individual part-tones of the stimulus were 25, 40, and 55 dB. The pitch of the complex tone becomes rather weak at levels below 25 dB, and the highest SPL was mainly imposed by the level of the distortion components of the matching signal. The sensation levels of the test signal and the matching signal were always equal.

Particular aspects of pitch perception of two-frequency stimuli appeared to emerge from conditions with different levels L_1 and L_2 of f_1 and f_2, respectively. In the situations (n,n+1)=(8,9),(9,10), and (10,11) investigated in this section, both observers judged the pitch of the complex tone to be most pronounced for $L_2<L_1$; for example: L_2=34 dB for L_1=40 dB. In that case, we got the impression that the individual part-tones corresponding to f_1 and f_2 contribute equally to the complex tone. The asymmetry became even more marked when the conditions L_1,L_2=

40,25 dB and L_1,L_2=25,40 dB were compared. In the first condition, a pitch of the complex tone was perceived about as pronounced as the pitch in the condition 40, 40 dB, but the pitch in the second condition was hardly detectable. The part-tone corresponding to f_2 dominated completely in the condition 25,40 dB. The amplitude variations of the time pattern are equal for both conditions. Hence, sounding more or less pronounced does not seem to be determined by the "modulation depth" of the acoustic waveform. The existence region for combination tones, however, is larger for the condition 40,25 dB because it is mainly determined by L_1. Here, as well, there is an interesting connection between pitch of the complex tone and combination tones.

The pitch was investigated for the conditions L_1,L_2=25,25; 40,40; 55,55 dB; and 40,25 dB for (n,n+1)=(9,10) (observer TH) and (n,n+1)=(10,11) (observer GS). The method of Sec. 4 was used, employing stimulus frequencies with a constant difference of 200 Hz, frequency shifts in steps of 50 Hz away from a central harmonic

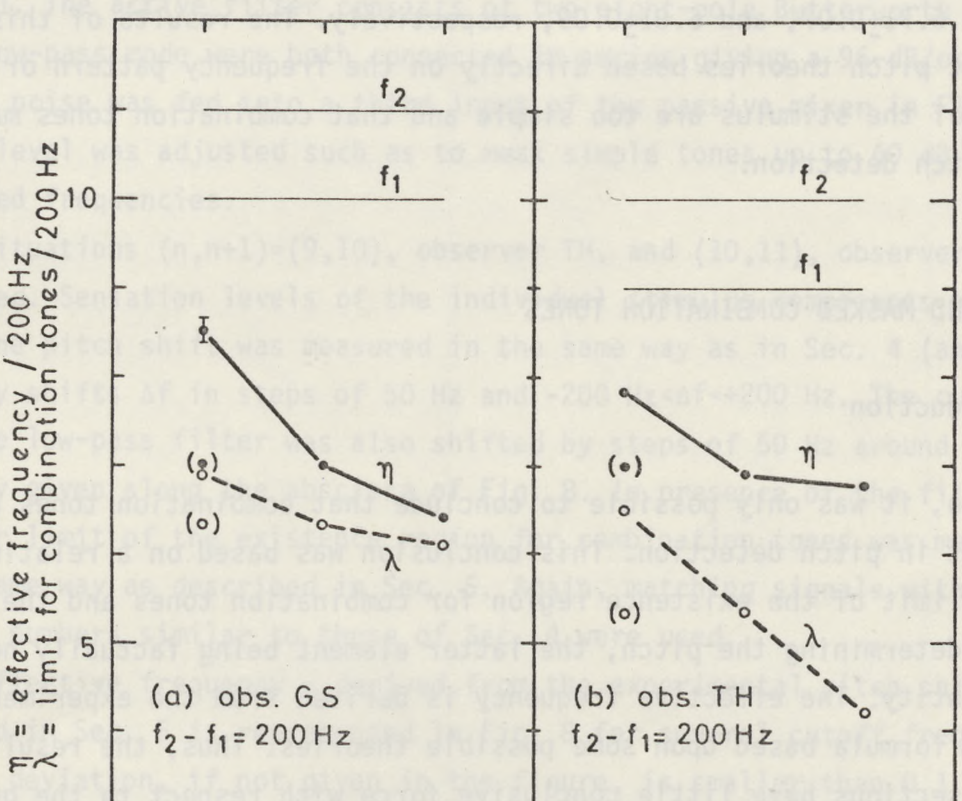


Fig. 7. Effective frequency η for pitch detection and lower limit λ of the existence region for combination tones as a function of the sensation level of individual stimulus components. Symbols o and •: $L_1=L_2$; symbols (o) and (•): L_1,L_2= 40,25 dB. Frequencies are expressed relative to the constant frequency difference f_2-f_1= 200 Hz; the stimulus frequencies f_1 and f_2 in the central harmonic situation are indicated.

situation up to +200 Hz or down to -200 Hz, and matching signals with different harmonic numbers similar to those of Sec. 4. The effective frequency η derived from experimental pitch shifts and the lower limit λ of the existence region for combination tones (measured as in Sec. 5) are both given in Fig. 7. The standard deviation of the effective frequency determined by 54 adjustments (six series of nine frequency settings) is typically 0.08 scale unit (harmonic numbers) except for the condition 25,25 dB of observer GS, as indicated. The standard deviation of the mean lower limit of the existence region is about 0.1 scale unit.

Again, the results show a relation between the effective frequency and the lower limit of the existence region. Generally, an equal decrease of both L_1 and L_2 gives an increase of the lower limit and an increase of the effective frequency. A decrease of L_2 with constant L_1=40 dB, which practically does not affect the lower limit of the existence region, results in a negligible change of the effective frequency (see Fig. 7, symbols within brackets). This was also found in an additional experiment with the situation (n,n+1)=(8,9) and observer GS; stimulus conditions L_1,L_2=40,25 dB; 40,35 dB; and 40,45 dB gave effective frequencies of η= 6.87±0.08, 6.71±0.07, and 6.81±0.09, respectively. The results of this section affirm that pitch theories based directly on the frequency pattern or on the time structure of the stimulus are too simple and that combination tones may play a part in pitch detection.

7. PITCH AND MASKED COMBINATION TONES

7.1. Introduction

Hitherto, it was only possible to conclude that combination tones are likely to play a part in pitch detection. This conclusion was based on a relation between the lower limit of the existence region for combination tones and the effective frequency determining the pitch, the latter element being factually not an experimental quantity. The effective frequency is derived from the experimental pitch shift by a formula based upon some possible theories. Thus, the results of the preceding sections have little conclusive force with respect to the question of whether the experimental deviations from the predicted pitch shift are really caused by combination tones. Suppose one of the proposed theories to be correct; even then, an effective frequency lower than the stimulus frequencies and always within the existence region for combination tones does not imply that pitch originates necessarily from combination tones. A casual rather than a causal relation may exist or, as a third possibility, if pitch is determined by the time structure, then there may be transformations of the time structure accounting for the large pitch shift independent of the presence of combination tones while they both are introduced by a common nonlinearity.

Owing to the property that these combination tones can be masked selectively, the question can be examined more directly. The lower limit of the existence region for combination tones can be increased in a certain situation by adding to the stimulus low-pass-filtered noise that masks the combination tones up to a new higher limit. If a pitch originates from certain combination tones, then we may expect that a proper masking of these combination tones affects the pitch. Moreover, it is interesting to check whether the effective frequency derived from the pitch shift remains within the constricted existence region for combination tones.

7.2. Experiment

A noise generator producing pink noise with a slope of -3 dB/oct (Peekel R 230) and a digitally tuned active filter (Krohn-Hite 3342) were added to the apparatus of Fig. 1. The active filter consists of two eight-pole Butterworth networks which in the low-pass mode were both connected in series giving a 96-dB/oct slope. The filtered noise was fed into a third input of the passive mixer in Fig. 1. Its masking level was adjusted such as to mask simple tones up to 40 dB SL at the unattenuated frequencies.

The situations (n,n+1)=(9,10), observer TH, and (10,11), observer GS, were investigated. Sensation levels of the individual stimulus components amounted to 40 dB. The pitch shift was measured in the same way as in Sec. 4 (and Sec. 6) with frequency shifts Δf in steps of 50 Hz and $-200\ \text{Hz}<\Delta f<+200\ \text{Hz}$. The cutoff frequency f_c of the low-pass filter was also shifted by steps of 50 Hz around the cutoff frequency given along the abscissa of Fig. 8. In presence of the filtered noise, the lower limit of the existence region for combination tones was measured exactly in the same way as described in Sec. 5. Again, matching signals with different harmonic numbers similar to those of Sec. 4 were used.

The effective frequency η derived from the experimental pitch shifts in a way discussed in Sec. 5 is represented in Fig. 8 for several cutoff frequencies. The standard deviation, if not given in the figure, is smaller than 0.1 scale unit (harmonic numbers). The affected lower limit λ of the existence region for combination tones is also given in Fig. 8. The standard deviation of the mean limit appeared to be smaller than 0.2 scale unit. The measurements at $f_c=0$ were performed without noise and were intermingled with the other conditions.

During the pitch matchings, the observers noticed already that the noise affects the clarity of the pitch. The pitch becomes less pronounced for higher cutoff frequencies. The adjustments at $f_c=1400$ Hz and $f_c=1700$ Hz for observer GS and at f_c=1400 Hz for observer Th were difficult, and it was impossible for observer TH to make pitch adjustments at f_c=1700 Hz. This is remarkable because the noise hardly affects the region of the stimulus frequencies in these cases; only at the highest cutoff frequencies f_c=1700 Hz for observer GS and f_c=1400 Hz for observer TH does the noise give some masking at f_1 of 10 dB and 6 dB above hearing threshold, respectively. Thus, the less pronounced pitch is likely to be caused by masking of the combination tones and not by masking of the time structure or of the frequency pattern, both corresponding to the stimulus. Moreover, the increase of the effective frequency with increasing cutoff frequency is in agreement with this assumption. The effective frequency found remains within the constricted existence region for combination tones. Assuming that the nonlinearity itself is not affected by introducing the noise, which is yet most likely, it can be concluded that, for higher stimulus frequencies, the pitch of the complex tone originates from the combination tone region. As in Sec. 5, the effective frequency seems to

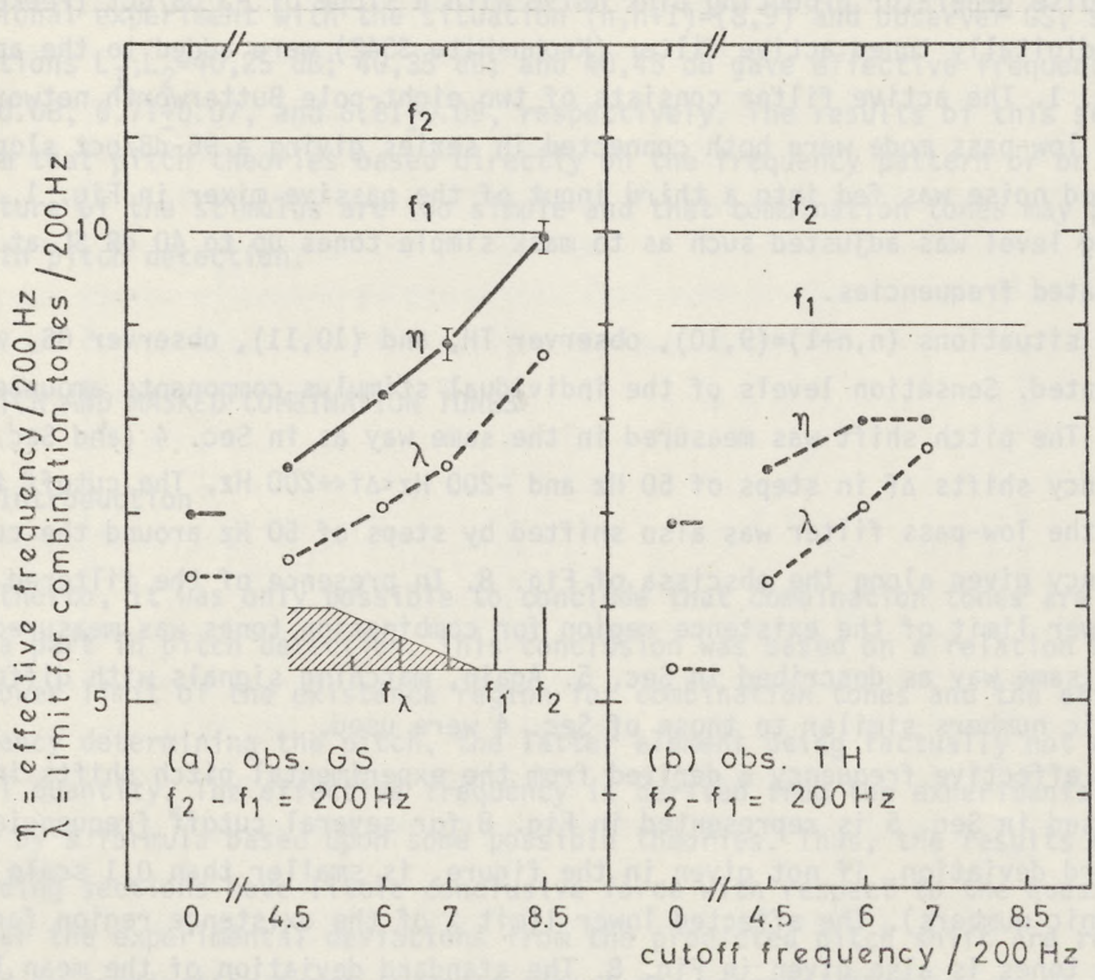


Fig. 8. Effective frequency η for pitch detection and lower limit λ of the existence region for combination tones as a function of the cutoff frequency of low-pass-filtered noise with a −96 dB/oct slope. In the passband, the noise masks simple tones up to 40 dB sensation level; the sensation levels of stimulus components are also 40 dB. Frequencies are expressed relative to the constant frequency difference f_2-f_1=200 Hz; the stimulus frequencies f_1 and f_2 in the central harmonic situation are indicated.

be limited to $\eta<8$ for observer TH. For observer GS, however, higher values were found, but together with a progressive increase of the standard deviation.

Noise-induced pitch shifts have been reported more than once. The pitch of a simple tone may change if a band of noise is introduced in a frequency region adjacent to the stimulus component. Comparable pitch shifts, which would show up as a parallel translation of the linear approximations, have not been found in these experiments. The noise affected only the slopes.

8. GENERAL DISCUSSION

The results of the preceding experiments on the pitch of inharmonic stimuli have shown that pitch theories based directly upon physical properties of the stimulus are inadequate. Both theories departing from either time structure or frequency pattern and Walliser's intermediate theory predict a smaller pitch shift than actually was found for stimuli of higher harmonic numbers, above a certain SL. This finding turns up through all literature available. A nonsymmetrical weighting of the spectral components cannot so much as account for all experimental pitch shifts. This paper demonstrates that an incorporation of combination tones of the type $f_1-k(f_2-f_1)$ into the theory can account for the pitch shifts larger than predicted by former theories. For the effective frequency determining the pitch appeared to lie always within the existence region for combination tones. In addition, several effects which could not be explained by the unextended theory can now be explained. The pitch shift depending on stimulus SPL (while time structure and frequencies were not changed ($L_1=L_2$)) can be explained by an SPL-dependent existence region for combination tones. Also, the pitch shift independent of limited variations of L_2 can be explained by an unaffected lower limit of the existence region for combination tones. Another effect to be mentioned is that a two-frequency stimulus with partials in the region of the 10th harmonic and with successive component levels L_1,L_2=40 dB,25 dB has a pronounced pitch of the complex tone, whereas a stimulus consisting of the same partials but with levels L_1,L_2=25 dB,40 dB has a hardly detectable pitch. A mutual masking, if it would impede a frequency-pattern recognition asymmetrically, is expected to give rather the reverse result and, with respect to time structure, in both conditions the envelopes are similar. The finding may be explained by a smaller existence region for combination tones in the condition L_1,L_2=25 dB,40 dB and masking from f_2 in the combination tone region. Finally, it has been found that a selective masking of the combination tones affects the pitch shift when there is hardly any masking at the stimulus frequencies. The effective frequency stayed above the lower limit of the existence region.

Whether the mechanism of pitch perception is based upon temporal information, spectral information, or a combination of both, has not been specified with the introduction of combination tones as a possible means to explain the large pitch shifts. The experimental deviations from predictions may be explained equally well in terms of time or frequency as Schroeder (1966) has shown in particular for the results by Schouten et al. (1962). Schroeder postulated an internally generated sinusoidal phase modulation of the SAM signal antiphasic to the amplitude modulation. The experimental results may be predicted rather well if, in addition to his suggestion, the phase-modulation index is assumed to increase with the SPL and the harmonic numbers. Then, the instantaneous frequency at the crest of the envelope of the time pattern becomes lower, which gives time intervals between pronounced deflections that can account for the experimental pitch values. However, this postulation leaves unexplained the specific dependence of the pitch shift on the SPL of f_1 (for higher harmonic numbers). Besides providing an explanation in the time domain, Schroeder also showed that such a phase modulation introduces new spectral components mainly at the low-frequency side of the stimulus. These frequency components might be identified with the combination tones. A pitch mechanism based upon frequency-pattern recognition is equally well possible.

The preceding reasoning is in terms of a model. Experimental facts have to be examined. As for a possible frequency-pattern recognition, there are no problems. Combination tones behave like tones that are mechanically present in the ear (Goldstein, 1967). Thus, a pitch mechanism based upon frequency patterns may be simply extended to the distortion components and an appropriate nonsymmetrical weighting will predict the pitch correctly. As for a possible time-structure detection, the electrophysiological results of Goldstein and Kiang (1968) are important. They reported nerve spikes in primary fibers of the cat to be time-locked to externally generated frequency components corresponding to combination tones. Although the results have to be interpreted carefully (de Boer, Kuyper, and Smoorenburg, 1969), they support the idea that the combination tones are recovered in the timing of nerve spikes in primary fibers of characteristic frequencies in the combination tone region. The results by Brugge et al. (1969) and Rose et al. (1969) show a timing of the nerve spikes closely related to unilateral elevations of the basilar membrane. They did not need to take into account nonlinearities producing combination tones; probably because of their particular choice of the two stimulus frequencies, one at each side of the characteristic frequency of the fiber. According to electrophysiological data, the timing of nerve spikes corresponds closely to the internal frequency pattern (combination tones included). Thus, the results of the present experiments can be explained as well on the basis of time-structure detection.

The large pitch shifts found in the present experiments can be understood on the basis of the afore-mentioned dominant frequency region in the perception of pitch of complex tones (Sec. 2). The effective frequency η determining the pitch will tend to the fourth harmonic. If the stimulus frequencies are all above this dominant region, then the pitch will be determined by the lowest frequency components present in the auditory system. These frequency components appeared to correspond often to combination tones. The partials in the dominant region are well analyzed by the human ear. This is illustrated by an aftermasking pattern of a complex tone found by Plomp (1964), which shows that harmonics up to the fifth are well separated. (Spectral aftermasking patterns are perhaps a good representation of the projection of the frequency spectrum upon the auditory system.) Tentative investigations into the aftermasking pattern of two-frequency stimuli revealed also a fast deterioration of component separation above the fifth harmonic. Thus, a pitch mechanism based upon frequency-pattern recognition, which requires a good separation of components, can explain satisfactorily the dominance of low part-tones and consequently the large pitch shifts. Frequency-pattern recognition can also account for the upper limit of the frequency region in which these two-frequency stimuli produce a low pitch (the region of tonality). The aftermasking patterns of the two-frequency stimuli show a deterioration of component separation in accordance with this upper limit. Plomp (1964) found that the ear is able to analyze two adjacent harmonics up to the 10th, which suggests $\eta<10$ while $\eta<8$-9 was found experimentally. In the present experiments, the observers were able, in determining the lower limit of the existence region for combination tones, to perceive individually the combination tones near the effective frequency. The upper limit of the region of tonality coincided with a merging of these combination tones. All these findings suggest that the region of tonality is limited at the high-frequency side by the frequency-analyzing power of the human ear.

However, a simple frequency-pattern recognition supposed to underlie the perception of pitch of complex tones cannot account for all experimental results. It was reported in Sec. 3 that subjects perceiving a pitch jump of individual part-tones in case of successive presentation of two different two-frequency stimuli did not perceive the low pitch of the complex tone. Yet, the two-frequency stimuli were best analyzed by these subjects. Thus, perception of the pitch of the complex tone seems to require more than information about individual frequency components present in the auditory system. The correlation between subjects perceiving the low pitch of the complex tone and subjects having a relatively large existence region for combination tones (Sec. 5.5) may contain a cue to this question. It is but an author's guess that the perception of the low pitch of the complex tone may be determined by the presence of a perceptually distinct low rattle and that the origin of this rattle may be linked with the nonlinearity generating combination tones.

The perceptual rattle is likely to be related to the time structure. The upper limit of the region of tonality is likely to be related to the frequency-analyzing power. Thus, it is impossible to decide on whether the pitch mechanism operates in the time domain or in the frequency domain. In this connection, Walliser's (1968, 1969c) functional model is interesting, because both temporal and spectral information play a part in it. The model can account for the results of the present experiments if this rule: the pitch of a complex tone is determined by *that subharmonic of the pitch of a part-tone corresponding to the lowest partial in the stimulus* which has a pitch corresponding as close as possible to the envelope frequency, is modified for the part in italics into: *that subharmonic of the pitch of an appropriate part-tone, combination tones included.* The rule predicts the effective frequency (*re* $f_2-f_1=200$ Hz) to be restricted to integers. This prediction could not be tested conclusively because the standard deviation for individual frequency settings was too large. (Values of η given in the figures of this paper were obtained by calculations based upon more than one frequency setting, which means that these values may deviate from integers if individual settings do correspond to integer, but different, values of η.) Walliser's functional model for pitch perception of complex tones is difficult to understand, however, in terms of a physiological mechanism.

The results of the present experiments support the view that frequency analysis plays an essential part in pitch detection, whereas Schouten et al. (1962), Ritsma (1962), and Ritsma and Engel (1964) concluded that pitch detection of three-frequency stimuli is based entirely upon time structure. It may be argued that these two conclusions do not necessarily contradict each other because of the different number of frequency components. The low pitch of the complex tone produced by three harmonics is more pronounced than the pitch evoked by two-frequency stimuli and phase effects are perceived only for stimuli with three or more frequency components. Yet, a closer examination of the arguments pleading for time-structure detection is interesting. The limited frequency resolution of the ear has been brought up as one of the arguments against a pitch mechanism based upon spectral information. Ritsma (1962) reported a region of tonality for SAM signals exceeding harmonic numbers of 20, which is rather high with respect to the frequency resolution, even if in that case combination tones corresponding to harmonics down to about the 14th play a part. However, Ritsma's criterion was not based upon pitch matchings by adjusting frequencies but upon recognizing a pitch similarity of two complex tones produced by harmonic SAM signals with the same fundamental frequency. One signal certainly had a tonal residue, and the modulation depth of the second signal had to be adjusted such that the tonal residue was just noticeable. Pitch matchings are only reported for harmonic numbers up to 12 (e.g., Schouten et al., 1962). A further study of just-noticeable pitch differences of complex tones and of the relation between these differences and just-noticeable pitch differences of individual part-tones is necessary, especially for high harmonic numbers. Results of another experiment by Ritsma and Engel (1964) on pitch perception of QFM signals (see Sec. 2) seem to support the theory in which pitch is related to time intervals between pronounced deflections of the waveform. However, it is uncertain whether it was permitted that they derived the time intervals from the acoustic waveform rather than from the actual cochlear waveforms. The cochlear waveforms are difficult to predict because these narrow-band signals are subject to a nonlinearity which plays an important part in pitch detection, as this paper has shown. This nonlinearity introduces combination tones whose levels depend on the phase relation (SAM or QFM) of the three-frequency stimuli (Goldstein, 1970) such that an influence of phase upon pitch might be explained also by a pitch theory based upon a simple frequency-pattern recognition.

Evidently, the results do not allow a decision on whether the mechanism of pitch detection operates in the time domain or in the frequency domain. More likely it is a combination of both. In this respect, it is noteworthy that the dominant frequency region in pitch perception of complex tones centered around the fourth to fifth harmonic also offers temporal information. According to Zwicker et al. (1967), stimulation patterns of the basilar membrane in the dominant region, although relatively widely spaced, are likely to overlap so much at normal sensation levels that they produce temporal information about the periodicity of the stimulus. This information is mainly determined by the addition of two frequency components. At SLs of 10-15 dB, the temporal information may just arise in the dominant region. Therefore, the supposition that the dominant frequency region is the optimal region for pitch detection of complex tones because it offers both temporal and spectral information is attractive. Experimental evidence, however, is necessary.

9. CONCLUSIONS

- There is a reproducing disparity among subjects in perceived pitch shift of a tone burst consisting of the frequencies 1750 and 2000 Hz followed by another burst of 1800 and 2000 Hz. Subjects perceive consistently either the pitches of individual part-tones or the complex tone as a whole with a pitch corresponding to the fundamental frequency.

- The pitch shift for inharmonic signals of higher harmonic numbers is significantly larger than any prediction from theories based directly upon physical properties of the stimulus.

- The large pitch shifts can be explained on the basis of time-oriented as well as frequency-oriented pitch theories if the nonlinearity generating combination tones of the type $f_1-k(f_2-f_1)$ is taken into account.

- Also, the dependence of the pitch shifts on SPL can be explained by taking into account this nonlinearity.

- Masking of the combination tones shows that, indeed, the large pitch shifts are connected with pitch detection in the frequency region of the combination tones.

- The results of these experiments suggest strongly that detection of the low pitch of the complex tone requires spectral information.

REFERENCES

American Standards Association. (1960). *Acoustical Terminology, S1.1-1960* (American Standards Association, New York).

de Boer, E. (1956). "On the 'Residue' in Hearing," Doctoral dissertation, Univ. of Amsterdam, The Netherlands.

de Boer, E., Kuyper, P., and Smoorenburg, G.F. (1969). "Proposed Explanation of Synchrony of Auditory-Nerve Impulses to Combination Tones," J. Acoust. Soc. Amer. *46*, 1579-1581(L).

Brugge, J.F., Anderson, D.J., Hind, J.E., and Rose, J.E. (1969). "Time Structure of Discharges in Auditory Nerve Fibres of the Squirrel Monkey in Response to Complex Periodic Sounds," J. Neurophysiol. *32*, 386-401.

Fischler, H. (1967). "Model of the 'Secondary' Residue Effect in the Perception of Complex Tones," J. Acoust. Soc. Amer. *42*, 759-764.

Fischler, H., and Cern, L. (1968). "Simulation of the Secondary Residue Effect by Digital Computer," J. Acoust. Soc. Amer. *44*, 1379-1385.

Flanagan, J.L. (1962). "Computational Model for Basilar-Membrane Displacement," J. Acoust. Soc. Amer. *34*, 1370-1376.

Flanagan, J.L., and Guttman, N. (1960a). "On the Pitch of Periodic Pulses," J. Acoust. Soc. Amer. *32*, 1308-1319.

Flanagan, J.L., and Guttman, N. (1960b). "Pitch of Periodic Pulses without Fundamental Components," J. Acoust. Soc. Amer. *32*, 1319-1328.

Fletcher, H. (1924). "The Physical Criterion for Determining the Pitch of a Musical Tone," Phys. Rev. *23*, 427-437.

Goldstein, J.L. (1967). "Auditory Nonlinearity," J. Acoust. Soc. Amer. *41*, 676-689.

Goldstein, J.L., and Kiang, N.Y.S. (1968). "Neural Correlates of the Aural Combination Tone $2f_1-f_2$," Proc. IEEE *56*, 981-992.

Goldstein, J.L. (1970). "Aural Combination Tones," In *Frequency Analysis and Periodicity Detection in Hearing*, R. Plomp and G.F. Smoorenburg, Eds. (Sijthoff, Leiden, The Netherlands).

Guttman, N., and Flanagan, J.L. (1964). "Pitch of High-Pass Filtered Pulse Trains,"

J. Acoust. Soc. Amer. *36*, 757-765.

Helle, R. (1969/1970). "Amplitude und Phase des im Gehör gebildeten Differenztones dritter Ordnung," Acustica *22*, 74-87.

von Helmholtz, H.L.F. (1863). *Die Lehre von den Tonempfindungen als physiologische Grundlage fur die Theorie der Musik* (F. Vieweg & Sohn, Braunschweig), 1st ed., Chap. 4.

Ohm, G.S. (1843). "Ueber die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen," Ann. Phys. Chem. *59*, 513-565.

Ohm, G.S. (1844). "Noch ein Paar Worte über die Definition des Tones," Ann. Phys. Chem. *62*, 1-18.

Plomp, R. (1964). "The Ear as a Frequency Analyzer," J. Acoust. Soc. Amer. *36*, 1628-1636.

Plomp, R. (1965). "Detectability Threshold for Combination Tones," J. Acoust. Soc. Amer. *37*, 1110-1123.

Plomp, R. (1967). "Pitch of Complex Tones," J. Acoust. Soc. Amer. *41*, 1526-1533.

Plomp, R. and Mimpen, A.M. (1968). "The Ear as a Frequency Analyzer II," J. Acoust. Soc. Amer. *43*, 765-767.

Ritsma, R.J. (1962). "Existence Region of the Tonal Residue. I," J. Acoust. Soc. Amer. *34*, 1224-1229.

Ritsma, R.J., and Engel, F.L. (1964). "Pitch of Frequency-Modulated Signals," J. Acoust. Soc. Amer. *36*, 1637-1644.

Ritsma, R.J. (1967a). "Frequencies Dominant in the Perception of Pitch of Complex Sounds," J. Acoust. Soc. Amer. *42*, 191-198.

Ritsma, R.J. (1967b). "Frequencies Dominant in Pitch Perception of Periodic Pulses of Alternating Polarity," IPO Ann. Progr. Rep. *2*, 14-24. IPO, Eindhoven, the Netherlands.

Ritsma, R.J. (1970). "Periodicity Detection," in *Frequency Analysis and Periodicity Detection in Hearing*, R. Plomp and G.F. Smoorenburg, Eds. (Sijthoff, Leiden, the Netherlands).

Rose, J.E., Brugge, J.F., Anderson, D.J., and Hind, J.E. (1969). "Some Possible Neural Correlates of Combination Tones," J. Neurophysiol. *32*, 402-423.

Rosenberg, A.E. (1965). "Effect of Masking on the Pitch of Periodic Pulses," J. Acoust. Soc. Amer. *38*, 747-758.

Schaefer, K.L., and Abraham, O. (1904). "Zur Lehre von den sogenannten Unterbrechungstönen," Ann. Phys. *13*, 996-1009.

Schodder, G.R., and David, E.E., Jr. (1960). "Pitch Discrimination of Two-Frequency Complexes," J. Acoust. Soc. Amer. *32*, 1426-1435.

Schouten, J. F. (1938). "The perception of Subjective Tones," Proc. Kon. Nederl. Akad. Wetensch. *41*, 1086-1093.

Schouten, J.F. (1940a). "The Residue, a New Component in Subjective Sound Analysis," Proc. Kon. Ned. Akad. Wetensch. *43*, 356-365.

Schouten, J.F. (1940b). "The Residue and the Mechanism of Hearing," Proc. Kon. Ned. Akad. Wetensch. *43*, 991-999.

Schouten, J.F. (1940c). "The perception of Pitch," Philips Tech. Rev. *5*, 286-294.

Schouten, J.F., Ritsma, R.J., and Cardozo, B.L. (1962). "Pitch of the Residue," J. Acoust. Soc. Amer. *34*, 1418-1424.

Schouten, J.F., and 't Hart, J. (1965). "De slagtoon van klokken" Publication No. 7, Ned. Akoest. Genootschap, Stieltjesweg 1, Delft, pp. 20-49.

Schroeder, M.R. (1966). "Residue Pitch: A Remaining Paradox and a Possible Explanation," J. Acoust. Soc. Amer. *40*, 79-81.

Seebeck, A. (1841). "Beobachtungen über einige Bedingungen der Entstehung von Tönen," Ann. Phys. Chem. *53*, 417-436.

Seebeck, A. (1843). "Ueber die Sirene," Ann. Phys. Chem. *60*, 449-481.

Seebeck, A. (1844a). "Ueber die Definition des Tones," Ann. Phys. Chem. *63*, 353-368.

Seebeck, A. (1844b). "Ueber die Erzeugung von Tönen durch getrennte Eindrücke, mit Beziehung auf die Definition des Tones," Ann. Phys. Chem. *63*, 368-380.

Walliser, K. (1968). "Zusammenwirken von Hüllkurvenperiode und Tonheit bei der Bildung der Periodentonhöhe," Doctoral dissertation, Technische Hochschule München.

Walliser, K. (1969a). "Zusammenhänge zwischen dem Schallreiz und der Periodentonhöhe," Acustica *21*, 319-328.

Walliser, K. (1969b). "Zur Unterschiedsschwelle der Periodentonhöhe," Acustica *21*, 329-336.

Walliser, K. (1969c). "Ueber ein Funktionsschema für die Bildung der Periodentonhöhe aus dem Schallreiz," Kybernetik *6*, 65-72.

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

Zwicker, E., and Feldtkeller, R. (1967). *Das Ohr als Nachrichtenempfänger* (S. Hirzel Verlag, Stuttgart).

Zwicker, E. (1968). "Der kubische Differenzton und die Erregung des Gehörs," Acustica *20*, 206-209.

AUDIBILITY REGION OF COMBINATION TONES

INTRODUCTION

A stimulus consisting of two frequency components f_1 and f_2 ($f_1<f_2$) often gives rise to tones with pitches that do not correspond to f_1 and f_2 but to frequencies such as f_2-f_1, $2f_1-f_2$, and $3f_1-2f_2$. These tones are called combination tones. Among them the members of a distinct class are given by $f_1-k(f_2-f_1)$, where k is a small integer. Combination tones of this category are audible only in a restricted frequency region below f_1. This audibility region is the subject of the present paper.

Combination tones were first described more than two centuries ago. In 1714, Tartini reported the appearance of a third tone when two tones sounded simultaneously (Jones, 1935). Since that time a number of experiments on the audibility of combination tones have been performed. An extensive historical review by Plomp (1965) shows that the most frequently reported combination tones corresponded to f_2-f_1 and $2f_1-f_2$. Also frequently reported were the combination tones corresponding to $3f_1-2f_2$ and $4f_1-3f_2$. Plomp's experiments following his historical review show that the significant combination tones are given by f_2-f_1, $2f_1-f_2$, and $3f_1-2f_2$.

The first ideas about combination tones originated with theories of acoustic vibration. These theories proposed that the pitch of a tone was determined by the rate of vibration. When two tones are presented simultaneously beats are perceived (providing the frequency difference is small) which correspond to the amplitude modulations of the stimulus. The beat rate corresponds to the frequency difference between the two tones. At large frequency differences the beats blend and Romieu

---

This paper has been submitted for publication in The Journal of the Acoustical Society of America.

(1751; see Jones, 1935) assumed that these blending beats give rise to a difference tone. In 1843 Ohm introduced his acoustical law which states that, in a compound tone, a certain pitch can be distinguished only if the spectrum of the stimulus includes the corresponding frequency component. The perception of the difference tone produced by a simple two-frequency stimulus is difficult to reconcile with this law since the spectrum does not contain a component at the difference frequency. Von Helmholtz clearly recognized this problem.

Von Helmholtz (1863) introduced the concept that combination tones arise from a nonlinear stimulus transformation of the ear prior to the ear's frequency analysis. Consequently, the waveform would be distorted and cochlear frequency analysis would reveal the frequency components introduced by the nonlinearity. According to von Helmholtz the generation of combination tones was a high-level phenomenon. He suggested that at high stimulus levels the tympanic membrane does not respond linearly because of its asymmetric attachment. This was described mathematically by a quadratic term in the membrane's equation of motion. The quadratic term becomes progressively important at higher levels. A perturbation solution of the quadratic equation of motion yielded not only the second order components but a complete series of combination components. This is easily seen if one realizes that the quadratic equation of motion corresponds to a square-root-like transfer function. The perturbation solution by von Helmholtz is similar to a series expansion of this transfer function with frequency-dependent coefficients. The amplitude of, for example, $2f_1-f_2$ is described principally by the third power term of this series expansion. Von Helmholtz's theory became widely accepted although some investigators disagreed with his view. One of their objections was that difference tones could be perceived at low stimulus levels and sometimes were rather intense. These opponents preferred the alternative explanation based upon temporal aspects of the stimulus.

When one listens to the sound produced by two sine waves it will be noticed that the beats perceived at small frequency differences change into a rattle-like sensation if the frequency difference is increased. This rattle can be perceived as an independent attribute of the complex tone. As such it might be called a difference tone (f_2-f_1) because this tone is related to the amplitude variations of the stimulus which rate is given by the frequency difference of the two sinusoids. The rattle can be perceived even at low stimulus levels. It has probably played a part in the formulation of the beat-tone theory. However, this rattle is not a difference tone according to the current definition. According to this definition only a tone which sounds like a simple tone produced by one frequency component (a part-tone, see terminology Smoorenburg, 1970) and which is not present in the stimulus is called a combination tone. Only the combination tones satisfying this definition will be examined in the next paragraphs.

Recent research has shown that the difference tone, f_2-f_1, is audible only at stimulus levels above about 50 dB (Plomp, 1965). Moreover, the cancellation experiments of Zwicker (1955) and Goldstein (1967, 1970) suggest both that the difference tone is rather weak, and that the relation between its level and the stimulus level is quadratic. This means that, in accordance with von Helmholtz's theory, the generation of the difference tone can be described by a quadratic term with a small coefficient in the transfer function. The site of the nonlinearity is not yet completely known; the difference tone may be generated in the middle ear, in the cochlea, or in both.

In marked contrast with the difference tone, the combination tone corresponding to $2f_1$-f_2 and the higher order combination tones (f_1-$k(f_2$-$f_1)$, $k>1$) violate von Helmholtz's theory. These combination tones can be perceived at low stimulus levels and they do not seem to follow a (2k+1)th-power relation (Zwicker, 1955, 1968; Goldstein, 1967; Helle, 1969/1970). Moreover, the combination tones appear only in a restricted frequency region below f_1 and their levels increase markedly with decreasing f_2/f_1 (Goldstein, 1967; Zwicker, 1968). This strong frequency dependence suggests that these combination tones are generated in the cochlea, because their production is subject to a frequency selectivity that is not found in the middle ear. Goldstein (1967, 1970) and Goldstein and Kiang (1968) have demonstrated that the combination tone $2f_1$-f_2 resembles in many respects a part-tone which has a corresponding frequency component in the stimulus. Therefore, Goldstein suggested that von Helmholtz's idea that combination tones are produced by a nonlinearity of the ear prior to its frequency analysis can still be valid, but that the nature of the nonlinearity should be reconsidered.

This paper reports a systematic investigation of the frequency region in which the combination tones f_1-$k(f_2$-$f_1)$ can be heard. The restricted audibility region is a fundamental property of these combination tones. A subsequent paper will discuss the nature of the nonlinearity and its origin (Smoorenburg, 1972). A few experiments on the audibility region have been reported before (Smoorenburg, 1970). The audibility region was then called the existence region for combination tones and the concept was used to demonstrate that combination tones may play a role in pitch perception of two-frequency stimuli.

1. AUDIBILITY REGION OF $2f_1$-f_2

1.1. Method

The measurements were directed principally to a determination of the lower limit of the audibility region. For this purpose the method of adjustment was used.

Before each measurement the lower stimulus component was fixed at a certain frequency f_1 and the higher component was set at $f_2=2f_1$. Then the subject was instructed to turn the dial of the oscillator producing f_2 downward until a new part-tone just became audible. With this technique we took advantage of the particular properties of the combination tones under investigation. The appearance of a new part-tone (a combination tone entering the audibility region) was easily noticed. Also, it was easily recognized as a combination tone (of the type $f_1-k(f_2-f_1)$) since its pitch rose as f_2 was decreased.

Usually, the combination tones described by k= 1, 2, 3, etc. became audible successively. The order, k, was determined by matching the pitch of a simple tone to the pitch of the combination tone. (If the order of the combination tone is defined as the lowest power term in a transfer function which produces $f_1-k(f_2-f_1)$ the order should be 2k+1 in stead of k.)

The stimuli were produced by two free-running sine-wave generators. After attenuation the two sine waves were mixed and the result was presented monaurally through a headphone (Beyer DT 48 S). Distortion components corresponding to combination tones were more than 75 dB down in the acoustic signal.

1.2. Audibility Region as a Function of SL

In order to be able to study the frequency selectivity involved in the generation of the combination tone $2f_1-f_2$, the lower limit of the audibility region was measured as a function of the levels, L_1 and L_2, of the stimulus components, f_1 and f_2. These measurements, then, give a relation between the stimulus frequencies and stimulus levels for which the combination tone has a certain level, the threshold of audibility. This method of studying the frequency selectivity supplements studies which include cancelling the combination tone with an externally introduced component of appropriate frequency,level and phase (Goldstein, 1967; Zwicker, 1968). Different assumptions are required for an interpretation of the results obtained by the two methods. The present method requires an assumption about the threshold level of the combination tone whereas an interpretation of the cancellation data requires an assumption about the relation between the level of the cancellation tone and the level of the combination tone, a relation in which an essential nonlinearity is involved.

Each session was started with a rough measurement of the absolute threshold at the frequencies 300 Hz, 700 Hz, 1400 Hz, 2800 Hz, and 5600 Hz. For each session independently the average threshold found for 700 Hz, 1400 Hz, and 2800 Hz was defined as 0 dB SL_0 and L_1 and L_2 were adjusted accordingly. At the end of each session the thresholds were determined for the frequencies 500 Hz, 1000 Hz, 2000 Hz,

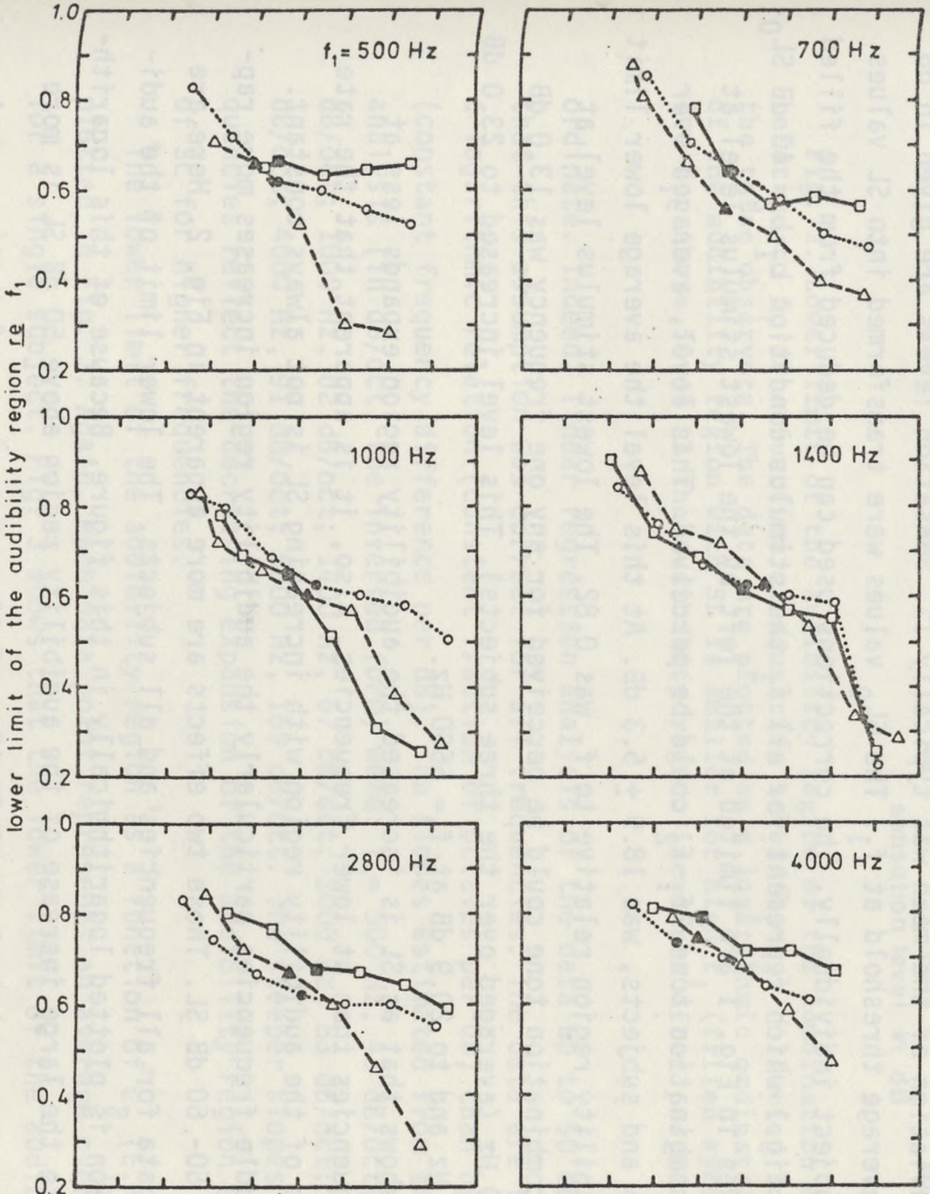


Fig. 1. Lower limit of the audibility region of the combination tone $2f_1-f_2$ relative to f_1 as a function of sensation level, where $L_1=L_2$. The sensation levels L_1, L_2 are given relative to the threshold at f_1, the corrections applied to transform SL_0 into SL can be deduced from the filled symbols which represent 40 dB SL_0 for all f_1.

4000 Hz, and 7000 Hz. The threshold values averaged over all sessions were used to transform sound pressure levels into sensation levels. The standard deviation of these average threshold values was typically 0.6 dB.

Using the method described above, the lower limit of the audibility region was determined as a function of L_1, L_2, with $L_1=L_2$, and as a function of L_2, with L_1= 40 dB SL_0. The levels ranged up to 70 dB SL_0 in steps of 10 dB. In addition, the lowest levels L_1, L_2 ($L_1=L_2$) and L_2 (L_1= 40 dB SL_0) at which the combination tone could just be perceived were measured in steps of 1 dB. Measurements were made at 8 frequencies of f_1 ranging from 500 Hz up to 5600 Hz in steps of half an octave. In one daily session four conditions L_1, L_2 (either $L_1=L_2$ or L_1= 40 dB) were measured for all f_1 in a random order. Three trained subjects participated.

The lower limit of the audibility region, defined here relative to f_1, was derived from the relation $(2f_1-f_{2,max})/f_1$ where $f_{2,max}$ is the highest frequency of the component f_2 at which the combination tone could be perceived. The results of measurements of this lower limit by the three subjects are given in Fig. 1 for the conditions $L_1=L_2$. Each data point represents the average of 5 adjustments. The standard deviation of the mean was typically 1%. Sensation levels are given in dB above the average threshold at f_1. The SL_0 values were transformed into SL values for each subject individually. The corrections used can be deduced from the filled symbols in Fig.1 which represent for all f_1 the stimulus condition $L_1=L_2$= 40 dB SL_0.

The curves in Fig. 1 are limited at the left by the lowest stimulus level at which the combination tone $2f_1-f_2$ could be perceived. This level, averaged over frequencies and subjects, was 18.9 ± 5.3 dB. At this level the average lower limit of the audibility region relative to f_1 was 0.82. The lowest stimulus level at which the combination tone could be perceived for any one frequency was 13.0 dB at f_1= 1400 Hz (averaged over the three subjects). This level increased to 23.0 dB at f_1= 500 Hz and to 20.9 dB at f_1= 5600 Hz.

Fig. 1 shows that as SL is increased the audibility region expands less at higher frequencies than at lower frequencies. Also, it is apparent that the rate of increase of the audibility region with increasing SL is not always constant. For the middle frequencies particularly the audibility region increases more rapidly above 50- 60 dB SL. These two effects are more apparent in Fig. 2. Here are shown the data for all frequencies and all subjects. The lower limit of the audibility region is plotted logarithmically in this figure. Because of this logarithmic ordinate the large increase of the audibility region above 50 dB SL is more marked.

There is no apriori basis for predicting the change in audibility region with SL. But a closer look at Fig. 1 shows that the slopes of the curves tend to decrease for increasing SL up to 50- 60 dB. This suggested that up to this level a better linear relationship would be found for a logarithmic rather than a linear

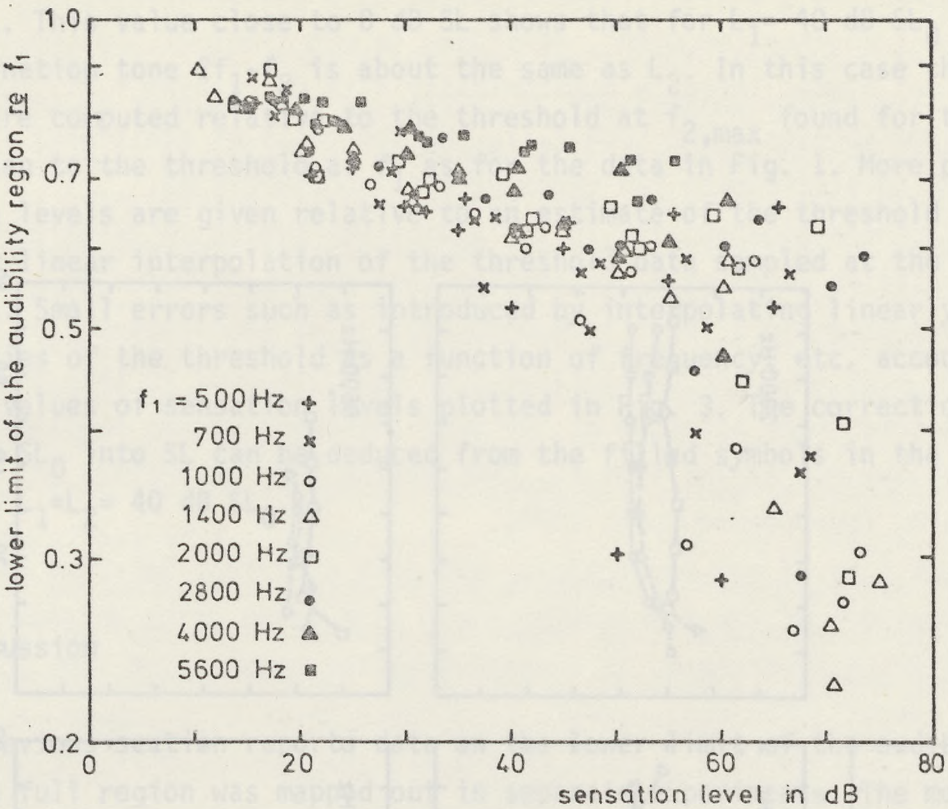


Fig. 2. Lower limit of the audibility region of the combination tone $2f_1 - f_2$ relative to f_1 as a function of sensation level ($L_1 = L_2$) collected for all f_1 and the three observers. The data are plotted on a logarithmic ordinate. The increase of the audibility region with SL is smaller for high f_1 (filled symbols) than for low f_1. Above 50 dB SL the audibility region increases often progressively.

ordinate. Indeed, linear regression analysis of the data up to 50 dB SL proved that this assumption was correct for all frequencies. The data are fit better by a logarithmic relation (constant fraction of octaves per dB) than by a linear one (constant frequency difference per dB). The slopes estimated from the regression analysis (in dB/oct, the inverted form) were: f_1 = 500 Hz, 43 dB/oct; 700 Hz, 38 dB/oct; 1000 Hz, 62 dB/oct; 1400 Hz, 67 dB/oct; 2000 Hz, 83 dB/oct; 2800 Hz, 84 dB/oct; 4000 Hz, 91 dB/oct; 5600 Hz, 169 dB/oct. The steeper slopes at higher frequencies reflect the decreasing expansion of the audibility region as a function of SL for higher frequencies.

The lower limit of the audibility region as a function of L_2 (L_1 = 40 dB SL_0) is plotted in Fig. 3. Again, each data point represents an average over 5 adjustments for a single subject. Fig. 3 shows that the lower limit of the audibility region is approximately constant for all L_2 > 20 dB SL, and tends to rise somewhat for L_2 < 20 dB SL. The results presented in Fig. 1 and Fig. 3 suggest that the audibility region depends mainly on L_1.

The lowest L_2 at which the combination tone could be perceived averaged over all frequency conditions f_1 = 500- 5600 Hz and over the three subjects was 4.2 ±

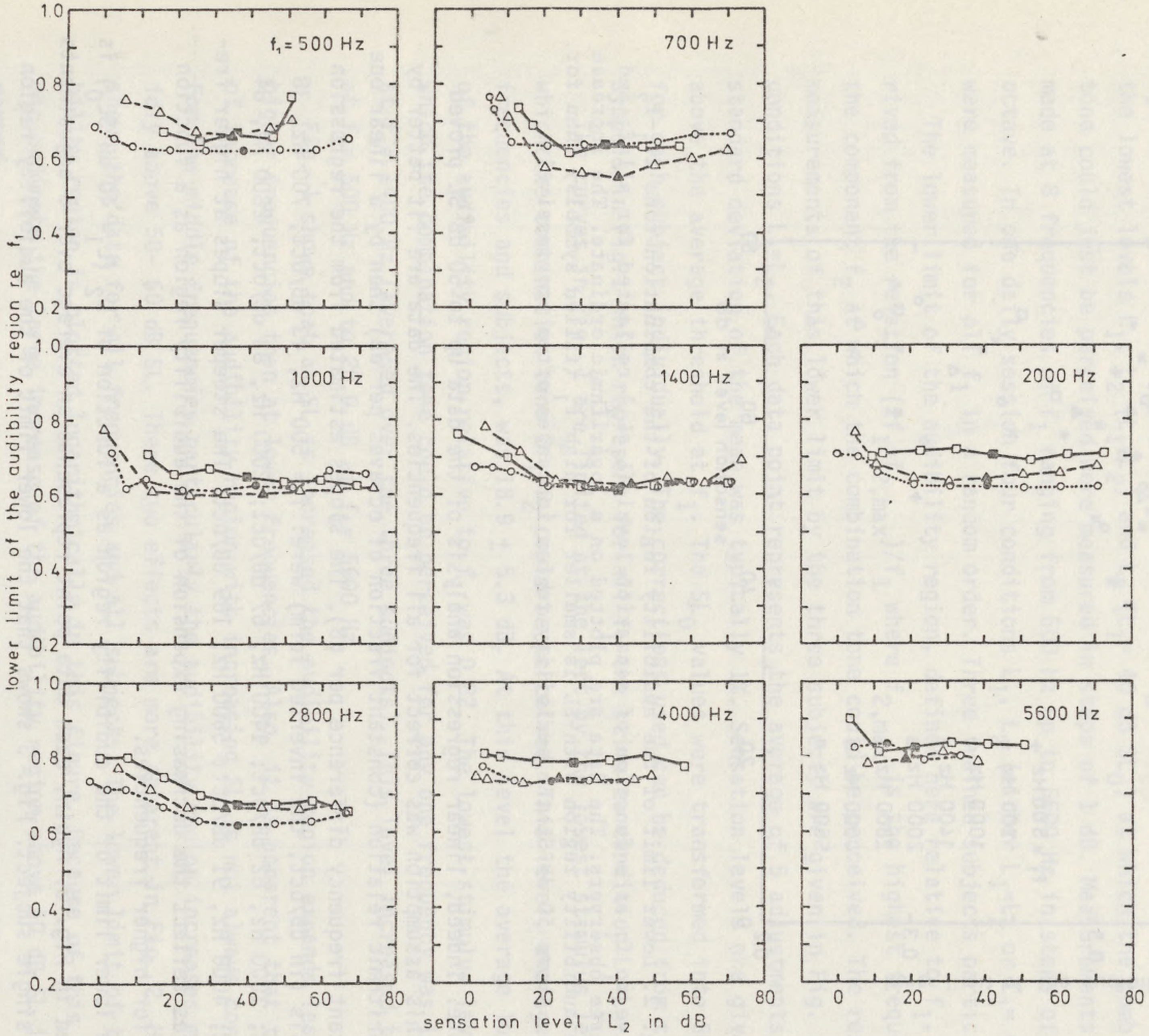


Fig. 3. Lower limit of the audibility region of the combination tone $2f_1-f_2$ relative to f_1 as a function of L_2 where $L_1 = 40$ dB SL_0. The sensation level L_2 is given relative to the threshold at that f_2 which corresponds to the data point at the lowest L_2 of each curve. The lowest L_2 approaches 0 dB SL. Note that 0 dB SL is off the left-hand boundary of each subfigure.

5.0 dB SL. This value close to 0 dB SL shows that for L_1= 40 dB SL_0 the level of the combination tone $2f_1-f_2$ is about the same as L_2. In this case the sensation levels were computed relative to the threshold at $f_{2,max}$ found for the lowest L_2 rather than to the threshold at f_1 as for the data in Fig. 1. More precisely, the sensation levels are given relative to an estimate of the threshold at $f_{2,max}$ obtained by linear interpolation of the threshold data sampled at the half octave intervals. Small errors such as introduced by interpolating linearly, by small irregularities of the threshold as a function of frequency, etc. account for the few negative values of sensation levels plotted in Fig. 3. The corrections applied to transform SL_0 into SL can be deduced from the filled symbols in the figure which represent L_1=L_2= 40 dB SL_0.

1.3. Discussion

The previous section reports data on the lower limit of the audibility region only. The full region was mapped out in separate experiments. The most important results of these experiments will be summarized here.

The combination tone $2f_1-f_2$ could be perceived from the lower limit of the audibility region up to about $0.9f_1$. The factor 0.9 was slightly subject dependent. It is conceivable that the fact that the combination tone was inaudible between $0.9f_1$ and f_1 was a result of the limited frequency resolution of the ear. This 10% frequency difference was also implied by Plomp's data on the ear's frequency resolving power (Plomp, 1964). The upper limit of the audibility region declined for L_2 low with respect to L_1. Masking of the combination tone by f_1 can account for this, since the combination tone level decreases with L_2 ($L_2<L_1$). This decrease of the combination tone level with L_2 was found by the cancellation method (Goldstein, 1967; Zwicker, 1968; Helle, 1969/1970) as well as by different techniques (Smoorenburg, 1972). At very high L_2 ($L_2>L_1$) the upper limit of the audibility region declined probably because of masking by f_2. It is not likely that a decrease of the combination tone level below absolute threshold causes either decline since at the same level of L_2 the combination tone is perceived for larger f_2/f_1 and as a rule the combination tone level increases with decreasing f_2/f_1.

The level of the combination tone in the audibility region may also be of some interest. Goldstein (1967, 1970) reported a monotonic increase of the cancellation level with combination tone frequency (decreasing f_2/f_1). He suggested that the combination tone level in dB simply increases linearly with decreasing f_2/f_1 from 0 dB at the lower limit of the audibility region up to L_1 (L_1=L_2) for f_2/f_1= 1. Zwicker (1968) also reported an increasing cancellation level for decreasing f_2/f_1. He related the increasing level for $L_1<L_2$ and $L_1>L_2$, respectively, to the low-frequency slope and the high-frequency slope of a presumed excitation pattern produced by one frequency component. According to our findings the combination tone level does not always increase monotonically with decreasing f_2/f_1. Occasionally, it drops below threshold for some f_2/f_1. On the basis of additional investigations we were able to distinguish two types of non-monotonicities.

One type of non-monotonicity in the audibility region is revealed by the fact that the sensation level of a combination tone may be affected by a small dip (a threshold elevation in a small frequency region) at the frequency of the combination tone or even that the combination tone may disappear in this dip. Higher order combination tones were affected at the same frequency of the combination tone, $f_1-k(f_2-f_1)$, and not at the same f_2. Combination tone generation was also affected by a dip at f_2 (see Smoorenburg, 1972).

The second type of non-monotonicity was more puzzling. Relative minima in the loudness functions of the combination tones were found which depended critically upon the stimulus conditions. In a case where the combination tone disappeared completely cancellation measurements as a function of L_1,L_2 showed a phase jump of the cancellation tone of π radians at the minimum (Smoorenburg, 1972). These results suggest that more than one nonlinearity contributes to the combination tone. If such were the case there is no reason to expect that the combination components produced by each nonlinearity would always add in-phase. For example, the individual contributions may cancel each other. Then a phase jump of π radians may be expected. The minima could be related to out-of-phase addition of contributions to the combination tone from different places along the cochlear partition. Similar sudden drops of the cancellation level as a function of L_1 and L_2 have been reported by Helle (1969/1970).

The large increases of the audibility region above 50 dB SL (Fig. 1) may also be the result of the action of an additional nonlinearity. A nonlinearity at a more peripheral stage would probably generate combination tones in a much less frequency selective manner. If such a nonlinearity becomes effective above 50 dB SL then it could account for the large increases of the audibility region found above this level. As mentioned before the distortion components in the acoustic signal were more than 75 dB down. Thus, equipment distortion could not be the basis of the second nonlinearity.

At the lowest stimulus level for which the combination tone $2f_1-f_2$ could just be perceived the average lower limit of the audibility region was $0.82f_1$ while the upper boundary was typically $0.9f_1$. A few detailed measurements confirmed that at the lowest stimulus level the combination tone was audible in a small frequency region below $0.9f_1$. Thus, the audibility of the combination tone extends over a region of frequencies if the sensation level is just above a minimum level. This may suggest that the combination tone is generated only if a certain threshold is exceeded. But masking of the combination tone by f_1 also accounts for this result. For decreasing f_2/f_1 the combination tone level increases but masking of the combination tone by f_1 increases too. Thus, if the combination tone is masked for some f_2/f_1 then it is likely to be masked in a frequency region around this value of f_2/f_1. If both masking and combination tone level were proportional to L_1 then it would not be the case that at low SL the combination tone is masked and at higher SL not masked. We found that masking of a component at the combination tone frequency increases less than proportionally at these low sensation levels. Also, we found that the initial rise of the cancellation tone level is occasionally more than proportional. If tone-on-tone masking and tone-on-combination-tone masking are comparable, then the threshold for the combination tone is probably determined by the masking effect of f_1.

The limited audibility region of the combination tone suggests that the level of the combination tone may depend on the amount of interaction of the cochlear stimulation patterns produced by the two stimulus components. Suppose the combination tone is generated by a nonlinearity operating uniformly, at each place in the cochlea, on the stimulation produced by f_1 and f_2. Suppose also that the important contributions to the build up of the combination tone come from those places where the stimulation by f_1 and f_2 is about equal. (This last supposition refers to studies of the combination tone level as a function of L_1,L_2; Goldstein, 1967; Smoorenburg, 1972). Thus, the same amount of overlap of the stimulation patterns would produce the same combination tone level. This means that for a constant combination tone level, the high-frequency skirt S_h of the stimulation pattern of f_1 and the low-frequency skirt S_l of f_2 intersect at a constant level. Thus, an increase of the audibility region with SL should follow a $1/S_h$ relation if L_1 is increased, a $1/S_l$ relation if L_2 is increased, and a $1/S_h+1/S_l= 1/S$ relation if both levels are increased, S in dB/oct.

The data of Figs. 1-3 can be compared with the previous predictions only if it is assumed that the threshold of audibility of the combination tone at the lower limit of the audibility region implies a constant combination tone level along this lower limit. Of course, this assumption neglects small frequency-dependent threshold differences. Also, this assumption does not allow for the possible masking effects of f_1 and f_2.

Estimates of S derived from the present measurements increased from 40 dB/oct at 500 Hz up to more than 100 dB/oct above 4000 Hz. But measurements of the movements of the basilar membrane by von Békésy (1960), by Johnstone et al. (1970a, 1970b), and by Rhode (1971) suggest that the shallow slope (S_h) of the stimulation pattern is only 6 or 12 dB/oct. It might be argued that these measurements of basilar membrane movements are not representative of the actual stimulation patterns because, for example, they are done at high levels. Masking experiments often reveal a steeper slope S_h at low levels than at high levels (Zwicker and Feldtkeller, 1967). In addition, the nonlinearity may operate at a more central stage than the stage of mechanical stimulation and the stimulation pattern at this higher stage may be narrower because of an intermediate sharpening process. Values of S derived from the psychophysical measurements reported by Zwicker and Feldtkeller (1967) are 30 dB/oct at 250 Hz, 50 dB/oct at 1000 Hz, and 45 dB/oct at 4000 Hz. There is no important difference between these values of S and our estimates at low frequencies. At higher frequencies, however, Zwicker and Feldtkeller do not estimate an increase in one of the slopes S_l or S_h whereas we do. On the other hand, Goldstein's (1967) cancellation data also suggest a smaller audibility region for higher frequencies. Also, the measurements of the movements of the basilar membrane with the Mössbauer technique (Johnstone et al., Rhode) and by direct observation (von Békésy) suggest that the mechanical filtering may improve at higher frequencies. Thus, the concept of a nonlinearity operating upon overlapping mechanical stimulation patterns seems suggestive but it is necessary to postulate an additional filtering of the combination tone in order to account for the increase in the audibility region with SL.

A good estimate of the slope of the low-frequency skirt S_l of the stimulation pattern is 100 dB/oct (Mössbauer measurements and psychophysical measurements). This slope would imply a decrease of the lower limit of the audibility region with L_2 of 30% over a 50 dB range. However, no clear decrease of the lower limit with L_2 was found (Fig. 3). This important inconsistency also suggests that the width of the audibility region cannot be explained simply on the basis of the amount of interaction of the stimulation patterns.

The stimulation patterns of f_1 and f_2 overlap at a place which is not maximally sensitive to the frequency of the combination tone but to frequencies between f_1 and f_2. If the combination tone is generated at this place of overlap then the additional filtering of the combination tone may be identified with the frequency selectivity at the place of overlap (cf Goldstein, 1967, 1970). The distinction between the two frequency selectivities will be discussed more extensively in a subsequent paper (Smoorenburg, 1972).

1.4. Interpersonal Differences

The lower limit of the audibility region appeared to be subject dependent. Therefore, we decided to determine the audibility region for a large group of subjects. The measuring procedure described above appeared to be well suited for this purpose. For all subjects the lower limit of the audibility region of the combination tone $2f_1-f_2$ was measured with f_1= 2000 Hz and both components at the same fixed sound pressure level (SPL). The SPL used was about 40 dB above "normal" threshold (i.e. 40 dB SL). Each subject made three adjustments on each of two different days. The measurements were preceded by a rough threshold determination. Forty subjects were used who had thresholds not more than about 10 dB above "normal" in the frequency range 1000- 3000 Hz.

The 6 adjustments were averaged for each subject. Table 1 gives the number of subjects having an audibility region (defined here as $(f_1-(2f_1-f_{2,max}))/f_1 = (f_{2,max}-f_1)/f_1$) between 0.00 and 0.05, 0.05 and 0.10, etc. The results show that the audibility region of the combination tone $2f_1-f_2$ for $f_1 = 2000$ Hz and $L_1 = L_2 =$ 40 dB SPL varies among subjects between 0.1 and 0.4.

Table I. The width of the audibility region of the combination tone $2f_1-f_2$ for 40 subjects. Stimulus conditions: $f_1 = 2000$ Hz and $L_1 = L_2 = 40$ dB SPL. The table gives the number of subjects out of the group of 40 for whom the width of the measured audibility region re f_1, $(f_1-(2f_1-f_{2,max}))/f_1$, lies between 0.00 and 0.05, 0.05 and 0.10, etc.

audibility region re f_1	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40
number of subjects	0	1	0	5	9	14	7	4	

These interpersonal differences cannot be explained by interpersonal threshold differences. Data presented in Fig. 1 show that the audibility region re f_1 decreases at most 0.1 between 40 dB and 30 dB SL, and the audibility region is not likely to be much smaller for a subject with a threshold 10 dB above "normal" than for a subject with a normal threshold and a stimulus level reduced by 10 dB. The following two arguments support the idea that the large interpersonal differences in audibility region are a result of large differences in the amount of interaction of stimulus components in the ear.

(1) There is a positive correlation within subjects between the width of the audibility region and the perceptibility of the low pitch which corresponds to the (absent) fundamental of a stimulus consisting of two high adjacent harmonics.

In a previous paper (Smoorenburg, 1970) we reported a disparity among subjects in their pitch judgments of two 200 msec tone bursts, one consisting of the components $f_1, f_2 = 1800, 2000$ Hz and the other of $f_1, f_2 = 1750, 2000$ Hz. When these two stimuli were presented successively about half of the 42 subjects perceived a pitch change corresponding to the change in fundamental frequency (200 Hz and 250 Hz, respectively). The other subjects perceived a pitch change corresponding to the change in the frequency of the lowest component (1800 Hz and 1750 Hz, respectively). The audibility region for $2f_1-f_2$ was measured for these same subjects. The average audibility region found for the group which had perceived the pitch change corresponding to the fundamental frequency change appeared to be significantly wider than the audibility region of the group which had perceived the pitch change corresponding to the frequency change of the lower component.

(2) Within subjects 'phase susceptibility' and the width of the audibility region appear to be positively correlated.

Plomp and Steeneken (1969) studied the effect of phase on the timbre of complex tones. They used stimuli consisting of 10 successive harmonics with a -6 dB/oct amplitude weighting. The harmonics were added in various phase relations. By means of the method of triadic comparisons the largest timbre difference between any two stimuli was found between a stimulus made up of sine (or cosine) components and one made up of alternating sine and cosine components. They also studied the magnitude of this timbre difference in proportion to the effect on timbre of changes of 1 dB/oct in amplitude weighting pattern. The ratio of the phase effect and the slope effect was clearly subject dependent. Assuming that this was largely due to a subject-dependent phase effect they defined 'phase susceptibility' as that change in slope which had an effect on timbre of the same magnitude as the change in phase. In Fig. 4 the phase susceptibility of 8 subjects (in dB/oct) is compared with the width of their audibility regions (re f_1). The audibility regions were measured again in the condition f_1= 2000 Hz and $L_1=L_2$= 40 dB SPL. The symbols used in Fig. 4 correspond to the symbols used by Plomp and Steeneken in Fig. 5 of their paper. The inverted triangle in Fig. 4 represents a subject with no measurable phase effect. The audibility regions were compared with the phase susceptibility for the 292 Hz fundamental condition because for this fundamental frequency the higher harmonics contributing most to the phase effect are situated at about 2000 Hz, which is the frequency for which the audibility region was determined. Fig. 4 shows clearly that subjects with a large audibility region (for $2f_1-f_2$) have a relatively high phase susceptibility.

A high phase susceptibility and a large audibility region may be related to a strong nonlinearity and/or to a low frequency-resolving power. A weaker nonlinearity and a higher frequency-resolving power would suggest a lower phase susceptibility and a smaller audibility region. The interpersonal differences in pitch perception of two-frequency stimuli are not yet quite understood, but in so far as they are determined by peripheral factors the perception of the low pitch may be attributable to relatively pronounced fused beats produced by the combined action of frequency components. On the other hand the perception of individual part-tones may be related to a relatively high resolving power in the subject's ear. The correlations found within subjects among the perceptibility of a low pitch in the com-

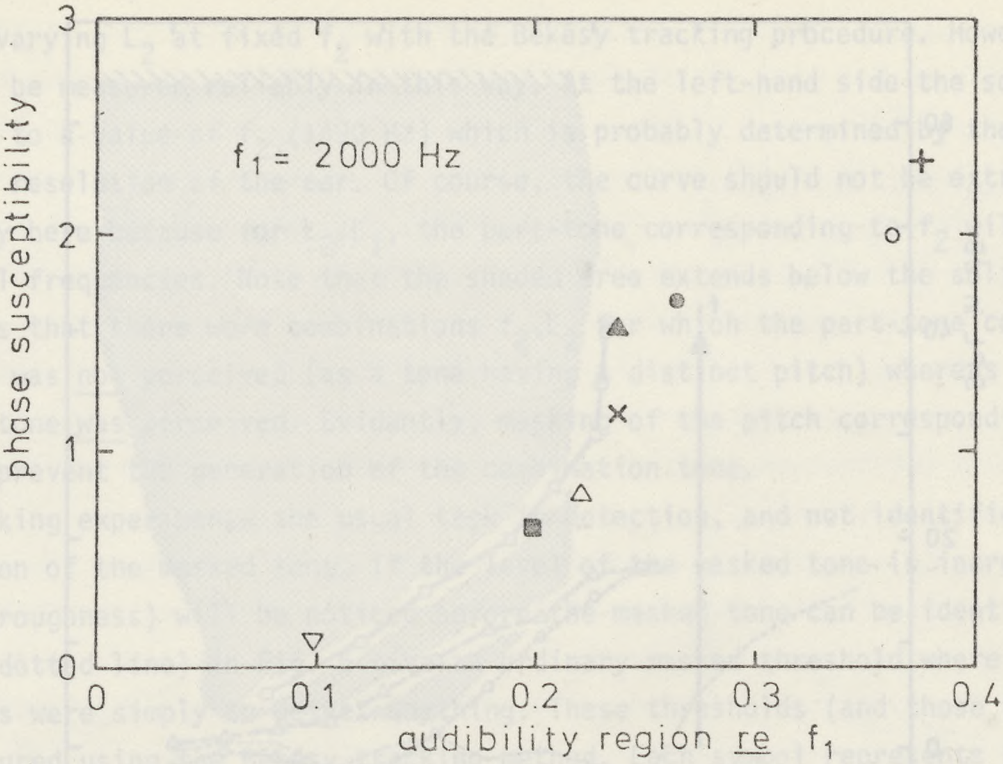


Fig. 4. A within subjects comparison of the width of the audibility region of the combination tone $2f_1-f_2$ (expressed relative to f_1; $(f_1-(2f_1-f_{2,max}))/f_1$) and the 'phase susceptibility' as reported by Plomp and Steeneken (1969). The phase susceptibility is expressed as that change in the slope of the amplitude pattern (in dB/oct) which produced a timbre difference of the same magnitude as changing the phase relation of the stimulus components from all cosine into an alternating sine and cosine relation.

plex tone, the phase susceptibility, and the width of the audibility region, support the view that the interpersonal differences are due to subject-dependent resolving power and/or a subject-dependent nonlinearity. The direct mutual relationships between the three phenomena are still unclear.

1.5. Pure Tone Masking and Combination Tones

The measurements of the audibility region have demonstrated that for $L_1 = 40$ dB SL the combination tone $2f_1-f_2$ is audible at very low levels of f_2. Because of the likely asymmetric stimulation pattern of f_1, f_2 may be masked by f_1 before the combination tone is masked. Thus, at least partly, a masked threshold may be determined by the detection of a combination tone. This was investigated in an experiment reported here.

In the first part of the experiment the audibility region for one combination tone was mapped out in detail. Fig. 5 shows the results. The shaded area gives the various f_2,L_2 combinations for which the combination tone $2f_1-f_2$ could be perceiv-

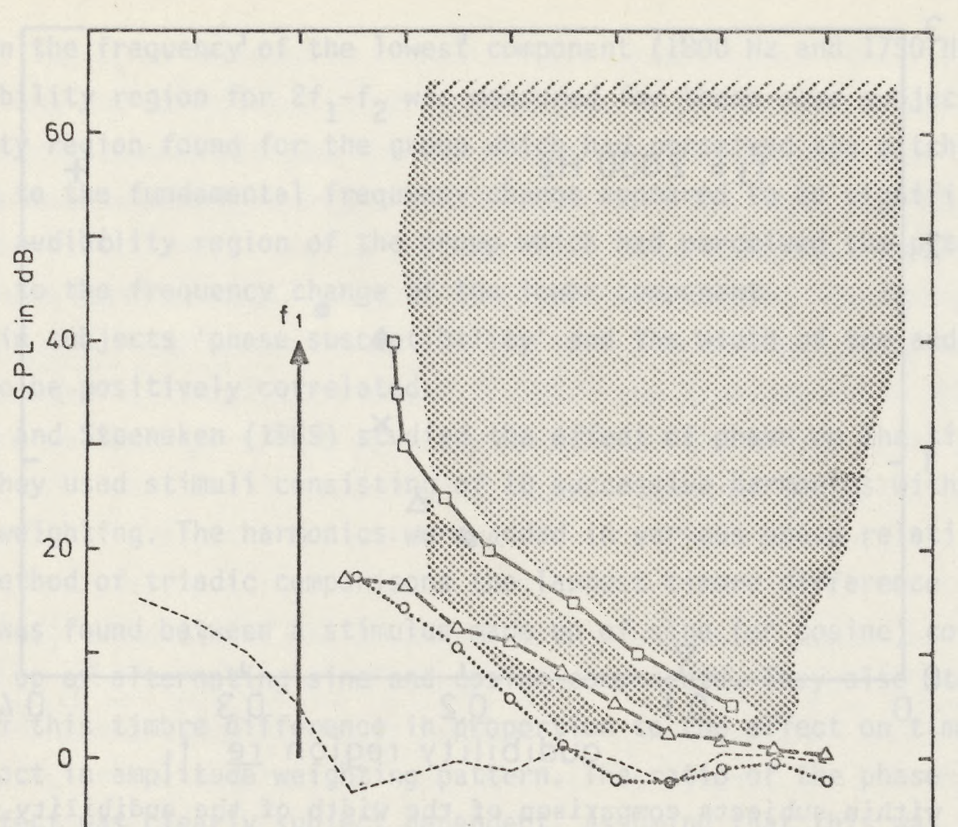


Fig. 5. Tone-on-tone masking in relation to combination tones. Masker is f_1, test tone is f_2. The shaded area gives the combinations f_2, L_2 for which the combination tone $2f_1-f_2$ could be perceived. Above the solid line connecting the squares the part-tone with a pitch corresponding to f_2 was perceived. The threshold of detection of anything and not in particular the pitch corresponding to f_2 is given by circles connected by dotted lines. Coincidence of this curve with the lower boundary of the shaded area suggests that the detection threshold is determined by detection of the combination tone. After addition of noise which masked the combination tones (masking level indicated by thin dashed line) an increase of the detection threshold to the dashed curve connecting triangles proved that the threshold of detection was determined by the detection of combination tones.

ed (f_1= 1400 Hz and L_1= 40 dB SL). The measuring procedure was the same as before. The boundaries of the audibility region were determined by 5 adjustments. The standard deviation of the mean was 6 Hz. The right-hand boundary, $f_{2,max}$, corresponds to the lower limit of the audibility region, $(2f_1-f_{2,max})/f_1$, as given in Fig. 3.

Next, we determined the region in which the part-tone corresponding to f_2 itself could be perceived. This was measured by fixing L_2 and decreasing f_2 until the <u>pitch</u> corresponding to f_2 disappeared. This amounts to a <u>threshold of identification</u> for f_2. For all f_2, L_2 combinations above this threshold a pitch corresponding to f_2 was heard. The squares (connected by the solid line) in Fig.5 represent the average of 5 adjustments of this sort; the standard deviation of the mean was 10 Hz. We also tried to measure the threshold of the pitch corresponding to f_2 by varying L_2 at fixed f_2 with the Békésy tracking procedure. However, it could not be measured reliably in this way. At the left-hand side the solid line converges to a value of f_2 (1490 Hz) which is probably determined by the limited frequency resolution of the ear. Of course, the curve should not be extrapolated vertically here because for $L_2 > L_1$, the part-tone corresponding to f_2 will be audible at all frequencies. Note that the shaded area extends below the solid curve. This means that there were combinations f_2, L_2 for which the part-tone corresponding to f_2 was not perceived (as a tone having a distinct pitch) whereas the combination tone was perceived. Evidently, masking of the pitch corresponding to f_2 does not prevent the generation of the combination tone.

In masking experiments the usual task is detection, and not identification or recognition of the masked tone. If the level of the masked tone is increased interference (roughness) will be noticed before the masked tone can be identified. The circles (dotted line) in Fig. 5 give an ordinary masked threshold where the instructions were simply to detect anything. These thresholds (and those to follow) were measured using the Békésy tracking method. Each symbol represents the average threshold obtained in 5 series of 50 trials. The standard deviation of the mean was typically 0.8 dB. Note that the detection threshold and the identification threshold are parallel and are separated by about 13.5 dB. Note also that the detection threshold coincides with the lower boundary of the shaded area. This lower boundary represents the lowest $L_2(f_2)$ at which the combination tone could be perceived. Thus, the detection threshold may be determined by the detection of the combination tone.

In a second masking experiment, directed specifically at this issue, low-pass noise was added to the tonal masker f_1. The cut-off frequency of the noise was just below f_1 and the noise was intense enough to mask any combination tones which might be produced in these stimulus conditions. With the added noise the threshold for f_2 was significantly higher (triangles and dashed line in Fig. 5). This could not have been a result of the masking effect of the noise at f_2. With the noise alone (f_1 absent) the threshold is given by the thin dashed line (no symbols) in Fig. 5. This result strongly suggests that combination tones (not necessarily only $2f_1-f_2$ but possibly also the higher order combination tones $f_1-k(f_2-f_1)$) have played an important part in this masking experiment.

The possible role of combination tones in masking was discussed previously in the literature (Wegel and Lane, 1924; Egan and Hake, 1950; Ehmer, 1959). But these discussions concern irregular high-frequency skirts of masking patterns measured at high masker levels. The high-frequency skirts often showed relative minima. These minima were explained by the difference tone being detected in stead of the test tone. Chistovich (1957) distinguished between the detection threshold (of anything) and the identification threshold for the masked tone, and reported that the irregular masking behavior was not found when the subjects were instructed to identify the masked tone. Zwicker and Feldtkeller (1967) also show that the masking irregularities disappear if the subject is instructed not to listen to the combination products. Recently, Greenwood (1970) has shown that the minima disappear if low-pass noise is added below the masker frequency in order to mask the combination tones. In contrast to these previous experiments this section deals with a low level masking experiment in which only monotonic masking patterns were found. Even in this case the results clearly demonstrate the influence of combination tones. Of course, our data are based on only one observer and only one condition. The role of combination tones in masking experiments should obviously be investigated more extensively.

It might be argued that the addition of the low-pass noise to the tonal masker increased the detection threshold because the noise prevented the detection of splattered energy introduced by the transients of the test tone. Detection of energy off the test tone's frequency is a nonneglegible phenomenon ("off-frequency listening"; Leshowitz and Wightman, 1971). However, the frequency splatter of the test tone used in the present experiment was sufficiently small to avoid off-frequency detection. But, in a general sense, off-frequency listening was responsible for the detection threshold since combination tones were detected. Both, frequency splatter and combination tones may account for off-frequency detection.

2. AUDIBILITY REGION OF $f_1-k(f_2-f_1)$, $k>1$

2.1. Experiment

In a separate experiment the lower limit of the audibility region for higher order combination tones was measured. The method was similar to that described in the previous section. Generally, the higher order combination tones became audible successively as f_2 was decreased. In sec. 1.3 it was reported that in some cases the combination tone $2f_1-f_2$ vanished and then reappeared as f_2 was decreased. Since such reappearance could easily be mistaken for the appearance of a higher order combination tone (such as $3f_1-2f_2$) every time any combination tone became audible, its pitch was matched with the pitch of a simple tone. This allowed us to determine the order of the combination tone (k) with some certainty.

The experiment was done with the same three subjects as before and at $L_1=L_2=$ 40 dB SL_0 for all frequencies. The results have been plotted in Fig. 6. Each data point gives the average of three adjustments made in different sessions. Data points connected by a dotted line to the lower order data point represent the combination tones that were perceived only twice. A combination tone perceived only

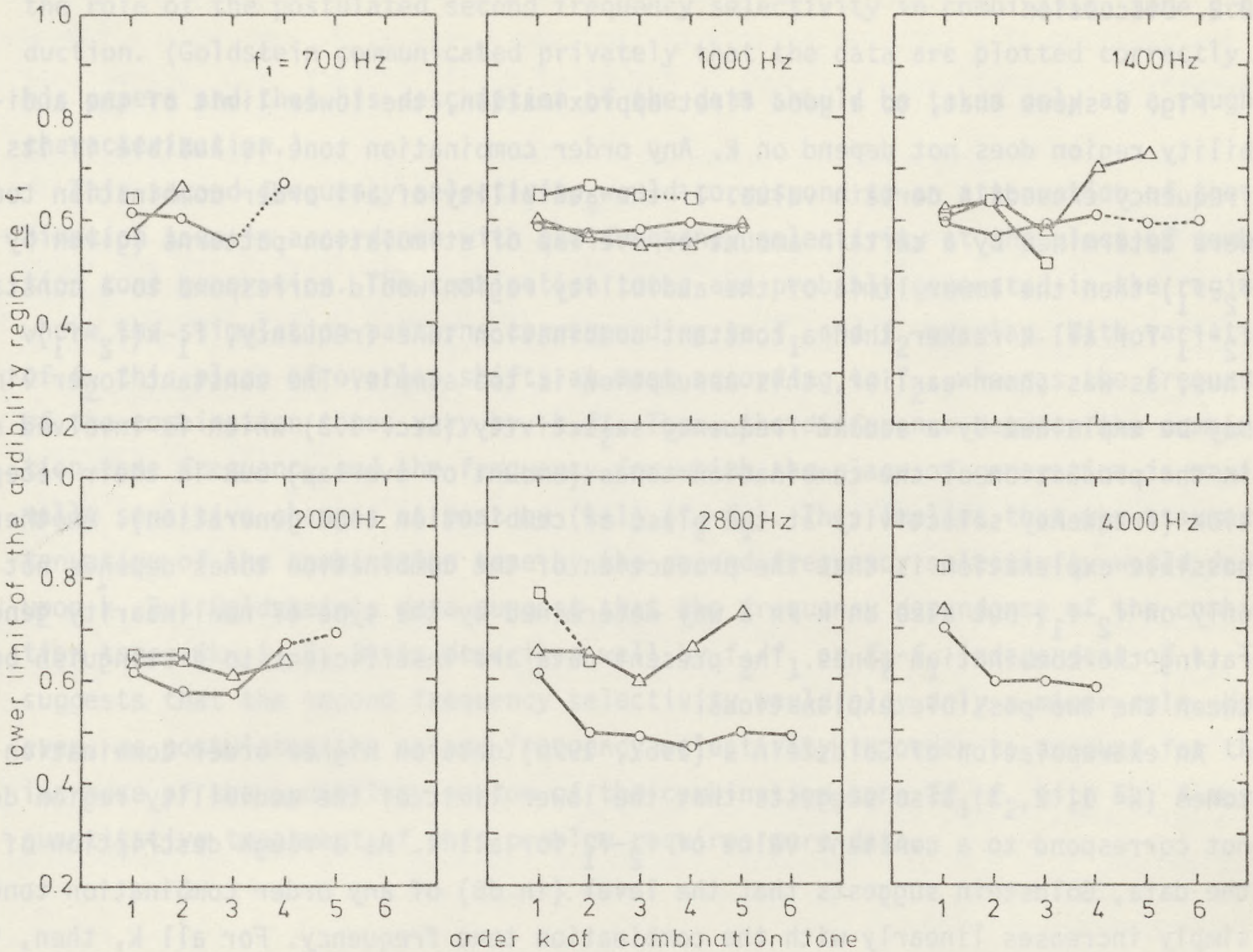


Fig. 6. Lower limit of the audibility region relative f_1, $(f_1-k(f_{2,max}-f_1))/f_1$, as a function of the order k of the combination tone at 40 dB SL_0. The symbols denoting subjects are the same as in Figs. 2 and 4. Each data point represents the average of three adjustments. If the combination tone was perceived only twice the corresponding data point was connected by a dotted line to the lower order data point.

once is not represented. This experiment was preceded by many pilot experiments. The results of these pilot experiments suggested that three adjustments in each condition would suffice. For f_1= 500 Hz only one subject (circles) perceived higher order combination tones (data not shown). In this case the lower limit re f_1 as a function of k was: k= 1, 0.61; k= 2, 0.61; k= 3 (perceived only twice), 0.62. For f_1= 5600 Hz no higher order combination tones were perceived. (At 5600 Hz, 40 dB SL_0 corresponds to about 23 dB SL.)

The higher order combination tones were not always perceived from the lower limit of the audibility region up to, say, $0.9f_1$. With increasing order their sensation levels decreased and became so low that small threshold differences might have easily affected the audibility of these combination tones.

2.2. Discussion

Fig. 6 shows that, to a good first approximation, the lower limit of the audibility region does not depend on k. Any order combination tone is audible if its frequency exceeds a certain value. If the audibility of all order combination tones were determined by a certain amount of overlap of stimulation patterns (given by f_2-f_1) then the lower limit of the audibility region would correspond to a constant f_2-f_1 for all k rather than a constant combination tone frequency, $f_1-k(f_2-f_1)$. Thus, as was shown earlier, this assumption is too simple. The constant lower limit may be explained by a second frequency selectivity (Sec. 1.3) which is involved not in the production of the combination tones (amount of overlap) but in their reception (frequency selectivity at the place of combination tone generation). Another possible explanation is that the production of the combination tones depends not only on f_2-f_1, but also on k in a way determined by the type of nonlinearity generating the combination tones. The present data are insufficient to distinguish between the two possible explanations.

An extrapolation of Goldstein's (1967, 1970) data on higher order combination tones (k= 1, 2, 3) also suggests that the lower limit of the audibility region does not correspond to a constant value of f_2-f_1 for all k. As a rough description of the data, Goldstein suggests that the level (in dB) of any order combination tone simply increases linearly with the combination tone frequency. For all k, then, the combination tone levels equal 0 dB SL at a certain combination tone frequency below f_1 and increase up to L_1 ($L_1=L_2$) at f_1. This implies that the lower limit of the audibility region is determined by the combination tone frequency and not by f_2-f_1. However, in the limit, this implies that the levels of an infinite number of combination tones converge to L_1 as f_2 approaches f_1. This is obviously absurd. Moreover, Goldstein's data suggest that as f_2 goes to f_1 the cancellation levels go to about L_1, L_2-20 dB, and L_1-32 dB for k= 1, 2, and 3, respectively. Thus, the cancellation level clearly decreases for increasing k. For increasing f_2 extrapolations to 0 dB cancellation level give f_2/f_1= 1.41, 1.26, and 1.18 for k= 1, 2, and 3. These values of f_2/f_1 correspond to combination tone frequencies re f_1 of 0.59, 0.48, and 0.46, respectively. The latter numbers are estimates of the lower limit of the audibility region based upon Goldstein's cancellation measurements. Together with the cancellation levels as a function of k for $f_2=f_1$, given above, they suggest that with increasing order there is less increase in the cancellation level with combination tone frequency. Indeed, Goldstein's data show that the frequency dependence of the cancellation level ($\Delta L(f_2)$) is described well by f_2/f_1 or f_2-f_1 ($\Delta L/(f_2-f_1)$= const) for k= 1, 2, and 3. Thus, the dependence of the cancellation level on combination tone frequency diminishes by about 1/k for higher order combination tones. This result has an important theoretical implication with respect to the role of the postulated second frequency selectivity in combination tone production. (Goldstein communicated privately that the data are plotted correctly in his papers and that his description of the data should be taken only as a rough characterization.)

This second frequency selectivity would correspond to an attenuation of the combination tone in accordance with the frequency selectivity at the place of combination tone generation. The combination tones are probably generated in the region where the stimulation patterns corresponding to f_1 and f_2 overlap. With variations of f_2 this place of overlap shifts at most according to f_2, whereas the frequencies of the combination tones vary by $-k.f_2$. Thus, the difference between the combination tone frequency and the frequency for which the place of generation is maximally sensitive changes at most by $(k+1).(f_2-f_1)$. This implies that the assumed attenuation of the combination tone by the second frequency selectivity would depend upon k. But Goldstein's data suggest that the frequency dependence of the combination tones (k= 1, 2, 3) is described well by f_2/f_1 or f_2-f_1 independent of k. This suggests that the second frequency selectivity would play only a minor role. However, we postulated the second frequency selectivity in order to account for the increase of the audibility region of the combination tone $2f_1-f_2$ with SL. A more quantitative treatment of this problem requires more data.

3. CONCLUSIONS

- The combination tone $2f_1-f_2$ is audible only in a restricted frequency region below f_1. The lower boundary of this region is determined primarily by L_1.

- The width of the audibility region differs significantly among subjects.

- The lowest stimulus level at which the combination tone $2f_1-f_2$ can be perceived is $L_1=L_2=$ 15-20 dB SL. The lowest level of the higher frequency component for which the combination tone can be perceived with $L_1=$ 40 dB SL is $L_2=$ 4 dB SL.

- The lower limit of the audibility region is approximately the same for all audible combination tones of the type $f_1-k(f_2-f_1)$ at $L_1=L_2=$ 40 dB SL_0. Higher order combination tones can be perceived up to k= 5 or 6.

- The restricted audibility region cannot be explained completely by the amount of interaction of the mechanical stimulation patterns corresponding to the stimulus components. The combination tones seem to be subject to an additional frequency selectivity.

• The detection threshold of a tone masked by another tone may be determined by the detection of a combination tone of the type $f_1-k(f_2-f_1)$, even with a low-level masker.

REFERENCES

von Békésy, G. (1960). *Experiments in Hearing* (Mc Graw-Hill Book Company Inc., New York), Ch. 11.

Chistovich, L.A. (1957), as quoted by Small, A.M. (1959). "Pure-Tone Masking", J. Acoust. Soc. Amer. *31*, 1619-1625.

Egan, J.P., and Hake, H.W. (1950). "On the Masking Pattern of a Simple Auditory Stimulus", J. Acoust. Soc. Amer. *22*, 622-630.

Ehmer, R.H. (1959). "Masking Patterns of Tones", J. Acoust. Soc. Amer. *31*, 1115-1120.

Goldstein, J.L. (1967). "Auditory Nonlinearity", J. Acoust. Soc. Amer. *41*, 676-689.

Goldstein, J.L. (1970). "Aural Combination Tones", In *Frequency Analysis and Periodicity Detection in Hearing*, R. Plomp and G.F. Smoorenburg, Eds. (Sijthoff, Leiden, The Netherlands), 230-247.

Goldstein, J.L., and Kiang, N.Y.S. (1968). "Neural Correlates of the Aural Combination Tone $2f_1-f_2$", Proc. IEEE *56*, 981-992.

Greenwood, D.D. (1970). Discussion Report in *Frequency Analysis and Periodicity Detection in Hearing*, R. Plomp and G.F. Smoorenburg, Eds. (Sijthoff, Leiden, The Netherlands), 436-444.

Helle, R. (1969/1970). "Amplitude und Phase des im Gehör gebildeten Differenztones dritter Ordnung", Acustica *22*, 74-87.

von Helmholtz, H.L.F. (1863). *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik* (F. Vieweg und Sohn, Braunschweig), 6th edition (1913), Chap. 7 and Appendix 12.

Johnstone, B.M., Taylor, K.J., and Boyle, A.J. (1970a). "Mechanics of the Guinea Pig Cochlea", J. Acoust. Soc. Amer. *47*, 504-509.

Johnstone, B.M., and Taylor, K.J. (1970b). "Mechanical Aspects of Cochlear Function", In *Frequency Analysis and Periodicity Detection in Hearing*, R. Plomp and G.F. Smoorenburg, Eds. (Sijthoff, Leiden, The Netherlands), 81-93.

Jones, A.T. (1935). "The Discovery of Difference Tones", The American Physics Teacher *3*, 49-51.

Leshowitz, B., and Wightman, F.L. (1971). "On-Frequency Masking with Continuous Sinusoids", J. Acoust. Soc. Amer. *49*, 1180-1190.

Ohm, G.S. (1843). "Ueber die Definition des Tones nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildenden Vorrichtungen", Ann. Phys. Chem. *59*, 513-565.

Plomp, R. (1964). "The Ear as a Frequency Analyzer", J. Acoust. Soc. Amer. *36*, 1628-1636.

Plomp, R. (1965). "Detectability Threshold for Combination Tones", J. Acoust. Soc. Amer. *37*, 1110-1123.

Plomp, R., and Steeneken, H.J.M. (1969). "Effect of Phase on the Timbre of Complex Tones", J. Acoust. Soc. Amer. *46*, 409-421.

Rhode, W.S. (1971). "Observations of the Vibration of the Basilar Membrane in

Squirrel Monkeys Using the Mössbauer Technique", J. Acoust. Soc. Amer. *49*, 1218-1231.

Smoorenburg, G.F. (1970). "Pitch Peception of Two-Frequency Stimuli", J. Acoust. Soc. Amer. *48*, 924-942.

Smoorenburg, G.F. (1972). "Combination Tones and Their Origin" , to be submitted to the J. Acoust. Soc. Amer.

Wegel, R.L., and Lane, C.E. (1924). "The Auditory Masking of One Pure Tone by Another and Its Probable Relation to the Dynamics of the Inner Ear", Phys. Rev. *23*, 266-285.

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

Zwicker, E. (1968). "Der kubische Differenzton und die Erregung des Gehörs", Acustica *20*, 206-209.

Zwicker, E., and Feldtkeller, R. (1967). *Das Ohr als Nachrichtenempfänger* (S. Hirzel Verlag, Stuttgart), 2nd revised ed., Ch. 3.

The detection threshold of a tone masked by another tone may be determined by the detection of a combination tone of the type $f_1-k(f_2-f_1)$, even with a low-level masker.

REFERENCES

von Békésy, G. (1960). *Experiments in Hearing* (McGraw-Hill Book Company Inc., New York), Ch. 11.

Chistovich, L.A. (1957), as quoted by Small, A.M. (1959). "Pure-Tone Masking", J. Acoust. Soc. Amer. 31, 1619-1625.

Egan, J.P., and Hake, H.W. (1950). "On the Masking Pattern of a Simple Auditory Stimulus", J. Acoust. Soc. Amer. 22, 622-630.

Ehmer, R.H. (1959). "Masking Patterns of Tones", J. Acoust. Soc. Amer. 31, 1115-1120.

Goldstein, J.L. (1967). "Auditory Nonlinearity", J. Acoust. Soc. Amer. 41, 676-689.

Goldstein, J.L. (1970). "Aural Combination Tones", In *Frequency Analysis and Periodicity Detection in Hearing*, R. Plomp and G.F. Smoorenburg, Eds. (Sijthoff, Leiden, The Netherlands), 230-247.

Goldstein, J.L., and Kiang, N.Y.S. (1968). "Neural Correlates of the Aural Combination Tone $2f_1-f_2$", Proc. IEEE 56, 981-992.

Greenwood, D.D. (1970). Discussion Report in *Frequency Analysis and Periodicity Detection in Hearing*, R. Plomp and G.F. Smoorenburg, Eds. (Sijthoff, Leiden, The Netherlands), 436-444.

Helle, R. (1969/1970). "Amplitude und Phase des im Gehör gebildeten Differenztones dritter Ordnung", Acustica 22, 74-87.

von Helmholtz, H.L.F. (1863). *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik* (F. Vieweg und Sohn, Braunschweig), 6th edition (1913), Chap. 7 and Appendix 12.

Johnstone, B.M., Taylor, K.J., and Boyle, A.J. (1970a). "Mechanics of the Guinea Pig Cochlea", J. Acoust. Soc. Amer. 47, 504-509.

Johnstone, B.M., and Taylor, K.J. (1970b). "Mechanical Aspects of Cochlear Function", in *Frequency Analysis and Periodicity Detection in Hearing*, R. Plomp and G.F. Smoorenburg, Eds. (Sijthoff, Leiden, The Netherlands), 81-93.

Jones, A.T. (1935). "The Discovery of Difference Tones", The American Physics Teacher 3, 49-51.

Leshowitz, B., and Wightman, F.L. (1971). "On-Frequency Masking with Continuous Sinusoids", J. Acoust. Soc. Amer. 49, 1180-1190.

Ohm, G.S. (1843). "Ueber die Definition des Tones nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen", Ann. Phys. Chem. 59, 513-565.

Plomp, R. (1964). "The Ear as a Frequency Analyzer", J. Acoust. Soc. Amer. 36, 1628-1636.

Plomp, R. (1965). "Detectability Threshold for Combination Tones", J. Acoust. Soc. Amer. 37, 1110-1123.

Plomp, R., and Steeneken, H.J.M. (1969). "Effect of Phase on the Timbre of Complex Tones", J. Acoust. Soc. Amer. 46, 409-421.

Rhode, W.S. (1971). "Observations of the Vibration of the Basilar Membrane in

COMBINATION TONES AND THEIR ORIGIN

INTRODUCTION

This paper reports the results of psychophysical experiments on the combination tone $2f_1-f_2$. Such a tone is often audible during stimulation by the two frequency components f_1 and f_2, $f_1<f_2$. The combination tone $2f_1-f_2$ is the most important member of a category of combination tones corresponding to $f_1-k(f_2-f_1)$ with k a positive integer. The main features of these combination tones are that they can be heard at low stimulus levels and that they are found only in a restricted frequency region below f_1, the 'audibility region of combination tones' (Smoorenburg, 1972). These properties indicate that the generation of these combination tones in the ear cannot be described by a simple frequency-independent, overloading type of nonlinearity. This in marked contrast with the difference tone, f_2-f_1, the level of which is rather independent of the stimulus frequencies and which is audible only at stimulus levels of 50 dB above threshold (Zwicker, 1955; Plomp, 1965; Goldstein, 1967). The combination tone $2f_1-f_2$ will be called the cubic difference tone, CDT, since a cubic term in a transfer function introduces a difference component $2f_1-f_2$.

The present research is directed at the type of nonlinearity generating the CDT and at the site of CDT generation. Four experiments are reported followed by a general discussion and some speculations. The first part describes how the audibility of the CDT is affected by a dip (a local elevation of the absolute threshold). The second part gives measurements of the CDT using the cancellation method. The level and phase of the cancellation tone (CT) are determined as a function of

---

This is a preliminary version of a paper to be submitted for publication in The Journal of the Acoustical Society of America.

stimulus levels and frequencies. In the last part the cancellation method is reconsidered. Cancellation measurements are compared with measurements of the effect of the CDT in gap masking (comparable with forward masking), loudness matches, and measurements of the CDT's 'pulsation threshold' (a type of loudness balance in which the stimulus f_1,f_2 and the reference tone $2f_1$-f_2 alternate at 4 Hz).

1. EFFECT OF A THRESHOLD ANOMALY ON CDT GENERATION

1.1. Introduction

In many respects the CDT behaves like a tone produced by a stimulus component $2f_1$-f_2, though this component is not actually included in the stimulus. The CDT sounds similar to a tone produced by a stimulus component, it beats with a stimulus component of slightly different frequency and its pitch can be described by the frequency relation $2f_1$-f_2. This suggests that the CDT is generated by a nonlinear stimulus transformation in the ear and that the distortion product is subsequently analyzed just like a stimulus component. However, the CDT is perceived only at small frequency differences f_2-f_1. This suggests that there also is a frequency selectivity which precedes the nonlinearity generating the CDT. Thus, the nonlinearity generates a component $2f_1$-f_2 only if the frequency difference is small enough that both stimulus components f_1 and f_2 are passed by the same filter. If, for large frequency differences, the stimulus components are separated into different channels the nonlinearity in each channel operates upon either f_1 or f_2. In this case, there is no channel in which f_1 and f_2 occur simultaneously at the input of the nonlinearity and, thus, combination components cannot be generated.

According to present knowledge, there is only one frequency analyzing element in the ear and that is the basilar membrane in the cochlea. Consequently, it appears contradictory to assume that there must be a frequency selectivity which precedes the nonlinearity as well as a frequency selectivity which succeeds the nonlinearity. The most plausible inference is that the frequency analysis and the nonlinearity cannot be separated. This would mean that the nonlinearity operates at the same stage of auditory processing as the frequency analysis and, thus, that the CDT is generated at a cochlear level.

Physiological experiments have frequently revealed cochlear nonlinearities. Von Békésy (1960) and Tonndorf (1970) observed fluid eddies in cochlear models. This nonlinear hydrodynamic phenomenon may account for several combination tones. However, the eddies occur only above a certain sound pressure level (SPL), and this level probably is greater than the stimulus level at which the CDT can be perceived already. Recently, Rhode (1971) reported a nonlinearity of the mechanics of the cochlear partition measured with the Mössbauer technique. The amplitude of the partition's vibration (for maximally effective frequencies) appeared to increase less than linearly with stimulus level. This nonlinearity may account for a combination component such as $2f_1-f_2$ but the nonlinearity has not yet been demonstrated for the low SPL at which the CDT is audible. The combination component $2f_1-f_2$ has been found also in cochlear microphonics (Newman et al., 1937; Wever et al., 1940a, 1940b). However, Dallos (1969, 1970) pointed out that the behavior of this microphonic component does not agree with Goldstein's (1967) psychophysical data on the CDT. Finally, Goldstein and Kiang (1968) have shown that acoustical nerve fibers may fire synchronously with $2f_1-f_2$ during acoustic stimulation by f_1 and f_2. Quantitatively, these data closely resemble Goldstein's psychophysical data. The presence of the CDT in the firing of primary nerve fibers suggests that the CDT is generated in the cochlea.

1.2. Aim and Method

The introduction indicates that, most probably, the CDT is generated in the cochlea. But the precise site of CDT generation is not known. We studied this problem by means of a psychophysical experiment with a subject who had a threshold elevation in a narrow frequency region (a dip in the audiogram). Such a threshold elevation is probably due to a defect in tonotopic processing at some stage of the auditory pathway. If the nonlinearity generating the CDT acts after this defect, then the defect might prevent a stimulus component from reaching the nonlinearity and consequently no CDT would be generated. On the other hand, if the nonlinearity would precede the defect and if the defect does not include the CDT frequency, then even for stimulus frequencies in the region of the threshold elevation no effect of the threshold anomaly on the CDT is expected. In that case the combination component would bypass the defect. These considerations suggested the present experiment in which the effect of a threshold anomaly on the CDT was investigated.

The audibility of the CDT was measured as a function of the frequency f_2. The frequency f_1 and the levels L_1 and L_2 were fixed. The procedure was similar to one used in previous experiments (Smoorenburg, 1972). Each measurement was started with $f_2=2f_1$ (CDT at zero frequency). Then f_2 was decreased to f_1 and the range of frequencies of f_2 for which the subject perceived the CDT was recorded. The CDT was identified by means of pitch matchings with a simple tone.

The stimuli were produced by two free-running sine-wave generators. After attenuation the two sine waves were mixed and the result was presented monaurally through a headphone (Beyer DT 48 S). Distortion components in the acoustical signal were more than 75 dB down. The subject and the experimenter were seated together in a sound-insulated room.

1.3. Experiment and Results

First, the hearing threshold was measured with the Békésy tracking procedure. This was repeated at the end of the experiment. The threshold values were sampled at 50 Hz intervals. An average threshold curve, fitted by eye to the data points, is given in Fig. 1 (solid line). The data points showed a spread of 1 or 2 dB around this curve. The stimulus levels are given relative to the subject's threshold "normal" at 1000 Hz. This is indicated by SL_0.

Next recruitment and diplacusis were measured. A complete recruitment (horizontal isophones) was found within 20 dB above the top of the elevated threshold and a diplacusis was found which is described by a pitch shift corresponding to -60 Hz at the steep skirt (between 2200 Hz and 2400 Hz) and to +105 Hz at the shallow

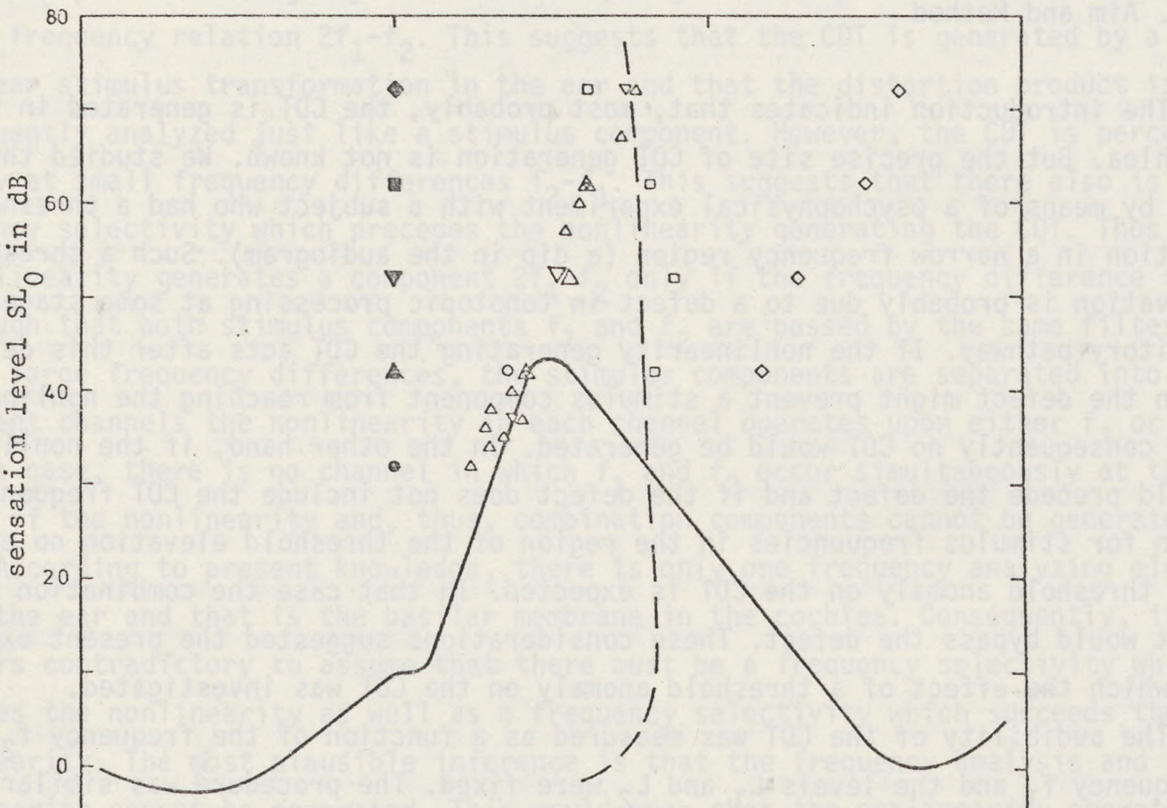


Fig. 1. Audibility of the CDT in relation to an anomalous threshold. The solid curve represents the absolute threshold. At a certain level of f_1 given by the filled symbols, the CDT could be heard if f_2 was in the frequency region below the same unfilled symbol and above about 2200 Hz. The dashed line indicates the upper limit of this region (L_1 = 47 dB SL_0) for the subject's normal ear. The standard deviation per adjustment was typically 25 Hz. Small symbols represent the average of 1-3 adjustments, large symbols the average of 4-7 adjustments.

skirt (between 2800 Hz and 3300 Hz). Both pitch shifts were rather level-independent. These were familiar results for dips.

After the measurements of the threshold anomaly we studied the effect of the threshold elevation on the generation of the CDT. The stimulus frequencies were chosen in such a way that only the stimulus components were in the region of the threshold anomaly and not the frequency component corresponding to the CDT. This determined our choice of f_1=2000 Hz. Five levels L_1 of f_1 were used ranging from 32 dB SL_0 to 72 dB SL_0 at 10 dB intervals. These conditions are indicated in Fig. 1 by the different filled symbols. The audibility of the CDT as a function of f_2 was measured for various L_2 at each L_1. The results of the adjustments to the highest f_2 for which the CDT could be perceived at a certain L_2 are given by the unfilled symbols in Fig. 1. The type of symbol indicates L_1 in that condition. The measurements were completed in four sessions. The number of adjustments was not equal for all conditions. Small symbols in Fig. 1 represent the average of 1-3 adjustments, large symbols the average of 4-7 adjustments. The standard deviation of the individual adjustments was of the order of 25 Hz.

The range of f_2 for which the CDT was perceived extended from about 2200 Hz to the unfilled symbols. At frequency differences smaller than f_2-f_1= 200 Hz the CDT was not heard, probably because of the imperfect frequency resolution of the ear. For L_1= 32, 62, and 72 dB SL_0(circles, squares, and diamonds respectively) L_2 was set at 10 dB intervals. Absence of a data point indicates that the CDT was not perceived at that L_2. For L_1= 42,52 dB smaller intervals of L_2 were used. In this case, the lowest symbol (triangle or inverted triangle) gives, within a few dB, the lowest L_2 at which the CDT was perceived. In a few conditions the CDT was not audible for some f_2 between 2200 Hz and the unfilled symbols: At L_1,L_2= 62,52 dB SL_0 and L_1,L_2= 62,62 dB SL_0 the CDT was not heard for f_2 in a small frequency region above the top of the threshold curve. At L_1,L_2= 62,42 dB SL_0 and L_1,L_2= 72,42 dB SL_0 the CDT was perceived only for about f_2>2700 Hz.

The audibility of the CDT as a function of f_2,L_2 was also measured for the subject's other ear. For this apparently normal ear the highest f_2 as a function of L_2 for which the CDT could be perceived (L_1= 47 dB SL_0) is given by the dashed line in Fig. 1. This result shows that the highest frequency is almost independent of L_2. This agrees with previous findings (Smoorenburg, 1972).

1.4. Discussion

The results presented in Fig. 1 show clearly the effect of the threshold anomaly. The CDT was perceived only when the higher stimulus component (f_2) exceeded the elevated threshold. The f_2,L_2 region for which the CDT was audible differed considerably from the region measured in the normal ear. Additional results on the combination tones $f_1-k(f_2-f_1)$ with k= 2,3 showed that these tones as well were perceived only if f_2 exceeded the elevated threshold. The same was found for f_1= 1600, 1700, and 1800 Hz at different levels. These results are not represented because the influence of the elevated threshold is less marked in these cases.

The results described above suggest that the defect which causes the elevated threshold precedes the nonlinearity generating the CDT. Thus, stimulus components below the elevated threshold do not reach the nonlinearity and, consequently, no combination components are generated. It is possible to extend this interpretation of the results. It might be assumed that the elevated threshold is caused simply by a complete lack of stimulus processing between about 2000 Hz and 3600 Hz. Further, it might be assumed that the slopes of the stimulation pattern (in dB/oct) produced by one frequency component are approximately constant. The threshold pattern then reflects the stimulation pattern. All stimulus components at the left of the steep skirt of this pattern and of the skirt's linear extrapolation would stimulate the low-frequency side of the gap and all stimulus components at the right side of the shallow skirt and its extrapolation would stimulate the high-frequency side of the gap (Fig. 2). Stimulus components in the region below the elevated threshold would not stimulate either of the two sides. According to this assumption f_1 at L_1= 32, 42, and 52 dB SL_0 would stimulate only the low-frequency side of the gap. Combination components could be assumed to be generated only if both stimulus components reach the same nonlinear channel. Thus, the CDT would be generated only if f_2 also stimulates the low-frequency side of the gap. This implies that the CDT would be

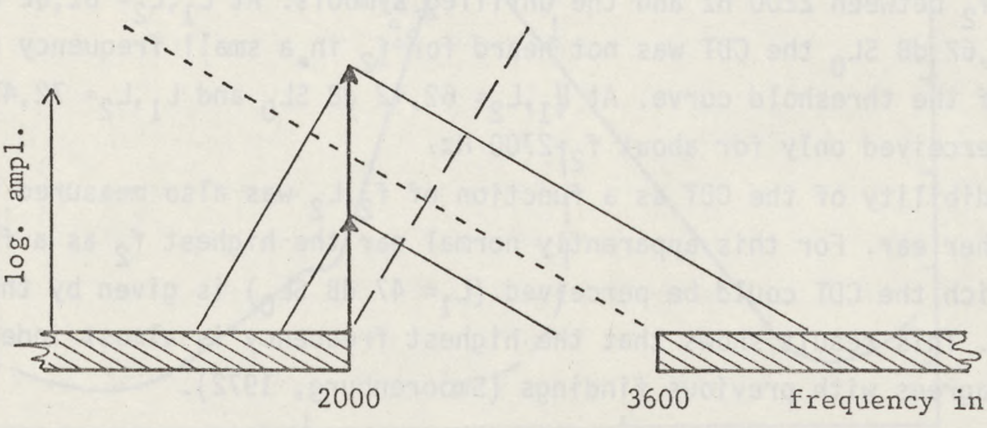


Fig. 2. The proposed influence of a gap in tonotopic stimulus processing by the ear. The slopes of the stimulation pattern produced by one frequency component are assumed to be constant. Stimulus components at the left side of the dashed line would stimulate only the low-frequency side of the gap. Stimulus components at the right side of the dotted line would stimulate only the high-frequency side of the gap. Thus, a stimulus component at 2000 Hz and at a moderate level would stimulate only the low-frequency side of the gap whereas the same stimulus component at a high level would stimulate both sides.

audible only for frequencies of f_2 below the steep skirt and its extrapolation. In this way the experimental results are explained. At L_1= 62 and 72 dB SL_0 f_1 would stimulate also the high-frequency side of the gap. In that case the CDT may be audible also for higher frequencies of f_2. This again is in agreement with the experimental results. (Since the shallow slope of the stimulation pattern seems to decrease for increasing stimulus levels (Zwicker and Feldtkeller, 1967) the shallow slope of the threshold curve should really not be extrapolated linearly but it should decline for higher levels. Therefore, it may be assumed that the component f_1 at L_1= 62 dB SL_0 (filled square in Fig. 1) stimulates the high-frequency side of the gap although the square lies just below the linear extrapolation of the shallow skirt of the threshold curve.)

Having deduced that the defect must precede the nonlinearity, we may speculate on the site of the nonlinearity. Goldstein (1967) distinguished between nonlinear basilar-membrane mechanics and inherently linear basilar-membrane mechanics subject to a nonlinear hair-cell coupling of basilar membrane and tectorial membrane. The threshold anomaly is most plausibly not caused by a defect of cochlear hydrodynamics or basilar-membrane mechanics but by a defect at a higher stage of auditory processing. This assumption is supported by the fact that the threshold elevation is very frequency selective. There is histological evidence that threshold anomalies of this sort are caused by a sensorineural deficiency. An example of such a defect is broken stereocilia on the hair-cells. Thus, we might speculate that the nonlinearity resides at a hair-cell or higher level. According to Hoogland (1953) the generation of the difference tone, f_2-f_1, is not affected by a threshold elevation at the frequencies of the stimulus components. Inaudible stimulus components may produce a difference tone at normal sensation levels. Hoogland's result supports the hypothesis that the origin of the difference tone is more peripheral than the origin of the CDT and the higher order combination tones.

Our conclusion that the nonlinearity resides at a stage of auditory processing higher than cochlear hydrodynamics or basilar-membrane mechanics raises the question of what frequency-resolving mechanism analyzes the products of the nonlinearity (see also the introduction to this section). The only known frequency analyzer is the basilar membrane. This question might be answered by Goldstein's hypothesis of nonlinear hair-cell coupling. This implies that the distortion components are fed back to the level of the basilar membrane and thus subsequently spectrally resolved. This idea was examined in an additional experiment in which the effect of the anomalous threshold on the CDT itself was investigated. The result was clear. Normally audible CDTs were not audible if their frequency was in the region of the elevated threshold. For sensation levels below 70 dB the CDT was not perceived for f_1= 4000 Hz or lower frequencies but the CDT was perceived for f_1= 4500 Hz (the next sample) and higher frequencies. This finding that the anomalous threshold affects also the CDT and the results of the first experiment which suggested that the combination component is generated after the gap support the view that there is a bidirectional coupling between the frequency-resolving mechanism and the nonlinearity. However, this finding does not give conclusive evidence for the assumption of bidirectional coupling. It may be argued that there is a second frequency-resolving mechanism at a higher stage of the auditory pathway which resolves the CDT normally but which is defective in this anomalous case.

Since the results of the present experiment are based upon only one subject the conclusions must be tentative. Similar effects were observed for other subjects but in those cases systematic studies have not yet been accomplished.

2. CDT LEVEL AND PHASE BEHAVIOR

2.1. Introduction

The CDT is audible only for small frequency differences of the two stimulus components f_1 and f_2. In a previous paper this frequency selectivity was studied by measuring as a function of the stimulus levels L_1,L_2 the range of frequency differences for which the CDT was perceived (Smoorenburg, 1972). In this section the frequency selectivity is investigated in more detail. This is done by cancelling the CDT with an externally generated combination component $2f_1-f_2$ of appropriate level and phase. The externally generated combination component is called the cancellation tone, CT.

The cancellation method has been applied before to study the frequency selectivity. Goldstein (1967) and Zwicker (1968) measured the cancellation level and phase as a function of the frequency difference of the stimulus components. In our measurements L_2 was varied with several fixed frequency differences. The reason we chose this approach was as follows: In the previous section it was pointed out that two frequency selectivities may be distinguished in the generation of the CDT. One frequency selectivity is involved in the <u>generation</u> of the CDT, and may be related to the amount of interaction of the stimulation patterns produced by f_1 and f_2. The second frequency selectivity is involved in the <u>reception</u> of the produced combination component. The region of overlap of the stimulation patterns in which the combination component is probably generated is maximally sensitive to frequencies between f_1 and f_2. As the difference between this frequency region and the combination frequency increases the combination component might be increasingly attenuated. The two successive frequency selectivities cannot be separated when cancellation measurements are made as a function of f_2-f_1. For frequency variations the combination component would depend on both frequency selectivities.

2.2. Apparatus

In order to cancel the CDT a stimulus component $2f_1-f_2$ (CT) was produced with adjustable level and phase. The CT was derived from the stimulus components by means of analog multipliers. First, the component f_1 was squared. This gives the frequency component $2f_1$ plus a DC component. The DC component was removed by AC coupling of this output to a second multiplier. The stimulus component f_2 was connected to the other input of this second multiplier. Thus, the frequency components $2f_1-f_2$ and $2f_1+f_2$ were obtained. The component $2f_1+f_2$ could be easily removed by subsequent low-pass filtering. The remaining component was led successively through an adjustable phase shifter and an adjustable attenuator. Finally, it was mixed with the individually attenuated stimulus components f_1 and f_2. The resultant three-component signal was presented monaurally through a headphone (Beyer DT 48 S). Acoustic distortion components were sufficiently down.

The adjusted CT level was measured simply by means of an electronic AC voltmeter. Much attention was paid to a correct measurement of the adjusted phase. The phase of the CT should be specified with respect to the stimulus components. As the phase shift of the analog multipliers was negligible, it was sufficient to measure the phase of the CT with respect to the output of the second multiplier. This was done by comparing on an oscilloscope screen the superposition of $2f_1-f_2$ and $2f_1+f_2$ at the output of the second multiplier with the position of the waveform of the component $2f_1-f_2$ tapped just before it was mixed with the stimulus components. Of course, the phase shift could have been measured in a simpler way if the component $2f_1+f_2$ had been removed first by filtering. However, we preferred our method since the filtering may introduce phase drifts which then would not be included in the phase measurement.

2.3. Method and Procedure

2.3.1. *Cancellation*

The subject was seated in a sound-insulated room. He was able to adjust the level and the phase of the CT simultaneously with two hands. His task was simply to remove the pitch of the CDT. In order to help the subject identify the CDT he could switch off the stimulus components f_1 and f_2. Then the CT was perceived alone; this gave the subject a cue for the pitch of the CDT. The CT could also be switched off. At the point of cancellation this made the CDT reappear.

Cancellation measurements are accurate. Assuming linear superposition of the CDT and the CT, a CDT of 20 dB above threshold would be inaudible for CT levels between 19.1 and 20.8 dB. For a CDT level of 40 dB these limits are 39.9 and 40.1 dB. However, the accuracy decreases markedly at very low levels. The cancellation level for a CDT at 6 dB above threshold could lie between 0 and 9.5 dB. (These limits show that adjustments with a logarithmic attenuator will give too low estimates of the CDT level.) We used a more accurate method for measuring very low CDT levels.

2.3.2. *Mirroring*

The inaccuracy of cancellations at low CDT levels means that loudness matches may be more accurate. For low levels this suggested to change the cancellation procedure into a loudness match. However, we discarded matching the loudness of a simple tone of frequency $2f_1-f_2$ to the loudness of the CDT. Such a match may be biased by the absence of the components f_1, f_2 in the matching stimulus. Therefore, the following method was chosen.

The subject started with an ordinary cancellation. The phase was adjusted as accurately as possible, for example, by minimizing the CDT level as a function of the CT's phase with a CT level somewhat lower than the cancellation level. Then the CT level was raised above the cancellation level and adjusted such that the loudness of the CDT plus the CT was equal to the loudness of the CDT alone.

The loudness match implies that the amplitude of the CDT produced by f_1 and f_2 alone, a_{CDT}, equals the amplitude of the CDT plus the CT, $a_{CDT,CT}$. If linear superposition holds at these low levels, then $a_{CDT,CT} = a_{CDT} - a_{CT}$. (The phase of the CT and the CDT are opposite.) Thus, $-a_{CDT} = a_{CDT} - a_{CT}$ or $a_{CT} = 2a_{CDT}$. This means that, in case of matched loudnesses, the level of the CT is 6 dB higher than the cancellation level at which $a_{CT} = a_{CDT}$. Since the loudness match implies opposite phases of the matched components it is called the mirroring method. In the transition region where both methods were used (CDT levels of 10 to 20 dB above threshold) the results differed consistently by 6 dB. Thus, we decided to use the mirroring method for CDT levels below 10 to 20 dB above threshold. The resulting data points were appropriately corrected by 6 dB.

2.4. Experiments and Results

In all experiments the cancellation level and phase were measured as a function of L_2. The levels are expressed relative to the absolute threshold at f_1, and the phases are given relative to the phase at which the CT is produced by the multipliers. (This phase with respect to the stimulus components $+\cos(2\pi f_1 t+\phi_1)$ and

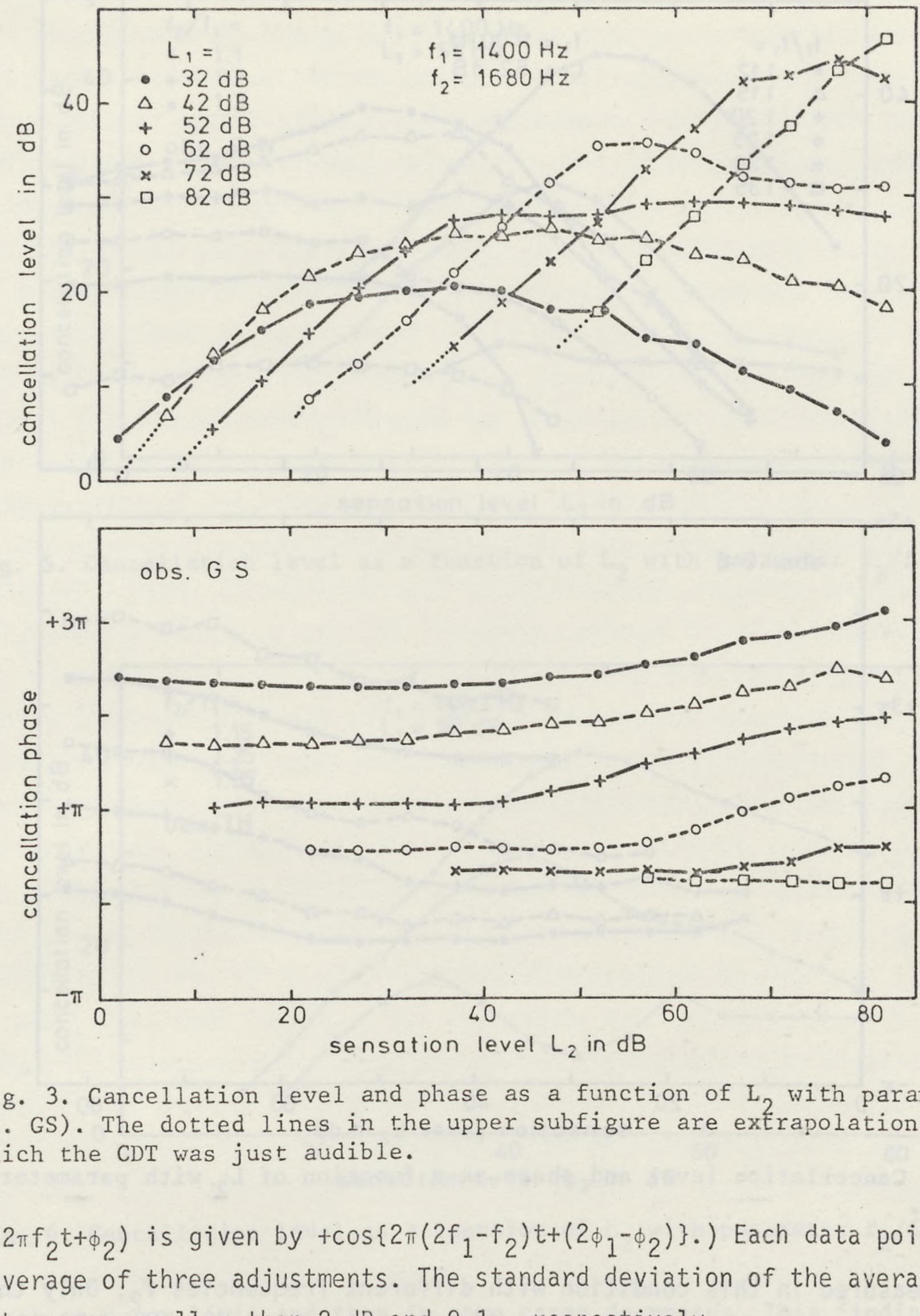


Fig. 3. Cancellation level and phase as a function of L_2 with parameter L_1 (subj. GS). The dotted lines in the upper subfigure are extrapolations to that L_2 at which the CDT was just audible.

$+\cos(2\pi f_2 t+\phi_2)$ is given by $+\cos\{2\pi(2f_1-f_2)t+(2\phi_1-\phi_2)\}$.) Each data point represents the average of three adjustments. The standard deviation of the averages of level and phase are smaller than 2 dB and 0.1π, respectively.

Figure 3 represents the results from one subject for fixed stimulus frequencies f_1= 1400 Hz and f_2= $1.2f_1$= 1680 Hz. L_1 is the parameter. The dotted lines in the upper subfigure are extrapolations of the curves to that L_2 at which the CDT could just be perceived. The results show that a CDT at the threshold of audibility does not always imply a cancellation level of 0 dB. Masking of the CDT by f_1 and f_2 may account for this.

Figure 4 represents the same subject's results for a fixed lower stimulus component f_1= 1400 Hz and L_1= 52 dB SL. In this case f_2 is the parameter. Two other

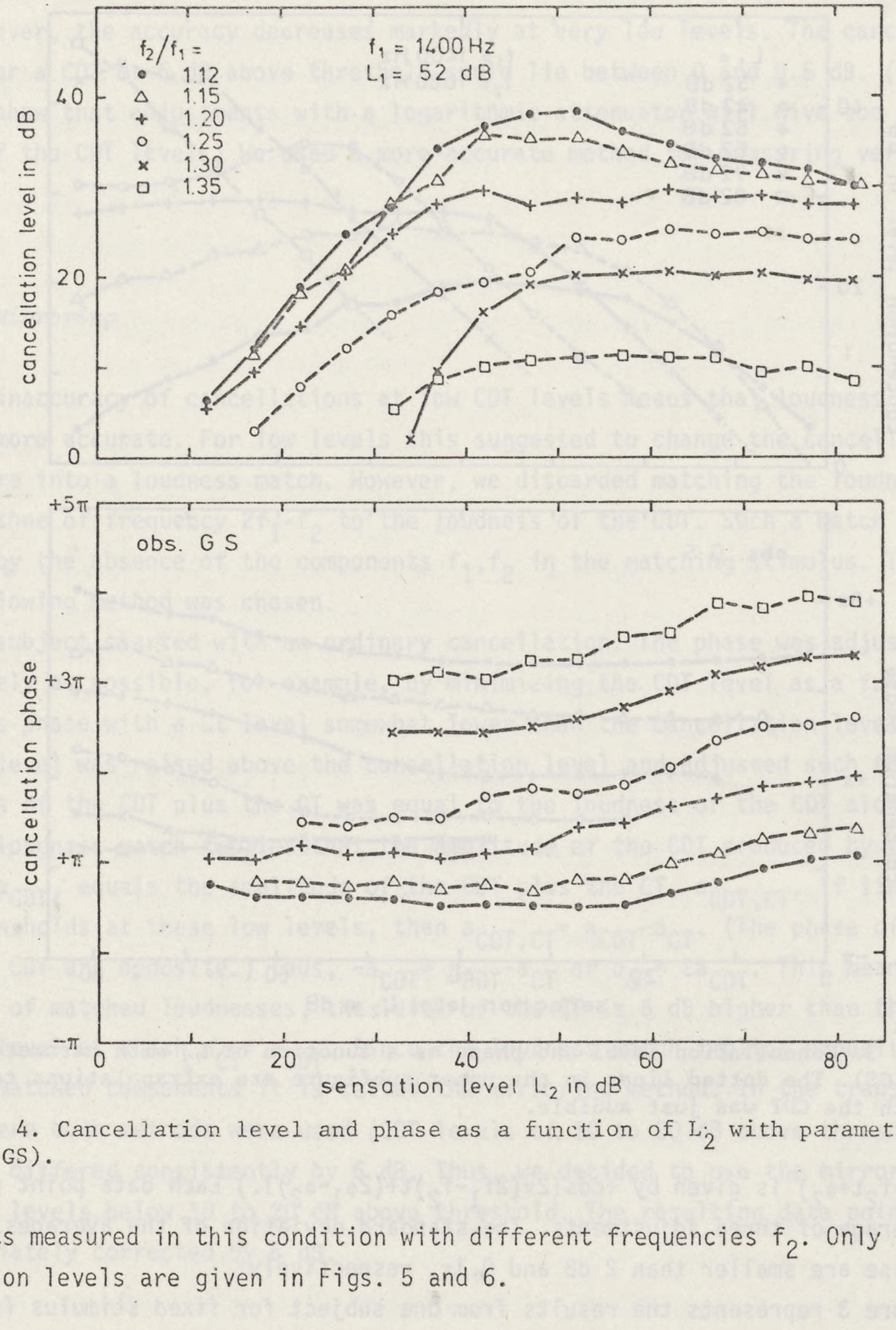


Fig. 4. Cancellation level and phase as a function of L_2 with parameter f_2/f_1 (subj. GS).

subjects measured in this condition with different frequencies f_2. Only the cancellation levels are given in Figs. 5 and 6.

2.5. Discussion

2.5.1. *The frequency selectivity in CDT generation*

Figures 3 through 6 show an initial increase of the cancellation level which is proportional to L_2. With a continued increase of L_2 the cancellation level appears

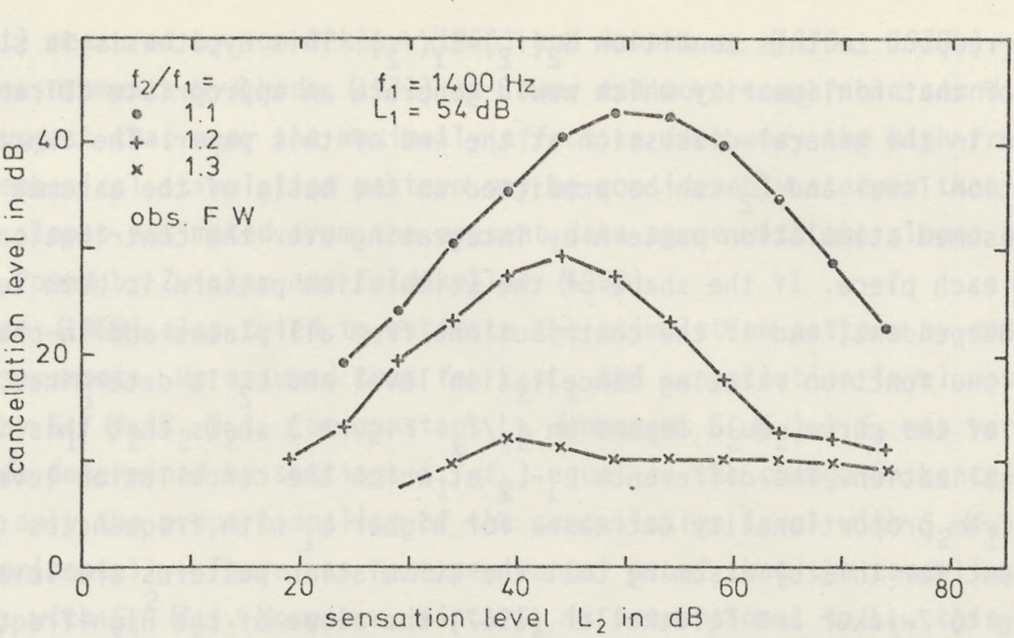


Fig. 5. Cancellation level as a function of L_2 with parameter f_2/f_1 (subj. FW).

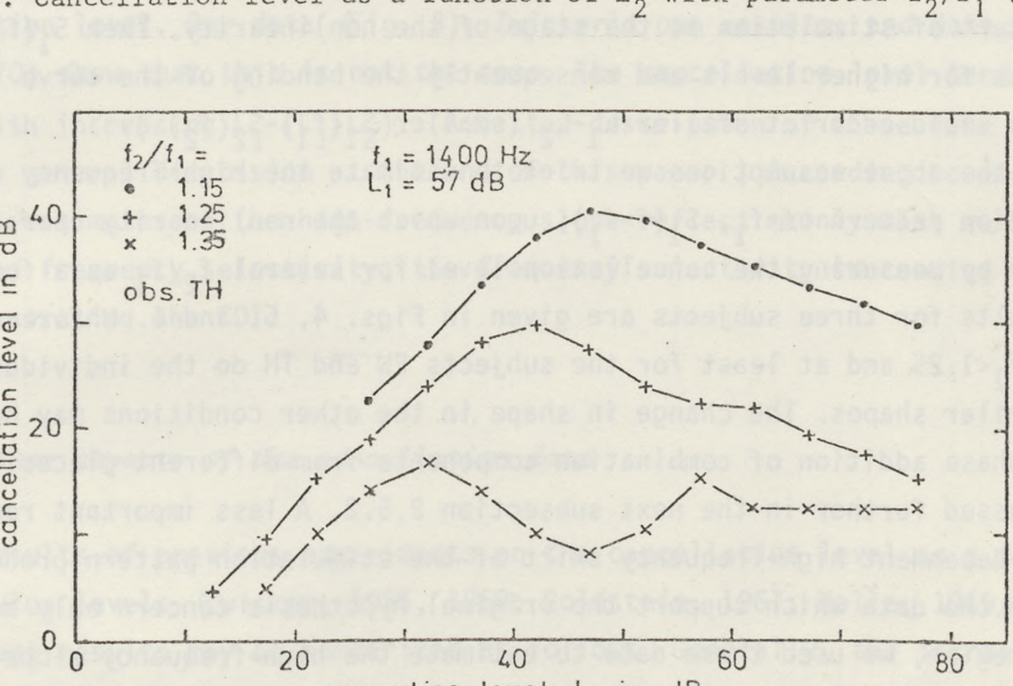


Fig. 6. Cancellation level as a function of L_2 with parameter f_2/f_1 (subj. TH).

to reach an upper limit and then in some cases decreases. This finding agrees with results reported by Goldstein (1967), Zwicker (1968), and Helle (1969/1970).

Figure 3 shows that the cancellation level increases proportionally to L_2 until L_2 approaches L_1. This result may be interpreted as follows. The proportional increase is found for $L_2<L_1$. For $L_2<L_1$ we might hypothesize that the stimulus components interact most at a place maximally sensitive to f_2. Then, as L_2 is raised, a critical point will be reached if the stimulation level of f_2, $S_2(f_2)$, equals the stimulation level of f_1 at the place of f_2, $S_1(f_2)$. (Note that the parenthesized frequency stands for the place of stimulation.) The bending of the curves

might correspond to this condition $S_2(f_2) \approx S_1(f_2)$. This hypothesis is supported by studies of that nonlinearity which would generate an appropriate CDT and which are presented in the general discussion at the end of this paper. The function relating cancellation level and L_2 can be predicted on the basis of the assumed nonlinearity and an assumed stimulation pattern by integrating over the contributions to the CDT from each place. If the shape of the stimulation pattern is both level and frequency independent, and if the contributions from all places add in-phase, then the shape of the function relating cancellation level and L_2 is determined. Only the position of the curve would depend on f_2/f_1. Figure 3 shows that this is too much of an idealization. The difference L_1-L_2 at which the cancellation level behavior departs from proportionality decreases for higher L_1 with frequencies fixed. We can account for this by assuming that the stimulation patterns are level dependent. According to Zwicker and Feldtkeller (1967) the slope of the high-frequency skirt of masking patterns decreases with increasing masker levels. The same may hold for the pattern of stimulation at the stage of the nonlinearity. Then $S_1(f_1)$- $S_1(f_2)$ decreases for higher levels and consequently the bending of the curve ($S_2(f_2) \approx S_1(f_2)$) would occur at smaller L_1-L_2 (smaller $S_1(f_1)$-$S_2(f_2)$).

With the above assumptions we tried to estimate the high-frequency skirt of the stimulation pattern of f_1, $S_1(f>f_1)$, upon which the nonlinearity operates. This was done by measuring the cancellation level for several f_2/f_1 as a function of L_2. The results for three subjects are given in Figs. 4, 5, and 6. Unfortunately, only for $f_2/f_1<1.25$ and at least for the subjects FW and TH do the individual curves have similar shapes. The change in shape in the other conditions may be due to out-of-phase addition of combination components from different places. This will be discussed further in the next subsection 2.5.2. A less important reason may be a level-dependent high-frequency skirt of the stimulation pattern produced by f_2. Although the data which support the original hypothesis concern only a small frequency region, we used these data to estimate the high-frequency slope. The bending of the curves measured by subject FW (TH) for f_2/f_1= 1.1 (1.15) and f_2/f_1= 1.2 (1.25) occurs at different L_2. The difference $\Delta L_2(f_2=1.1f_1, f_2=1.2f_1)$ equals $S_1(1.1f_1)$-$S_1(1.2f_1)$ since $L_2=S_2(f_2) \approx S_1(f_2)$. The difference ΔL_2 was determined by parallel shifting of one curve such that the best fit of the two curves was obtained. The result was $\Delta L_2(1.1f_1, 1.2f_1)$= 7.3 dB (subj. FW) and $\Delta L_2(1.15f_1, 1.25f_2)$= 7.7 dB (subj. TH). (The amount of shift in the vertical direction, 14 dB for subj. FW and 11.7 dB for subj. TH, was not used since this shift is determined also by the second frequency selectivity succeeding the nonlinearity.) Thus, $S_1(1.1f_1)$-$S_1(1.2f_1)$= 7.3 dB (subj. FW) and $S_1(1.15f_1)$-$S_1(1.25f_1)$= 7.7 dB (subj. TH). This gives estimates of the high-frequency slope $S_1(f>f_1)$ of 59 dB/oct and 64 dB/oct, for the two subjects respectively. These values are considerably higher than the high-frequency slopes of the stimulation patterns of the cochlear partition (6 to

12 dB/oct) measured by von Bêkésy (1960) and Johnstone (1970a, 1970b). However, recent measurements by Rhode (1971) at lower and more comparable sound pressure levels suggest that near the maximally effective frequency the high-frequency slope of the mechanical stimulation pattern may be considerably steeper than 6 to 12 dB/oct. The slopes estimated from the present data agree with the slopes of masking patterns found by Zwicker and Feldtkeller (1967).

Zwicker (1968) also tried to estimate the stimulation pattern by combination tone measurements. He assumed that for $L_2<L_1$ the cancellation level would be described by $S_1^2(f_2)S_2(f_2)$. For constant L_2 (constant $S_2(f_2)$) f_2 was varied and $S_1(f_2)$ was determined by settings $S_1^2(f_2)$ equal to the change in cancellation level. However, only the proportionality of the cancellation level with $S_2(f_2)$ was verified experimentally. There was no experimental proof that the cancellation level increases with $S_1^2(f_2)$. Moreover, since S_1 is proportional to L_1, this implies that an increase of L_1 should be followed by a two-fold increase (in dB) of the cancellation level. Our data (Fig. 3), Zwicker's own results, and data from Helle (1969/1970) show that this is not the case. The cancellation level tends to decrease with increasing L_1 ($L_1>L_2$). (A better description of these data is presented in the general discussion at the end of this paper.) According to our concept Zwicker's estimates of the high-frequency slope reflect mainly what we have called the second frequency selectivity, the frequency selectivity succeeding the nonlinearity generating the CDT.

2.5.2. *Other aspects of the cancellation data*

The results of previous experiments on the cancellation level as a function of the stimulus levels (Zwicker, 1955, 1968; Goldstein, 1967; Helle, 1969/1970) and the present results may be summarized as follows. Generally, the increase of the cancellation level with stimulus level (L_1=const+L_2) is nearly proportional at low and moderate stimulus levels and less than proportional at high levels (>70 dB SL). However, with $f_2/f_1>1.2$ the increase is often much less than proportional for various levels.

An example from our results of a nonproportional increase of the cancellation level with L_1,L_2 are the data in Fig. 3. These data for the condition $L_1=L_2$ are replotted in Fig. 7 (plusses). Further studies of the cancellation level as a function of L_1,L_2 ($L_1=L_2$) revealed a highly irregular behavior at $f_2/f_1=1.3$ (crosses in Fig. 7). At a certain stimulus level the CDT disappeared completely. In the region of disappearance of the CDT there was a rapid phase shift of π radians within 1 dB. This is represented in the lower graph of Fig. 7. Because of the obvious ambiguity of additive multiples of 2π in phase measurements, the sign

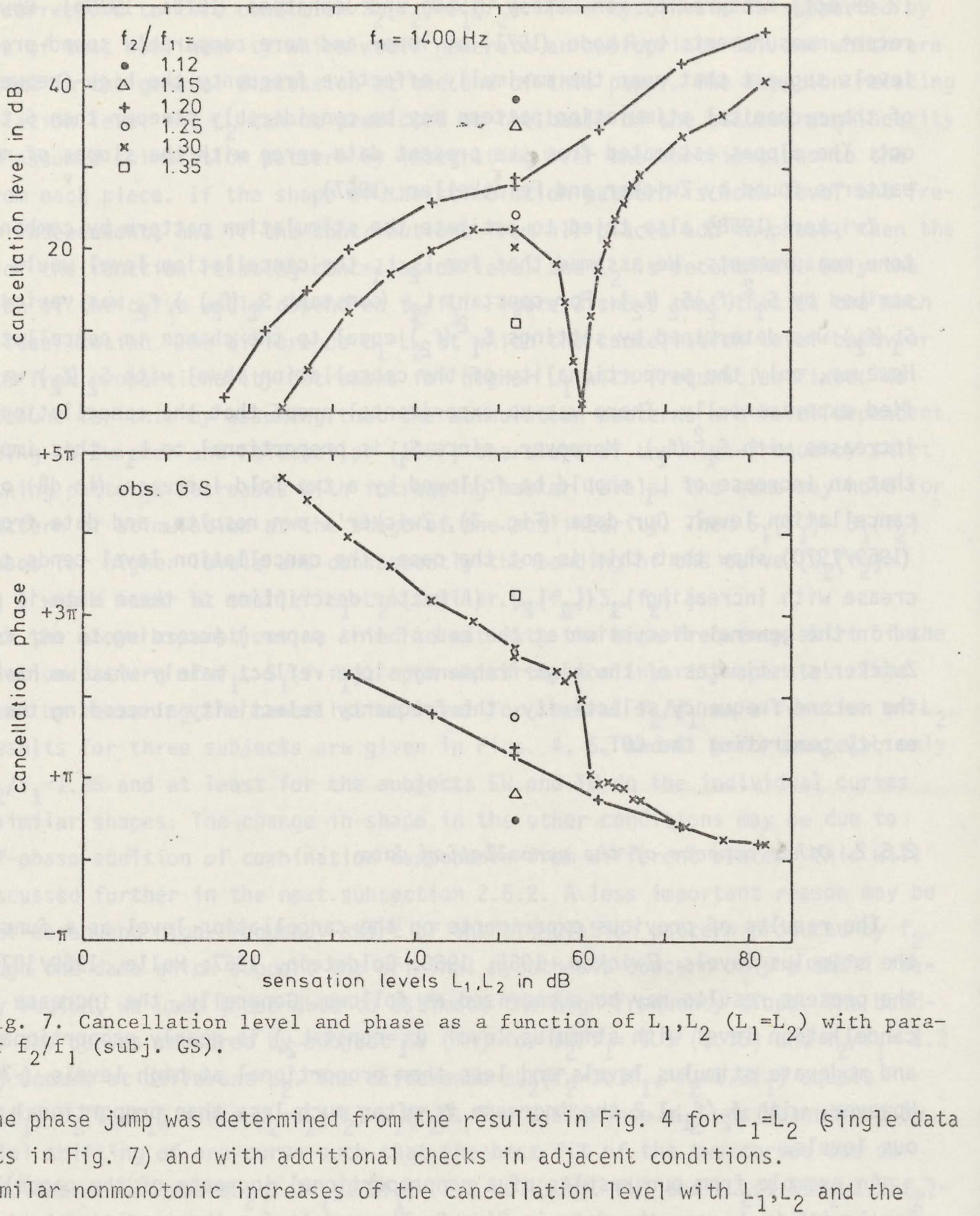


Fig. 7. Cancellation level and phase as a function of L_1, L_2 ($L_1=L_2$) with parameter f_2/f_1 (subj. GS).

of the phase jump was determined from the results in Fig. 4 for $L_1=L_2$ (single data points in Fig. 7) and with additional checks in adjacent conditions.

Similar nonmonotonic increases of the cancellation level with L_1, L_2 and the accompanying rapid phase shifts, were observed with three other subjects, FW and TH included. Helle (1970) has also described this phenomenon.

The irregular cancellation behavior for $f_2/f_1>1.2$ may be explained if we assume that combination components generated at adjacent places along the cochlear partition do not add in-phase. The disappearance of the CDT in a particular stimulus condition would thus imply that the contributions from all places cancel each other. In such a case the vector sum goes most probably through zero. Then the polarity of the vector sum changes. The jump of the cancellation phase by π radians reflects this polarity change. Because of the possibility of out-of-phase addition one should be careful in deducing the frequency selectivity involved in CDT generation directly from the behavior of the cancellation phase with changes of f_2/f_1 (Schroeder, 1969).

The cancellation phase as a function of L_1, L_2 depends on f_2/f_1. This suggests that probably this phase behavior cannot be explained by one particular type of nonlinearity. This is true even though certain nonlinearities may account for a polarity change, and a nonlinearity which includes hysteresis may account for some continuous phase shifts. Thus, the cancellation phase must be determined mainly by the properties of the filters involved in the generation of the CDT. The dependence of the cancellation phase on stimulus level implies that the characteristics of this filter are level dependent. The decreasing influence of f_2/f_1 on the cancellation phase at higher stimulus levels suggests that the filter is broader at these higher levels. This also is suggested by masking patterns which show a decreasing high-frequency slope at high levels (Zwicker and Feldtkeller, 1967). The apparent broadening of the filter may be caused by an intimate coupling between the filter and the nonlinearity which generates the CDT. The idea of a broadening filter is supported by recent Mössbauer measurements of the mechanics of the cochlear partition (Rhode, 1971). It is puzzling that Goldstein and Kiang (1968) did not find a phase shift with stimulus level for nerve discharges time-locked to the CDT frequency (see also Goldstein, 1971).

As the stimulus level increases, the cancellation phase seems to converge to one value independent of f_2/f_1 (Fig. 7). This suggests that the phase shifts introduced by the filters involved in the CDT generation become small. In that case the phase limit, increased by π radians because of opposite polarity of CT and CDT, gives the phase at which the CDT is generated by the nonlinearity. Common (single-valued) nonlinearities generate combination components only with phases of 0 or π radians. The results in Fig. 7 suggest that the cancellation phase converges to zero, the phase at which the cancellation tone was produced by the analog multipliers. Thus, the CDT is generated with a negative sign, $-\cos(2\omega_1-\omega_2)t$, with stimulus components $+\cos\omega_1 t$ and $+\cos\omega_2 t$ (cf. Sec. 2.4). The same result was reported by Goldstein (1967) and Schroeder (1969). They measured the cancellation phase as a function of f_2/f_1 and extrapolated the data to $f_2=f_1$ assuming a quadratic increase of the cancellation phase with f_2-f_1. Similar extrapolations of our data from Fig. 4 for L_1, L_2= 52,27; 52,52; and 52,77 dB SL resulted in estimates of the cancellation phase of 0.36π; 0.19π; and 0.60π, respectively. The data did not quite follow the quadratic relation.

3. CDT MEASUREMENTS WITH DIFFERENT TECHNIQUES

3.1. On the Cancellation Method

Essentially, one assumes linear superposition of the externally generated cancellation tone (CT) and the internally generated CDT if at the point of cancellation the CT level is identified with the CDT level. However, the high cancellation levels found experimentally for the CDT indicate that the CDT must be generated by an important nonlinearity of the ear. Therefore, the equalization of the cancellation level to the CDT level is not obvious. The addition of the CT and the CDT was studied in the following experiment.

The experiment was started by finding the cancellation phase. At this phase the cancellation tone was added with various levels to the stimulus f_1, f_2. A simple tone of frequency $2f_1 - f_2$ preceded and succeeded this stimulus. The subject was instructed to match the loudnesses of both simple tones individually to the loudness of the internally generated CDT changed by the externally added CT. Two silent intervals of 0.4 sec were programmed between the three stimuli to avoid temporal interactions. The threefold sequence was repeated periodically. The adjustments of

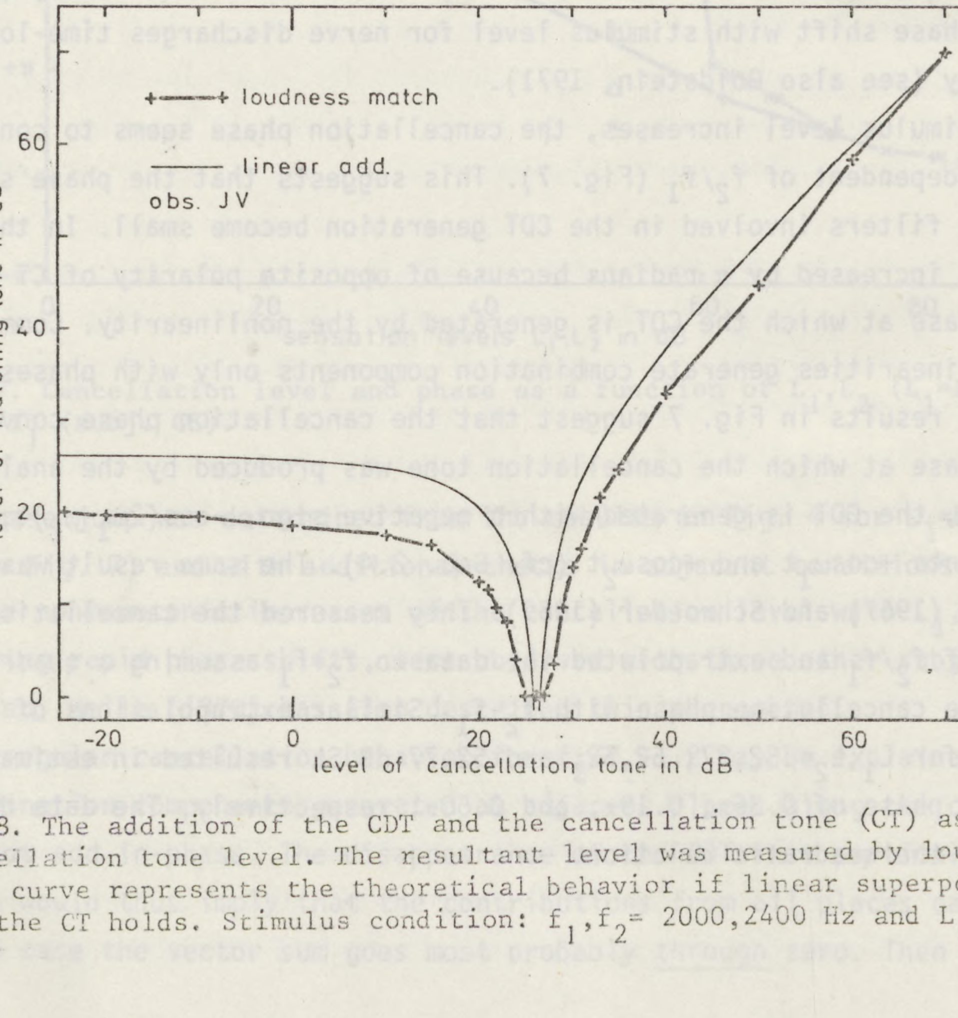


Fig. 8. The addition of the CDT and the cancellation tone (CT) as a function of the cancellation tone level. The resultant level was measured by loudness matching. The thin curve represents the theoretical behavior if linear superposition of the CDT and the CT holds. Stimulus condition: $f_1, f_2 = 2000, 2400$ Hz and $L_1, L_2 = 45, 45$ dB SL.

the first and the second matching tone never differed more than 2 dB, which suggested that temporal interactions were indeed unimportant.

An example of the results is given in Fig. 8 (plusses) for the condition f_1,f_2= 2000,2400 Hz and L_1,L_2= 45,45 dB SL. Each data point represents the average of 5×2 adjustments. The standard deviation of the mean is less than 1.5 dB. The thin curve is a theoretical one based upon the assumption of linear superposition of the CT and the CDT.

Figure 8 shows that the shape of the experimental curve found for loudness matching is similar to the theoretical curve. But the data points are all below the predicted values. A loudness match to the unaffected CDT (absent CT) revealed a value of 19.7 dB whereas the cancellation level was 26.2 dB. However, the low loudness of the CDT does not necessarily imply a low level of the CDT. The loudness may be related to the amount of activity above the masked threshold. (A completely masked tone of arbitrary level has zero loudness.) Thus, masking by f_1 at the place of the CDT may at least partly account for the low loudness of the CDT. Therefore, we conducted new experiments with two other measuring techniques. These techniques were introduced by Houtgast (1972).

3.2. CDT in Gap Masking

3.2.1. *Stimuli and Procedure*

The temporal stimulus configuration is shown in Fig. 9. The masker was switched off every 200 msec during about 50 msec. Four successive times the probe tone was presented in the inserted gap after which it was left out twice. By leaving out the probe tone periodically the subject was given a reference of how the stimulus

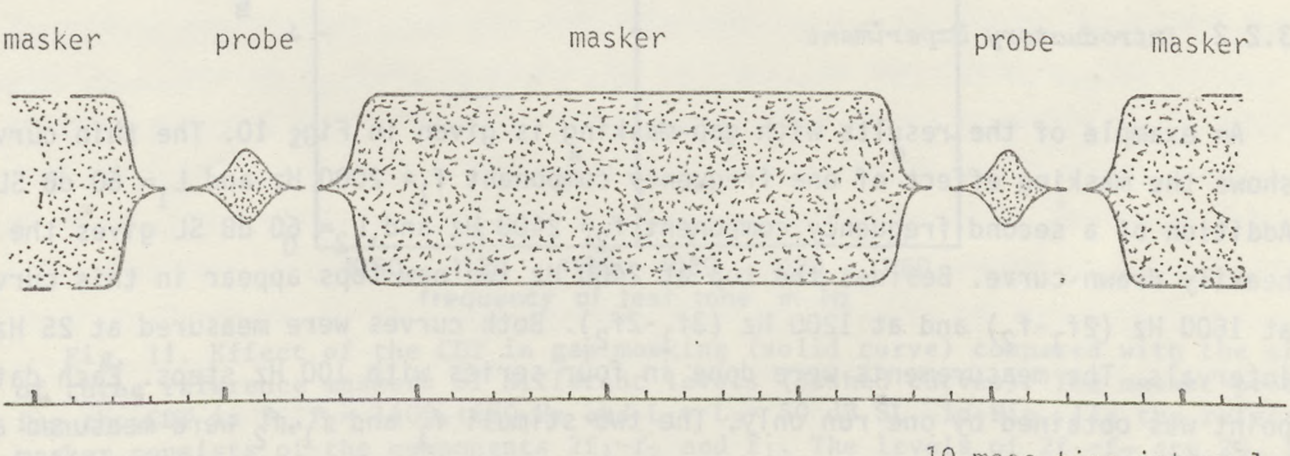


Fig. 9. Temporal stimulus configuration for gap masking. The masker is presented repeatedly.

sounded without the probe tone which made the probe tone easier to detect. The masked threshold of the probe tone was measured with a simple Békésy tracking procedure. After four successive probe tone presentations the level of the probe tone was decreased by 2 dB when the subject had perceived the probe tone and increased by 2 dB when he had not. Each run consisted of 50 presentations of probe-tone foursomes.

The temporal envelope of the probe tone was gaussian. If time spread and frequency spread are defined as the second moment of the temporal and spectral power distribution, respectively, then the frequency spread of the probe tone at a given time spread (effective duration) is minimal for a gaussian envelope of the probe tone (Gabor, 1946). However, this does not imply that in a masking experiment the optimal envelope is gaussian. The primary concern in masking experiments is to avoid detection of the probe tone at another place than the place corresponding to the probe's frequency. Therefore, the frequency spread down to threshold level should be considered with respect to the internal filters. But this requires knowledge of the filter characteristics. For general purpose the gaussian envelope suits well. It produces a nice parabolic spectral probe (ordinate in dB). In our case the duration of the probe tone taken between the -1σ and $+1\sigma$ points of the envelope was 9.5 msec. The bandwidth between the 1σ points of the amplitude spectrum of the probe tone was 67 Hz. Mathematically, this bandwidth is in good agreement with the effective duration. At -50 dB the bandwidth was 230 Hz. Of course, the gaussian envelope did not extend to infinity. The probe tone was switched on and off at about ½% of the maximal amplitude. The spectral distortion components corresponding to deviations from the ideal gaussian envelope were more than 48 dB down. The masker had the same transients as the probe tone. Both were led alternately through the same gate.

3.2.2. *Introductory Experiment*

An example of the results with gap-masking is given in Fig. 10. The thin curve shows the masking effect of one frequency component f_1= 2000 Hz and L_1= 60 dB SL. Addition of a second frequency component f_2= 2400 Hz and L_2= 60 dB SL gives the heavily drawn curve. Besides the top at 2400 Hz two new tops appear in this curve at 1600 Hz ($2f_1-f_2$) and at 1200 Hz ($3f_1-2f_2$). Both curves were measured at 25 Hz intervals. The measurements were done in four series with 100 Hz steps. Each data point was obtained by one run only. The two stimuli f_1 and f_1,f_2 were measured alternately.

The original reason to use gap masking in stead of simultaneous masking was to avoid beats between the masker and the probe tone. The results in Fig. 10 show an

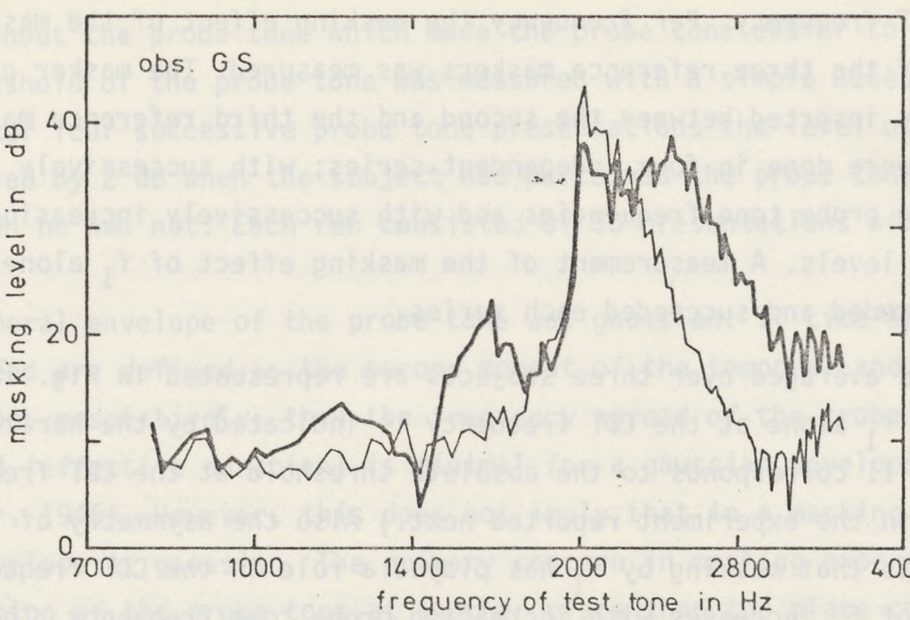


Fig. 10. Comparison of the effects of maskers consisting of one and of two frequency components in gap masking. The thin curve represents the masked threshold of a one-component masker f_1= 2000 Hz, L_1= 60 dB SL. The heavy curve represents the masked threshold if the component f_2= 2400 Hz and L_2= 60 dB SL is added to the masker. Note the two peaks at about 1600 Hz ($2f_1-f_2$) and 1200 Hz ($3f_1-2f_2$).

additional advantage of this method. The low-frequency slope of the masking pattern produced by f_1 appears to be very steep in gap masking. Consequently, there is only minor masking by f_1 at the place of the CDT. This allows for an accurate estimation of the CDT level by measuring its masking effect with respect to the masking effect of reference maskers.

3.2.3. *Experiment with reference masker $2f_1-f_2, f_1$*

Of course, the CDT level could be estimated by determining the level of a reference masker of the CDT's frequency for which the amount of masking equals the masking effect of the CDT. However, in this first experiment the frequency component f_1 was included in the reference maskers in order to get a good similarity between the reference maskers and the masker f_1, f_2 producing the CDT. In both types of maskers the components f_1 had equal levels. One advantage of including f_1 in the reference maskers was that this yielded for both maskers a similar residual masking by f_1 at the place of the CDT. Another advantage was that in this case both maskers had about equal loudness.

The stimulus components were f_1= 1400 Hz and f_2= $1.2f_1$= 1680 Hz with levels L_1= L_2≈ 50 dB SL. Three levels were used for the component $2f_1-f_2$ (1120 Hz) of the reference masker; these 'reference levels' were $L_{2,-1}$= 20, 25, and 30 dB SL. The masked threshold was measured with five equally spaced probe tone frequencies at and around the CDT frequency. Per frequency the masking effect of the masker producing the CDT and of the three reference maskers was measured. The masker producing the CDT was always inserted between the second and the third reference masker. The measurements were done in four independent series; with successively increasing and decreasing probe tone frequencies and with successively increasing and decreasing reference levels. A measurement of the masking effect of f_1 alone at the CDT frequency preceded and succeeded each series.

The results averaged over three subjects are represented in Fig. 11a. The masking effect of f_1 alone at the CDT frequency is indicated by the horizontal bar. (0 dB in Fig. 11 corresponds to the absolute threshold at the CDT frequency which was measured in the experiment reported next.) Also the asymmetry of the masking curves suggests that masking by f_1 has played a role in the CDT frequency region. An influence of f_1 increases with increasing probe tone frequency. The similarity of shape of the masking curve of the CDT and the curves of the three reference

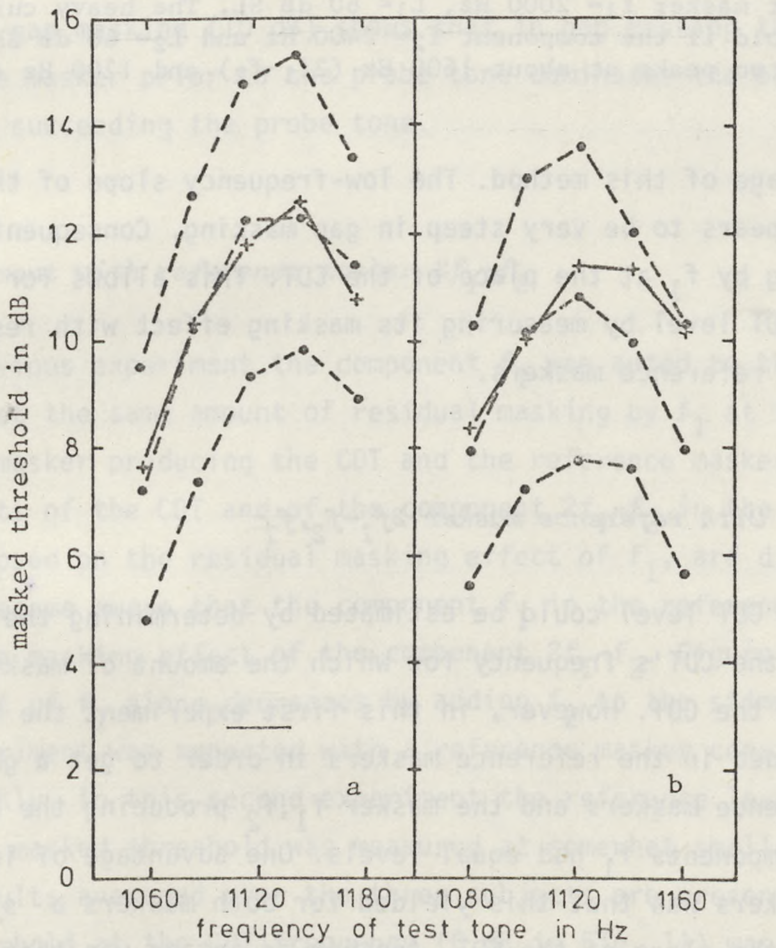


Fig. 11. Effect of the CDT in gap masking (solid curve) compared with the effect of three reference maskers of different levels (dashed curves). The masker producing the CDT is f_1, f_2= 1400,1680 Hz and L_1= L_2≈ 50 dB SL. In Fig. 11a the reference masker consists of the components $2f_1-f_2$ and f_1. The levels of $2f_1-f_2$ are 20, 25, and 30 dB SL, respectively. The level of f_1 equals 50 dB SL. The horizontal bar indicates the masking effect of f_1 alone. 0 dB on the ordinate corresponds to the absolute threshold at the CDT frequency. In Fig. 11b the reference masker is the component $2f_1-f_2$ only and its levels are 10, 15, and 20 dB SL, respectively.

maskers suggests that the role of f_1 at the CDT frequency was probably similar for both types of maskers. Thus, the CDT level can be estimated from the position of the CDT masking curve with respect to the reference curves.

The estimation of the CDT level was actually realized by interpolating the data of each set of three reference maskers for the masking effect of the CDT measured in the same set. Thus, four series with five probe tone frequencies gave twenty estimates of the CDT level. The average and standard deviation of these estimates for individual subjects are: GS, 24.1±0.4 dB; FW, 21.5±0.7 dB; and TH, 29.4±0.5 dB SL.

A similar measurement was made for a masker which was turned off 60 msec after its onset. This was the only difference from the previous experiment. Thus, in this case the masker followed the probe tone and the backward-masking effect was measured. Although we found a masking effect of only 2 dB the twenty estimates of the CDT level gave an average value of 23.6±0.9 dB SL (subj. GS). This value is close to the gap-masking estimate. The small masking effect in this case with respect to the effect found in gap masking (10 dB) shows that in gap masking the effect of forward masking of the masker prior to the probe tone dominates the backward-masking effect of the masker succeeding the probe tone.

3.2.4. *Experiment with reference masker $2f_1-f_2$*

In the previous experiment the component f_1 was added to the reference maskers in order to get the same amount of residual masking by f_1 at the place of the CDT for both the masker producing the CDT and the reference maskers. In that case the masking effects of the CDT and of the component $2f_1-f_2$ in the reference masker, both superimposed on the residual masking effect of f_1, are directly comparable. However, we became aware that the component f_1 in the reference masker possibly suppresses the masking effect of the component $2f_1-f_2$. Figure 10 shows that the masking effect of f_1 alone decreases by adding f_2 to the stimulus. Therefore, the previous experiment was repeated with a reference masker consisting of the component $2f_1-f_2$ only. In this second experiment the reference levels were 10, 15, and 20 dB SL. The masked threshold was measured at somewhat smaller frequency intervals. The results averaged over the three subjects are presented in Fig. 11b. The absolute threshold at the CDT frequency (0 dB in Fig. 11) was measured before and after each series of this second experiment.

Figure 11b shows that with absence of f_1 in the reference maskers the level of the component $2f_1-f_2$ can be about 7.5 dB lower in order to give the same masking effect. The masking curves of these reference maskers appear to be more symmetric. This supports the view that masking by f_1 at the place of the CDT accounts for the asymmetric masking curves. An estimation of the CDT level in this case requires a correction of the CDT masking effect for the masking by f_1. However, we did not study how masking adds. Estimates of the CDT level based upon the lowest three probe-tone frequencies for individual subjects are: GS, 13.1±0.6; FW, 17.5±0.5; and TH, 18.1±1.3 dB SL. Comparison of these values with the results of the previous experiment shows that the levels in this second experiment are significantly lower and that the difference is subject dependent. If corrections for the masking effect of f_1 should have been taken into account also for these lowest three probe-tone frequencies, then the difference would have been even greater. Before discussing the difference between the results for the two reference maskers another experiment will be described.

3.3. Pulsation Threshold of the CDT

In this experiment the stimulus f_1,f_2 producing the CDT was alternated with the reference stimulus $2f_1-f_2$. The alternation cycle was 4 Hz (on-time of each stimulus 125 msec). The transients were cosinusoidal and, from 0% to 100% amplitude, 15 msec in duration. The transients of the stimulus f_1,f_2 and the reference stimulus $2f_1-f_2$ were complementary such that an alternation of the same stimuli produced a continuous signal.

The idea behind this method is that the CDT and the reference tone together would be perceived as a continuous tone if the level of the reference tone equals the level of the CDT. An increase of the level of the reference tone, then, pro-

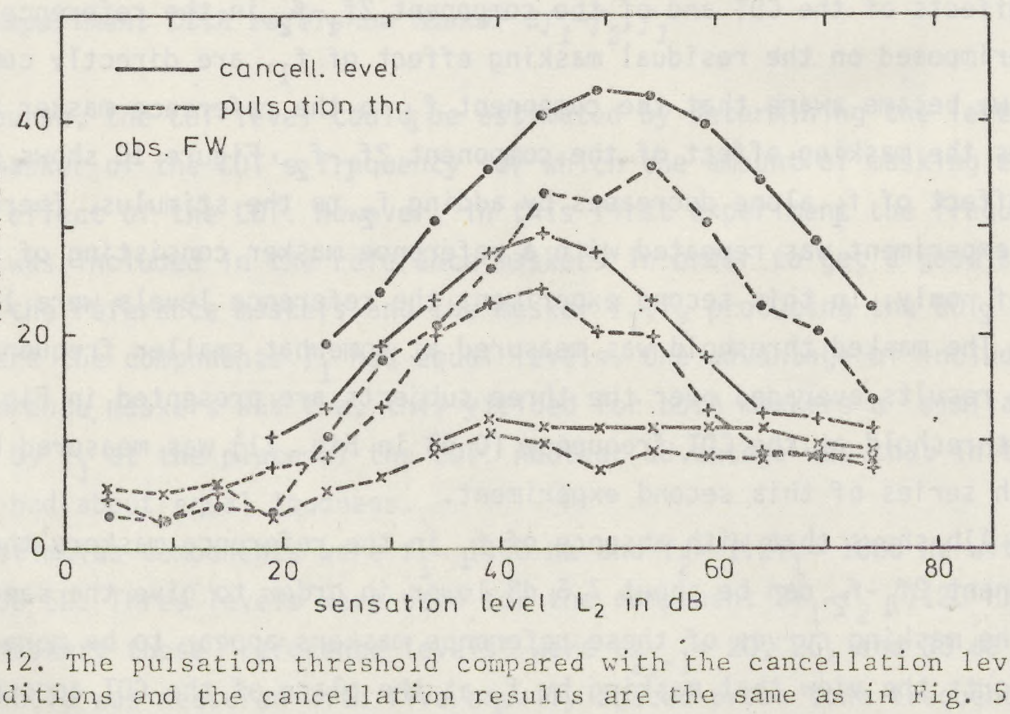


Fig. 12. The pulsation threshold compared with the cancellation level. The stimulus conditions and the cancellation results are the same as in Fig. 5.

duces a 'pulsation' of this tone. The subject had to adjust the level of the reference stimulus to the highest value at which the CDT did just not pulsate; 'the pulsation threshold'. Before each adjustment to the pulsation threshold a cancellation measurement was made. In the succeeding pulsation threshold measurement the phase of the reference stimulus was opposite to the cancellation phase and, thus, probably in-phase with the CDT. After a pulsation threshold adjustment a reversal of the polarity of the reference stimulus introduced audible clicks at the transients.

In Fig. 12 the pulsation threshold is compared with the cancellation levels presented before in Fig. 5. The measurements were made as a function of L_2 for f_2/f_1= 1.1, 1.2, and 1.3 with f_1= 1400 Hz and L_1= 54 dB SL (subject FW). It is apparent that the pulsation threshold is significantly lower than the cancellation level.

In a separate experiment the cancellation level and the pulsation threshold were measured for the condition used in the previous gap-masking experiments (f_1,f_2= 1400,1680 Hz and $L_1= L_2\approx$ 50 dB SL) and for the same three subjects. The results together with the data from the gap-masking experiments for both reference maskers are given in Table I. Results of the same experiments for a fourth subject and the condition used in Sec. 3.1 (f_1,f_2= 2000,2400 Hz and $L_1= L_2$= 45 dB SL) are added to this table. For each experiment and subject the standard deviation of the mean was less than 1.5 dB.

Table I. Comparison of estimates of the CDT level obtained by different techniques. For the subjects GS, FW, and TH the stimulus condition was f_1,f_2= 1400,1680 Hz and $L_1= L_2\approx$50 dB SL. The condition for subject JV was f_1,f_2= 2000,2400 Hz and $L_1= L_2$= 45 dB SL. The parenthesized value is the result of a loudness match (Sec.3.1).

subject	cancellation level	gap masking, ref. masker $2f_1-f_2,f_1$	gap masking, ref. masker $2f_1-f_2$	pulsation threshold	
GS	24.2	24.1	13.1	17.3	
FW	20.4	21.5	17.5	15.2	
TH	28.4	29.4	18.1	16.4	
JV	26.2	26.8	20.9	21.7	(19.7)

3.4. Discussion

Averaged over subjects the pulsation-threshold estimates of the CDT level are about equal to the estimates revealed by gap masking with the reference masker $2f_1-f_2$. But the estimates based upon gap masking with f_1 included in the reference masker are considerably higher. This difference may be explained by assuming that the component f_1 suppresses the weaker component $2f_1-f_2$ (cf. Houtgast, 1972). Then, the component $2f_1-f_2$ has to be presented at a higher level in order to produce the same amount of masking. The cancellation levels are also considerably higher than the pulsation thresholds and the gap-masking data for the reference masker $2f_1-f_2$. But the cancellation data are close to the gap-masking data with f_1 included in the reference masker. This suggests that the difference between the cancellation levels on the one hand and the pulsation thresholds and the gap-masking data obtained with the reference masker $2f_1-f_2$ on the other hand may also be explained by suppression effects. (The suppression effect of f_2 is much smaller than the effect of f_1 at the CDT frequency.) In the pulsation-threshold or gap-masking experiments the reference tone or probe tone is presented nonsimultaneously with the stimulus f_1,f_2. Therefore, this reference tone or probe tone will not be suppressed if the suppression stops immediately after the termination of the stimulus f_1,f_2. (If the suppression does not stop immediately and the probe tone is also somewhat suppressed, only part of the suppression effect would be measured.)

At this stage the question arises whether the CDT or the CT is suppressed. If the suppression precedes the generation of the CDT, then the CT would be suppressed before it cancels the CDT. But if the suppression succeeds the generation of the CDT, then the cancellation level is a good estimate of the CDT level and the CDT would be suppressed after its generation. This question cannot be answered by comparing the data in Table I. We studied this problem by measuring the addition of the CDT and the CT as was done in Sec. 3.1. But this time pulsation thresholds were determined accurately as a function of the CT level. On the basis of tone-on-tone-suppression measurements slightly different results were expected for the two cases distinguished. If the suppression precedes CDT generation, then only the CT would be suppressed. If the suppression succeeds CDT generation, then the CDT plus the CT would be suppressed. Different results were expected since the suppression effect is somewhat dependent on the level of the suppressed component. However, two accurate experiments gave no conclusive results. We think it is probable that the generation of the CDT and the suppression are two phenomena which originate with one and the same nonlinear mechanism. This is further discussed in the next section.

A difference between the cancellation level and the apparent CDT level was found also for electrophysiological measurements of the CDT (Goldstein, 1970). The time locking of nerve discharges to the CDT was removed with an external frequency component $2f_1-f_2$ of appropriate level and phase. The activity produced by this cancellation tone alone was higher than the activity produced by the CDT. Goldstein attributed this difference to a partial suppression by the primary tones f_1,f_2 of the response to the CDT.

4. GENERAL DISCUSSION AND SPECULATIONS

4.1. The Site of the Nonlinearity

The fact that the CDT can be cancelled with an external component of frequency $2f_1-f_2$ and of appropriate level and phase indicates that the CDT is temporally coded in the auditory system. Probably, the CDT is derived from the stimulus f_1,f_2 by a nonlinearity of the ear. This nonlinearity must be an important one since the combination tones of the type $f_1-k(f_2-f_1)$ are audible even at low stimulus levels. Is there any physiological evidence for such a nonlinearity?

Of course, the initiation of action potentials is a highly nonlinear process. This nonlinearity introduces many combination components in the neural signal. But how would these combination components be analyzed? There is no evidence for a frequency analyzer in the neural system of the ear. However, temporal analyzers detecting time intervals between action potentials have been proposed in models of pitch perception (Schouten *et al.*, 1962). But time intervals corresponding to the frequencies $f_1-k(f_2-f_1)$ are probably not introduced by the mechanism initiating the action potentials, although there are many time intervals between the action potentials which do not correspond to the stimulus frequencies (Rose *et al.*, 1969). Moreover, the time intervals between the nerve discharges depend on the amplitude and phase relation of the stimulus components whereas the pitch of a combination tone does not.

Inevitably, the nonlinearity must be followed by a frequency analyzer in order to reveal the combination components. The only known frequency analyzing device in the ear is the basilar membrane. Thus, the nonlinearity of the initiation of action potentials may account for the generation of combination tones only if the nonlinearity is coupled to the basilar membrane. For example, one might assume that the compliance of the stereocilia on the hair cells changes with the initiation of an action potential. However, the action potentials are most probably initiated outside the cochlear duct (Spoendlin, 1970; and private communication). Therefore, this model does not seem to be very realistic. The compliance of the stereocilia may be affected also by the cochlear potentials such as the CM or by the depolarization and hyperpolarization of the hair cell. If the speculation of a chemico-electric influence on the compliance of the stereocilia is wrong, then it is difficult to see how the known chemico-electric nonlinearities would account for the generation of the combination tones. In that case mechanical nonlinearities may play a role such as a possible nonlinear hair-cell coupling of the basilar membrane to the tectorial membrane. The idea that the nonlinearity is imposed on the basilar membrane by a non-constant compliance of the stereocilia or by nonlinear hair-cell coupling was supported somewhat by the results of the exepriment described in

Sec. 1 in which the CDT generation was studied in case of a threshold anomaly.

All these speculations on the origin of the nonlinearity were based upon the assumption that the nonlinearity must be coupled to the basilar membrane for frequency analysis of the distorted waveform to be possible. Recently, measurements of the vibrations of the basilar membrane with the Mössbauer technique gave direct experimental evidence for a nonlinear relation between the movements of the basilar membrane and the acoustic stimulus (Rhode, 1971). However, these measurements could not be done at SPL's lower than 70 dB in spite of the high sensitivity of the Mössbauer technique. Moreover, only one place of the basilar membrane was investigated. It is puzzling that Dallos (1969, 1970) did not find combination components in the CM which correspond to the psychophysical combination tones. The psychophysical data suggest that the combination components are present in the cochlea as ordinary traveling waves and in that case corresponding CM components are expected.

4.2 The Nature of the Nonlinearity

The nature of the nonlinearity generating the CDT is best studied with small frequency differences between the stimulus components. In that case the contributions to the CDT from different places along the cochlear partition probably add in-phase. This implies that only the amplitude patterns of the stimulations by f_1 and f_2 have to be considered in order to relate the CDT data to the nonlinearity operating at each place. For these small frequency differences our cancellation results (Sec. 2.4) are in agreement with the results reported by Zwicker (1955, 1968), Goldstein (1967), and Helle (1969/1970). The cancellation level increases proportionally to L_1, L_2 with a simultaneous increase of L_1, L_2 and also proportionally to L_2 if $L_2 < L_1$. For $L_2 > L_1$ the cancellation level decreases with an increase of L_2 according to a slope between 0 and -1. These results for small frequency differences will be discussed here. The irregular data for large stimulus-frequency differences (Fig. 7), attributed to out-of-phase addition of the contributions to the CDT from different places, are not discussed any further.

The authors who studied the CDT previously with the cancellation technique tried to describe their results with a nonlinear transfer function f(x) including a third power term, $f(x) = c_1 x + c_3 x^3$. According to this function the amplitude of the CDT is given by $\frac{3}{4} c_3 a_1^2 a_2$ if the stimulus components are $a_1 \cos 2\pi f_1 t$ and $a_2 \cos 2\pi f_2 t$. They concluded that the cancellation results disagree with this relation. For a simultaneous increase of a_1, a_2 the cancellation level does not increase proportionally to the third power of the stimulus amplitude and as a function of a_2 only, the cancellation level does not increase monotonically with an increase of a_2 but it reaches an upper limit at about $a_2 = a_1$. The mathematical description was modified by normalizing the nonlinearity with respect to the peak amplitude of the stimulus (a_1+a_2). After this modification the amplitude of the CDT was described by $ca_1^2a_2/(a_1+a_2)^2$ (c a constant value). This relation is in better agreement with the cancellation data. (The word 'level' always refers to the logarithm of the amplitude.)

The introduced normalization may represent an adaptive nonlinearity. We investigated the nonlinearity for adaptiveness by measuring the forward-masking effect of the CDT as a function of the duration of the masker f_1,f_2. Similar to the procedure used in Sec. 3.2.4 the level of the CDT was expressed in that level of a reference masker of frequency $2f_1-f_2$ which gave the same amount of masking. The duration of the reference masker was always equal to the duration of the masker f_1,f_2. Three subjects participated in this experiment. For two subjects the estimated CDT level with a 20 msec masker was about 3.5 dB lower than the estimated CDT level with a 150 msec masker. For the third subject there was no difference with these two durations. These results do not suggest that the nonlinearity is adaptive. In case of an adaptive nonlinearity rather an increase than a decrease of the CDT level is expected at shorter durations since the normalizing factor would not rise to its top value in a short time interval. Thus, we did not find evidence for a mechanism which is described by the normalized nonlinearity.

So far in this discussion, the cancellation level was identified with the CDT level. Implicitly, it was assumed that the nonlinearity does not affect the cancellation tone. This is true if the linear term of the transfer function is the most important one. However, the high cancellation levels revealed experimentally suggest that the nonlinear term must be very important. In that case also the cancellation may be affected by the nonlinearity. One might better understand the proportionality of the increase of the cancellation level with simultaneously increased L_1,L_2 if the nonlinearity operates upon both the stimulus components f_1,f_2 and the cancellation tone. This will be illustrated on the basis of a νth-law device described by the odd transfer function $f(x)=x^\nu$ if $x>0$ and $f(x)=-|x^\nu|$ if $x<0$; $\nu>0$, $\nu\neq 1$. Suppose that the CDT is cancelled with the cancellation tone amplitude a_{CT} and the cancellation tone phase ϕ_{CT}. A cancelled CDT means that the Fourier component $2f_1-f_2$ at the output of the νth-law device equals zero:

$$\int_{-\infty}^{+\infty} (a_1\cos 2\pi f_1 t + a_2\cos 2\pi f_2 t + a_{CT}\cos\{2\pi(2f_1-f_2)t+\phi_{CT}\})^\nu . \cos\{2\pi(2f_1-f_2)t+\psi\}dt = 0$$

(ψ arbitrary)

If both the stimulus amplitudes a_1,a_2 and the cancellation tone amplitude a_{CT} are multiplied by A, then:

$$\int_{-\infty}^{+\infty} (a_1A\cos 2\pi f_1 t + a_2A\cos 2\pi f_2 t + a_{CT}A\cos\{2\pi(2f_1-f_2)t+\phi_{CT}\})^\nu . \cos\{2\pi(2f_1-f_2)t+\psi\}dt =$$

$$A^\nu \int_{-\infty}^{+\infty} (a_1\cos 2\pi f_1 t + a_2\cos 2\pi f_2 t + a_{CT}\cos\{2\pi(2f_1-f_2)t+\phi_{CT}\})^\nu . \cos\{2\pi(2f_1-f_2)t+\psi\}dt = 0.$$

Thus, the amplitude of the Fourier component $2f_1-f_2$ at the output of the νth-law device remains zero if the cancellation tone level is increased proportionally to a_1,a_2. This example shows that for a large class of essential nonlinearities the cancellation level increases proportionally to a_1,a_2 if the cancellation tone is considered properly as an input signal which is used to eliminate the combination component $2f_1-f_2$ at the output of the nonlinear device. Moreover, in this case the cancellation amplitude as a function of a_1 and a_2 individually is described rather closely by the emperical relation mentioned above $ca_1^2a_2/(a_1+a_2)^2$. According to this relation the cancellation level would increase proportionally to a_2 if $a_2<a_1$ and would decrease inversely proportionally to a_2 if $a_2>a_1$. The cancellation data mostly show a less than inversely proportional decrease for $a_2>a_1$. This discrepancy can be attributed to the fact that the CDT is built up by contributions from different places along the cochlear partition. If $a_2>a_1$, then not at each place the stimulation by f_2 is higher than the stimulation by f_1.

The νth-law device generates the combination component $2f_1-f_2$ with an amplitude proportional to $a_1^2a_2^{\nu-2}$ if $a_1<a_2$ and proportional to $a_1^{\nu-1}a_2$ if $a_2<a_1$. (See note.) The loudness of the CDT diminishes with increasing a_1 ($a_1>a_2$). This indicates that $\nu-1<0$ or $\nu<1$. For $\nu<1$ the combination component is generated by the νth-law device with a polarity opposite to the polarity of the stimulus components. This is in agreement with the experimental results of Sec. 2.5.2. The results of the Mössbauer measurements of the vibration of the cochlear partition also suggest $\nu<1$ (Rhode, 1971).

If the input to the νth-law device consists of two frequency components, one with a low amplitude a and one with a high amplitude A, then the output amplitude of the weak component is given by $caA^{\nu-1}$. There is a linear relation between the input amplitude a of the weak component and its output amplitude. This linear input-output relation holds generally for weak input components. It is an important property of nonlinearities called 'linearization'. The linearization effect validates the mirroring procedure introduced in Sec. 2.3.2. The output amplitude of the weak component is determined also by the input amplitude A of the strong component. If $\nu<1$, then the strong component suppresses the weak component. This is a general property of signal-compressing types of nonlinearities. The decrease of the CDT level, if the amplitude of one of the stimulus components is raised above the other, can be interpreted as suppression of the CDT by the stronger stimulus

---

Note. The amplitude of the combination component $2f_1-f_2$ generated by the νth-law device can be derived analytically by Laplace transformation of the transfer function and a Jacobi-Anger expansion of the transformed function (Davenport and Root, 1958). The terms of the expansion give the distortion components. Calculation of the amplitudes of the distortion components requires the evaluation of an integral over a product of two Bessel functions; the Weber-Schafheitlin integral (Abramowitz and Segun, 1968, p. 487). See also Spekreyse (1970). Blachman (1964) gives another evaluation for $a_2<<a_1$.

component. Probably, the different estimates of the CDT level reported in Sec. 3 are also due to this suppression effect. The cancellation tone would be suppressed by the stronger stimulus components f_1 and f_2 whereas the nonsimultaneously presented probe or reference tones in gap-masking, loudness-matching, and pulsation-threshold experiments would not be suppressed. Also the tone-on-tone suppression effects measured by Houtgast (1972) might be explained on the basis of this type of nonlinearity.

An important property of the suppression mediated by a signal-compressing type of nonlinearity is that a strong component suppresses a weak component instantaneously. On the contrary, time constants are involved in inhibition mediated by neural networks. The inhibition effects found in primary nerve fibers are instantaneous (Nomoto *et al.*, 1964; Arthur *et al.*, 1971). Thus, these results support the idea of suppression mediated by a nonlinearity. Pfeiffer (1970) proposed the same signal-compressing type of nonlinearity as proposed here to account for the suppression effects found in the primary neurons. Hind (1970) showed that the nonlinear function relating the probability of firing to the stimulus accounts already for suppression effects.

The νth-law device was introduced in this discussion merely as an example. The essence of the discussion is that a signal-compressing type of nonlinearity probably accounts for both the generation of the combination tones of the type $f_1-k(f_2-f_1)$ and suppression effects. New experiments are started which hopefully will contribute to a more quantitative description.

5. CONCLUSIONS

o The generation of the CDT is affected by a dip. Since a dip is probably due to a sensorineural deficiency this result suggests that the nonlinearity generating the CDT resides at a hair-cell or higher level.

o Results of cancellation experiments suggest that the high-frequency slope of the pattern of stimulation upon which the nonlinearity generating the CDT operates is comparable with the slope revealed in masking experiments.

o The CDT level found with experiments in which the reference tone is presented nonsimultaneously with the stimulus f_1,f_2 producing the CDT is significantly lower than the cancellation level. The difference is probably due to suppression effects.

o Both the generation of the CDT and the suppression effects might be explained by a signal-compressing type of nonlinearity such as $f(x)= x^\nu$ if $x>0$, $= -|x^\nu|$ if $x<0$; $0<\nu<1$.

REFERENCES

Abramowitz, M. and Segun, J.A. (1968). *Handbook of Mathematical Functions* (Dover Publications Inc., New York).

Arthur, R.M., Pfeiffer, R.R., and Suga, N. (1971). "Properties of 'Two-Tone Inhibition' in Primary Auditory Neurones", J. Physiol. *212*, 593-609.

von Békésy, G. (1960). *Experiments in Hearing* (Mc Graw-Hill Book Company Inc., New York), Ch. 11.

Blachman, N.M. (1964). "Band-Pass Nonlinearities", IEEE Trans. Inform. Theory *II-10*, 162-164.

Dallos, P. (1969). "Combination Tone $2f_1-f_h$ in Microphonic Potentials", J. Acoust. Soc. Amer. *46*, 1437-1444.

Dallos, P. (1970). "Combination Tones in CM Potentials", in *Frequency Analysis and Periodicity Detection in Hearing*, R. Plomp and G.F. Smoorenburg, Eds. (Sijthoff, Leiden, The Netherlands), 218-229.

Davenport, W.B., and Root, W.L. (1958). *An Introduction to the Theory of Random Signals and Noise* (Mc Graw-Hill Book Company Inc., New York), Ch. 13.

Gabor, D. (1946). "Theory of Communication", J. Inst. Electr. Eng. *93*, 429-457.

Goldstein, J.L. (1967). "Auditory Nonlinearity", J. Acoust. Soc. Amer. *41*, 676-689.

Goldstein, J.L. (1970). "Aural Combination Tones", In *Frequency Analysis and Periodicity Detection in Hearing*, R. Plomp and G.F. Smoorenburg, Eds. (Sijthoff, Leiden, The Netherlands), 436-444.

Goldstein, J.L. (1971). In MIT Q.P.R. *100*, 202-203.

Goldstein, J.L., and Kiang, N.Y.S. (1968). "Neural Correlates of the Aural Combination Tone $2f_1-f_2$", Proc. IEEE *56*, 981-992.

Helle, R. (1969/1970). "Amplitude und Phase des im Gehör gebildeten Differenztones dritter Ordnung", Acustica *22*, 74-87.

Helle, R. (1970). Discussion Report in *Frequency Analysis and Periodicity Detection in Hearing*, R. Plomp and G.F. Smoorenburg, Eds. (Sijthoff, Leiden, The Netherlands), 246-247.

Hind, J.E. (1970). "Two-Tone Masking Effects in Squirrel Monkey Auditory Nerve Fibers", In *Frequency Analysis and Periodicity Detection in Hearing*, R. Plomp and G.F. Smoorenburg, Eds. (Sijthoff, Leiden, The Netherlands), 193-203.

Hoogland, G.A. (1953). "The Missing Fundamental", Doctoral Dissertation, University of Utrecht.

Houtgast, T. (1972). "Psychophysical Evidence for Lateral Inhibition in Hearing", accepted for publication in J. Acoust. Soc. Amer.

Johnstone, B.M., Taylor, K.J., and Boyle, A.J. (1970a). "Mechanics of the Guinea Pig Cochlea", J. Acoust. Soc. Amer. *47*, 504-509.

Johnstone, B.M., and Taylor, K.J. (1970b). "Mechanical Aspects of Cochlear Function", In *Frequency Analysis and Periodicity Detection in Hearing*, R. Plomp and G.F. Smoorenburg, Eds. (Sijthoff, Leiden, The Netherlands), 81-93.

Newman, E.B., Stevens, S.S., and Davis, H. (1937). "Factors in the Production of Aural Harmonics and Combination Tones", J. Acoust. Soc. Amer. *9*, 107-118.

Nomoto, M., Suga, N., and Katsuki, Y. (1964). "Discharge Pattern and Inhibition of Primary Auditory Nerve Fibers in the Monkey", J. Neurophysiol. *27*, 768-787.

Pfeiffer, R.R. (1970). "A Model for Two-Tone Inhibition of Single Cochlear-Nerve Fibers", J. Acoust. Soc. Amer. *48*, 1373-1378.

Plomp, R. (1965). "Detectability Threshold for Combination Tones", J. Acoust. Soc. Amer. *37*, 1110-1123.

Rhode, W.S. (1971). "Observations of the Vibration of the Basilar Membrane in Squirrel Monkeys Using the Mössbauer Technique", J. Acoust. Soc. Amer. *49*, 1218-1231.

Rose, J.E., Brugge, J.F., Anderson, D.J., and Hind, J.E. (1969). "Some Possible Neural Correlates of Combination Tones", J. Neurophysiol. *32*, 402-423.

Schouten, J.F., Ritsma, R.J., and Cardozo, B.L. (1962). "Pitch of the Residue", J. Acoust. Soc. Amer. *34*, 1418-1424.

Schroeder, M.R. (1969). "Relation between Critical Bands in Hearing and the Phase Characteristics of Cubic Difference Tones", J. Acoust. Soc. Amer. *46*, 1488-1491.

Smoorenburg, G.F. (1972). "Audibility Region of Combination Tones", submitted to J. Acoust. Soc. Amer.

Spekreijse, H., and Oosting, H. (1970). "Linearizing: A Method for Analysing and Synthesizing Nonlinear Systems", Kybernetik *7*, 22-31.

Spoendlin, H. (1970). "Structural Basis of Peripheral Frequency Analysis", In *Frequency Analysis and Periodicity Detection in Hearing*, R. Plomp and G.F. Smoorenburg, Eds. (Sijthoff, Leiden, The Netherlands), 2-40.

Tonndorf, J. (1970). "Nonlinearities in Cochlear Hydrodynamics", J. Acoust. Soc. Amer. *47*, 579-591.

Wever, E.G., Bray, C.W., and Lawrence, M. (1940a). "The Origin of Combination Tones", J. Exptl. Psychol. *27*, 217-226.

Wever, E.G., Bray, C.W., and Lawrence, M. (1940b). "A Quantitative Study of Combination Tones", J. Exptl. Psychol. *27*, 469-496.

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

Zwicker, E. (1968). "Der kubische Differenzton und die Erregung des Gehörs", Acustica *20*, 206-209.

Zwicker, E., and Feldtkeller, R. (1967). *Das Ohr als Nachrichtenempfänger* (S. Hirzel Verlag, Stuttgart), 2nd revised ed., Ch. 3.

Abramovich, S.I. [illegible] Fundamental Tone [illegible] Auditory Nonlinearity [illegible]

Arthur, R.M., Pfeiffer, R.R., and Suga, N. (1971). "Properties of 'Two-Tone Inhibition' in Primary Auditory Neurons", J. Physiol. 212, 593-609.

von Békésy, G. (1960). *Experiments in Hearing* (Mc Graw-Hill Book Company Inc., New York).

Blachman, N.M. (1964). "Band-Pass Nonlinearities", IEEE Trans. Inform. Theory IT-10, 162-164.

Dallos, P. (1969). "Combination Tone $2f_1$-f_2 in Microphonic Potentials", J. Acoust. Soc. Amer. 46, 1437-1444.

Dallos, P. (1970). "Combination Tones in CM Potentials", in *Frequency Analysis and Periodicity Detection in Hearing*, R. Plomp and G.F. Smoorenburg, Eds. (Sijthoff, Leiden, The Netherlands), 218-227.

Egan, J.P. [illegible]

Gabor, D. (1946). "Theory of Communication", J. Inst. Electr. Eng. 93, 429-457.

Goldstein, J.L. (1967). "Auditory Nonlinearity", J. Acoust. Soc. Amer. 41, 676-689.

Goldstein, J.L. (1970). "Aural Combination Tones", in *Frequency Analysis and Periodicity Detection in Hearing*, R. Plomp and G.F. Smoorenburg, Eds. (Sijthoff, Leiden, The Netherlands), 230-244.

Goldstein, J.L. (1971). In MIT Q.P.R. 102, 201-203.

Goldstein, J.L., and Kiang, N.Y.S. (1968). "Neural Correlates of the Aural Combination Tone $2f_1$-f_2", Proc. IEEE 56, 981-992.

Helle, R. (1969/1970). "Amplitude und Phase des im Gehör gebildeten Differenztones dritter Ordnung", Acustica 22, 74-87.

Helle, R. (1970). Discussion Remark in *Frequency Analysis and Periodicity Detection in Hearing*, R. Plomp and G.F. Smoorenburg, Eds. (Sijthoff, Leiden, The Netherlands), 246-247.

Hind, J.E. (1970). "Two-Tone Masking Effects in Squirrel Monkey Auditory Nerve Fibers", in *Frequency Analysis and Periodicity Detection in Hearing*, R. Plomp and G.F. Smoorenburg, Eds. (Sijthoff, Leiden, The Netherlands), 193-203.

Hoogland, G.A. (1953). "The Missing Fundamental", Doctoral Dissertation, University of Utrecht.

Houtgast, T. (1972). "Psychophysical Evidence for Lateral Inhibition in Hearing", accepted for publication in J. Acoust. Soc. Amer.

Johnstone, B.M., Taylor, K.J., and Boyle, A.J. (1970a). "Mechanics of the Guinea Pig Cochlea", J. Acoust. Soc. Amer. 47, 504-509.

Johnstone, B.M., and Taylor, K.J. (1970b). "Mechanical Aspects of Cochlear Function", in *Frequency Analysis and Periodicity Detection in Hearing*, R. Plomp and G.F. Smoorenburg, Eds. (Sijthoff, Leiden, The Netherlands), 81-93.

Newman, E.B., Stevens, S.S., and Davis, H. (1937). "Factors in the Production of Aural Harmonics and Combination Tones", J. Acoust. Soc. Amer. 9, 107-118.

Nomoto, M., Suga, N., and Katsuki, Y. (1964). "Discharge Pattern and Inhibition of Primary Auditory Nerve Fibers in the Monkey", J. Neurophysiol. 27, 768-787.

Pfeiffer, R.R. (1970). "A Model for Two-Tone Inhibition of Single Cochlear-Nerve Fibers", J. Acoust. Soc. Amer. 48, 1373-1378.

SAMENVATTING

Een belangrijke eigenschap van het gehoororgaan is frekwentieanalyse. Bij aandachtig luisteren naar een toon voortgebracht door een muziekinstrument is het meestal mogelijk een aantal deeltonen waar te nemen. Deze deeltonen korresponderen met frekwentiekomponenten van de door het muziekinstrument voortgebrachte trilling. Soms blijken er echter deeltonen waarneembaar te zijn die geen overeenkomstige frekwentiekomponenten in de stimulus bezitten. Zo kan een stimulus bestaande uit de twee frekwentiekomponenten f_1 en f_2 ($f_1<f_2$), indien voldoende luid, een deeltoon voortbrengen die korrespondeert met de frekwentiekomponent f_2-f_1. Deze deeltoon wordt de verschiltoon genoemd. Het ontstaan van de verschiltoon wordt toegeschreven aan vervorming van het signaal door overbelasting van het gehoororgaan. Men neemt aan dat deze vervorming in het gehoororgaan voorafgaat aan de frekwentieanalyse zodat de door de vervorming geintroduceerde frekwentiekomponenten worden geanalyseerd en vervolgens hoorbaar worden als deeltonen. Deeltonen die slechts hoorbaar zijn bij een gelijktijdige aanbieding van twee of meer frekwentiekomponenten worden kombinatietonen genoemd.

In dit proefschrift worden kombinatietonen bestudeerd die korresponderen met de frekwentiekomponenten $f_1-k(f_2-f_1)$, waarbij k een klein positief geheel getal is. Deze kombinatietonen blijken reeds te worden waargenomen bij de zeer lage geluiddrukniveaus van 15 à 20 dB boven de drempel. Gezien deze lage niveaus kan in dit geval niet gesproken worden van vervorming als gevolg van overbelasting. Aangenomen wordt dan ook dat deze kombinatietonen de produkten zijn van een meer essentiële niet-lineariteit van het gehoororgaan. De experimenten hadden tot doel deze niet-lineariteit langs psychofysische weg te bestuderen. De resultaten worden gepresenteerd in de vorm van drie artikelen. In het eerste artikel wordt de toonhoogteperceptie bestudeerd van een signaal dat uit slechts twee opeenvolgende harmonischen bestaat. Voor hogere harmonischen blijkt de toonhoogteperceptie gebaseerd te zijn op de kombinatietonen van het type $f_1-k(f_2-f_1)$. Het tweede artikel laat zien dat deze kombinatietonen slechts voorkomen in een beperkt frekwentiegebied onder f_1. De niet-lineariteit blijkt gekoppeld te zijn aan de frekwentieanalyse. Het laatste artikel gaat dieper in op de plaats waar deze kombinatietonen worden gegenereerd en op de aard van de niet-lineariteit.

CURRICULUM VITAE

De schrijver doorliep het Coornhert Lyceum te Haarlem en voltooide de middelbare schoolopleiding aan de Rijks hbs te Amersfoort. In 1961 werd de studie in de wiskunde en natuurwetenschappen begonnen aan de Rijks Universiteit te Utrecht. Het doktoraalexamen experimentele natuurkunde met groot bijvak toegepaste wiskunde werd in 1967 afgelegd. Daarna was de schrijver wetenschappelijk medewerker bij de afdeling medische en fysiologische fysica van de Rijks Universiteit te Utrecht op een aanstelling van de Stichting voor Zuiver Wetenschappelijk Onderzoek ZWO. Sinds 1 januari 1971 is de schrijver in dienst van het Instituut voor Zintuigfysiologie RVO-TNO te Soesterberg.