Chapter 13

On the "Residue" and Auditory Pitch Perception *

E. de Boer, Amsterdam, Netherlands

With 40 Figures

Contents

A. Trigger (Introductory Remarks) 481

B. Impulses of Interest (Outline of Problems, Fourier Theory) 483
1. Introduction 483
2. Fourier Analysis 485
3. Linear and Nonlinear Transformations 487
4. Pitch, Intervals, Aural Detection of Components 489
5. Terminology 490

C. Fourier's Proper "Place" (Early History, General Concepts) 491
1. Training to Hear Partials 491
2. Ohm's Law 492
3. Acoustic Siren 493
4. Seebeck's Crucial Experiment 494
5. Ohm's Reply; Helmholtz 495
6. Fletcher, the Missing Fundamental, Distortion 496
7. Place Theory 497

D. The Schouten "Period" (the Residue Theory) 497
1. Two Place Principles 497
2. The Optical Siren 498
3. Schouten's Principal Experiment: Cancellation of the Fundamental; the Residue Hypothesis 499
4. New Formulation of Ohm's Law 500
5. The Period as the Cue for Pitch 501
6. Seebeck Revisited 502
7. Mechanism of Period Extraction 502
8. Corollaries; Inharmonic Signals 506
9. Introspection, Slow Acceptance 507

E. Wave Crests (the Impact of Periodicity Pitch) 508
1. Outline of this Part 508
2. Licklider's Demonstrations 508
3. Chopped Noise — Direct Evidence of Periodicities (?) 509
4. Licklider's Autocorrelation Theory 509
5. Seebeck Re-Revisited 512
6. De Boer's Experiments on Inharmonic Signals 513

* To J. F. Schouten who "placed" his mark on a "period".

7. First and Second Effect, the Pseudo-Period 516
8. De Boer's Phase Rule 517
9. The Pseudo-Fundamental Theory 519
10. Two Types of Residue 519
11. Multiple Modes of Pitch Perception 520
12. Small's Work, Time Separation Pitch 521
13. Averaging of Pseudo-Periods and Pseudo-Fundamentals 523
14. Pitch Ambiguity; the Relation between Phase Effects and Aural Resolution . . 524

F. The Next "Cycle": Forebodings of a New Way of Thinking 526
1. Introduction 526
2. The Existence Region of the Residue (Ritsma) 526
3. The Principle of Dominance I 528
4. The Principle of Dominance II 530
5. Some Reflections, the Second Effect 530
6. The Residue in a New Definition, Resolution of Components 531
7. Audibility of High Partials 533
8. Repetition Pitch 535

G. The Missing Link (Towards a Unified Framework) 537
1. Nonlinear Processes, Combination Tones 537
2. Amplitude Functions 538
3. The Difference Tone (f_2—f_1) 539
4. The Cubic Difference Tone ($2f_1$—f_2) 540
5. Essential Nonlinearity, Goldstein's Work 542
6. Residue and Combination Tones — Smoorenburg 544
7. Existence Region for Combination Tones 546
8. Physiological Considerations 548
9. Single-Fiber Responses 550

H. Return to Place (?) (Increasing Importance of Spectral Concepts) 551
1. Doubts About Relevance of Periodicity 551
2. Houtsma and Goldstein's Experiments 552
3. Dichotic Pitch Recognition 554
4. Conclusions — Wightman's Theory 556
5. Wightman's Theory, Continued 558
6. Goldstein's Theory — Basic Constraints 560
7. Goldstein's Theory — Verification 563
8. General Conclusions 563
9. Dichotic Pitch Phenomena — Dichotic Repetition Pitch 564
10. Relation with "BMLD" and the "EC" Theory 566
11. The Internal Spectrum 567
12. Phase Effects — Buunen's Work 568
13. Binaural Diplacusis and the Residue (van den Brink) 569
14. Spectral Aspects of Time Separation Pitch and Monaural Repetition Pitch . . . 572

I. "Time-Out" (General Reflections) 574
1. The Importance of "Set" 574
2. Guide to Related Literature 575
3. Conferences 576
4. Place and Period Coding in Nerve Fibers 577
5. The One-Tone Residue 578

References 579

A. Trigger (Introductory Remarks)

The present chapter on *the residue phenomenon* is set up in an unorthodox way. Instead of giving a cursory review it presents the historical development in great detail. It is more a textbook on the psychophysics of pitch than a review of recent research. There are several reasons for this, as listed below:

a) Very few scientists, even those working on pitch perception, are aware of all the intricacies involved.

b) Textbooks on psychophysics or physiology of hearing in general are either outdated – the development has been very fast the last five years – or not authorative on this subject.

c) The older literature on the residue is inaccessible; as a consequence, an overview is difficult to obtain.

d) The publications include many subtleties in psychophysical experiments or advanced mathematical details of signal generation which make them difficult to understand for physiologists, etc.

e) There exist many misconceptions about the relevance of certain physiological experiments for pitch perception (and vice versa).

f) The problem of the residue is of prime importance for speech as well as for music, but a comprehensive account of the subject does not exist.

g) In some instances psychophysical or physiological studies are reported that seem to indicate that the "residue" is about to be discovered again. Such situations may be prevented by the existence of a general text.

These reasons led the present author to compile a description of the development of residue theory within a unifying framework. The account is constructed with a minimum of reference to the mathematical description of signals; as a consequence, there are only a few formulas, and those included, are very simple. In this way it is intended to reach the largest possible number of potential readers.

The treatise is rather lengthy, and the description is concentrated on only a few of the great many possible topics. The length is dictated by the wish to be instructive and clear. The choice as to which topics should be treated at length, which ones mentioned only in passing, and which ones should be skipped entirely, was difficult. The ultimate choice has been made with the greatest possible caution and after much deliberation. The strongest emphasis was put on the development of concepts; hence, many details of a procedural or methodological nature are completely left out of the discussion. There will be several subjects that seem to be "forgotten" entirely; the choice of subjects treated remains entirely the author's responsibility. The opinions ventured represent the author's own (1975) convictions. They are believed to represent also the communis opinio of the most noted scientists, but exceptions are possible, of course. The author would like to know whether his approach has been successful, in other words, whether his work can be understood by physiologists, psychologists and other scientists, and he welcomes feedback from the readers.

Many discussions with other scientists have contributed to a clear delineation of the most important ideas. Ample opportunity for discussion was provided by the conferences listed in Section I. 3. The author is especially grateful for the opportunity provided by Bell Telephone Laboratories for extensive discussions

with the groups of scientists involved in speech analysis and synthesis and in hearing research. These discussions have contributed greatly to the form of the present treatise.

The general outline of the chapter is as follows. The "residue story" is described in seven sub-chapters. The intention was that every one of these can be read and understood more or less independently of the others. That is the reason why many repetitions can be found in the text. Part B contains a general introduction to methodology and terminology of psychophysical research. A short outline of FOURIER analysis is included. The earliest history on pitch research provided the path leading to the "place theory". The period encompassing the second part of the 19th century and the first few decades of the 20th century is described in Part C. The "residue theory" was conceived and described around 1940 by the Dutch scientist SCHOUTEN. Part D contains a comprehensive description of SCHOUTEN'S experiments and the conception of the residue theory. That theory stressed the importance of waveform aspects of the signals employed, in particular, the periodicity. Many studies were devoted to the study of "periodicity pitch", as it was called, and Part E of the chapter describes the most important ones. The description is centered around the study of the pitch of inharmonic signals, and it appeared that a generalization of the definition of "residue" was needed to encompass all findings.

A new twist was given to the study of pitch perception when it was discovered that the main contributions to the pitch arise from low-numbered harmonics – in the original conception the main contribution was thought to arise from the unresolved high-numbered harmonics. This finding and its subsequent elaboration entailed a gradual shift of emphasis from waveform to spectrum aspects of the signals employed, and it took some time before this was established. The period in which these studies were performed is described in Part F. The relation between residue signals consisting of unresolved, closely spaced, components and those consisting of resolved, widely spaced, components remained somewhat unclear. The existence of aural combination tones (particular forms of products of nonlinear distortion) proved to be the factor that resolved this matter. Part G of the chapter gives a more detailed description of nonlinear distortion in general, and the properties of aural combination tones in particular. The newer residue theories are described in Part H; residue pitch is presently recognized to be almost entirely determined by the spectral content of the sound stimulus. This part describes the crucial experiments underlying this notion, several theories based upon it, and a few related phenomena. Some general reflections and warnings about mis-interpretations are described in the concluding Part I. The end of the story may read exactly like the begin: is it "place", or "period"? The meaning of these terms has changed completely, however.

A few words of explanation are in order on physiological aspects of the residue phenomenon. In the course of the present treatise it becomes clear why it is not likely that we will discover any clear-cut physiological mechanism responsible for the formation of residue pitch. Hence there is very little opportunity to describe physiological aspects of the residue phenomenon. An exception is formed by the topic of auditory nonlinearity (Part G).

The temporal features of stimuli appear to have only little to do with pitch. These features do play a part in perception: they are certainly related to the perception of "flutter", "rattle", "roughness", and the like. There will be physiological correlates to these phenomena, no doubt. It was decided, however, to leave these physiological aspects out of the discussion entirely since the main emphasis was to be put upon pitch phenomena.

B. Impulses of Interest (Outline of Problems, Fourier Theory)

1. Introduction

The "residue" in hearing theory is a concept that is intimately connected with the perception of *pitch*. Pitch – or as it can be called, in a slightly different context: tone height – is an attribute of sound that has been recognized as something special since the earliest documented times. Pitch plays a very important part in music and many observations and theories have been focused on this attribute during the long history of music. Scientific study of it (in the modern sense) was only possible after the development of acoustic instruments that permitted the generation of tones and signals with well-determined physical properties. Nevertheless, a most important and crucial question on pitch could already be formulated – and partly answered – on the basis of experiments with what we consider as the most primitive acoustic instrument possible, the siren. That is the question as to what physical parameter of the acoustic signal determines the pitch of the perceived sound; we will see in the sequel how this problem was attacked with unequalled ingenuity in a period where virtually no quantitative experiment could be performed.

We are not yet in a position to give a definition of what is the "residue" in hearing. However, the uninitiated reader may be satisfied at this stage by the following annotations. *Tones* are sounds that generally have a clearly determined musical pitch. Physically, tones can be regarded as combinations of a number of sinusoidal vibrations, all with different frequencies. These constituent vibrations are called *components*. Most of the tones used in music consist of combinations of sinusoidal vibrations of which the frequencies form an arithmetical sequence (*e.g.* 250, 500, 750, 1000, 1250, ... Hz). The pitch of such tones is usually associated with the lowest of these components. However, when the lowest component happens to be very weak or inaudible, the tone still has the same pitch (as if the lowest component were present). This pitch is produced by the cooperation of components of higher frequencies. Such a combined action of components which gives rise to a percept with a clear musical pitch is called a *residue*. The concept will be explained in detail later on. The most remarkable property for the moment is that a residue can have a pitch.

The study of the relation between physical properties of signals and attributes of perception belongs to the realm of *psychophysics*. In the context of a Handbook on Sensory Physiology it is necessary to devote some space to the nature of this field of scientific study in general. One principal problem in this field is the quantification of a sensation. In general, sensations cannot be measured in a numerical sense; this problem can, however, often be circumvented by the use of special techniques, such as the measurement of just noticeable differences in perception. By measuring the physical counterpart of a just noticeable change in a signal the main point of what a sensation really is, by-passed, and the limit of perceptual analysis is measured instead. By this means, it remains possible to study the properties of the analyzing mechanism without actually specifying the output. In several other branches of psychophysics, it is attempted to quantify attributes of sensations more directly; the logical consistency of experimental results is then the guideline along which the relevance of certain signal parameters is confirmed. In the field of audition this applies, *e. g.*, to the measurement of loudness.

To come back to the pitch problem: pitch is, in a sense, a directly measurable attribute because there already exists a musical scale which uniquely assigns a number (*e. g.* a frequency value) to a pitch. Conversely, we may state that this pitch scale derives its structure from the properties of human pitch perception. Pitch can be measured directly along the musical pitch scale by persons possessing "absolute pitch". Other persons may compare pitches of two sounds and so arrive at a numerical value for the pitch of a sound; this procedure is even possible with unmusical persons. In these respects, pitch has many of the properties of a "Gestalt".

In the realm of the "residue" theory, where the residue pitch is a central issue, a tangible and practical definition of pitch is called for. A *pure tone*, *i. e.*, the tone perceived from a sinusoidal sound signal, seems to be the clearest bearer of pitch as it is devoid of a specific timbre. For this reason, the operational definition of pitch involves the comparison of the unknown sound with a pure tone and adjusting the frequency of the latter until the two are judged to have equal pitch. This definition is almost exclusively used whenever numerical measurement of pitch is desired. Most of the sounds used in residue research have a timbre that is so widely different from that of a pure tone that the observers have great difficulties with the comparison. This is partly reflected in the results of such experiments. In some studies, a complex signal with a definite pitch is used as the comparison signal. Pitch matches are then easy to make, but the disadvantage is that the comparison signal is a multi-parameter signal, and the decision to use only one parameter as the bearer of pitch is an arbitrary one which requires justification.

The basic duality of psychophysics – direct measurement *versus* comparison experiments – causes several inherent problems. This is also the case for the pitch attribute. In the present author's opinion, this accounts for many troubles that occurred in pitch research and for the grave problems of connecting results of all research projects together in a meaningful way. This is especially true since many of the signals that have been used in pitch research – and, especially, in "residue" research – are much less tonal than most of the sounds used in music. Hence, it is easy to lead an observer astray so that, in fact, he does not use his pitch sense at all. Only very few scientists have employed listeners that were sufficiently well versed in music that they could be expected to have a well-developed sense of pitch. Pitch has many attributes of a "Gestalt", and observations on pitch require a good deal of introspection. Both aspects do not lend themselves easily to positivistic scientific experimentation based on the repeatability and reproducibility of measurements. However, if one wishes to demonstrate quite general properties of the auditory system, it is necessary to prove that the phenomena observed can be measured in what are called "naive listeners" as well. It should not be done the other way around. Even worse is to do the tests exclusively with naive listeners and to base far-reaching conclusions about perception aspects on these persons' responses.

2. Fourier Analysis

It is assumed that the reader is at least vaguely acquainted with the formalism of Fourier analysis. In order to avoid misinterpretations and to provide a unified basis for forthcoming descriptions, the principles are briefly summarized in this section. In this chapter, we shall mainly consider signals, written as a function $f(t)$ of time t, that are deterministic. That means the waveform of $f(t)$ can be predicted when only part of it is known. An important sub-class of these signals are the *periodic signals*. A periodic signal $f(t)$ with period T can be written as the sum of a number of sinusoidal signals $s_i(t)$, all with different frequencies f_i (i is an integer index). Because of the periodicity the frequencies f_i have the peculiar property that they all are integral multiples of the *fundamental frequency* $f_0 = 1/T$. The waveforms of all the constituent sinusoids $s_i(t)$, when added together, form exactly the waveform $f(t)$. The following formula expresses this mathematically (Σ denoting a summation):

$$f(t) = \sum_{i=1}^{N} s_i(t) \tag{1}$$

(the index i is an integer and in this context always limited to cover a finite range, from 1 to N).

Each of the signals $s_i(t)$ is called a *component*; it is a sinusoidally varying signal and hence can be written in the form

$$s_i(t) = a_i \cos(2\pi f_i t + \varphi_i). \tag{2}$$

For a periodic signal $f(t)$ all frequencies are, as stated, integral multiples of the frequency f_0; this is most simply expressed by using the integer index i as the multiplier:

$$f_i = i \cdot f_0 \tag{3}$$

with $f_0 = 1/T$.

Components of which the frequencies are integral multiples of a frequency f_0 are called *harmonics* of f_0. The rank number of harmonics corresponds to the index i in Eq. (3). Formerly, such components were also called *overtones*, with a rank number one lower than i. For example, the second overtone is the same as the third harmonic ($i = 3$). The lowest harmonic, with i equal to 1, is commonly referred to as the *fundamental*; to avoid ambiguity we shall often refer to it as the fundamental component. The frequency f_0 is the fundamental frequency; as stated, this is the inverse of the period T. It is confusing that a series of frequencies conforming to Eq. (3) is often called a *harmonic series*. Because of the wide acceptance of this terminology, it will be difficult to insist upon one consistent description, andit will occur sometimes that different terms are used for the same idea or concept.

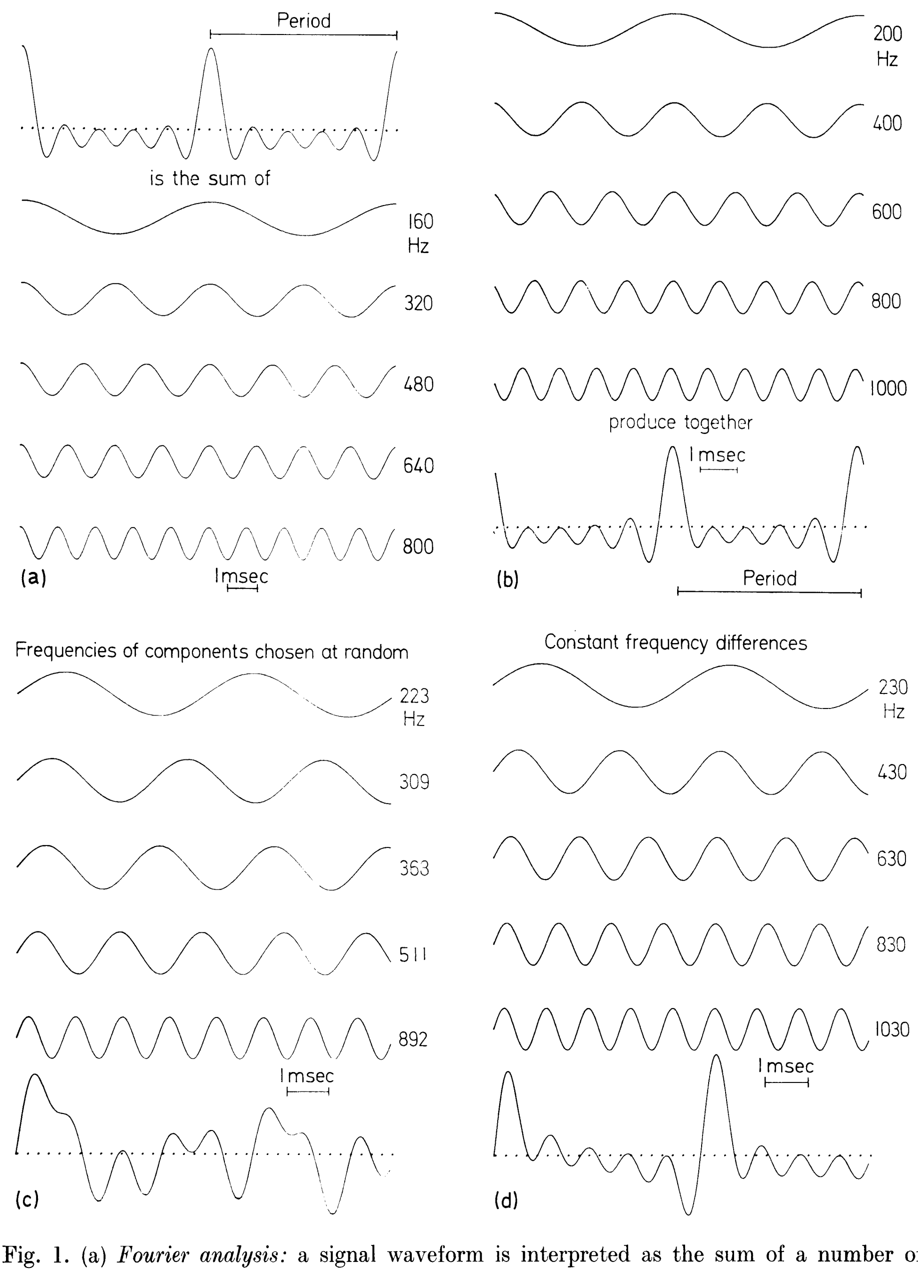

Fig. 1. (a) *Fourier analysis:* a signal waveform is interpreted as the sum of a number of sinusoidal waveforms. (b) *Fourier synthesis:* a number of sinusoidal signals is added to form a resultant waveform which is not sinusoidal. In the case shown, the component frequencies form a harmonic series; as a consequence, the resultant waveform is periodic (note that the component phases are different here). (c) *Fourier synthesis:* in this case, the component frequencies are random. (d) *Fourier synthesis:* in this case the frequencies form an arithmetical sequence

In order that relation (1) is obeyed, the amplitudes a_i and phases φ_i must have particular values for a particular signal $f(t)$. These values are found by applying the mathematical formalism of *Fourier analysis*. This process is illustrated by Fig. 1a for a periodic signal. Conversely, one may generate a particular set of sinusoidal signals $s_i(t)$, with chosen values of amplitudes and phases, and obtain the signal $f(t)$ by addition *(superposition)*. This is *Fourier synthesis*, and it will be used quite often in this chapter. See Fig. 1b.

For the waveform of the sum signal $f(t)$ the component amplitudes a_i are as important as the component phases φ_i. For auditory perception, the amplitudes are by far more important. The graph in which component amplitudes are plotted as a function of frequency is commonly referred to as the *spectrum*. Furthermore, Fourier analysis is also known as *spectral analysis*. However, it should always be remembered that, mathematically, phases are as important as amplitudes and that only certain properties of auditory analysis permit the suppression of phase data. In residue research, phase effects have played and continue to play an important part as we shall see.

For Fourier synthesis, there is no principal restriction in the choice of the component frequencies f_i. When these frequencies are not chosen to be integral multiples of a value f_0, the resulting sum signal will not be periodic, but it will still be deterministic. Such signals are usually not produced directly as specific waveforms and then analyzed, but they are produced by *synthesis* from a finite number of constituent sinusoidal components (see Fig. 1c for an illustration). In research of the residue phenomenon, the component frequencies are not random, but they are usually chosen to form an arithmetical series, *i. e.*, they have equal differences. The resulting signals, although having no simple periodicity, show a particular regularity in their waveform which has played an important part in residue theory. This will be described in the sequel. An illustration is provided by Fig. 1d.

If a part of a deterministic signal is analyzed, always the same components are found, no matter which part is taken (at least theoretically). This does not hold true for a stochastic signal like random noise. Every section of a noise signal has a different composition. For every frequency region, one can only specify the average spectral content; hence, spectral analysis of random signals can only be carried out in the "mean" sense. Moreover, the number of constituent components is not finite as in all the signals considered above, and that is the reason we cannot speak of the "amplitude of the component at a particular frequency" but are forced to use always the "average spectral content over a certain frequency region". Noise signals will only briefly be touched upon in this chapter; they are described here just for the sake of completeness.

3. Linear and Nonlinear Transformations

The theory of Fourier analysis is more than a mathematical tool. In physics, it is used a great deal to simplify the description of phenomena. This can be done wherever the system under study is a *linear* system. Put in the simplest possible terms, a linear system reacts to any disturbance in a way that is independent of the occurrence of other disturbances. This implies that when at any point the sum of two disturbances – here represented as two signals, $x_1(t)$ and $x_2(t)$ — is present, the reaction at another point is the sum $y_1(t) + y_2(t)$ of the reactions $y_1(t)$ and $y_2(t)$ brought about by each of the two x-signals separately. The notion of linearity implies also that the "output" signal (the reaction) is proportional to the "input" signal (the disturbance), *i. e.*, if $x(t)$ becomes, *e. g.*, 3 times as large, $y(t)$ will also become 3 times as large. A peculiarity of a linear system is that the reaction to a (sustained) sinusoidal signal is again a sinusoidal signal. Amplitude and phase may have been changed by the linear system, but the frequency has remained the same, of course. This property explains the utility of Fourier analysis in the domain of linear systems theory. Let us consider an input signal $x(t)$ and the resulting output signal $y(t)$, and let us decompose the former into a sum of sinusoidal signals:

$$x(t) = \sum_{i=1}^{N} {}_x s_i(t). \tag{1-a}$$

Each of the s-signals is transformed into another sinusoidal signal ${}_y s_i(t)$, in which process only the amplitude and the phase can be changed, and the output signal $y(t)$ must be the sum of these y-components:

$$\sum_{i=1}^{N} {}_y s\ (t) = y(t). \tag{1-b}$$

As a consequence, the transformation that the system performs for any input signal $x(t)$, can be described entirely by the transformations that *sinusoidal* input signals undergo. Frequency is then the only independent parameter. Hence the transformation is completely specified by the *amplitude* and *phase* characteristics describing the changes in amplitude and phase as a function of frequency.

Many transformations that take place in the peripheral auditory organs conform to the principle of linearity. For some transformations linearity fails significantly and we must then explicitly account for the *nonlinear* properties of the system. For a linear system, all Fourier components of the output signfy must already be contained in the input signal, since the system can only modial sinusoidal components, not create them. For a nonlinear system this is no longer true: a nonlinear system often reveals its nonlinearity by way of newly created sinusoidal components in the output signal. The newly created components are called *distortion products*. Moreover, in a nonlinear system, there usually is no proportionality between the amplitude of output and input signals, not even for sinusoidal signals. Nonlinear systems are extremely difficult to specify or to analyze. Let it be sufficient here to mention just the type and nature of distortion products that are relevant in the present context.

Let us start with a nonlinear system subjected to a single sinusoidal input signal $s(t)$ with frequency f_0. The output signal will not be sinusoidal, but it will have the same period as the input signal. Hence, it can be described as a sum of sinusoids like Eq. (1) and the newly created components will have frequencies that are integral multiples of f_0 [compare Eq. (3)]. In other terms, the components created are harmonics or overtones of the input signal. The amplitudes of these harmonics do not vary in direct proportion to the amplitude of the sinusoidal input signal. When the input signal is a composite signal, the situation is much more complicated. Let us consider the case where the input signal is composed of two sinusoidal signals $s_1(t)$ and $s_2(t)$, with frequencies f_1 and f_2, respectively. In this case, the frequencies are entirely arbitrary; there is no relation between them, except that f_2 is assumed to be larger than f_1. One may now expect in the output signal two series of harmonics: those of which the frequency is an integral multiple of f_1 and the same for f_2. This expectation is met indeed, but that is not all. A second group of distortion products have frequencies that are specific combinations of the frequencies f_1 and f_2. Quite appropriately, these are called *combination tones*. A few examples: the *difference tone* with frequency f_2-f_1, the *sum tone* with frequency f_1+f_2, other combination tones have frequencies like $2f_1-f_2$, $2f_2-f_1$, $3f_1-2f_2$, $2f_1+f_2$, etc. Several of these distortion products will turn out to be important for the theory of the residue phenomenon as we shall see. Moreover, combination tones have played an important part in the history of music (*cf.* Plomp, 1965) and have been the cause of much confusion; the Tartini pitch, for instance, may either be a simple difference tone or a higher-order combination tone.

When the input signal consists of more than two sinusoidal components, the situation becomes so complicated that it is hard to sort out everything. The principal point is that distortion products arise having frequencies which are always linear combinations of at least two of the input frequencies:

$$f_d = k_1 f_1 + k_2 f_2 + \dots .$$

In this formula, f_d is the frequency of a distortion product, and f_1, f_2 ... are the frequencies of the input signal. The coefficients $k_1, k_2, \dots$ are all (positive or negative) integral numbers. It is recalled that there is no general rule that controls the amplitude of any distortion product (*cf.* Section G. 2).

4. Pitch, Intervals, Aural Detection of Components

Long before Fourier theory was established, musicians knew that musical instruments produce multiple tones. In many musical instruments, notably bowed instruments and organ-pipe-like instruments, a tone is produced of which the waveform is periodic but definitely not sinusoidal. Such a tone can put a resonator (*e. g.* a string) into vibration whenever the resonance frequency of the resonator corresponds to one of the component frequencies of the tone. Furthermore, two tones can blend beautifully when the fundamental frequency of one corresponds to one of the component frequencies of the other. The harmonics of a tone with a fundamental frequency of about 125 Hz are shown in musical notation by Fig. 2. This figure indicates a peculiarity that has to do with the relation between the physical parameter: frequency and the subjective attribute,

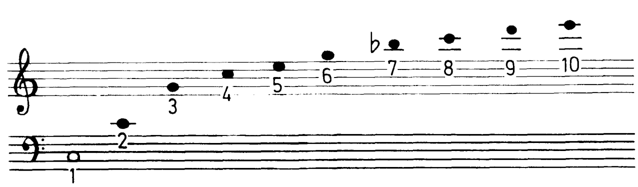

Fig. 2. The harmonic series in musical notation. This figure shows the pitches of each of the harmonics of a tone of 125 Hz (approx.). The numbers indicate the rank numbers of the harmonics

pitch. The series of frequencies f_i according to Eq. (3) form an arithmetical series, *i. e.*, their differences are constant. In musical notation, the differences are not constant: the pitch difference between the lowest two harmonics is the largest, and all subsequent pitch differences become successively smaller. The difference between the 1st and 2nd harmonics amounts to one octave, that between the 2nd and 3rd harmonic, a fifth, between the 3rd and 4th a fourth, etc. Our subjective pitch scale, in general terms, is laid out in such a way that a certain *musical interval*, *i. e.*, a difference of two pitches, corresponds to a specific *ratio of frequencies*. The ratio's 2:1, 3:2, 4:3, etc. successively become smaller, and hence the musical intervals become smaller too.

The various components of a tone can be identified (objectively) with the help of resonators. Up to a point, however, they can also be identified by the ear. That is, with a certain amount of training, a musical listener can aurally isolate the components of a complex signal. Each component is then recognized as what it actually is: a sinusoidal vibration (simple tone) that has an almost colourless tone quality. With a little bit of training, one can learn to identify at least the 1st to the 5th harmonic. Now, when one listens to a tone, two attitudes are possible. One is to listen to the sound as a whole, appreciating its pitch and perceiving the timbre as a characteristic quality of the entire sound. The other attitude is one of subjective analysis: one tries to break up the sound into constituent sounds (which happen to correspond to sinusoidal components), and the qualities of the sound as a whole are lost. In the latter case, we may allude to the situation by saying that the auditory system is attempting to carry out a Fourier analysis. That this has its limitations is self-evident. However, the possibility of this process has tremendous consequences for auditory theory.

5. Terminology

In the previous sections, several terms and their interrelations were described. There are more terms in common use which do not always conform to a strict definition. It seems appropriate, therefore, to group the terms into different categories and to try to separate them according to the categories to which they belong. The main categories will be: *descriptors of signals* in physical/mathematical terms and *descriptors of sensations*. A provisional grouping might look like Table 1. The terms are grouped according to the main category of usage. Corresponding terms in the other category are indicated within parentheses. Obvious omissions in the table can be filled in by resorting to a consistent terminology as introduced by SMOORENBURG (1970, 1972). The new term, part-tone, has been underlined in the table. Following these suggestions, we should adhere to a convention. The word *tone* and combinations of it, like simple tone and complex tone, should be used only in the perceptual sense. The corresponding physical concept should be described with the word *signal*. This will be fairly difficult since the term tone has acquired already many connotations. The reader will readily acknowledge this when he tries to re-read Section B. 4 and, noting how often the word tone is used in a dualistic sense, mentally tries to substitute another word for it. Furthermore, it is very difficult to replace the word tone whenever it is used in a musical context.

The word *partial* can be reserved exclusively for physical descriptions as soon as we can resort to a corresponding term like part-tone for its perceptual counterpart. The auditory Fourier analysis described in Section B. 4 can then be summarized by stating that a complex tone (*i. e.* a tone that is not a simple tone) can either be perceived as a tone or analyzed into part-tones. Each part-tone happens to correspond to a partial and, when attention is drawn to it, is perceived as a simple tone. Not all components can be recognized as part-tones; the remaining components form together a tone which cannot be analyzed any more. (These nonanalyzable partials constitute what has been called a "residue".)

Table 1. Terms in common use

Term, concept	Physical/mathematical meaning	Perceptual meaning
Sound	+	(+)
Tone[a]		+
Pitch		+
Musical interval	(frequency ratio)	+
Simple tone, pure tone	(sinusoidal signal)	+
Complex tone	(multi-component signal)	+
Frequency	+	
Frequency ratio	+	(interval)
Linearity	+	
Nonlinearity	+	
Distortion	+	+
Distortion product	+	+
Combination tone	(+)	+
Sinusoidal signal	+	(simple tone)
Frequency component	+	(part-tone)
Fourier component		
Partial	+	(part-tone)
Part-tone	(component, partial)	+
Harmonic	+	
Overtone	+	+
Fundamental frequency	+	
Fundamental (component)	+	+
Fundamental tone		+

[a] This concept is used in music in a dualistic sense.

C. Fourier's Proper "Place" (Early History, General Concepts)

1. Training to Hear Partials

In Sections B. 4 and B. 5, it has been described how the auditory system is capable of carrying out an incomplete type of Fourier analysis. The physical presence of harmonics can be proven by the use of resonators. However, as Helmholtz (1862, 1896, 1954) described, these can also be used to train an observer in auditory analysis and to make him realize that the part-tones he can hear correspond to the *partials* that can be proven to exist by physical means. The strings of a piano can be used as resonators in this type of experiment. The first test shows the presence of partials. A key is depressed which corresponds to one of the harmonics (2nd, 3rd, 4th, etc.) shown in Fig. 2. This action frees the corresponding string from its damper. While the key of, *e. g.*, the tone g_1 is depressed, the tone c is struck vigorously. When the c key is released subsequently, the tone c ceases but its 3rd harmonic g_1 is clearly heard which proves that this tone was contained in the previous sound.

By small variations of this experiment, the listener can convince himself that the partial g_1 is not only contained in the c sound, but that he can hear it in the tone c and that he can mentally isolate it. A very useful trick to train one-self to do this is to let the harmonic under question be heard previously and to direct attention to that particular pitch. With the help of a piano (or another string instrument), it is easy to hear the part-tone corresponding to the 5th or 6th harmonic. HELMHOLTZ reports that he was able to isolate aurally up to the 16th harmonic. However, he could not do this without artificial means of enhancing a partial – he used resonators in the form of hollow spheres provided with a small open tube, the Helmholtz spheres. PLOMP (1966) gives a short historical review of the attempts to determine the upper limit of aural isolation of harmonics. There is very little correspondence between the results of different experiments, mainly because of the vagueness of criteria used. By the use of very strict criteria and a very rigorous measurement procedure, PLOMP established that auditory analysis of components in a multi-component signal is well possible to the 5th harmonic, but that the power of resolution decreases quickly in the range of the 5th to the 9th harmonic. This should not rule out, however, that analysis of higher components is not possible, only that it requires a special situation. The limit found by PLOMP is related to the limit of aural resolution as it is expressed by the concept of the *critical bandwidth*. See other chapters of this volume for a description of this concept.

2. Ohm's Law

The relation between Fourier theory and auditory analysis was first formulated by OHM (1843). This was known later as *Ohm's acoustical law*. In its original version it read:

α) "... this is my formulation of the old definition of a 'tone':

a) To produce a tone of frequency m a series of periodic impressions with a period equal to $1/m$ is necessary; the form $a \cdot \sin 2\pi(mt + p)$ must be present in each period either purely or as a real component.

b) These forms must have the same value of p in successive periods.

c) The value of a must be the same in successive periods."

β) "... as the means to assess whether or not a given impression contains the form $a \cdot \sin 2\pi(mt + p)$, I use the theorem of Fourier which has become so famous because of its numerous and important applications, and which reads as follows...."

3. Acoustic Siren

It is seen that this formulation by Ohm is more strict than can be substantiated by experiments, as we have seen above. We must remember, though, that the law was formulated in a time in which hardly any quantitative acoustic measurement could be carried out; a time also, in which Fourier's theorem had been formulated only two decades earlier. As Ohm put it, "... I tried to prove whether the definition of the 'tone', as it was given to us by our forbears, would contain everything necessary and sufficient to explain the newer facts or not." And he concluded, "... that right was done to the old endowned definition...."

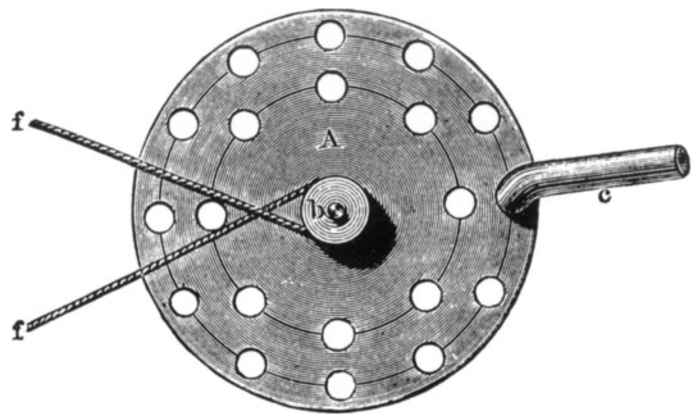

Fig. 3. The acoustical siren as used in Seebeck's time. After Helmholtz (1896)

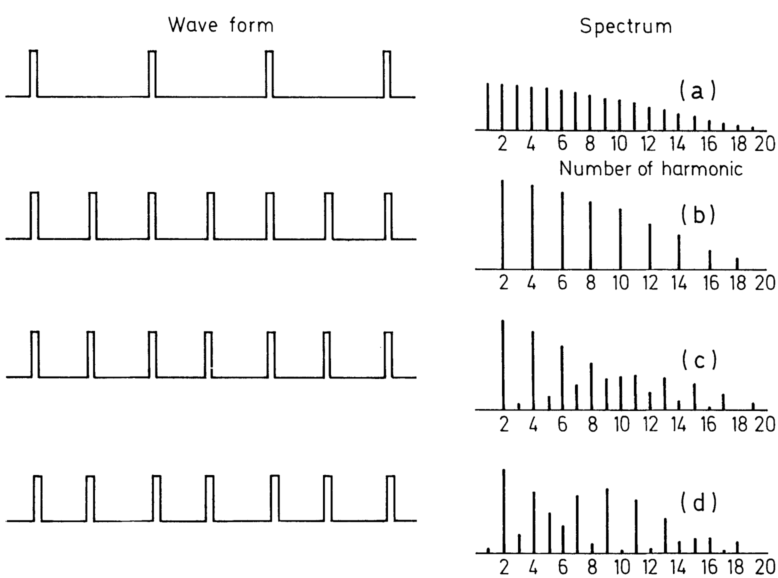

Fig. 4a—d. "Modern" interpretation of Seebeck's principal experiment. In (a) and (b,) all pulses are equidistant. In (c) and (d), their distances alternate (see text). After Schouten (1940a)

Ohm's definition was formulated in such strict terms as a strong foothold against hypotheses put forward by SEEBECK two years earlier (1841). SEEBECK used an acoustic siren as sound generator. Figure 3, taken from HELMHOLTZ' book, shows its construction. A disc A, provided with a series of holes is rotated by way of a string attachment f. Air is blown through a tube c so that a puff of air can come out whenever one of the holes is opposite the end of the tube. When the disc rotates with constant speed and the holes are equally spaced around the circumference, the resulting signal consists of a periodic series of puffs. In a stylized form, the waveform can be considered as a periodic series of rectangular pulses (see the upper trace of Fig. 4). Such a waveform has many Fourier components as the spectrum (also shown in Fig. 4) illustrates. Theoretically, the lower harmonics have almost equal amplitudes. (It goes without saying that the actual observation and analysis of acoustical waveforms could not be carried out until much later.) The corresponding part-tones can be analyzed by the ear and their loudnesses are not very different. SEEBECK reported that the fundamental dominated.

4. Seebeck's Crucial Experiment

Now, SEEBECK constructed a signal in which the harmonics were of highly unequal strength; he did this in two stages. His disc contained four concentric rows of holes. In the first row, the angular distance between the holes was 20°; in the second, the distance was 10°. It is clear that the tone will sound one octave higher when the tube of the siren is moved from the first to the second row. The resulting signal and its associated spectrum (highly idealized, of course) are shown in the second panel of Fig. 4. The components can again be regarded as a harmonic series but with a fundamental frequency that has become twice as high.

SEEBECK's crucial experiment – the second stage – was carried out with the third and the fourth rows of holes. In these rows, the holes are not equidistant. In the third row, they are alternatingly 9.5° and 10.5° apart. The basic period of repetition thus encompasses not two but four holes, and the fundamental component is the same as for the first signal. However, since the waveform (see the third panel of Fig. 4) resembles so much that of the second row, it comes as no surprise that the second, fourth, sixth, etc. harmonics dominate and that the odd harmonics are rather weak. This is especially true for the fundamental component; the latter is so weak as to be barely visible in the figure. The fourth row of holes has distances of 9° and 11°, respectively. Waveform and spectrum, as shown by the fourth panel in the figure, should be self-evident now. The great surprise of the experiment is that the step from the second to the third row is accompanied by a change of the pitch one octave downward. If the pitch were associated exclusively with the part-tone (partial) at the fundamental frequency, it would be very weak. SEEBECK concluded that the strength of (what we now call) a part-tone, is not determined by one (objective) Fourier component. Furthermore, he posed the statement that the pitch of a musical sound is not determined by the frequency of the lowest Fourier component but by the period of the signal's waveform.

It is to be noted that the change in period between the second and the third row of holes could also have been brought about by using only one row of holes and two tubes through which compressed air is blown. When the distance between the pipes is varied, all situations depicted in the figure can be obtained. Perhaps SEEBECK used this method when he discovered the effect. The principal point challenged by SEEBECK's experiments was the alleged quantitative connection between the strength of a partial and the loudness of the corresponding part-tone. When we realize how crude the experiments were in those times, we marvel at the ingenuity with which SEEBECK's attack was launched.

OHM's reply (1843) contained the aforementioned apodictical formulation of what later became known as Ohm's acoustical law, a great number of calculations proving, amongst other things, that the fundamental component must have been contained in SEEBECK's signals and no reference to any conclusive experiment that would contradict Seebeck's contentions. SEEBECK (1843) formulated his criticism a second time, devoting some attention to the effects of the width of the pulses (see Fig. 4) with respect to the period, and showing once more that in his experiments with the third and the fourth rows of holes, the subjective strength of the first and the second harmonics did not correlate with the objective strength. He concluded that each (subjective) part-tone was not simply the result of one Fourier component. He even went as far as to suggest that a number of higher harmonics might collaborate to enhance the strength of the first, or that these may produce a tone with the pitch of the fundamental (without the fundamental present). Real forebodings of the later residue theory!

5. Ohm's Reply; Helmholtz

OHM's final paper on the question was more authorative and witty than scientific. OHM admitted that he could not use his ears in experiments because "... nature had completely denied him a musical ear" (1844). He closed the argument by stating that SEEBECK's deviating observations were simply due to an auditory illusion. In his monumental book, *On the sensations of tone as a physiological basis for the theory of music*, HELMHOLTZ (1862, 1896, 1897, 1954) devoted close attention to this matter. On the basis of his arguments, he agreed with OHM, and it is most remarkable to notice that this argumentation just does *not* concern the points SEEBECK made. SEEBECK's observations led to the conclusion that some part-tones could be appreciably louder than would follow from their FOURIER counterparts. HELMHOLTZ' arguments, on the contrary, concerned mainly the reasons why part-tones could be judged to be weaker than would follow from the theory. In any event, OHM's and HELMHOLTZ' authoritative works remained the principal bases for the theory of the perception of pitch until about one century later.

One important reason for the impact of HELMHOLTZ' work was that he provided a possible physiological explanation for auditory analysis. He postulated that different parts of the organ of Corti in the cochlea are tuned to different resonance frequencies. In this way, each FOURIER component of a sound excites only a single stretch of the organ of Corti, and the nerve fibers originating from that stretch carry only information about that component. This theory also accounted in an elegant way for the fundamental law about tone quality (timbre) that HELMHOLTZ discovered. The law states that the timbre of a sound depends on the distribution of the Fourier components with respect to intensity and that timbre does not depend on the phases of the components (HELMHOLTZ was careful to leave the matter for components with a small pitch difference undecided).

The studies by VON BÉKÉSY (1960) proved that HELMHOLTZ' assertions were essentially right, with two exceptions. The first is that the apical part of the cochlea is tuned to the lower frequencies and the basal part to the higher frequencies, and not the other way around, as HELMHOLTZ supposed in his earliest works. The second point is that the mechanical damping in the cochlea is fairly high so that components will not be well resolved by the cochlear mechanism. Still later studies have shown that the situation is far better on the level of the activity of individual cochlear nerve fibers. According to the most modern findings, the frequency selectivity at this point is rather high and certainly sufficient for separating the lower harmonics (the explanation of these facts in terms of cochlear mechanics and physiology is quite another matter...). See the chapter on frequency selectivity of the cochlea by EVANS in Vol. V/2.

HELMHOLTZ also suggested a mechanism by which the impression of the fundamental pitch could possibly be enhanced. He described at great length the various types of distortion products, notably the difference tone, but he did not apply this directly to the problem brought forward by SEEBECK.

6. Fletcher, the Missing Fundamental, Distortion

From the time that vacuum tubes could be used to generate and analyze signals, research in physiological acoustics received a great boost. The problem of the pitch of complex tones received attention because of the fact that telephone exchange circuits have the property that frequencies corresponding to the fundamental of the human voice are attenuated quite heavily. Despite the absence of the fundamental (and of several of the lower components), the human voice via the telephone sounds with the natural pitch. In the time when FLETCHER wrote his first book, *Speech and Hearing* (1929), there was no doubt about this fundamental property; the phenomenon was commonly referred to as "the case of the missing fundamental". The general consensus was that aural distortion was the cause of the phenomen; the missing fundamental was restored as a distortion product. That this is quite feasible is evident when we realize that all pairs of consecutive harmonics give rise to a difference tone with the same frequency as the (objectively missing) fundamental. FLETCHER's quantitative estimates (1929, 1931) showed that this explanation could well be correct for the high levels of signal strength as used in telephone conversations. Although FLETCHER once described (1924) that the pitch of a multi-component signal without fundamental remains the same when the sound is made weaker and weaker toward the limit of audibility, no further mention is made of this property and the distortion hypothesis is considered adequate.

7. Place Theory

The "place theory" of auditory frequency analysis was established when the theory of HELMHOLTZ on resonance in the cochlea was formulated. According to the place theory, the components of a sound are directed to various locations in the cochlea, each component producing the largest vibration at the location that is tuned to its frequency. The process is referred to as; "spatial frequency analysis". In HELMHOLTZ' original conception this analysis was rather precise; each component stimulating a circumscript region of the cochlea. In later times amendments were put forward, as described above. Although we now know that at the level of responses of auditory nerve fibers, the analysis is fairly precise, this does not imply that we can explain all auditory phenomena of frequency resolution on the basis of place theory alone. In the sequel, we will have numerous opportunities to come back to the issue of the place theory; we will refine it, we will describe its limitations, etc. One point should be stressed here, however. In its purest form, the place theory assumes perfect isolation of all components in the cochlea. In this form, it is in excellent agreement with the fundamental law on timbre, as stated above. The actual analysis carried out by the cochlear mechanism will not be perfect, of course. Then, there will be components that are not analyzed perfectly in the cochlea. We may expect that the ear will not be "phase deaf" for the components that are incompletely resolved by the place mechanism in the cochlea. In this form, we may appreciate that there exists a fundamental relation between cochlear frequency analysis and phase sensitivity of the auditory system.

D. The Schouten "Period" (The Residue Theory)

1. Two Place Principles

The demonstration by VON BÉKÉSY of the spatial frequency analysis of sound in the cochlea gave rise to the formulation of the "place theory" of hearing. As a matter of fact, it is necessary to distinguish two principles in this theory. The *first* principle is that a sound of a particular frequency will be directed in the cochlea to a certain location where it will provoke the largest vibration amplitude. The location is specific for the frequency; high-frequency signals will stimulate the basal part of the cochlea, low-frequency tones, the apical part. This part of the theory is well substantiated by experiments (in point of fact, better by more modern experiments than by VON BÉKÉSY's studies which were directed at gross movement patterns). The *second* part of the place theory is founded on the possibility of the auditory analysis of part-tones corresponding to the partials of a sound signal. This principle states that stimulation of nerve fibers at a particular location gives rise to a tone sensation with a pitch corresponding to the frequency that is characteristic for that location. In short, the first place principle says: "frequency" gives "place"; the second principle says "place" gives "pitch". The second principle was tacitly assumed to be true until the Dutch scientist, SCHOUTEN, took up SEEBECK's experiments of one century earlier again and proved that the situation is really more complicated (SCHOUTEN, 1938). It is his study that led to the concise formulation of the "residue" theory (SCHOUTEN, 1940a).

2. The Optical Siren

SCHOUTEN used an optical siren to produce periodic signals of any desired waveform and to combine (superimpose) different signals. A disc with a series of narrow radial slits is rotated in front of an illuminated surface. A mask, in

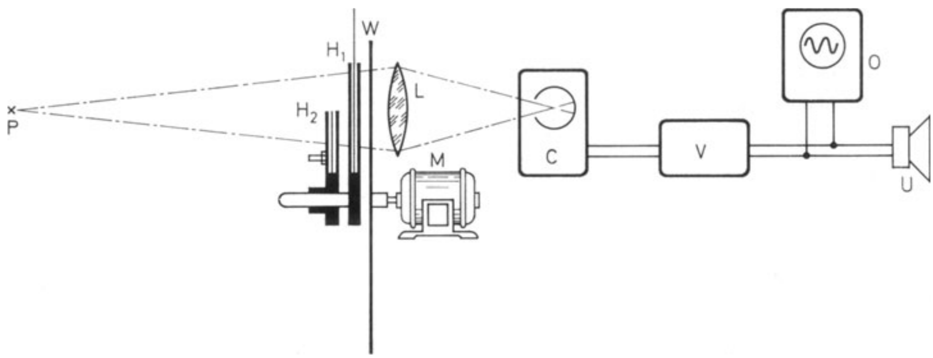

Fig. 5. Optical siren for producing synthetic sound. *P:* light source, H_1 and H_2: holders for masks, *W:* rotating disk with slits (driven by motor M), *L:* lens, *C:* photo-electric cell, V*:* amplifier, *O:* oscilloscope, *U:* loudspeaker. After SCHOUTEN (1939)

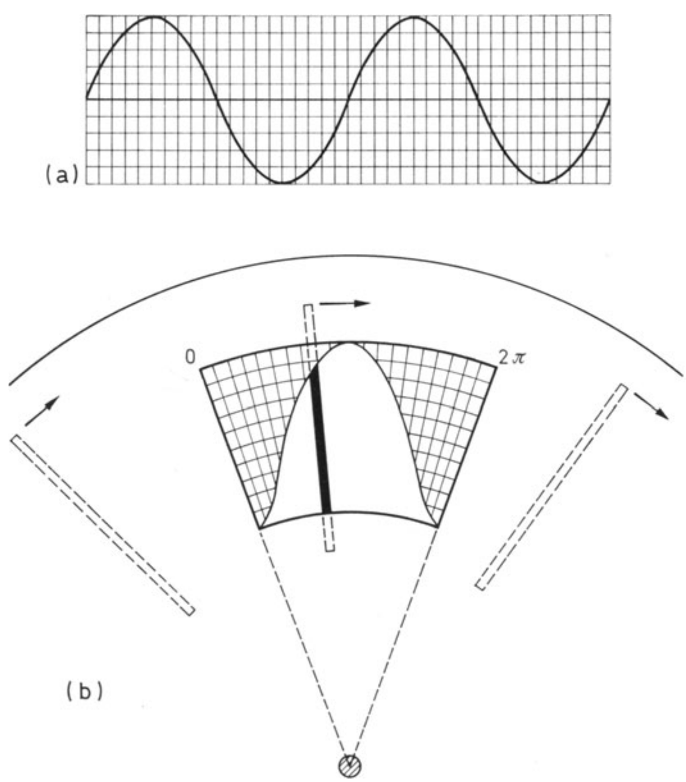

Fig. 6a and b. Sinusoidal waveform (a) transformed into a mask (b). The slits of the rotating disk are shown dotted. After SCHOUTEN (1939) the form of a sector, is put in front of the disc so that the slits pass it successively. As a result, the amount of light transmitted varies as a function of time t, as prescribed by the form of the mask. If the slits in the disc are equidistant, the resulting light signal will be periodic. A photosensitive cell collects the transmitted light and converts it into an electric signal. In its turn, the electric signal drives a loudspeaker via an amplifier. Figure 5 shows the arrangement in detail, and Fig. 6 shows how the mask should be cut in order to obtain a sinusoidal signal.

It is fairly easy to generate a periodic series of pulses with the instrument described in Figs. 5 and 6. The frequency of repetition was 200 Hz in SCHOUTEN's experiments. The spectrum contained many harmonics of which the intensity decreased slowly with increasing index i (see Fig. 7a). The signal was presented to the ear at a sound level of about 40 dB above the threshold of audibility. The tone quality was very sharp, the pitch corresponded to the frequency of the fundamental. With the proper amount of training several of the lower part-tones can be aurally isolated. We shall refer to this sound as sound *a*.

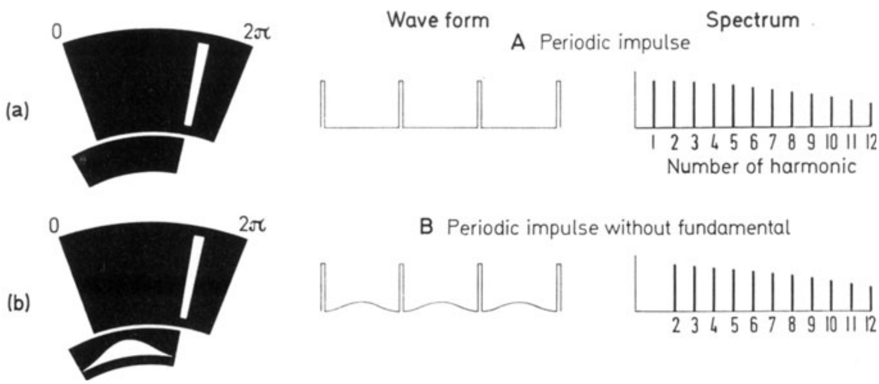

Fig. 7. (a) Periodic pulse series — waveform and spectrum. (b) Periodic pulse series with the fundamental component suppressed. The left-hand part of the figure shows the two masks used in the optical siren. After SCHOUTEN (1940a)

3. Schouten's Principal Experiment: Cancellation of the Fundamental; the Residue Hypothesis

By the addition of a second sector to the mask, the fundamental component can be removed from the signal. The added sinusoid just compensates for the corresponding component. The resulting signal and spectrum are shown by Fig. 7b. This situation is the purest form of the "case of the missing fundamental". The adjustment of amplitude and phase of the added signal is carried out with the help of a wave analyzer (an instrument which indicates the strength of the Fourier component at a specific frequency). The intensity of the fundamental component in the combined signal can thus be reduced to less than 0.5% of the original value. The sound (henceforth to be called sound *b*) looses some of its "body"; the timbre becomes somewhat sharper, but the pitch remains the same. After this, the adjustment of the compensation can be made subjectively. This proves to be fairly easy, and the part-tone, *i. e.*, the subjective component corresponding to the fundamental can be made to vanish completely. Quite contrary to expectation, the objective and the subjective settings coincide. This means that no appreciable distortion component of the fundamental frequency is formed in the ear. In other terms, the fundamental part-tone that we hear is exclusively due to the fundamental component in the sound. When the latter vanishes, so does the former, and vice versa. Schouten employed other means to verify that there was indeed no fundamental component generated in the ear, but we shall leave that aside. The surprising thing is that *the pitch ascribed to the sound b with the fundamental missing is the same as that of the original sound a.* It is definitely one octave lower than the frequency of the lowest Fourier component in sound *b*, 400 Hz. This experiment proves that the pitch of the sound is not the pitch of the lowest Fourier component present and that the fundamental is not reconstituted by nonlinear distortion.

4. New Formulation of Ohm's Law

Let us now return again to sound *a*. By concentrating the attention, the fundamental, the second and the third harmonic can be heard separately in the sound. As to actual perception, the sound consists of the pure tones of 200, 400, and 600 Hz[1] plus "something else". Sound *b* contains part-tones of 400 and 600 Hz plus "something else". In the latter case, the sound has a very sharp tone quality and this sharp tone has a pitch equal to that of the (missing!) fundamental. In the sharp tone, some part-tones, *e. g.* the 4th and 5th harmonics, may be distinguished, but the main body of it is due to the combined impression of the remaining higher-numbered components. The main point is that this sharp tone, this combined impression of unresolved components, has a pitch, in this case equivalent to the pitch of the fundamental, *i. e.* 200 Hz.

By successive removal of the next few harmonics, it can be ascertained that the sharp tone is indeed produced by the higher-numbered harmonics. Such a subjective tone is called by SCHOUTEN a "residue". Hence, *a "residue" is a subjective component of sound sensation that is the result of the combined impression of higher harmonics, namely those harmonics that cannot be resolved by the auditory organ.* In SCHOUTEN's own words (1940a), Ohm's law of subjective sound analysis may be extended as follows:

"1. The ear analyses a complex sound into a number of components each of which is separately perceptible.

2. A number of these components corresponds with the sinusoidal oscillations present in the inner-ear sound field. These components have a pure tone quality.

3. Moreover, one or more components may be perceived which do not correspond with any individual sinusoidal oscillation but which are a collective manifestation of some of those oscillations which are not or scarcely individually perceptible. Those components (residues) have an impure sharp quality."

If, in the light of this development, the difference between sounds *a* and *b* is considered again, we note that sound *a* contains *two* subjective components with the same pitch, 200 Hz. One has a pure-tone quality and is due to the presence of the (objective) fundamental component. The other, having a sharp tone quality and a rather large loudness, has a different origin, namely the mass

[1] Part-tones with pitches corresponding to 200, 400, and 600 Hz.

of unresolved components. Both subjective sound components, the fundamental part-tone and the residue, have the same pitch. In sound *b* only the residue is the bearer of the pitch – the *higher* components giving the lower *pitch*. Some part-tones may be recognized in it, with different pitches, but the pitch of the sound as a whole is determined by the residue. It should be clear from this description that proper recognition of the residue as a subjective sound component requires a good deal of abstraction and introspection. In spite of the careful and detailed description given by SCHOUTEN, the rules of the game have been followed only by very few of the later scientists working on the problem.

5. The Period as the Cue for Pitch

The next question which Schouten posed, was: what physical property of the set of higher harmonics determines the pitch of the residue? There are two obvious possibilities: all components in the sounds used have mutual frequency differences of 200 Hz, and the waveform has a repetition period of 1/200 sec. In order to solve the question, SCHOUTEN produced a signal in which these two parameters differ from one another. Consider Fig. 8a. Here a pulse series with *two* pulses per period of 1/200 sec is shown, together with its spectrum. Referred to the time interval of 1/200 sec and the fundamental frequency of 200 Hz, all components are even multiples, that is, they are integral multiples of 400 Hz. It is clear that the pitch of this sound will be 400 Hz. The second sound, Fig. 8b

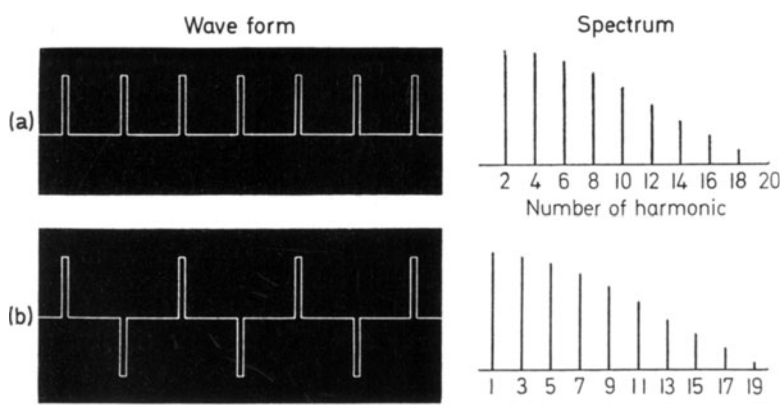

Fig. 8. (a) Periodic pulse series, fundamental frequency: 400 Hz. (b) Series of pulses with alternating signs, fundamental frequency: 200 Hz. After SCHOUTEN (1940a)

is a series of pulses with alternating signs. The components have frequencies that are odd multiples of 200 Hz, and the difference between any two successive harmonics is 400 Hz. The period of the waveform is, of course, 1/200 sec. SCHOUTEN found that the residue of this second signal, although being rather weak, had a pitch of 200 Hz. He concluded that:

"the ear thus ascribes a pitch to a residue by virtue of the periodicity of the total waveform of the harmonics which are responsible for this residue" (SCHOUTEN 1940a).

The notion of periodicity has been considerably refined by later studies and we shall come back to it later on. It has also been found that the signal of Fig. 8b is not the most suitable one for deciding on the physical parameter governing pitch, but, within the framework of an extended definition of the concept of "period", SCHOUTEN has been proved right.

6. Seebeck Revisited

The extension given by SCHOUTEN to OHM'S law is reminiscent of the extension given by SEEBECK almost one century earlier. SEEBECK (1843) wrote:

"... it follows that the tone of the siren, can at least be amplified by audible overtones. If this is the case..., then it can be supposed that it holds true for a tone that is not accompanied by harmonics, ... so that a tone with pitch m is not necessarily of the form $a \cos 2\pi(mt + \theta)$ but it may contain also terms of the form $a_s \cos 2\pi(smt + \theta_s)$ where s can be very large and a_s very small. It should not even be excluded that those latter terms, when taken together, can evoke a tone with pitch m when a term of the form $a \cos 2\pi(mt + \theta)$ is not present".

The case of the missing fundamental!

On the next page of SEEBECK'S remarkable paper, we find the germs of the later residue hypothesis:

"It appears to me irrefutable that the higher terms of the cosine series control the varieties of the tone. I consider it very likely that those terms of the series which are not individually perceptible take part in this, and that these terms by their common period length evoke a specific impression on the auditory organ."

SEEBECK'S experiments with the unequal pulse series (see Fig. 4) also receive a satisfactory explanation in the light of the residue hypothesis. The weakness of the fundamental component only implies that the corresponding fundamental part-tone is weak or inaudible. Of the higher harmonics, there are many that show the basic period in their combined waveform; hence, the residue will have a pitch corresponding to the frequency of the fundamental. With the optical siren, it is far easier to manipulate the parameters of the test sound and to study what happens than with the acoustical siren. SCHOUTEN reports that he repeated SEEBECK'S experiments also with an acoustic siren, only to find that the effects, though present, were extremely weak and variable and that much masking noise was produced by the instrument.

7. Mechanism of Period Extraction

In a third paper, SCHOUTEN described (1940b) a possible cochlear mechanism by which the period of a complex waveform can be detected. To understand his argument, it is essential to remember that the harmonics lie the closer together (in the musical sense) the higher their index. Compare again Fig. 2. A sinusoidal signal will produce vibrations in only a limited stretch of the basilar membrane, the location being a function of the frequency. This is the first part of the place principle. The second part ("place" giving rise to "pitch") is questionable in the light of the residue theory since high components can give rise to a low pitch. We shall see presently how this part should be modified. But the first part holds true unmistakenly. Only the extent of the excited region has been left unspecified, or, what is equivalent, the width of the response curve (resonance curve) at any particular location. In view of the relation between frequency ratios and musical intervals, it is logical to assume that the width of the response curve should be expressed as a musical interval, or, in other words, as a fraction of the center frequency. HELMHOLTZ was concerned about this measure; he reasoned that the response should be less than 0.1 of its maximum value for a frequency that deviates one whole tone (two semitones) from the frequency of maximum vibration. The difference of 3 dB corresponds to 0.2 semitone in his conception (see Fig. 9). At the time of SCHOUTEN's studies several sets of data about the so-called critical bandwidth (*i. e.* the bandwidth over which the auditory system integrates sound intensity) were known but these appeared to differ widely. SCHOUTEN assumed that the frequency response characteristic at any location had a half-value width of at least a semitone. That means that the lower harmonics, being an octave, a fifth, a fourth, etc. apart, will chiefly excite narrow

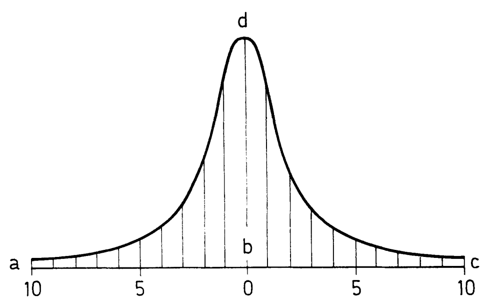

Fig. 9. Sharpness of resonance as estimated by HELMHOLTZ. Abscissa: 5 units are equivalent to a semitone. After HELMHOLTZ (1896)

stretches of receptors that are well isolated from one another by regions in which the response is comparatively small. Conversely, the receptors in those stretches will respond almost solely to isolated harmonics. Figure 10 shows the situation in SCHOUTEN's terms. Figure 11 shows the same situation in the light of more modern data on auditory frequency resolution; this figure is modified from PLOMP's doctoral dissertation (1966). Along the vertical axis, a number of response curves for selected equidistant locations along the length of the basilar membrane are plotted. The basal part of the cochlea corresponds to the top of the figure. The vertical axis is a logarithmic frequency axis so that equal musical intervals appear as equal distances along the basilar membrane. The width of the response curves appears as constant in this representation. (The response functions should be imagined to form a continuous set, only a few of them are drawn.) On the bottom of the figure, a periodic series of pulses is shown; this is the waveform with which the cochlea is supposed to be stimulated. This signal has a large number of Fourier components; for reasons of simplicity, these can all be assumed to be of equal amplitude. Their "proper" positions along the vertical axis are unequally distributed of course. In Schouten's figure, the positions for several of them are indicated. In Plomp's figure one can use the indicated numerical values for the frequency as a guideline. The remainder of these figures shows the waveforms that can be expected at selected positions along the vertical axis. The lowest three harmonics give rise to nearly sinusoidal vibrations at the positions corresponding to their frequencies. But note, what happens to the highest harmonics! At a certain position of the basilar membrane a number of harmonics

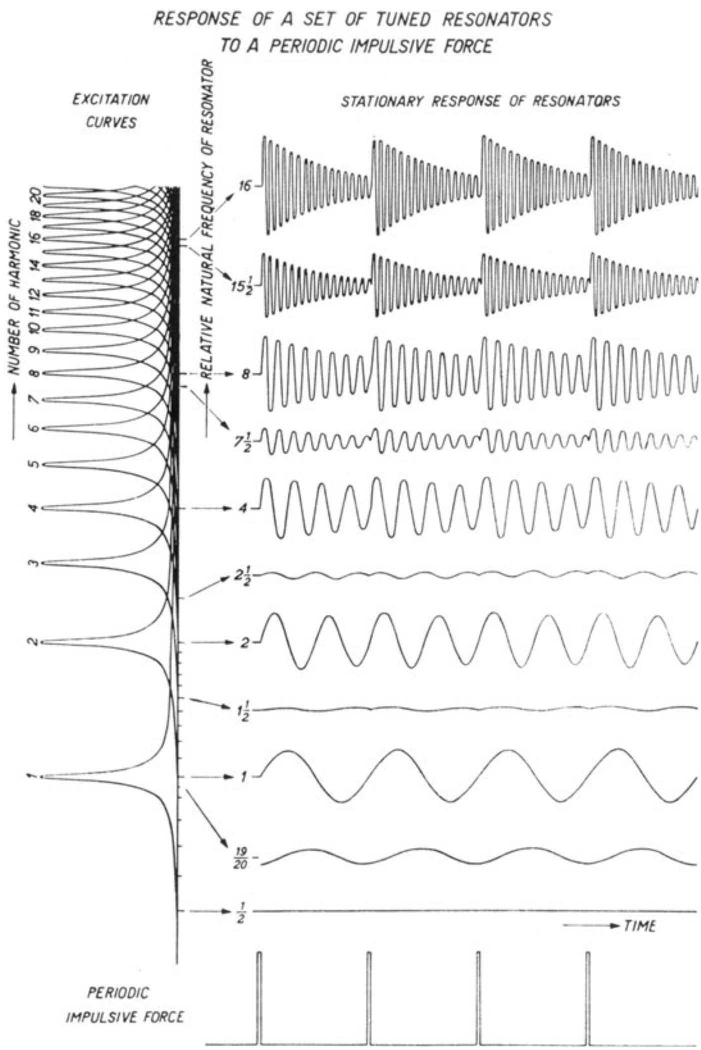

Fig. 10. Periodicity of signals and aural resolution of components: Schouten's explanation of the residue. After Schouten (1940b)

fall within the top of the response function and by interference they produce a pattern of vibration with a nonuniform amplitude. In fact, the combined excitation of such a set of higher harmonics gives rise to bursts of oscillation, and the period of these bursts is the same as the period of the pulse series that excites the cochlea. The model implied in these figures may not be quantitatively cor-

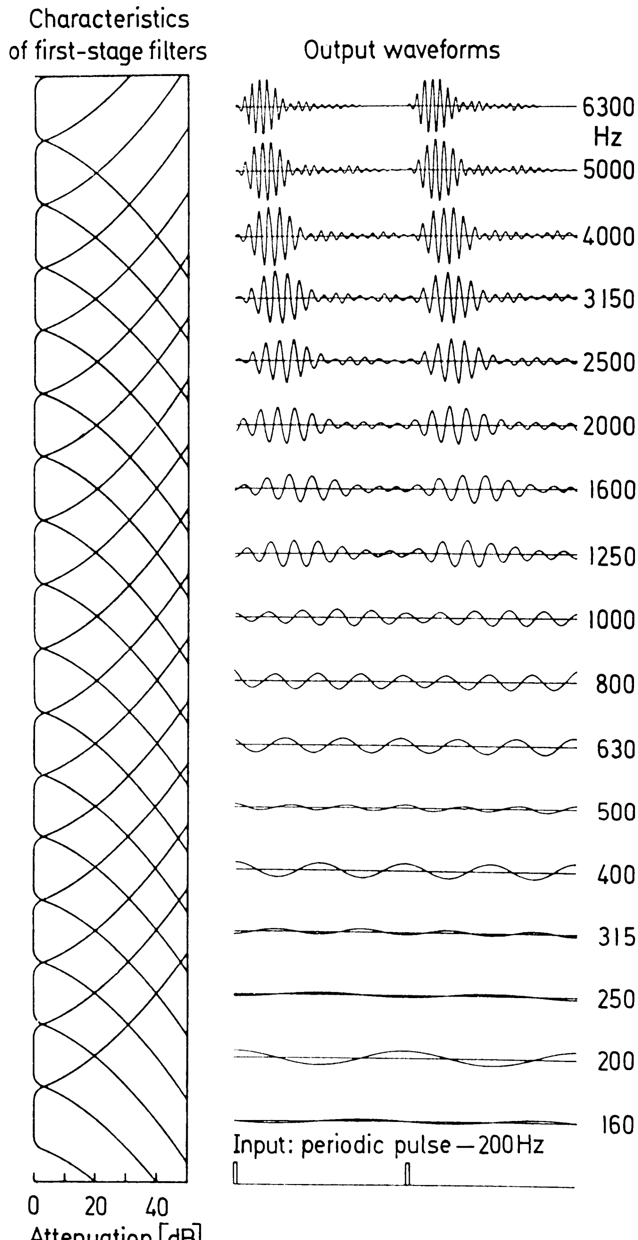

Fig. 11. Periodicity of signals and aural resolution of components: a more modern representation. After Plomp (1966)

rect, but it displays the properties to be expected from all resonating systems of analysis with a limited power of frequency resolution.

It is seen that the consideration of the first place principle, but with a limited resolving power built in, automatically leads to a response revealing the basic periodicity of the stimulus in those sections where harmonics are not resolved. That means that the second principle in the place theory should be replaced by a principle that states: when the stimulation of a particular location in the cochlea is of constant amplitude, a pitch is perceived corresponding to that location, but when the stimulation occurs in bursts, the ensuing residue has a pitch corresponding to the periodicity of the bursts, and no pitch corresponding to the stimulated location is perceived. In this theory, the nerve fibers are supposed to transmit both the quantity and the quality of the excitation of the receptors. The former is important in loudness and timbre; the latter, by virtue of the periodicity, for pitch.

8. Corollaries; Inharmonic Signals

Several phenomena acquire new perspectives in the light of the residue theory. The first is the "case of the missing fundamental" as it is encountered in the form of the telephone voice. It is clear now that the lack of the lowest harmonic, or of the lower few harmonics, will change the timbre somewhat, but there will be enough harmonics left to give a residue which lends its pitch to the sound. Whether or not the fundamental is re-instituted by nonlinear distortion is immaterial. The second point concerns the existence of more than one type of sound with the same pitch. In the signal of Fig. 7a two subjective components are present, a part-tone and a residue, both with the same pitch. In the signal of Fig. 7b, the former has disappeared, and the residue remains. The pitch is, again, 200 Hz. One can combine this second signal with a pure tone (sinusoidal signal) of *e. g.* 203 Hz; it is highly remarkable that there occur no beats (SCHOUTEN, 1940c). One hears a part-tone and a residue with slightly different pitches that coexist without causing beats. The conclusion is that beats can only occur when two signals, *e. g.*, sinusoids, stimulate one and the same receptor in the ear. Later on, we shall encounter examples of signals in which different, non-overlapping parts of the spectrum give rise to different residue pitches. The problem is, then, which of these pitches will be ascribed to the entire sound. The example cited above is the first of such signals with conflicting information (see Section F. 3).

The third phenomenon for which the residue theory contains an explanation, concerns the strike note of (carillon) bells. In general, none of the partials of a bell – and most of these are not harmonically related – corresponds to the pitch ascribed to the sound of the bell when it is struck. Furthermore, the strike note has a sharp timbre whereas each of the partials has the soft and mellow timbre characteristic for a sinusoidal signal. The hypothesis was advanced (SCHOUTEN, 1940c) that the strike note is a residue evoked by several of the partials that stand to one another in a harmonic relation. Closer inspection of the spectrum of the partials (Eigentones, as they are usually called in this case) reveals that several of these do have the required relation and that the residue pitch these partials can give corresponds to the pitch of the strike note. "Good" and "bad" bells would then be distinguished by the presence or absence of such special partials among the Eigentones.

In one of SCHOUTEN's papers (1940c), an experiment is described that proved to be extremely important for later studies. In a purely periodic signal, the harmonics form a *harmonic series*, *i. e.* a series in which the frequencies are integral multiples of a fundamental frequency f_0 [see Eq. (3)]. By a special procedure, borrowed from the technology of telephone carrier circuits, it is possible to modify a signal so that all harmonics are shifted by an equal amount of Δf. The frequencies

$$f_i = i \cdot f_0$$

then become

$$f_i' = i f_0 + \Delta f \tag{4}$$

and they no longer form a series of successive harmonics. We shall refer to signals with this property as *inharmonic* signals. For small values of Δf, *e. g.* 40 Hz, the 200 Hz periodic pulse series retains its tonal character, but the pitch is slightly different. This shows once more that the pitch is not determined by a distortion product such as the difference tone, since the frequency difference of the components remains the same after the shift. The same procedure was applied to a musical recording. The observed relative pitch shift was about equal to the relative frequency shift that the components in the 1000–2000 Hz region undergo. This observation can be understood with the help of the concept of dominance as it was developed later by PLOMP and by RITSMA (see Section F. 3). For the moment, we may conclude that these observations on bell tones and frequency-shifted signals indicate the plasticity of our auditory system: our ear will not only accept components forming a harmonic series but it will try to form a residue pitch when the series is nearly harmonic. Just what relations among the components are needed in order to do this remains to be seen.

9. Introspection, Slow Acceptance

Acceptance of SCHOUTEN's findings and interpretations by the scientific world was very slow, and, perhaps, it has not even been completed in the musical world now that, scientifically, things have already taken another turn. It appears that only few of the criticizers of SCHOUTEN's hypotheses really applied the method of introspection by which the residue is defined. So it happens that HOOGLAND (1953), when trying to disprove SCHOUTEN's theory, only succeeded in confirming the existence of aural distortion tones, in normal as well as in abnormal ears, something that nobody would have doubted at all. HOOGLAND experimented with a residue signal generated by the method of synthesis: a number of components were generated by separate tone generators. The difficulty with which this leads to perception of a residue with a clear pitch – which difficulty is no doubt due to the problem of keeping the phases of the components constant – suggests that there really are some constraints involved in the perception of residue pitch. It was extremely difficult at that time to envisage the real consequences of these problems.

E. Wave Crests (the Impact of Periodicity Pitch)

1. Outline of this Part

Perhaps one reason for the slow acceptance of SCHOUTEN's results was the need for introspection to elucidate the main aspects of the residue. There was a definite need for convincing demonstrations that could easily be turned into well-reproducible psychophysical experiments. On the other hand, a small number of studies were published that touched upon the problem of the residue pitch from various angles, and it was attempted several times to bring such widely divergent studies into a common focus. However, the number of experimental variables is quite formidable. In this part we shall first briefly describe the demonstrations of the residue effect as they were developed by LICKLIDER. These demonstrations, as well as a few other experiments to be described, served as a basis for a novel theory on pitch perception. The theory, LICKLIDER's duplex theory, provides for a possible neural mechanism that carries out its operations on the neural output of the cochlear analyzer and that succeeds in extracting information about the period contained in the excitation waveform. The second part of this sub-chapter will be concerned with the extensions and modifications that were given to the residue theory by the present author around 1956. The third part describes in condensed form the great interest that temporal phenomena in hearing received until the period 1960–1965 in which a change in attitude began to be noticeable.

2. Licklider's Demonstrations

LICKLIDER's demonstrations (1954, 1955) continue to be important because they are so ingenuously concentrated on the main aspects of the problem. Moreover, they can be repeated easily with modern equipment and they are most convincing. The frequency of the test signals was not fixed but was made to vary in order to see whether the pitch of the residue is the kind of pitch of which melodies can be made. The signals presented were, alternatingly, sinusoids and periodic pulse series (with the fundamental removed but that is not necessary). Each signal lasted about 0.5 sec and was switched on and off smoothly (the latter refinement is not essential either). In each pair of signals, the frequency of the sinusoid was the same as the repetition frequency of the pulse series, and the sinusoid was the louder signal of the two. The successive pairs progressed up and down the scale in frequency as controlled by the experimenter. While the sounds were presented in this way, a low-frequency random noise signal was turned on. This noise was produced by feeding white noise through a low-pass filter that passed only the frequency components below 1000 Hz. The low-frequency noise was sufficiently strong to mask the low-frequency channels of the auditory system. As a consequence, none of the sinusoids was audible any more as soon as the noise was turned on. However, the sharp sound produced by the high harmonics of the pulse series could be heard through the noise and it retained its low pitch. After the noise was left on for several pairs, it was turned off again, and this sequence could be repeated. Without the noise, the sinusoids and the pulse series showed both the impressed pitch course, but, after the noise was turned on, the "melody", so to speak, was only audible from the pulse series and not from the tones.

This demonstration proves conclusively that the second place principle should be modified in the sense that the receptors in the ear normally occupied with high-frequency components are really able to mediate a low pitch. In other words, there is not a one-to-one correspondence between "place" and "pitch", although there is a fairly precise projection from "frequency" (in the FOURIER sense) to "place".

3. Chopped Noise — Direct Evidence of Periodicities(?)

We have observed in the preceding sections that recognition of a temporal structure like periodicity is linked with insufficient spectral resolution. In most types of signals, it is impossible to vary the temporal structure without causing a corresponding variation in the spectral composition. There is one exception: interrupted white noise. The spectrum of noise can only be defined in statistical terms (see Section B. 2), and for white noise the average noise power content (per unit interval) is the same for all frequencies. If a white noise signal is repeatedly switched on and off at a rate of, say, 100 times per second, the spectrum, when defined in this sense, will remain the same. This is because switching a signal on and off can be regarded as a multiplication with a signal alternating between 0 and 1, this procedure causes many components to be generated that were not present in the original noise signal. Each frequency is equally likely for these new components; hence, the spectral content of interrupted white noise is the same for all frequencies, just as for uninterrupted white noise.

For high switching rates, interrupted noise cannot be distinguished from continuous noise by the ear. However, when the switching rate is low, *e. g.*, once per second, we can easily hear the temporal sequence, and the theoretical spectral content will be completely irrelevant for the ear. Somewhere in between, there is a transition. This phenomenon was studied by MILLER and TAYLOR (1948). A white noise signal was switched on and off with a duty cycle of 0.5 or more. The interruption rate could be varied between 0.1 and 5000 Hz. It was attempted to determine the pitch of the interrupted noise by comparison with a pure tone. Most listeners could produce a reliable match for interruption rates up to 200 or 250 Hz. Above 300 Hz, it was impossible to match the "buzz" of the interrupted noise to a pure tone. In a second experiment, the just noticeable variations of interruption rate were determined. (Of course this was only possible below 300 Hz.) For interruption rates below 100 Hz, the values were about twice as large as the just noticeable variations of the frequency of a pure tone. This suggests that detection of the temporal structure ("intermittency") is a major factor in the frequency discrimination of very low tones. It was not mentioned in the paper of MILLER and TAYLOR whether the "pitch" ascribed to interrupted noise is tonal enough so that "melodies" can be built from it. In the experience of the present author, the pitch is much less clear than the pitch of a residue, and not musical (DE BOER, 1956a).

4. Licklider's Autocorrelation Theory

Although auditory theories of the type of "telephone theories" had almost ceased to exist after VON BÉKÉSY's findings about the frequency-to-place transformations in the cochlea, they received some support after cochlear responses at the neural level could be studied. It soon became known that the fibers of the auditory nerve were able to follow to some extent the temporal pattern of the sound stimulus. The question is, of course, whether this frequency following plays a role in pitch perception. In his book, *Theory of Hearing*, WEVER (1949) strongly supports the hypothesis that it does. He assumed, on good grounds, that various neurons connected with a given location could follow a stimulus beyond the upper limit by frequency division. The firings of the neurons then rotate, as it were, between the members of the group of neurons; this is the famous "volley principle". At high frequencies, the volley principle may fail, whereas the place principle may fail at low frequencies. WEVER advanced the theory that both place and frequency determine pitch, the former being dominant at high frequencies and the latter at low frequencies. It is to be noted that this holds only for (sinusoidal) components and for the pitch of part-tones; the complexities of residue-like signals, as they are hypothized by SCHOUTEN, are left out. A similar attitude, the space time-pattern theory of hearing, was expressed by FLETCHER in his second book (1953), but it should be remembered that FLETCHER favoured the distortion hypothesis as the explanation of the phenomenon of the missing fundamental.

A theory directly aimed at *periodicity pitch* (under which term the residue pitch concept became known by English and American authors) was formulated by LICKLIDER (1951). It is specifically concerned with the mechanism necessary for extracting information about the period from the excitation. The first point of importance in this theory is to realize that the vibration at each location of the cochlea contains information about the frequency content of the stimulus as well as about the way the stimulus varies with time. When components are not resolved, the temporal pattern reveals the basic periodicity of the sound stimulus as Figs. 10 and 11 show. Neurons connected to this location will carry the same two types of information: by virtue of their point of origin about spectral content and by virtue of their frequency-following properties about the temporal pattern. The representation in neural terms is necessarily less accurate but that is not important at the moment.

In the theory, a network of neurons is assumed to be connected to each location of the cochlea. Figure 12 shows this network in a diagrammatical form. The principal element is a chain of delay elements (neurons) B_i; each one relays

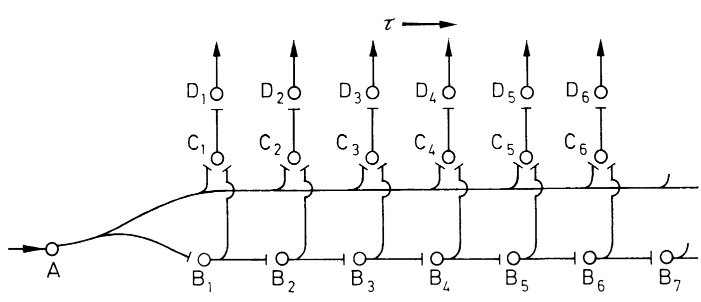

Fig. 12. Hypothetical network (neural autocorrelator), capable of measuring periodicity. See text. After LICKLIDER (1951)

the excitation at the input A after a certain delay. Each of the delay stages is connected to a "coincidence neuron" C_i which gives off an output pulse whenever it is excited nearly simultaneously at both its inputs. Consider first what happens if the system is excited by a stimulus without a definite time pattern. The delayed versions of the stimulus appearing at the neurons B_i have nothing specific in common with the original signal that is transmitted to all the coincidence neurons C_i. Hence, there will be no special pattern set up in the output neurons D_i.

Things are entirely different, however, when the excitation function is periodic. At one of the delaying stations, the output signal, being delayed by one period, is almost identical to the undelayed signal. As a consequence, the corresponding coincidence neuron is stimulated by two nearly identical input signals and that is the situation most apt to produce output pulses in this neuron. Somewhat further along the delay chain of neurons, the same situation arises; here, the delay amounts to two periods. At other places along the chain, there is no specific output. It is clear that a periodic signal evokes responses at particular locations along the neural network. In this way, the period of the stimulation is translated into the location (or, rather, locations) of the responding neurons; periodicity is encoded into a specific response pattern among the neurons.

In LICKLIDER's paper, it was shown that the analysis performed by this type of network is a form of what is mathematically known as *autocorrelation analysis*. As a matter of fact, the analysis is incomplete (a so-called running autocorrelation analysis) just as is required for the analysis of signals that can rapidly change their character.

The dimension along the delay chain may be labelled with the parameter τ. A similar arrangement is assumed to be connected to each point of the cochlea (see Fig. 13). Along the length of the cochlea, denoted by the dimension x, a limited-resolution spectral analysis is carried out. The resulting neural information can, thus, be labelled with two parameters, x and τ. Pure tones and pulse series each give rise to specific patterns of neural activity in these two dimensions.

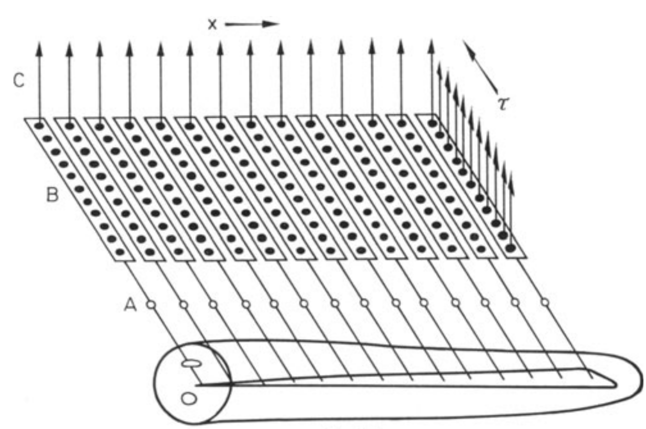

Fig. 13. Illustrating the duplex theory of hearing. After LICKLIDER (1951)

In both, the periodicity reveals itself as a succession of maxima along the τ-dimension; this may be responsible for the perceived pitch. From this two-dimensional representation, the theory derives its name, duplex theory of hearing. A related theory, in the form of a mathematical formalism, has been published by GOLDSTEIN (1957).

A later paper by LICKLIDER (1955) contains two extensions to the theory, the first of which is of considerable interest in neurophysiology. The pattern of activity, as it is analyzed by the set of networks described above, must transfer its information to higher centers in the auditory pathway of the brain. The projection may be unstructured at first, a random each-to-everyone projection. It may become structured through its own operation. The rule of modification may be that a functional connection is strengthened when it is often used and that connections which are not exercized will be weakened. When such a "plastic" system continues to operate with stimulations containing periodic signals, such as human vowels, the gradual reorganization may lead to projection of periodic points along the τ-dimension to a common point in an upper stratum of the brain. The main impetus for such a reorganization is the repeated occurrence of periodic sounds of low frequency or, rather, the occurrence of high-frequency signals with periodic envelopes. The principal result of such a process would be that periodicities detected by an autocorrelation-type of analysis could be brought into one sphere with the ordered series of excitations produced by cochlear analysis of sinusoidal signals. Pitch is not the only attribute to which such a process might apply; timbre may be another one. This would explain why listeners when presented with residue-like signals, especially those with only few Fourier components, have difficulty in separating pitch from timbre.

The second extension of the Duplex theory, which makes it into a Triplex theory, takes binaural phenomena into account. There are several types of pitch that can manifest themselves only by binaural presentation (in fact, by dichotic presentation), *i. e.*, different signals are led to the two ears. We will come to speak about these in a later section (Section H. 9).

5. Seebeck Re-Revisited

At a time that the residue theory was already more than ten years old and it continued to be challenged from several sides, the present author attempted to pinpoint once more the requirements for the existence of the phenomenon. Instead of an optical siren, like SCHOUTEN used, he built purely electronic equipment for the generation of signals. The first step consisted of rigourous application of SCHOUTEN's recipe: introspective realization that there are two possible bearers of pitch in a complex signal, the fundamental part-tone and the residue. This was considerably facilitated when the fundamental component was added in counter-phase to a periodic pulse series, just as in SCHOUTEN's central experiment. A repetition frequency of 200 Hz was confirmed to be optimal for such experiments. When the residue is easily recognized, repetition of SEEBECK's fundamental experiment (see Fig. 4b–d) once more confirms the independence of a residue and a part-tone. The experiment is performed by the addition of two periodic pulse series of 200 Hz with a variable delay between them. When pulses of one series fit exactly midway between those of the other, the pitch is the octave, 400 Hz, of the basic frequency. When they are slightly off-centre, the residue pitch jumps to 200 Hz, but the corresponding part-tone is inaudible. The lowest part-tone in the sound is still 400 Hz. An even more striking effect, which was tried but could not be demonstrated so easily in SEEBECK's days, is produced when three pulse series are combined. When all pulses are first equidistant and then one of the three series is shifted in timing, the pitch jumps down a duodecime or a twelfth (an octave plus a fifth). Other experiments were carried out with pulse series filtered by band-pass filters so as to attempt to "grow the residue as a culture". Because these experiments are also performed by other authors who appear to have exploited them more completely, they will not be described here (see Section E. 12).

6. De Boer's Experiments on Inharmonic Signals

The principal experiments of DE BOER (1956a, b) are extensions of SCHOUTEN'S endeavours with inharmonic tones described at the end of Section D. 8. The test signals were produced with a *modulator*, an electronic circuit that produces at its output a signal $z(t)$ which is the product of the two signals $x_1(t)$ and $x_2(t)$ at the two inputs:

$$z(t) = x_1(t) \cdot x_2(t). \tag{5}$$

This modulation process can also be considered in the frequency domain. Suppose that $x_1(t)$ contains a Fourier component of frequency f_1 and $x_2(t)$ a component of frequency f_2. Then the output, $z(t)$, will contain two Fourier components with frequencies $f_1 + f_2$ and $f_1 - f_2$, respectively. The modulator is a nonlinear circuit with the peculiarity that only the sum and difference frequencies are produced; the original components have disappeared completely. In DE BOER's experiments, the signal $x_1(t)$ was a sinusoidal signal, commonly referred to as the carrier signal. In accordance with later usage, the *carrier frequency* will be designated by the letter f. The second signal, $x_2(t)$, commonly called the modulation signal, was made to contain three components, with frequencies g, $2g$, and $3g$, respectively. This signal was derived from a pulse series with repetition frequency g by sharp low-pass filtering. Figure 14 gives a diagram of the setup; the figure indicates the nature as well as the spectral content of the various signals. When the modulation signal contains only the components g, $2g$, and $3g$, the output signal of the modulator contains the frequencies:

$$f - 3g,\ f - 2g,\ f - g,\ f + g,\ f + 2g,\ f + 3g. \tag{6}$$

Via a separate pathway, a fraction of the carrier signal (frequency: f) is added so that a full arithmetical series of frequencies arises. The series is harmonic only when f happens to be an integral multiple of g; in the other case, the series is characterized as an arithmetical series, *i. e.*, a series with constant frequency differences [cf. Eq. (4)]. In most of the experiments, f was sufficiently high so that the lowest Fourier component according to the series (6) had a positive frequency.

Let us start by assuming $f = 2000$ (Hz) and $g = 200$ (Hz). The resulting components then form a harmonic series:

1400, 1600, 1800, 2000, 2200, 2400, 2600 Hz.

Needless to say that the waveform is periodic. This output signal is presented to the ear of the listener via earphones. The signal constitutes a pure form of the residue; it is well-nigh impossible to distinguish aurally any of the constituent components. The residue has a clear pitch that is readily shown to be equivalent to 200 Hz, the fundamental frequency belonging to the series of component fre-

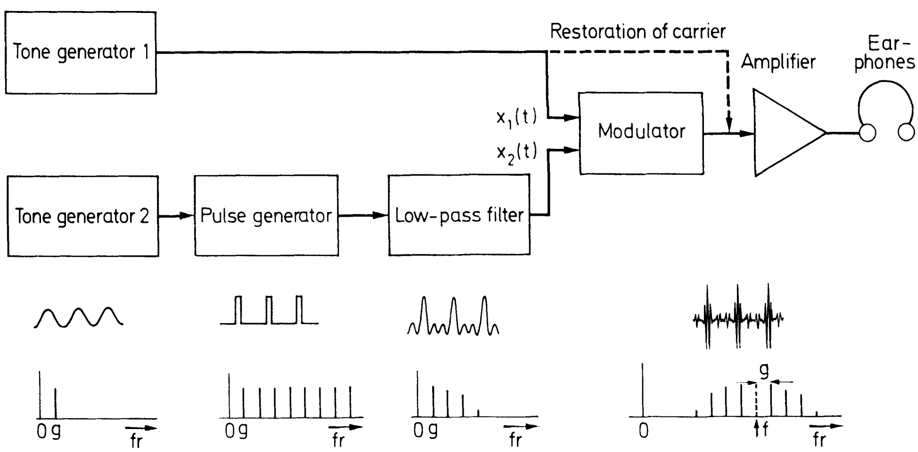

Fig. 14. Block diagram of DE BOER's signal generator. Waveforms and spectra of the signals at various points of the setup are illustrated. The spectral component shown by a dotted line is the carrier component (which is not passed by the modulator)

quencies. When instead of 2000 Hz, another integral multiple, *e. g.*, 2200 Hz, is chosen for f, the signal will again be periodic and the frequencies will again be a harmonic series, this time from 1600 to 2800 Hz. The signal again gives a residue pitch of 200 Hz, and the two signals sound much alike.

When, in going from 2000 to 2200 Hz, the carrier frequency passes intermediate values, the situation described by Eq. (4) arises. It is surprising that the sound does not reveal the prevailing inharmonicity at all; the residue does not sound much different. Only the pitch differs from 200 Hz. Consider the case $f = 2030$ Hz; the component frequencies will be (g is still 200 Hz):

$$1430, 1630, 1830, 2030, 2230, 2430, 2630 \text{ Hz}. \tag{7}$$

These frequencies do not have a common divisor in the neighbourhood of 200 Hz. Nevertheless, the residue is tonal and the pitch corresponds to approx. 205 Hz. When the carrier frequency goes upward from 2000 Hz, the pitch goes up too, a little more than proportionally since, *e. g.*, 205 Hz is higher than 203 Hz (one tenth of 2030 Hz). Similarly, when the carrier frequency goes down from 2000 Hz, the pitch goes down too. The same will hold true when the carrier frequency is decreased from 2200 Hz; the pitch will go down from 200 Hz. Some kind of paradox is indicated since a continuous variation of f from 2000 to 2200 Hz should lead to the same pitch again. Figure 15 shows the results of pitch matches as a function of the ratio f/g in this range. The matches were obtained not with a sinusoidal comparison tone but with a purely periodic residue signal of about the same (though harmonic) spectral composition. These measurements confirm that the pitch around a harmonic situation varies somewhat more than proportionally with the carrier frequency.

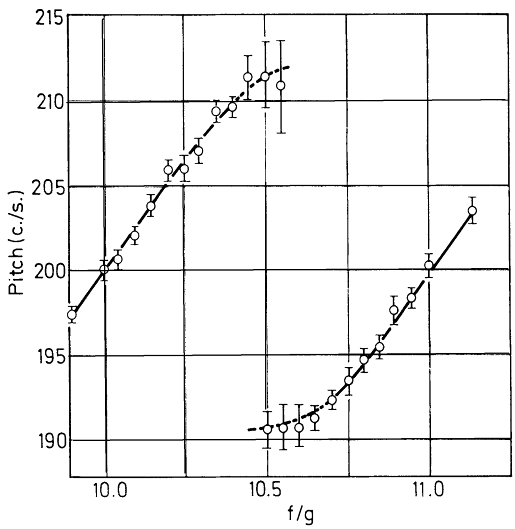

Fig. 15. Pitch of inharmonic signals as measured by DE BOER (1956b)

There are two pitch courses visible in the figure, and, somewhere in the centre of the interval, the attention of the listener seems to switch over from one to the other pitch. Clearly pitch is ambiguous in this region, and the mind can be focused on any one of the two possibilities. Both these pitches, however, are near 200 Hz. The point half-way between the harmonic situations is characterized by a peculiarity both in the waveform and in the spectrum. For instance, if $f/g = 10^1/_2$, the component frequencies are:

$$1500, 1700, 1900, 2100, 2300, 2500, 2700 \text{ Hz} \tag{8}$$

and these are all odd multiples of 100 Hz. It comes as no surprise that the waveform will be *anti*-symmetrical with a period of 1/100 sec. In other words, after 1/100 sec, modulation signal and carrier are in the same relative phase again. It is remarkable that, contrary to expectations, a pitch of 100 Hz is *not* heard. There are two possible pitches and both are near to 200 Hz. This observation constitutes the only substantial deviation from Schouten's results. Note, however, that SCHOUTEN utilized a signal with a full range of components and DE BOER a "cultured" signal containing only a few components.

7. First and Second Effect, the Pseudo-Period

The phenomena described in the previous section were investigated over the pitch range of 100 to 400 Hz. Provided the carrier frequency f is not too high with respect to the modulation frequency g, the pitch shifts can be easily observed. In all cases, the pitch varies a little more than proportionally with f and the amount of this deviation is large when the ratio f/g is low. The deviation does not tend to disappear when f/g increases up to a value of 10. Beyond that value the quality of the pitch diminishes rapidly. This suggests that it is appropriate to consider the actual pitch as being in essence proportional to f – the "first" effect of pitch – and to consider the deviations from proportionality as a perturbation – later called the "second" effect. The existence of the second effect implies that the pitch should go down when f is kept constant and g increases. This experiment is somewhat more difficult with the setup of Fig. 14, but the results confirm this expectation (DE BOER 1956a, b). Since the second effect has been a major tool in novel developments in this field by later authors – notably the recognition of the importance of combination tones – the actual measurement results will be left out of the present discussion.

We now turn to a possible explanation of the first effect. In the light of SCHOUTEN'S periodicity theory, it is logical to consider the waveform. Figure 16a shows the waveform in the harmonic case; this waveform is purely periodic. The dotted lines indicate the "envelope", imaginary boundaries within which all oscillations are confined. The envelope originates from the modulation signal $x_2(t)$ fed into the modulator. In this harmonic case, the fine structure repeats itself exactly in every lobe of the envelope. The arrows indicate the period of the waveform; this is, of course, equal to the period of the envelope. In Fig. 16b, the situation is drawn for an inharmonic signal obtained by the same procedure but with the carrier frequency shifted away from the harmonic position (Figure 1d shows another illustration of an inharmonic signal.) The envelope has remained the same, but the fine structure is not the same in successive lobes of the envelope. However, it is easy to mark a "pseudo-period" as the time distance over which the waveform repeats itself in an approximate way. This is indicated by the arrow in Fig. 16b. If the auditory analyzer measures the period in the case of a periodic signal, it is not far-fetched to hypothize that it will sort out the pseudo-period in an inharmonic signal. In the case drawn, the period as well as the pseudo-period encompasses ten oscillations of the carrier frequency. Hence, this concept causes the pitch to vary as a sub-multiple of the carrier frequency; this hypothesis is sufficient to explain the "first" pitch effect. The proportionality can be expressed by the following formula:

$$p = \frac{f_0}{\mathrm{n}}. \tag{9}$$

Here, f_0 is the central frequency (the carrier frequency of the complex signal); p is the inverse of the pseudo-period; and n is an integer number, namely, the rank order of the central component. The pseudo-period theory predicts that the measured pitch is equal to p. In the following discussion, this formula will repeatedly be referred to; the meaning of p can then be described as: the frequency equivalent to pitch.

The third part of Fig. 16 shows the situation in the odd-harmonic case. Here, the carrier frequency is midway between two harmonic positions; the waveform is *anti*-symmetrical in the sense that the phase of the fine structure alternates in successive lobes of the envelope. The true period is 1/100 sec, but the signal has *two* possible pseudo-periods as indicated by the horizontal arrows. It is readily understood why the pitch should be ambiguous at and around this situation. Even these pitches are preferred by the auditory system over the pitch (100 Hz) corresponding to the true period of this signal (presumably only for a narrow-band signal).

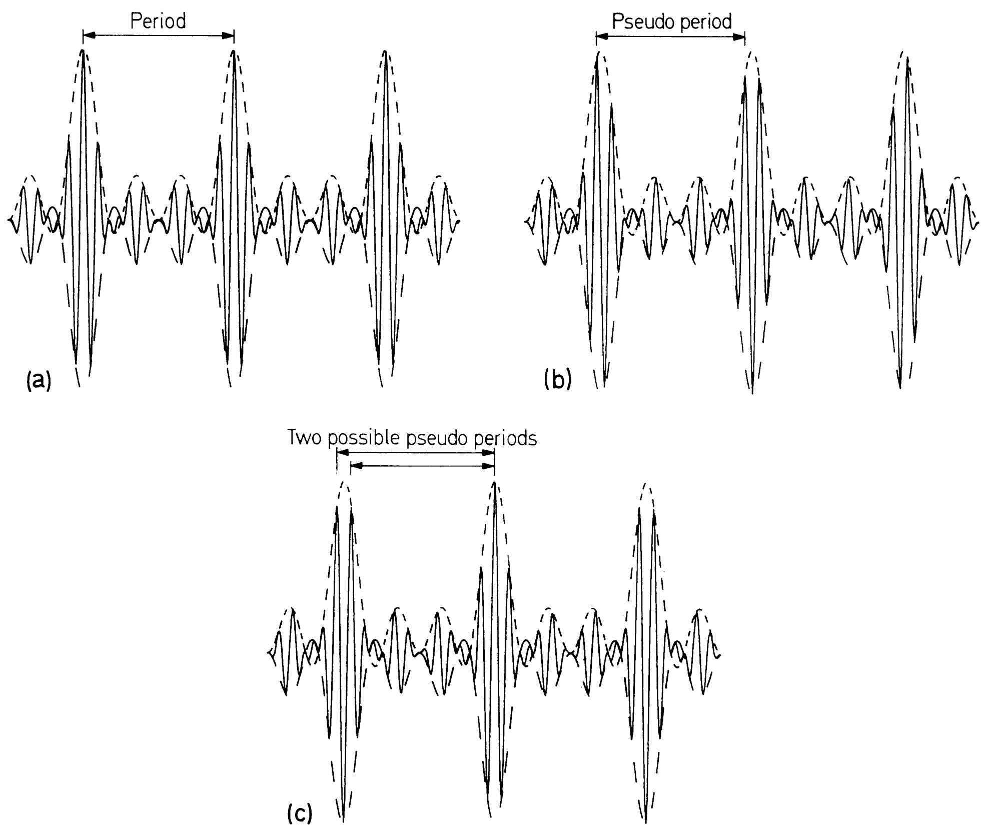

Fig. 16a—c. Generalization of the concept of "period": (a) Purely periodic signal (waveform). (b) Inharmonic signal: the envelopes (shown by dashed and dotted lines) are periodic but the signal is not periodic. The "pseudo-period" denotes the time distance over which the signal repeats itself approximately. (c) Signal containing only odd-numbered harmonics. Two pseudo-periods are possible

8. De Boer's Phase Rule

The fact that a slightly inharmonic signal sounds almost as tonal as a harmonic one is somewhat surprising. If the carrier frequency is very near to a harmonic position, the fine structure of the waveform drifts slowly through the lobes of the envelope (*cf*. Fig. 16b). We do not perceive a change of timbre related to this drifting. Hence, for these signals the timbre of the sound does not depend upon the phase of the fine structure with respect to the envelope. The invariance of timbre for this type of phase change is described by DE BOER's phase rule (1956a, 1961): *The phases φ_i of the components of a sound can be altered by an amount $\Delta\varphi$ that is a linear function of frequency:*

$$\left.\begin{aligned} \varphi_i &\to \varphi_i + \varphi_0 + c \cdot f_i \\ \Delta\varphi &= \varphi_0 + c \cdot f_i \end{aligned}\right\}^{2} \qquad (10)$$

without causing a change of timbre. [For the meaning of the symbol φ_i we refer to Eq. (2)].

The constant phase change φ_0 represents a phase change such as is caused by a very slight inharmonicity. For instance, a 1 Hz deviation of the carrier frequency causes φ_0 to vary so slowly as to complete one cycle per second. The phase changes described by the term φ_0 leave the envelope invariant and cannot be detected by the ear. The term $c \cdot f_i$, proportional to frequency, simply represents a shift of the origin of the time axis which is obviously imperceptible.

All other phase changes may lead to an observable change of timbre. Phase changes not obeying (10) can, for instance, be effected by changing the phase relations among the components of the modulating signal $x_2(t)$, see Eq. (5). Or else, the carrier component (shown dotted in Fig. 15) can be reintroduced in a different phase. Such phase changes do not affect pitch (in general) but they do affect timbre. By the term, "phase effects", is usually meant the dependence of timbre upon the phases of the components; hence, phase effects can only be demonstrated by introducing phase changes that do *not* obey DE BOER's phase rule.

Whether such phase changes are accompanied by detectable changes of timbre depends on two factors. The first factor has to do with the relative frequency separation of the components. When the components of the signal are spaced wide apart, each component stimulates more or less an isolated region of the cochlea. For signals with such a spectrum, no phase effects are audible; this case is described by HELMHOLTZ' fundamental law of timbre. When, on the other hand, the components are relatively close to one another, their patterns of excitation in the cochlea overlap a great deal. A change of phase, then, leads to a change of composite waveform and this may lead to a change of timbre. It appears that there is a close relation between perceptibility of phase changes and aural resolution of components. This relation was recognized much earlier (MATHES and MILLER, 1947), and the argument has been investigated repeatedly (see, *e. g.* LICKLIDER, 1955; GOLDSTEIN, 1967a).

The second factor that determines whether a phase change is perceptible concerns the magnitude of the phase change. If we think of a signal with narrowly spaced components, the phase change should cause a waveform variation exceeding a certain minimum variation in order to be perceptible. This waveform variation should pertain to a spectral width of approx. one critical band; hence, it would be logical to expect that the threshold for just noticeable phase changes can be expressed in terms of the phase deviations relative to a linear phase function [*cf.* Relation (10)] per critical band. As far as the author is aware, this line of reasoning has not been pursued systematically (probably because the

[2] $\to$ Means: is replaced by. $\Delta\varphi$ indicates a change of phase.

influence of sound intensity is so large). Phase effects of the nature considered here have continued to play an important part in residue theory, but other factors than those described above have been found to be important as well. However, the phase rule (10) has proven again to describe the phase changes for which the auditory system is "phase deaf". We shall come back to this point in Section H. 12.

9. The Pseudo-Fundamental Theory

The connection between perception of the residue and insufficient aural resolution of components was explicitly contained in SCHOUTEN's original definition. In this light, it is highly remarkable that DE BOER found that a residue pitch could be perceived quite well under conditions where the components are almost completely resolved. Such a residue shows clear changes of pitch when the signal is made inharmonic. Moreover, the pitch variations are larger than those expressed by Eq. (9), testifying to the fact that the "second effect" is quite large.

This result makes it impossible to advocate the pseudo-period theory as the *sole* explanation for pitch shifts due to inharmonicity. As an alternative, DE BOER proposed a theory operating on the frequencies of the individual (probably resolved) components. According to this theory, the auditory system tries to find a harmonic series of frequencies that corresponds in the best possible way with the frequencies of the components presented. For instance, the series of frequencies (7) can be approximated best by a harmonic series with 203 Hz as the fundamental frequency; the lowest components are just a bit too low and the highest components too high but on the average the deviations are minimal. The best-fitting fundamental frequency can be called the "pseudo-fundamental".

A more detailed formulation of this theory (DE BOER 1956a) shows that in first approximation the pseudo-fundamental is equal to an integral submultiple of the carrier frequency. Hence, the first effect of pitch shifts is again described by Eq. (9), the integer n being interpreted as the rank number of the carrier component. Actual pitch shifts are larger, of course, and the pseudo-fundamental theory has an inherent possibility to account for this fact. The determination of the best-fitting pseudo-fundamental must be made on the basis of some sort of criterion as to what is the "best" fit. In this process, the various components can be given different weights. For instance, the lower components can be considered as more important than the higher ones. In an inharmonic signal, the relative shift of the lower components is larger than that of the higher ones, and the resulting relative shift of the pseudo-fundamental will be larger when the lower components are given a larger weight. It will be seen that later developments in residue theory have utilized the same consideration. At the present stage of the description, we are not yet in a position to judge or appreciate these developments; the reader must abide his time: at least one major wave in scientific thinking is yet to come...

10. Two Types of Residue

According to DE BOER's findings, there are *two* types of residue. The first arises when components are relatively close together and cannot be resolved aurally. The pitch is probably derived from the waveform; the "pseudo-period" serves as the first-order determinant of the pitch for inharmonic signals. This type of residue is strongly phase sensitive – this term to be interpreted in the sense as described in the second part of Section E. 8. The second type of residue arises when the components are spaced relatively widely. This residue shows no phase effects, and the components are probably nearly completely resolved. The "pseudo-fundamental" can be proposed as a first approximation for the pitch of inharmonic residues of this type.

The two types of residue were named by DE BOER "intermittent" and "continuous" residue, respectively. These names are appropriate in an experimental setting in which phase effects are deliberately and continuously introduced: the intermittent residue then shows alternating periods of prominence and absence corresponding to the sharpness of timbre, and the continuous residue does not show any change with time. For the present, other names are to be preferred, names that have no connection with a particular experimental paradigm. We suggest short terms like *narrow* and *wide* residue; what is meant to express here is that the spectral components have narrow and wide spacings, respectively. These terms cannot easily be misinterpreted.

It might be objected that the word "residue" is a misnomer for the case of the wide residue since this is not the result of incomplete aural analysis. The term, "residue", is well known nowadays and there is no reason to abandon it now. Let us, therefore, continue to use this term, but let us use it in a wider sense. *The term "residue" is a universal descriptor for the joint perception of a number of components.* Whether the components are resolved or not, they give birth to a special percept by way of their combined action.

The two types of residue invoke different explanations of the observed pitch shifts due to inharmonicity. For a narrow complex, *i. e.* a signal yielding a narrow residue, the pseudo-period theory describes the pitch shifts in first approximation. For a wide complex, the pseudo-fundamental theory provides a more plausible explanation. There remain some difficulties in the explanation of the "second" effect. The pseudo-fundamental theory seems a bit more flexible, but we will see in the sequel that the pseudo-period theory must be amended also, and it finally receives about the same degree of flexibility as the pseudo-fundamental theory.

11. Multiple Modes of Pitch Perception

The present part of this chapter, in accordance with the title, wave crests, is mainly concerned with period-type aspects involved in residue perception. Hence, we continue with a short description of developments in the study of temporal phenomena, more specifically, the part played by the temporal order of pulses within one period. The experiments by the Bell group (FLANAGAN and GUTTMAN, 1960a, b; GUTTMAN and FLANAGAN, 1964) show that the question of periodicity perception is really somewhat more complex than previous work would suggest. Typical time sequences within one period are shown by Fig. 17. In the experiment, sequences of one type were matched with sequences of another type. When the repetition period of any of these signals is extremely low, *e. g.*, one second, one will hear one click per second for signal Number 1 but two clicks per second for signal Number 2, etc. When the repetition rate approaches 100 Hz (*i. e.* the period is repeated 100 times per second), the signals will be matched on the basis of their periodicity. This can be explained as the result of a pitch match, either of the fundamental's pitch or of the residue. For a signal like Number 4, the two modes differ by a factor of 4: for very low repetition rates, the signal is matched to a signal Number 1 with a repetition rate four times as high, whereas, at intermediate rates the best match is obtained when the rates are equal. There is a third mode, especially evident when the signals presented are devoid of their lower-order components. The match then occurs on the basis of the lowest component present; this mode is only evident for repetition rates above 500 Hz, hence, above the region which is important for periodicity pitch. An effort is made to correlate the findings with mechanistic events in the cochlea as reproduced by an analog model of the basilar membrane. All three "pitch modes" were found to be manifested in the mechanical operation of the cochlea. However, we think some caution is necessary. If we wish to explain the findings in terms of period detection, we do well to define exactly what we mean by period, and we must invoke specific assumptions as to what the auditory system will accept as a "periodic repetition" when in fact the repetition is only approximate (compare, *e. g.*, Signals 1 and 2).

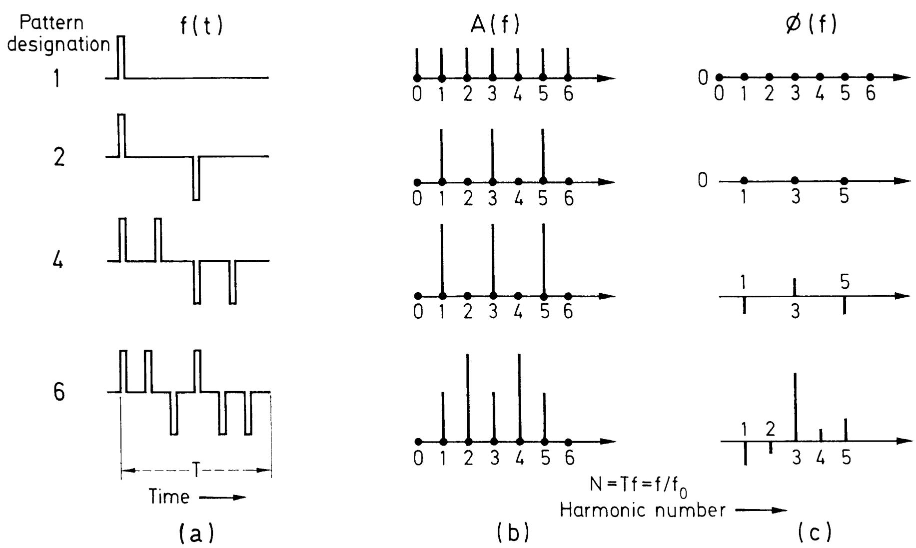

Fig. 17a—c. Waveform and spectrum of the signals used by FLANAGAN and GUTTMAN (1960a). (a) Waveforms $f(t)$ of the four signals used; T indicates the period of repetition. (b) Spectral amplitude $A(f)$ of the signals. (c) The phases $\Phi(f)$ of the spectral components, shown in a relative sense

12. Small's Work, Time Separation Pitch

That there are more subtleties involved in the perception of periodic multiple-pulse series is shown by the work of SMALL and collaborators. A signal most apt to give rise to periodicity pitch was produced by interrupting a pure tone, the carrier, repeatedly with a modulator. Typical signals were a carrier of 1000 Hz interrupted 100 times per second and a carrier of 5000 Hz interrupted 100 times per second. These signals are referred to as signals 1000/100 and 5000/100, respectively. The spectrum of such a signal is centered around the carrier frequency; to keep the spectral width down, each signal was filtered after the modulator by a band-pass filter centered around the carrier frequency. The pitch of the signal was determined by comparison with a pure tone. In contradistinction to usual methods, the matches were also carried out with a fixed tone of 100 Hz and a modulation rate which could be varied by the listener. The experiments confirmed that a low pitch of about 100 Hz is the main aspect of the signal; a match to a frequency in the region of the most prominent spectral components either could not be made or could be made with difficulty by the listeners (SMALL, 1955).

The most conspicuous finding is reported for a type of signal in which two of these pulse series are combined. Both series have nearly the same repetition frequency, *e. g.*, 100 Hz. At a certain moment, the sum of the two signals is a double pulse series as is schematically indicated by Fig. 18. (In this figure the pulses are drawn as rectangular, but, in fact, they are small sections of an oscillating waveform.) As time progresses, the pulses of one series drift with respect to those of the other series, to complete one period in something like 5 sec. THURLOW and SMALL (1955) reported that in this type of signal one can hear a pitch that is associated with this course of the time separation τ. When τ is very

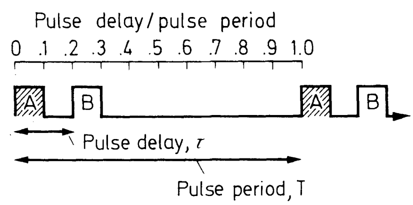

Fig. 18. Signal to demonstrate the "sweep pitch" (time-separation pitch). After SMALL and MCCLELLAN (1963)

small, this pitch is high. When τ gradually increases, the pitch drops downward. From the point that τ passes the center of the period, the pitch goes up again. The lowest pitch corresponds, of course, to the octave of the basic repetition frequency, in this case, to 200 Hz. The highest pitch was difficult to judge; in the 5000/100 condition it was considered to be higher than 1000 Hz. The effect was the most noticeable when the bandwidth of the spectrum was rather wide, a reason for experimenting with unfiltered pulse series later. In view of the experimental paradigm, this pitch was named the "sweep pitch". Since it appears to be closely associated with the separation in time between pulses, the later adopted term, "Time Separation Pitch (TSP), is a better one. It is to be noted that this type of pitch can best be heard in a situation just as the one described: a signal consisting of two pulse series with slightly different repetition frequencies. It is then recognized because of its sweeping character.

In a later study (SMALL and MCCLELLAN, 1963), it was experimentally confirmed that time separation pitch corresponds to the frequency $1/\tau$. In this study, the value of τ was constant in one experimental run. To concentrate the attention of the listener upon time separation pitch, one amongst three signals could be chosen: a single pulse series, the double pulse series, and a pure tone. The results were remarkably accurate; the largest deviations were 20 cents (*i. e.* 0.2 of a semitone; 100 cents is equivalent to one semitone). Some incidental deviations occurred but few of these were investigated further. In these experiments, the basic repetition frequencies were 25, 100, and 400 Hz, and the pulses were approximately rectangular. The relative pulse delay τ/T (see the figure) could be varied between 0.1 and 0.5. The resulting pitch had a range between 50 and 4000 Hz.

The results of SMALL and McCLELLAN are considered to confirm that the major cue for pitch determination consists in the timing aspects of the waveform. There is no mention made as to whether a melody can be made out of time separation pitches. The study of pitch perceptions related to time separations has in later times been taken up again under the heading, repetition pitch (BILSEN, 1966; see Section F. 8). The case described in the present section appears to be just an example of the great variety of signals for which some hidden type of repetition is heard as a pitch. A more general case is a noise signal to which a delayed version of the same noise signal is added; such a signal also acquires a pitch inversely proportional to the delay time. Since the philosophy of this type of experiment involves concepts which yet have to be developed, we must defer the description of this work until later.

13. Averaging of Pseudo-Periods and Pseudo-Fundamentals

From the findings described in the preceding sections, it is evident that even the most straightforward period theory runs into difficulties when confronted with experimental findings. The situation becomes quite complicated when we realize that the actual excitation at a particular location in the cochlea has the character of a band-filtered form of the stimulus. The time separations, which are so clear from the waveform of the sound stimulus as presented to the ear, are not nearly as evident from bandpass-filtered waveforms. Moreover, it is not certain that the time distance between peaks in the waveform of filtered signals is equal to the corresponding distance in the unfiltered waveform. Consider, for example, a wide-band inharmonic signal. For the waveform of the actual stimulus, the pseudo-period can readily be determined (see Fig. 16b). If we consider, however, a location on the basilar membrane tuned to a frequency in the lower part of the signal's spectrum, we observe a pseudo-period that deviates more from the period $1/g$. We may refer to Eq. (9) but note that the integer constant n is now equal to the rank order of the strongest component passed by the filter. For a wide-band inharmonic signal, there will exist an ensemble of pseudo-period values corresponding to different cochlea locations; these pseudo-periods will all be different. Consequently, when we hear but a single pitch, the pseudo-period values must be weighed and averaged in some way by the auditory system. The situation is still more complicated since the pseudo-period will not always be an integral submultiple of the local carrier. This is the result of phase modulation exhibited by the filtered signal. In fact, this process has been studied by FISCHLER as a possible means to explain the second effect (FISCHLER, 1967). We may conclude that although a residue theory on the basis of temporal phenomena may be attractive at first sight, the actual situation is by no means simple. Due appreciation of the ensemble of pseudo-periods, as described above, makes the pseudo-period theory considerably more flexible. In the process of averaging all pseudo-periods, we may consider giving different weights to different frequency regions. It is then easy to explain the existence of the "second" effect just as with the pseudo-fundamental theory. However, the magnitude of the second effect still presents difficulties.

14. Pitch Ambiguity; the Relation between Phase Effects and Aural Resolution

The studies reported in the period 1950–1960 have clearly shown the basic dilemma in residue theory. Most of the experimental findings fit well within the framework of a theory based on periodicity concepts. In this respect, the research reported in the preceding sections is a straightforward extension of SCHOUTEN's work. Two points are of special interest in this connection and are worthy of mention in a summary. The first point is that the auditory system appears to be able to utilize temporal information of greater subtlety than that contained in basic periodicity. The "sweep-note" effect discovered by THURLOW and SMALL proves that detailed repetitions contained in the signal's waveform can provide a cue for pitch. Later, the pitch associated with time separation was called, time-separation pitch; this name appropriately describes the most probable cue.

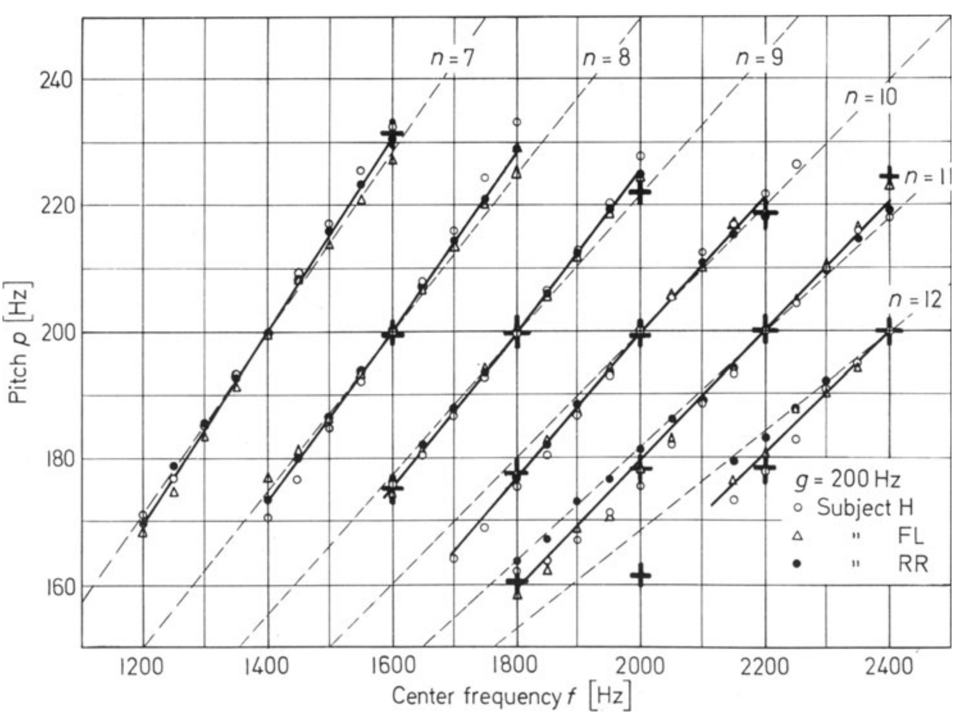

Fig. 19. Pitch measurements for three-component signals. Difference of component frequencies: 200 Hz. After SCHOUTEN *et al.* (1962)

The second point about temporal cues in a signal's waveform concerns ambiguity. In DE BOER's experiments on inharmonic signals, the situation midway between two consecutive harmonic situations led to two possible pitches. In terms of the signal waveform involved, this ambiguity is readily explained (see Fig. 16c). Which of the two possible pitches will be perceived depends on the state of attention, the signals heard previously, and so on. SCHOUTEN *et al.* (1962) repeated the experiments on inharmonic signals with complexes consisting of 3 components instead of 5 or 7. Their findings confirmed the existence of the "first" and "second" effects, but, moreover, they extended the concept of ambiguity of pitch. Figure 19 shows the results, pitch (p) as a function of the central frequency (f) of the complex. We observe that the lines connecting the data points can be extended over a wider range than is possible for DE BOER's results (Fig. 15). At any particular value of f, three or more pitches are possible. This does not necessarily imply that all of these pitches are equally prominent, of course.

We may note in passing that, apart from the second effect, the lines in Fig. 19 can be labelled on the basis of different values of n in Eq. (9). The existence of more than one pitch value then implies that the auditory system may utilize different values of n. The experimental procedure (a gradual increase of f, for instance) forces the subject to stick to a particular value of n but eventually his attention goes astray and he may pick up another pitch, corresponding to a different value of n.

We arrive at the remarkable conclusion that even in a purely harmonic complex a number of discrete pitches are acceptable to the subject. This finding was confirmed by a special experiment in which the listeners were instructed to try to find all possible pitches in a harmonic complex. The crosses in Fig. 19 show the results of this procedure; each cross stands for the centre of a cluster of data points.

Not all experimental results fit well within the framework of a "periodicity" theory. The experiments described in Section E. 11 clearly indicate that the auditory system possesses capabilities to adjust itself differently to the same situation. The clearest example is the finding of the existence of the "wide" residue, a residue formed out of a series of components that are probably almost completely resolved by the ear. This residue has all the properties of the "narrow" residue, except that it does not show phase effects. In particular, it is (at least) equally tonal. A pure temporal theory runs into difficulties when it is applied to a signal with widely spaced components. The pseudo-fundamental theory seems more fruitful in this case. It should be noted that the pseudo-fundamental theory can be regarded as an extension of the "old" place theory: the pseudo-fundamental frequency is determined on the basis of the frequencies of the components but it can equally well be based upon the "places" at which the components produce the largest excitations in the cochlea.

The last subject which should be mentioned in this summary concerns phase effects. It is well to remember at this stage the following properties of the residue:

a) The timbre of the residue is not affected by a phase change conforming to relation (10).

b) Equivalently, even the slightest amount of inharmonicity of the signal [of the type corresponding to Relation (10)] is inaudible.

c) Phase changes that do not agree with Relation (10) may cause changes of timbre; whether they do, depends greatly on the relative frequency spacing of the components. In this sense, phase effects are usually considered indicators of insufficient frequency resolution of signal components.

F. The Next "Cycle": Forebodings of a New Way of Thinking

1. Introduction

The period from 1960 to 1970 has witnessed a change in attitude toward the residue theory because a few new concepts have been formulated. Again, it has taken a long time before the findings were generally interpreted in what appeared to be scientifically the most advanced way. In the present part of this chapter, we shall describe this development. As has been explained in the introduction, there will be no room for an elaborate description of all of these newer studies; the relevant literature is easily accessible. We shall instead concentrate upon the conceptual development. Let it be repeated that the emphasis placed upon the various aspects of the work is the result of personal opinions of the present reviewer. Admittedly, other interpretations are possible; the one expressed here ties in with the newest developments and agrees with present views of most other experts in the field.

The major argument is concerned with the fact that there exists more than one type of residue. Furthermore, these residues are not equivalent to one another. Since one type of residue may be considered as "inferior" to another one, a relation like Eq. (9) for pitch shifts due to inharmonicity acquires a new perspective. In particular, the value of n in this equation need not be the rank number of the central component. It may be adjusted to be relevant to the more dominant part of the residue signal, as we shall see. In any event, the explanation of various properties of the residue has acquired a new dimension. The increased complexity of the matter has not lessened the trouble necessary for theoretical explanation of the experimental findings. For the moment, it seems better to stick to rather formal types of theorizing. Later developments may perhaps provide new insight that will serve to make simple what seemed quite complicated at first.

2. The Existence Region of the Residue (Ritsma)

It was Ritsma, one of Schouten's pupils, who made the first important contribution in this era (1962). He was intrigued by Hoogland's (1953) failure to detect a low-frequency pitch when the signal presented consisted of a number of components that were relatively close to one another ($f = 3000$, $g = 100$, for instance) and less than 60 dB SPL in level. Ritsma guessed that such sound complexes would indeed give a residue but one which is essentially without pitch. Consequently, he set out to explore the tonality of residue sounds for widely divergent values of the central frequency f and the component spacing g. He produced his 3-component signals with the modulation method (see Section D. 6 for an explanation). When the carrier frequency is f and the modulation frequency g, the resulting components will have the frequencies $f-g$ and $f+g$. A third component (carrier) with frequency f is added. In all signals f and g are harmonically related. One further experimental variable, m, is introduced by

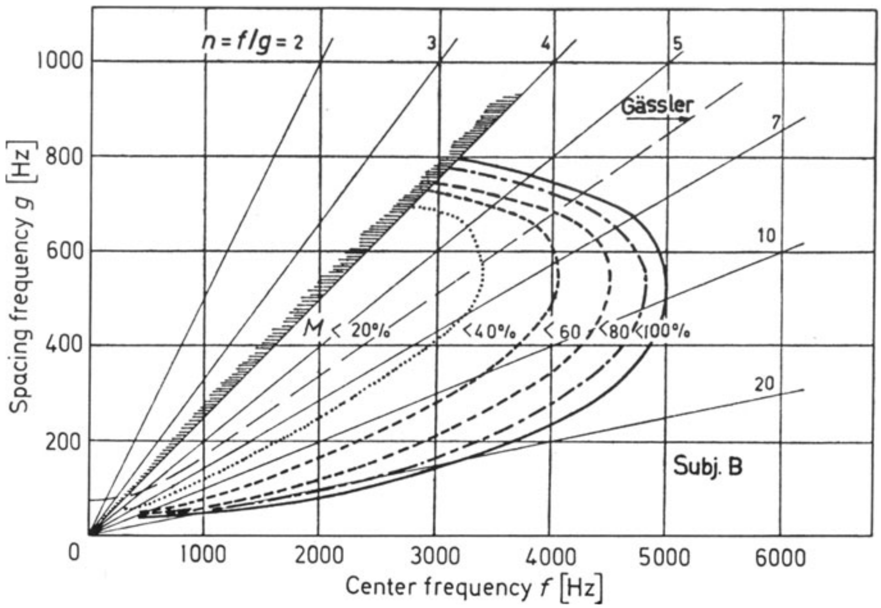

Fig. 20. Existence region for three-component residue signals. M is the modulation depth (called m in the text). The line marked "Gässler" indicates the critical bandwidth. After Ritsma (1962)

making the former two components variable in strength with respect to the latter. The amplitudes of these so-called sidebands are $^1/_2\, m$ as against unity for the carrier component. The constant m, the "modulation index", could be varied between 0.2 and a little over 1.0 (see Ritsma's paper for further details). In one experimental run, one of the parameters f, g, or m was varied while the others were kept constant.

Subjects were instructed to determine the limiting values for the variable parameter so as to find the boundary of the region of parameter values that leads to a *tonal* pitch. To aid in the judgement of tonality, a comparison signal was provided with a fixed modulation index m of 1.00, the same value of g and a considerably lower value of f. A typical result is shown in Fig. 20. The line $f/g = 4$ is a natural boundary because the value of 3 would lead to a signal of which the lowest component is just an octave above the residue pitch, a situation which would be too confusing for the listener.

Consider now the line for $m = 1.00$. This delineates the largest area in which residue signals of this type can be perceived as having a definite pitch, *the existence region of the tonal residue*. One of the most noteworthy features of this region is that it shows no evidence of being linked to the limits of auditory frequency analysis. The line which depicts the width of the critical band (*cf.* Zwicker *et al.*,

1957) corresponds to $f/g = 6$; this line (marked "Gässler" in Fig. 20) does not lie near the boundary of the existence region of the three-component residue. On the contrary, it runs right through it.

We may realize, however, that the resolution of components lying outside each other's critical bands is far from complete. A period-extracting procedure may then very well be postulated as an universal mechanism. RITSMA's (1962) paper gives further details about this mechanism; this development from earlier models of this kind (LICKLIDER, 1951) was made possible by the advances in our knowledge of the auditory system in the preceding period.

3. The Principle of Dominance I

The experiments that have led to the second important step were simultaneously, and apparently independently, carried out by RITSMA (1967) and PLOMP (1967a). In fact, the experiments were carried out in two different countries (USA and the Netherlands), and the manuscripts were received by the editor within one month. We shall describe mostly RITSMA's experiments. PLOMP's method was essentially similar and led to the same results. PLOMP's paper is, furthermore, recommended for its historical introduction.

In RITSMA's (as well as in PLOMP'S) experiments, the lower components were derived from a repetitive pulse series by low-pass filtering. The higher components were derived from another pulse series by high-pass filtering. The repetition frequencies of the two pulse series were not the same. One of the residue signals (A) consisted of the lower harmonics with fundamental frequency f_0, up to a cut-off frequency f_c; the components above f_c were harmonics belonging to a fundamental frequency $f_0 + \Delta f$ (see Fig. 21). In the comparison signal (B), the

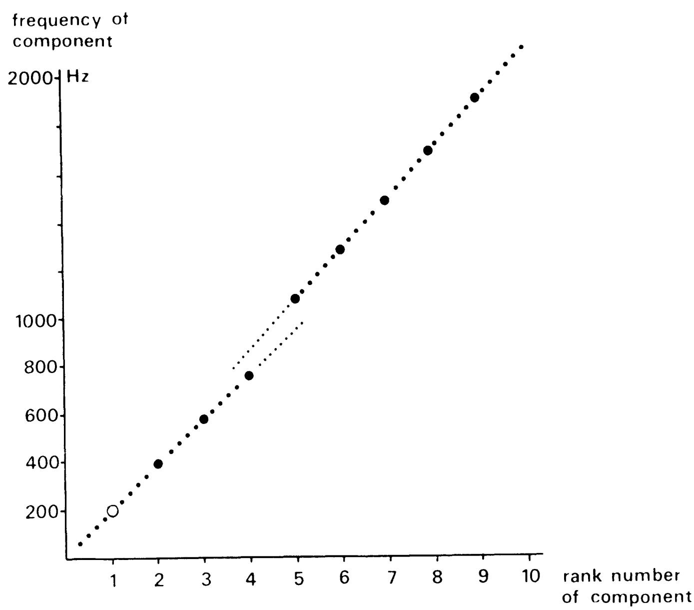

Fig. 21. Components of signal containing two conflicting types of information for pitch. The slopes of the dotted lines indicate the frequencies f_0 and $f_0 + \Delta f$. Only the A-signal is shown. The fundamental of the lower band is indicated by an open circle. Note that both dotted lines meet at the origin

roles of f_0 and $f_0 + \Delta f$ were interchanged; the subjects were to judge which of the tones had the higher pitch. The relative difference $\Delta f/f_0$ was 3 or 6%. For low values of f_c, Signal A was always judged to be higher in pitch; for high values of f_c Signal B had a higher pitch. The cross-over point depended somewhat upon the listener. For $f_0 = 100$, it was in the range 400–700 Hz; for $f_0 = 200$, it was

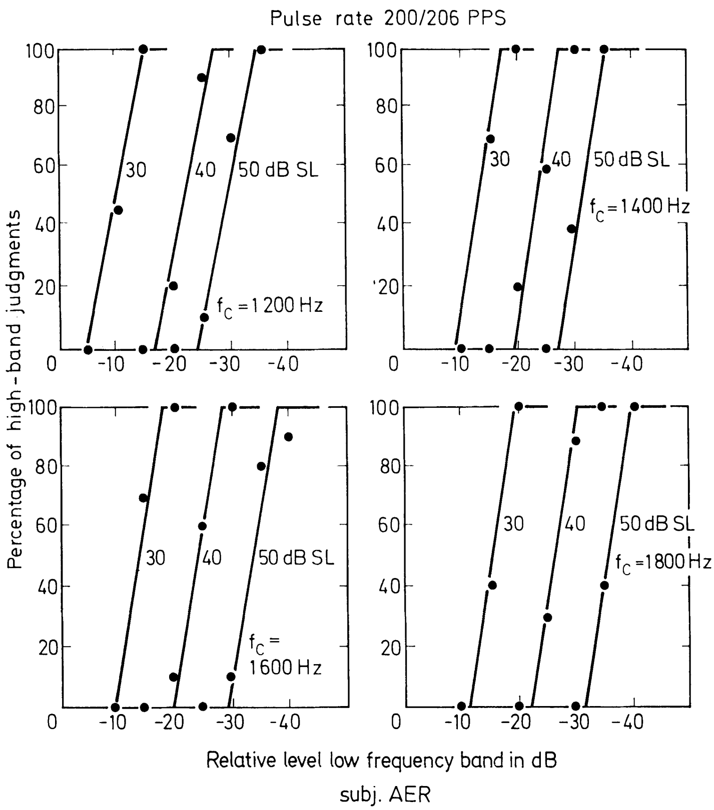

Fig. 22. Results of pitch judgements for dual-pitch signals in which the lower group of harmonics is attenuated. The terms, "low band" and "high band", refer to the two groups of harmonics shown in Fig. 21. After Ritsma (1962)

800–1400 Hz and for $f_0 = 400$, it was between 800 and 2000 Hz. In a way, it is surprising that these cross-over points are so low. The results signify to the fact that even a foursome harmonics of frequency f_0 against a multitude of components (all components above the fourth) with a fundamental of $f_0 + \Delta f$ force the total sound to have a pitch corresponding to f_0. The lower components are clearly *dominant*. Even though they are much less numerous, they outweigh the higher components.

To bring out this fact still further, Ritsma attempted to determine how strong these lower components should be to exercise this dominance. In his second experiment, f_c was fixed, and the lower components were attenuated by a variable amount. Again, the signals were presented in a two-alternative forced-choice test, and the subjects were required to judge which signal was higher in pitch. Some typical results, judgement percentage as a function of the attenuation level of the lower components, are presented in Fig. 22. It is this figure that gives the clearest demonstration of the principle of dominance: the lower components assume dominance within a level range of 5—10 dB. The threshold depends on the overall sound level in such a way that it seems that the lower components assume dominance once they surpass a certain threshold level, approximately 10 dB SL.

4. The Principle of Dominance II

In the third series of experiments, RITSMA attempted to reduce the number of components in the dominant part of the residue still further. The low-frequency part of the signal was filtered in such a way that only 2 or 3 components were left; all others were attenuated so much as to be inaudible. The lowest fundamental frequency was 200 Hz in this series. Again, there was a sharp transition in judgements when the level of this low-frequency part was varied. The general conclusion of this experiment is that at least two of the three components 600, 800, and 1000 Hz have to be present to exercise their dominance. The dominance becomes manifest as soon as these components are more than 10 dB above the subjective threshold.

Before we try to draw general conclusions, some minor points should be mentioned. Some listeners reported an ambiguity of pitch when the cut-off frequency f_c of the filters was in the dominant region. They reported no beats; apparently two neighbouring pitches can be present in a sound and this does not necessarily lead to beats. A second point is that in both RITSMA's and PLOMP's basic experiments the fundamental was present as one of the components. The results of the experiments indicate that the residue formed by higher components, all except the first, is dominant. In other words, the fundamental is *not* dominant, a point which is repeatedly stressed by PLOMP.

5. Some Reflections, the Second Effect

In the following sections, we shall encounter more examples of the principle of dominance. But before we describe further experiments, it is good to reflect on some previous findings and the possible connection with dominance. One of the most puzzling aspects has been the so-called "second effect" (see Section E.7). It is relatively easy to understand that the pitch of an inharmonic signal would be proportional to the frequency of the central component. That the relative pitch deviations actually are larger than the relative variations of the central component points to a more complex mechanism. In Section E. 13, it was pointed out that for an inharmonic signal the various pseudo-periods that can be associated with different locations in the cochlea are all different. Hence, if pitch is determined by the pseudo-period, it surely is an average of these pseudo-periods that should be considered. Conversely, if we want to assume a mechanism analogous to finding the pseudo-fundamental, it is again an average measure of this kind that should be envisaged. The averaging is performed over the domain of component frequencies, and soon it has been realized that the lower frequency regions carry the larger weight in the averaging process. There appear now to be *two* reasons for this. The first is related to the fact that our auditory frequency scale is a logarithmic one; in other words, the lower harmonics are wider apart, in a musical sense, than the higher ones. This differential weighing has been pointed out as a possible contribuant from the beginning period of the study of inharmonic signals on. In general, this type of weighing has been found insufficient.

RITSMA's and PLOMP's findings have contributed a second reason: the lower components are dominant with respect to pitch, and, hence, they should carry a far larger weight. The distribution of weights is very uneven indeed. The best agreement between experiment and theory can be obtained only by giving the lower components an overwhelmingly larger weight. Thus, WALLISER (1968, 1969a) could report that the pitch of the inharmonic residue is given by the frequency of the *lowest* component in the complex divided by its rank number. In other terms, the "centre of gravity" of the complex lies at the lowest component.

6. The Residue in a New Definition, Resolution of Components

The discovery of the principle of dominance has a profound effect on our thinking. If we recall that the residue was originally defined as the joint perception of the unresolved components, we are surprised to find that the dominant part of a residue signal resides in the region of lower components. Clearly, the sound that we perceive associated with the lower components has all the aspects of the residue. Hence, the definition should be modified so as to incorporate specifically this dominant part of the spectrum where components can be resolved almost perfectly by a trained ear. The new, less restrictive, definition reads:

A residue is the joint perception of a number of consecutive components.

It then becomes paramount to study the question as to which components in a residue signal can be resolved by the ear and which cannot. The dividing line between frequency resolution and spectral integration is provided by the "critical bandwidth". Abundant references can be found throughout the literature (for a short survey see DE BOER and BOUWMEESTER, 1974). For the frequencies of interest, the width of the critical band is approx. 16% of the frequency. This is represented, *e. g.*, by the line $f/g = 1/0.16$ in Fig. 20 (marked "Gässler").

Specific experiments on aural resolution in residue signals have been carried out by PLOMP (1964). The results of these experiments indicate the extent to which a component can be aurally separated from the complex. We shall describe only one of the experiments. A listener can choose, by way of a switch, to listen to a periodic residue signal or a pure tone. Actually, he has two possible pure tones at his disposal, one of which coinciding with one of the harmonics of the periodic signal and the other lying midway between two consecutive harmonics. The situation is symbolized in Fig. 23. The listener's task is to select from the two pure tones the one he considers to correspond to one of the audible harmonics of the periodic signal.

The experiment is conducted with the familiar two-alternative forced choice method. When the comparison tones are in the extreme high part of the spectrum, the probability of a correct answer is 0.50. Of course, the lower components can be detected with a probability of 1.0. Results were, in short, that the percentage of correct responses dropped from 100 to 50 in the range of n = 4 to n = 9.

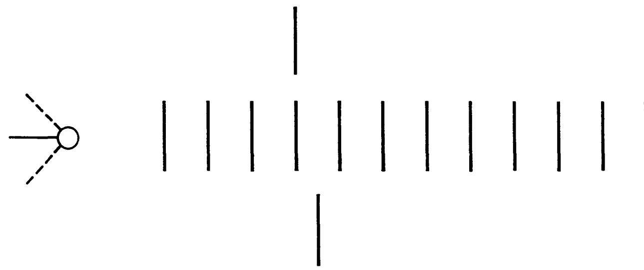

Fig. 23. The three signals from which the subject can choose in PLOMP's (1964) experiments on aural resolution. The lines indicate frequencies of components. The circle symbolizes the selection switch. After PLOMP (1964)

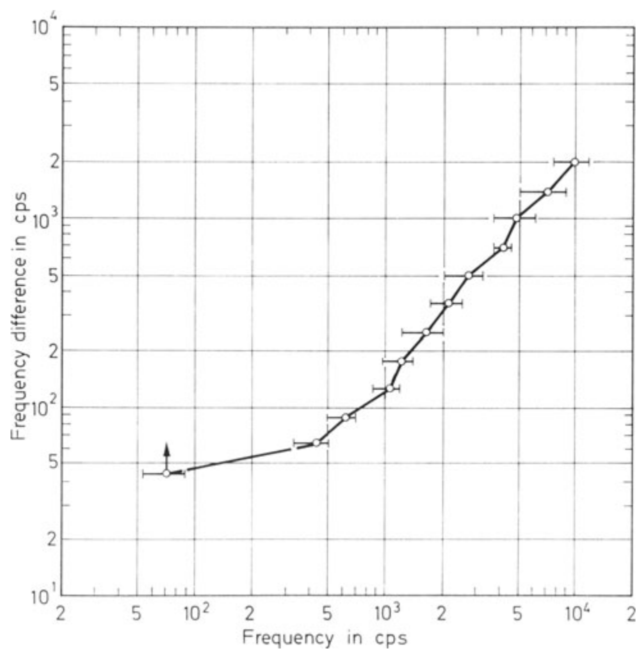

Fig. 24. Frequency difference between harmonics required to hear the part-tones separately. The course of the measured thresholds as a function of frequency corresponds closely to that of the critical bandwidth. After PLOMP (1964)

The results were essentially similar for inharmonic signals. The frequency difference of components that can be aurally resolved depends upon the central frequency in a way that is closely similar to the characteristics of the critical bandwidth. The relation is depicted in Fig. 24.

The results of PLOMP's study once more underline the fact that the dominant region of a residue signal is a region in which single components can be picked out of the signal with nearly perfect certainty. It should be noted, however, that this does not imply that the determination of the pitch of the residue cannot possibly be performed on a temporal basis. That a component can be aurally separated does not mean that components do not interact in the ear. The arguments provided by studies like this are only suggestive; for the dominant part of the residue, it is likely that interaction of components is minimal.

7. Audibility of High Partials

Although the higher components of a periodic pulse signal cannot be distinguished from one another, this does not mean that they cannot be detected under any circumstance. With a special procedure, even extremely high harmonics can be made audible. This effect has been studied by DUIFHUIS (1970, 1971). A brief description of the experiments follows. The basic signal is a periodic pulse series with fundamental frequency f_0. The repetition frequency f_0 was rather low, 25–100 Hz. A sinusoid of (exactly) the frequency $n \cdot f_0$ is generated separately and added to the pulse series. The phase of this extra component is fixed in such a way that the n-th harmonic in the sum signal can be present at any desired intensity and with positive as well as negative polarity.

The threshold of detection for the reconstituted n-th component can be determined as a function of its polarity. For values of n below 10, the threshold is the same for both polarities, as expected. For extremely high values, above 28, polarity has a profound effect. The result is that, when the n-th harmonic is added with the amplitude it normally has in the periodic pulse series, it cannot be detected. The *absence* of this harmonic can be detected very well. Similarly, the n-th harmonic can be detected when it has twice its normal amplitude. In the region with n between 10 and 28, a gradual transition appears between the two types of behaviour.

In DUIFHUIS' paper, a possible interpretation of the results for high n values is given. Consider the response of a high-frequency band-pass filter (tuned to the n-th harmonic) to a periodic pulse series, *e. g.*, from Fig. 11. The response to each pulse is extremely short. If we subtract the n-th component from the signal, it shows up in the response as a sinusoid filling up most of the period between the pulses. It can now be understood why we are able to hear the n-th component as a pure tone when it is actually absent in the spectrum. Furthermore, this explains why the complex sounds exactly the same when the n-th component is present with double its normal amplitude and when the n-th component is completely absent.

In the second paper (DUIFHUIS, 1971), the full consequences of the idea suggested earlier were pursued. Consider again the response of a high-frequency band-pass filter tuned to the n-th harmonic, and let the stimulus be a not-modified series of pulses. The response of the filter comes in bursts synchronous with the pulses. Between the pulses, there is little excitation, and it is likely that an extremely short tone burst can be detected when it is presented at just the right instant between the pulses. The threshold will be higher when the tone burst coincides with the peak of the filtered signal. This reasoning is borne out by the results of measurements. The short tone bursts were made to contain approximately 8 complete sinusoidal oscillations; the harmonic numbers used were 20,

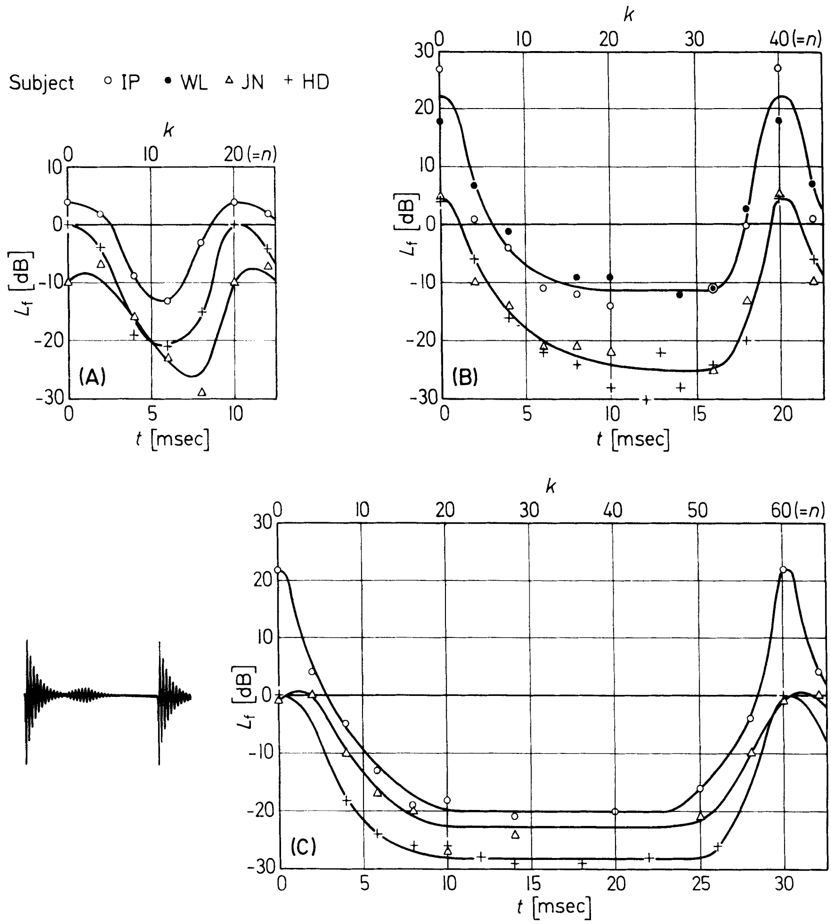

Fig. 25. Threshold of a tone pulse as a function of its temporal position in the period of a periodic pulse series. The tone frequency is 2 kHz; in *A*, *B*, and *C* the repetition frequencies are 100, 50, and $33^1/_3$ Hz (hence the harmonic numbers are 20, 40, and 60, respectively). The inset shows schematically how we conceive of the response of the "auditory filter" tuned to 2 kHz to the stimulus. After DUIFHUIS (1970)

40, 60, and 80. The threshold (not the threshold of audibility but the lowest level at which the added component could be recognized as a pure tone) was studied as a function of the delay t, between the pulse and the centre of the tone bursts. The results are presented as a function of the dimensionless variable k, defined as $k = nt/T$ where T is the time between successive pulses. Figure 25 shows the results for three different values of T and a fixed value of 2 kHz for the frequency of the n-th component. This figure shows that the threshold declines smoothly when the tone burst is moved away from the pulse. The course of this decline reflects the decay of activity of the auditory filter at this frequency after it has been excited by a pulse. DUIFHUIS was able to estimate the value of 10 as the best fitting value of the so-called "quality factor" of the pertinent resonant circuit. This value agrees well with similar values determined from responses of auditory nerve fibers (KIANG *et al.*, 1965). Corresponding values obtained from psychophysical measurements of masking patterns indicate similar or somewhat higher values.

8. Repetition Pitch

We conclude this part with the description of auditory phenomena that are not readily explained by either a temporal or a spectral analysis. "Time-separation pitch" (see Section E. 12) as studied by SMALL and associates seems at first sight a purely temporal phenomenon: pitch is closely associated with time separation. This simple-minded notion is reinforced when we realize that this pitch remains clear when the repetition rate of the pulse pairs is made very low or when the pulse pairs (always with the same interval) appear at a random rate (MCCLELLAN and SMALL, 1967). However, the pitch deviates when the second pulse of each pair is inverted (FOURCIN, 1965). As will be seen later, one possible explanation of this effect includes spectral filtering and the manifestation of a region of spectral dominance.

In the present section, we shall describe an effect that likewise suggests a purely temporal type of auditory processing but that turns out later to be more complex in nature. The effect is called "repetition pitch" (BILSEN, 1966) and we shall consider here only the monaural case. To the ear of the listener is presented the sum of two signals, a white-noise signal $n(t)$ and the identical white-noise signal delayed by a time τ. The waveform is then

$$x(t) = n(t) + pn(t - \tau), \tag{11}$$

where $p = +1$. This signal acquires a peculiar timbre due to the presence of the echo but by varying the delay τ it readily becomes apparent that there is a pitch associated with it. It comes as no surprise that the pitch corresponds to the frequency $1/\tau$ just as for time-separation pitch. The name "repetition pitch" (RP) seems quite appropriate since the phenomenon is manifest only when the delayed signal is highly correlated with the first signal. When the polarity factor p is made -1, the echo has the opposite polarity. There results a definite shift of the pitch of a little over two semitones (*cf.* FOURCIN, 1965). BILSEN (1966) made careful measurements of these pitch shifts, and he found that there were two pitches possible. These corresponded to the frequencies $1.14/\tau$ and $0.87/\tau$, respectively. Essentially the same values were obtained when the identical procedure was repeated but with randomly presented pulse pairs (Time Separation Pitch). Under the conditions tested, the accuracy of pitch matches appeared to be somewhat better for time-separation pitch than for repetition pitch.

The $p = -1$ condition is obtained by inverting the delayed version of the original input signal $n(t)$; or, in other words, shifting the phases of all components of the delayed signal over 180 degrees. A situation "halfway" between $p = +1$ and $p = -1$ can be obtained by shifting the phases of all components of the delayed input signal over 90° before adding the echo to the input. This case cannot be described by a specific value for p in Eq. (11) since the waveform of the "echo" is now quite different from that of the original signal $n(t)$. In fact, waveforms like these can only be obtained with sophisticated signal processing methods. The procedure resulted in pitch shifts about halfway between those discussed above. The pitch for the 90° condition was approximately $1.08/\tau$; this holds for both repetition pitch (RP) and time-separation pitch (TSP). Note that such a result cannot be accounted for by a simple temporal processing mechanism, a point to which we will return later.

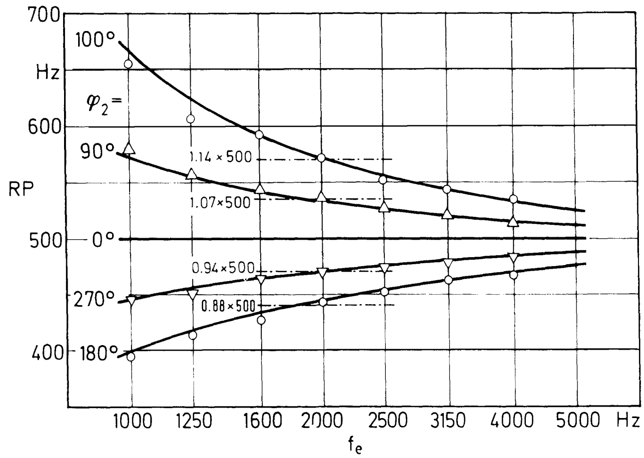

Fig. 26. The pitch of repetition pitch (RP) for four experimental conditions. Signals are filtered to a 1/3-octave bandwidth. Time delay τ: 2 msec. Abscissa: centre frequency of the filter. Central line: p = +1. Circles: p = —1 (two pitches). Triangles: echo phase shifted over +90° or —90°. Solid lines: computed according to RP = $1/(\tau \pm 1/2f_0)$ for p = — 1 and RP = $1/(\tau \pm 1/4f_0)$ for the ±90° conditions. Dashed lines: unfiltered signal with p = —1. After BILSEN and RITSMA (1969/1970)

In a later series of experiments, BILSEN and RITSMA determined the pitches for filtered signals. In view of the intimate relation between RP and TSP, they performed their experiments only with one type of signal, a series of pulse pairs with time interval τ. The (average) repetition rate was 20 times per second; the pulse pairs could be presented periodically or with random inter-pair intervals. The signal was filtered by a third-octave band-pass filter before being led to the ear. Figure 26 shows a representative result, the measured pitch for four conditions of the "echo" as a function of the central frequency of the band-pass filter. A theoretical estimate of expected pitch shifts can be obtained as follows. The response of the third octave filter with central frequency f_0 to the first pulse is a short-duration oscillation with frequency f_0 (the filter "rings" because it has relatively low damping). The response to the second pulse is similar but it differs in details depending upon the phase shift that the second pulse has undergone. For the case $p = +1$ the distance between the most prominent peaks will be an integral number of periods $T_0 = 1/f_0$. This distance will be close to τ, of course. For $p = -1$, the distance will be an odd number of half periods. This leads to the prediction that for the $p = -1$ condition the pitch will be $1/(\tau \pm 1/2f_0)$. A similar prediction can be made for the 90° and 270° conditions. These theoretical functions are drawn as solid lines in Fig. 26. It is seen that the measured points are all quite close to these lines.

The same reasoning can be applied backwards for the case in which the signal is not filtered. The measured pitch shifts then correspond to the pitch shifts for a value of f_0 which is about 4 times $1/\tau$. See the dashed lines in the figure. This can be formulated in terms of a dominant region; for wide-band RP signals, the pitch shifts are determined from the frequency region that is centered at 4 times $1/\tau$. The authors conclude that it is not possible to explain the properties of RP solely in the temporal domain. Neither is it possible to find a explanation in the spectral domain. We shall come back to this point later (Section H. 14). Only a combined approach, in which a temporal analysis is carried out from the signal as it appears filtered in the dominant frequency region, is fruitful. It remains puzzling that the auditory system does not employ a fixed dominance region but seems to adjust it on the basis of the time difference τ between signal and echo.

G. The Missing Link (Towards a Unified Framework)

1. Nonlinear Processes, Combination Tones

Several of the studies described in the preceding sections have put more emphasis on the spectral properties of residue signals than on the periodicity. This trend was initiated by the work on dominance (Section F. 4ff.) and it has been substantiated by later studies. In view of this aspect, it is necessary to give additional details about the internal representation of the spectrum. Two aspects of this have already been described, aural resolution (or critical-band filtering) and specific temporal effects induced by isolated high-frequency components. There is one aspect in which the internal representation of the spectrum differs from the objective spectrum, even when these two points are taken into account. Under special circumstances, our auditory system generates distortion products that contribute significantly to our perception. The presentation of a residue signal consisting of just a few components makes this effect especially noticeable. That is the reason why we have to be more specific about auditory nonlinearity than we were in Section B. 3.

The process of nonlinear distortion is usually studied in the case of a simple nonlinear transfer system. Consider a system in which an input signal $x(t)$ is transformed into an output signal $y(t)$ in such a way that each value of $y(t)$ is just a function of the value of $x(t)$ at the same instant of time:

$$y(t) = F\{x(t)\}. \tag{12}$$

Such a system is an instantaneous nonlinear transform system, often referred to as "memoryless". If the function $\mathrm{F}(a)$ which is the descriptive function of this system, were linear in a, no new Fourier components would be generated. Suppose now that $F(a)$ is a quadratic function of a, then the effect of Eq. (12) would be that each signal would appear to be multiplied by itself. Compare Eq. (5) of Section E. 6 when $x_1(t)$ and $x_2(t)$ refer to the same sinusoidal signal $x(t)$. Two new components will arise, one with the frequency 0 (which is immaterial in this context) and one with the double frequency. Hence, such a nonlinearity produces the second harmonic of a sinusoidal input signal. If the input signal $x(t)$ contains two components with frequencies f_1 and f_2, the output will consist of these harmonics and two further components having frequencies equal to $f_2 - f_1$ and

$f_2 + f_1$. Such a purely quadratic distortion is rarely encountered, however. It would be more appropriate to consider $F(a)$ as mixed:

$$F(a) = c_1 a + c_2 a^2. \tag{13}$$

Such a system transmits the original frequencies f_1 and f_2, and it generates distortion products with frequencies $2f_1$, $2f_2$, $f_2 - f_1$, and $f_2 + f_1$. A more complicated nonlinear system is described by

$$F(a) = c_1 a + c_2 a^2 + c_3 a^3. \tag{14}$$

The third-order term produces the following distortion products for a two-component stimulus:

a) the third harmonics of each component, frequencies $3f_1$ and $3f_2$,

b) third-order combination tones, frequencies $2f_1 \pm f_2$ and $2f_2 \pm f_1$.

Of these, the $2f_1 - f_2$ term turns out to be extremely important. Higher-order terms in a series like:

$$F(a) = c_1 a + c_2 a^2 + c_3 a^3 + c_4 a^4 + c_5 a^5 + \dots \tag{15}$$

produce higher-order harmonics of each input component and combination tones of corresponding complexity [see Eq. (6) of Section B. 3].

Let us consider now a residue signal consisting of a small number of consecutive harmonics of a fundamental frequency f_0. The lowest possible number of components is two, so let the frequencies be:

$$\begin{aligned} f_1 &= n f_0 \\ f_2 &= (n+1) f_0 \quad (n \text{ integer}). \end{aligned} \tag{16}$$

The third-order difference tones now have the frequencies:

$$\begin{aligned} 2f_1 - f_2 &= (n-1) f_0 \\ 2f_2 - f_1 &= (n+2) f_0. \end{aligned} \tag{17}$$

These are just the neighbouring components in the full harmonic series. If these combination tones are audible, it may well be that they play a role in residue perception.

2. Amplitude Functions

Before turning to experiments on the detectability of combination tones, we have to discuss one more topic of interest, the amplitude behaviour. We saw that each term of the series (15) gives rise to distortion components with frequencies $k_1 f_1 \pm k_2 f_2$, where $k_1 + k_2$ is equal to the order m of that term[3]. The creation of such components can be proven mathematically by substituting a two-component signal:

$$x(t) = A_1 \cos 2\pi f_1 t + A_2 \cos 2\pi f_2 t \tag{18}$$

into the corresponding term of (15). The reduction of the trigonometric formulae then shows one other important fact. The amplitude A of the distortion component with frequency $k_1 f_1 \pm k_2 f_2$, generated by the m-th term in the series (15),

[3] k_1 and k_2 are restricted to be positive numbers in this section.

depends in the following way upon the amplitudes A_1 and A_2 of the input components:

$$A_d = \text{const.}\ A_1^{k1} \cdot A_2^{k2} \tag{19}$$

with $k_1 + k_2 = m$. Hence, the third order-combination tone with frequency $2f_1 - f_2$ should have an amplitude proportional to the square of A_1 and to the first power of A_2. Note that this special amplitude dependence is dictated by the use of a power term in Eq. (15). Note also that the amplitude dependence will be different when more than one term in the series (15) contributes to a particular distortion product.

That the individual exponents of the factors in (19) are just k_1 and k_2 is difficult to see without mathematics. That the sum of the exponents should be equal to the degree m of the corresponding term is easily seen as follows. Consider the case where we first stimulate the system with a certain signal with amplitude A_0 (*e. g.*, $A_1 = A_2 = A_0$, but such a restriction is not necessary) and that we observe the combination tones generated by the term of the m-th degree. When the system is stimulated next with the same waveform but with the amplitude kA_0, this term will produce the same combination tones but the amplitude will be multiplied by k^m. Hence, the sum of the exponents in (19) must be equal to m, the degree of the term that produces the combination tones under study.

3. The Difference Tone ($f_2 - f_1$)

With this somewhat more detailed knowledge about distortion products, we turn to experimental evidence about the audibility of harmonics and combination tones. When one moderately loud purely sinusoidal signal is presented to the ear, the second harmonic can easily be distinguished; it sounds an octave higher than the primary tone. With various methods, the effective strength of the second harmonic can be measured as a function of the intensity of the primary tone. In general, the effective strength of the second harmonic behaves as the theory predicts: this distortion product can be described as if it were generated by the second term of the series (15). The primary tone has to be rather strong; when it is below 50 dB, no second harmonic is heard. Once the second harmonic is audible, it tends to have an effective intensity that rises as the square of the intensity of the primary tone. Or, in terms of decibel levels, for each 10 dB increase in level of the primary tone, the second harmonic increases by the double amount, 20 dB. Very little is known about the third aural harmonic; it probably is not significant.

There is a rich variety of distortion products when a signal is presented that contains two sinusoidal components. Let us call the frequencies of the two primary components f_1 and f_2, as before. Of these frequencies, f_1 is the lower. A quadratic term in the series (15) would produce (apart from the second harmonics of the two components) two combination tones, the so-called *difference tone* with frequency $f_2 - f_1$ and the *sum tone* with frequency $f_1 + f_2$. The difference tone has had a long history in music (*cf.* Plomp, 1965). Scientific investigation has shown that its behaviour can be described very well by the quadratic term in (15). In particular, the effective strength of the difference tone behaves as the square of the amplitude of the components, as the theory predicts (Zwicker, 1955). In the days of Helmholtz, the source for the underlying type of nonlinearity was sought in the middle ear. More recent experiments indicate that the frequency ratio f_2/f_1 has a larger influence on the difference tone than is compatible with a middle-ear origin (Plomp, 1965; Hall, 1972a). It is possible that both the middle and the inner ear contribute to the production of the difference tone. The sum tone is audible only under special circumstances; in most cases, it is masked by the primary components (masking extends more toward the higher than toward the lower frequencies) and the second harmonics of the two primary components. Hence, it is not known whether the strength of the sum tone is compatible with that of the difference tone as the theory requires.

4. The Cubic Difference Tone ($2f_1 - f_2$)

The cubic difference tone (CDT) with frequency $2f_1 - f_2$ is unusual in all respects. A distortion product with this frequency would be generated by the term of the third order in (15). As is pointed out in Section G. 1, the CDT could contribute to residue perception if it were audible. Many studies have been devoted to the properties of this particular combination tone, and it has turned out that this distortion product has bridged the gap between older theories of the residue and the newer developments that have been hinted at in Section G. 1. As a consequence, we shall devote special attention to the properties of the CDT.

First, a few words about measurement techniques. The primary signal consists of two (sinusoidal) components with frequencies f_1 and $f_2 (f_2 > f_1)$. When a distortion tone is audible because it is generated in the ear, it is generally quite difficult to estimate its loudness; the primary tones produce too much interference. The strength of the distortion product is, therefore, determined by an indirect method. The signal presented to the ear is made to contain an additional

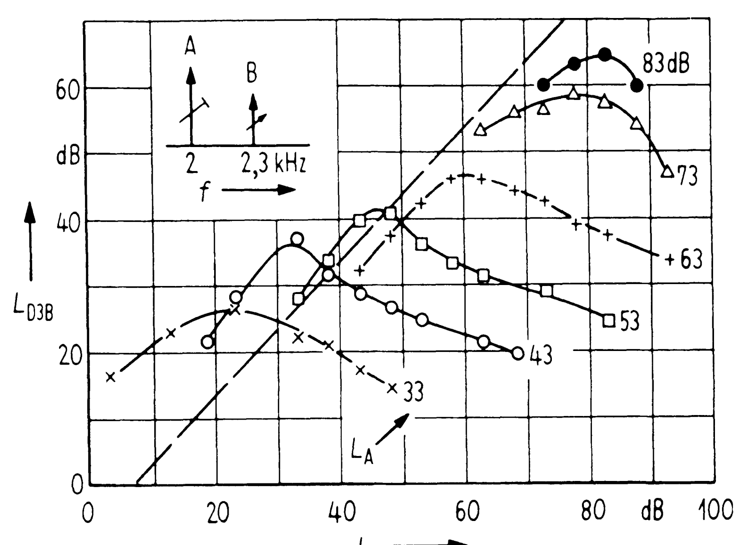

Fig. 27. Level of the cubic difference tone (CDT) as a function of the level of the higher component. Frequencies of the primary components: 2 and 2.3 kHz. The parameter is the level of the lower component. After Zwicker (1968)

component with (exactly) the frequency $2f_1 - f_2$ of which amplitude and phase can be adjusted. The additional tone is called, cancellation tone, and the experiment is aimed at compensation of the aural distortion product by the cancellation tone. In the experiment, amplitude and phase of the cancellation tone are adjusted in such a way that the distortion product becomes completely inaudible. The adjustment is carried out by the subject himself.

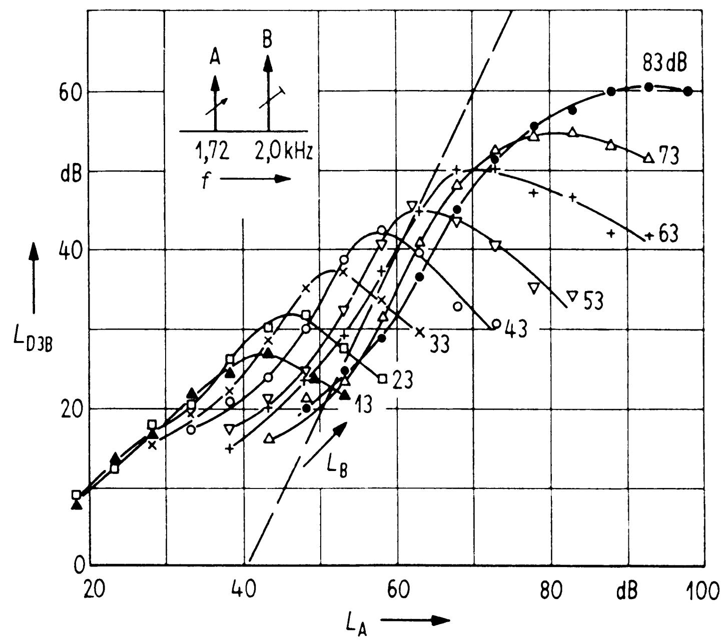

Fig. 28. Level of the cubic difference tone (CDT) as a function of the level of the lower component. Frequencies of the primary components: 1.72 and 2.0 kHz. The parameter is the level of the higher component. After ZWICKER (1968)

For the moment, we shall be concerned primarily with the amplitude of the CDT. ZWICKER, in 1955, was the first to report the unusual amplitude behaviour of the CDT with frequency $2f_1 - f_2$. Figures 27 and 28 taken from a later paper by ZWICKER (1968), serve to illustrate the findings. For Fig. 27, two tones, labelled A and B, were presented. The lower one had a fixed amplitude; the higher one was varied in level as the experimental variable. Theoretically, the level of the CDT should be varying precisely as much as the level of the higher primary tone in this situation (see the dashed line in the figure). It is seen that only the extreme left-hand parts of the experimental amplitude functions agree with theory. The main parts of the functions are quite different. In Fig. 28, the upper component is held at a fixed level and the lower one is varied. For this situation, the theory predicts a quadratic behaviour; the level of the CDT should vary over 20 dB for each 10 dB of variation of the lower-tone level (see the dashed line). It is seen that only a very small part of the curves has this theoretical slope. And, again, the main part is quite different.

One other feature of these findings is remarkable. The combination tone is audible when the components are rather weak. This is completely unlike the properties of, *e. g.*, the difference tone described above and is in complete disagreement with theory. When both primary components are varied in level simultaneously, the CDT should theoretically vary three times as much (30 dB for every 10 dB). The facts are quite different. For levels up to 60 dB, the CDT level varies at the same rate as the primary components. Hence, in this range, the *relative* CDT level is almost constant. Or, in other words, the percentage of distortion is almost constant. For primary-tone levels above 60 dB, the CDT level does not grow as much; it seems to saturate and reaches a maximum of 50 dB for primary levels of 100 dB. In this range, the percentage of distortion decreases with increasing level.

5. Essential Nonlinearity, Goldstein's Work

These findings indicate that the distortion underlying the CDT is not limited to high intensities as is the case for harmonic distortion and production of the difference tone. On the contrary, the distortion is manifest over almost the entire range of intensities. For low intensities, the distortion tone level may lie 15–20 dB below the level of the primary tones; this holds true almost down to threshold. Hence, even for very low levels the distortion is not at all negligible. Figure 29 illustrates the relative level of the CDT (relative to the level of the primaries). This figure is taken from a comprehensive paper on auditory nonlinearity by GOLDSTEIN (1967b). The cancellation-tone method was somewhat refined by an expedient intended to help the subject decide when the CDT disappeared completely. The stimulus, consisting already of three components (the two primary tones and the cancellation tone) was made to include a fourth tone. This tone,

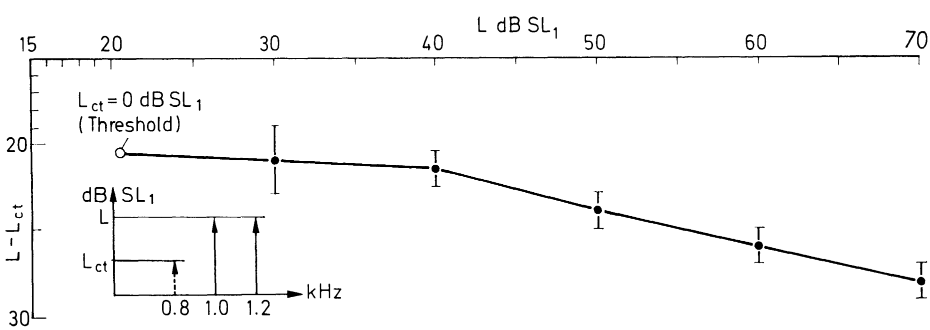

Fig. 29. Relative level of the CDT as a function of the level of the primary components. The inset shows the stimulus configuration in spectral terms (subject: JLG). After GOLDSTEIN (1967b)

termed probe tone, had a frequency differing a few Hz from that of the cancellation tone and was presented at a rather low intensity. As long as the compensation of the aural CDT by the cancellation tone is not perfect, the probe tone produces beats and these can serve to facilitate the adjustment for perfect compensation.

Figure 29 shows GOLDSTEIN's measurements of the relative CDT level as a function of the sensation level of the two primary tones. For low levels, the relative CDT level is almost constant which means that the relative amount of distortion is almost constant. For higher levels, the CDT level tends to saturate with the result that the relative CDT level begins to decrease.

Theoretically, the relative amount of distortion should always increase with increasing level. Hence, it is clear that the type of distortion that underlies the production of the CDT is completely unlike any mechanism that leads to a series development as (15). The nonlinearity that is responsible for CDT production has been termed an "essential nonlinearity" by GOLDSTEIN (1967 b), and later authors have continued to use this term (see, *e. g.*, SMOORENBURG, 1972 b). The theoretical aspects of essential nonlinearities have not been studied exhaustively although a few scattered results which are pertinent to the problem at hand are available (SCHROEDER, 1975; DE BOER, 1975; DUIFHUIS, 1975).

Two other aspects of the CDT findings are of interest. The first is that the CDT level is strongly dependent upon the frequency separation of the two primary tones. The CDT is strongest when the two tones are very near to one another, *e. g.*, when $f_2/f_1 = 1.1$. The CDT becomes much weaker when the tones are separated. The limiting point is a frequency ratio f_2/f_1 of 1.2–1.3 for sound levels of 40 dB; the CDT becomes rapidly weaker when the frequency ratio is increased. When we realize that cochlear mechanics causes each component to excite only a limited part of the organ of Corti, this strong frequency dependence is suggestive of a cochlear interaction process as the origin of the CDT. Moreover, no frequency selectivity of comparable sharpness is found in the middle ear. The problem of aural distortion, thus, appears to be very important for the study of cochlear physiology, a point to which we shall return later.

Up to this point, no mention has been made of the phase data that are produced by the experiments discussed. For difference tones and aural harmonics, the phase of the cancellation tone necessary to achieve compensation is almost independent of level. The phase data for the CDT, however, are strongly level dependent, especially in the region of low intensities. This property makes it difficult to give a description in general terms. A complicating factor is that recent studies have borne out that, in special situations, the level dependence of the CDT is not always as smooth as shown in Figs. 27 and 28. Any irregularity in the amplitude values is accompanied by irregularities in the phase. Representative papers have been written by HELLE (1969, 1970) and HALL (1972 a, b); SMOORENBURG (1972 a, b) has encountered the same phenomenon. The relevance of the CDT phase data for revealing the properties of aural frequency resolution has been discussed by SCHROEDER (1969).

The cubic difference tone (CDT) with frequency $2f_1 - f_2$ is not the only combination tone with the remarkable amplitude behaviour described above. There are combination tones with frequencies

$$f^{(k)} = f_1 - \mathrm{k}(f_2 - f_1) \quad (\mathrm{k} = 1, 2, 3\ldots) \tag{20}$$

that have similar properties. The case k = 1 corresponds to the CDT. For k = 2, we encounter a combination tone of the fifth order; for k = 3, of the seventh, etc. Note that all these combination tones, whenever they would be audible, can act as extensions of the spectrum of a residue signal [compare Eq. (16) and (17)]. More will be said about these higher-order combination tones later.

Finally, we must mention the fact that most studies have concentrated on combination tones that lie below the frequency region of the primary tones. In contrast to the CDT with frequency $2f_1 - f_2$, the CDT with frequency $2f_2 - f_1$ lies above the primary tones. This CDT is somewhat elusive; although it tends to be masked by the primary tones, this is by no means certain, yet its existence has not been described.

For the sake of completeness, we note that some confusion has arisen concerning the basic properties of the difference tone with frequency $f_2 - f_1$. Depending upon the detection criterion used, the difference tone can show the same type of amplitude behaviour as the CDT (see, *e.g.*, HALL, 1972b). The reasons why different experiments can yield such widely divergent results are not clear at the moment.

6. Residue and Combination Tones — Smoorenburg

After this excursion to the domain of aural distortion, we return to the perception of the residue. Successive parts of the present chapter have been concerned with residue signals with an ever decreasing number of components. SCHOUTEN used stimuli with a large, indefinite number of components. DE BOER reduced the number to 7 and 5; RITSMA (in his principal experiments) to 3. SMOORENBURG (1970) took the next step; he studied the perception of signals with only *two* frequency components (the ultimate step of the perception of single-component signals related to residue phenomena has also been taken; we shall come to speak about this in the concluding section). In his first experiment, SMOORENBURG used 42 test subjects; they were required to judge which of two signals had the higher pitch. One of the signals had components of 1800 and 2000 Hz, the other of 1750 and 2000 Hz. The former signal thus had a fundamental frequency of 200,

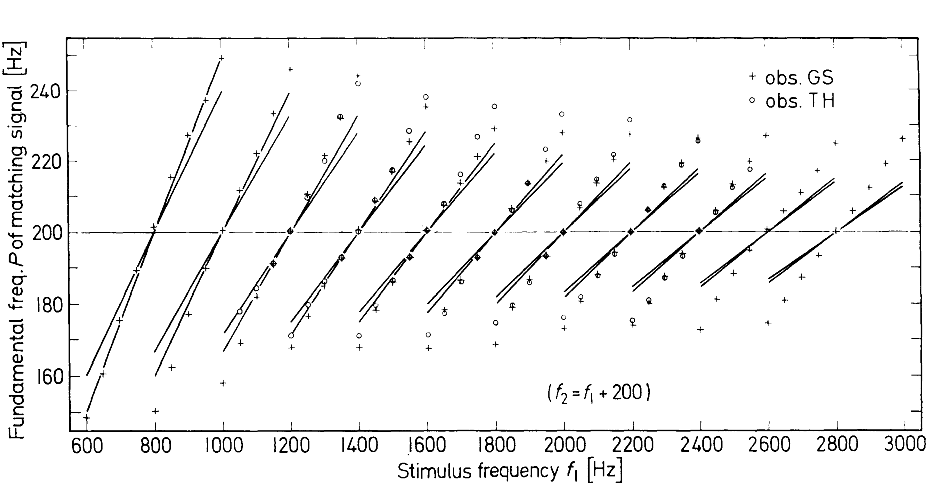

Fig. 30. Results of pitch measurements for two-component signals. Frequency difference of components: 200 Hz. Abscissa: frequency of the lower component. Ordinate: measured pitch value. The thin lines correspond to pitch shifts according to Eq. (9). After SMOORENBURG (1970)

the latter of 250 Hz. If a listener perceives a residue pitch, he will judge the second signal as having the higher pitch. If the listener's first impression of pitch is mediated by the components (in his perception, the part-tones; see the discussion on terminology in Section B. 5), his judgement will be the opposite one. From the results it appears that one of the two possible criteria is used consistently by each subject; there seems to be no rivalry between the two ways of judging the pitch. In about half of the subjects, the primary pitch impression was based on the complex tone being perceived as a whole and not on the part-tones.

Two of the subjects of the latter group participated in an extensive study of the pitch of two-tone complexes. The two components were produced by two tone generators; the components were kept 200 Hz apart; for this value of the frequency difference a residue pitch is audible when the stimulus frequencies are varied over a wide range. Another two-component signal was used as the comparison signal. This second signal was always harmonic; it was generated with special electronic equipment. To avoid the possibility that pitch matches were inadvertently based on part-tones, the ratio n of the lower component (frequency f_1) to the frequency difference was chosen so as to be different from n of the test signal. It is to be noted that the residue pitch for these two-component signals is quite weak, especially when $f_1/200$ is high. Nevertheless, reliable pitch matches were obtained for values of $f_1/200$ well over 10 (compare this with the limit of 20 reported by Ritsma in 1962 for three-component signals).

Figure 30 shows the results of the pitch measurements for the two observers. The abscissa is the frequency f_1 of the lower component; the frequency of the second component is 200 Hz higher. The ordinate is the "pitch" of the comparison signal, *i. e.*, the fundamental frequency of the comparison signal. When the test signal is harmonic, the pitch is observed to be equivalent to 200 Hz. For inharmonic signals, the pitch goes up and down around this value just as for signals with more than two components (compare Fig. 30 with Figs. 15 and 19). The full-drawn lines in Fig. 30 indicate the pitch course in the sense of the "first effect of pitch" (Section E. 7) computed according to Eq. (9) with n equal to $f_1/200$ and $f_2/200$, respectively. Remember that $f_2 - f_1 = 200$ (Hz) throughout these measurements. It is observed that the actually observed pitch shifts of inharmonic signals are larger, in other words, that the "second effect" is quite substantial.

The result is surprising. If we recall that in the pseudo-fundamental theory (see Section E. 9) the second effect can be explained by weighing the lower component more heavily than the higher one, we realize that for a two-component signal there would be very little space for a second effect. The same holds true when different pseudo-periods, measured at different locations of the basilar membrane, are considered (see Section E. 13). The largest pitch shift is obtained when the constant n in Eq. (9) is taken equal to $f_1/200$. However, the actual pitch shifts are much larger. Or, stated in a different way, the centre of gravity (*cf.* Section F. 5) lies completely outside the complex!

This point raises the question, do combination tones play a part in pitch perception? Combination tones with frequencies as given by Eq. (20) with low integral values of k would extend the given spectrum at the low-frequency side with equidistant components. If the two primary frequencies are shifted upwards over Δf Hz, all combination tones shift by the same amount. Hence, if these combination tones are audible, the aural spectrum is much richer in components than the objective one. This idea was investigated by SMOORENBURG in great detail and with several methods. Only part of this work will be described here: the remainder can be found in SMOORENBURG's papers (1970, 1972a, b).

7. Existence Region for Combination Tones

The first question is: under what conditions are the combination tones audible? This question was answered by a systematic search directed at the *existence region* for combination tones. First, f_2 was chosen twice the value of f_1, and then f_2 was gradually lowered in frequency while f_1 was kept constant. The subjects were instructed to notice the appearance of successive combination tones. The appearance of a new combination tone was noticed easily; recognition is facilitated because the pitch of the combination tone changes in the direction opposite to that of f_2. By comparison with a sinusoidal signal, the pitch of the combination tone was determined and this immediately yielded the appropriate value of k [see Eq. (20)]. In this way, the audibility of combination tones of different order (3, 5, 7..., corresponding to k = 1, 2, 3...) was determined for all possible combinations of f_1 and f_2. Figure 31 shows the results, interpolated for the situation where $f_2 - f_1 = 200$ (Hz). The result is expressed in the form of a variable λ, which is to be interpreted as follows. At a specific combination of f_1 and f_2 the

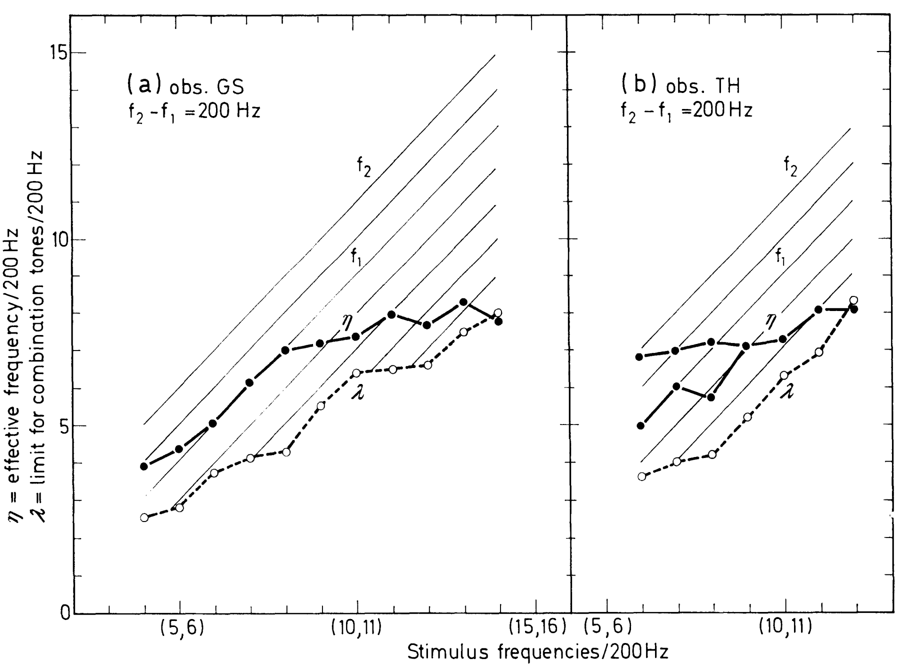

Fig. 31. Lower limit λ of the existence region of combination tones. Effective rank number η corresponding to measured pitch shifts for inharmonic two-component signals. Both functions are plotted against frequency (shown in a relative sense). The solid lines indicate the components of the "internal spectrum". Modified from SMOORENBURG (1970)

frequency f_λ of the lowest audible combination tone is measured. This frequency, divided by the difference $f_2 - f_1$ (200 Hz for this figure), yields λ. Hence, λ can also be regarded as the rank number of the lowest audible combination tone when all components are considered as harmonics (note that this interpretation holds only when the complex is harmonic). That λ turns out to be non-integral is the result of averaging over a number of experiments.

The upper two straight lines in Fig. 31 denote the primary components themselves (note that the abscissa is labelled in a relative sense). To improve clarity, several other lines have been drawn, each corresponding to one combination tone. These lines are all parallel and terminate on the function depicting λ. The figure, thus, gives an accurate account of the richness of the spectrum as it is perceived by the ear.

The second function depicted in Fig. 31 results from the pitch matches shown by Fig. 30. For inharmonic signals, the pitch deviates more from 200 Hz than is accounted for by Eq. (9). Let us rewrite this equation for deviations of pitch Δp rather than the pitch itself, and let us introduce an arbitrary constant η instead of n:

$$\Delta p = \frac{\Delta f}{\eta}. \tag{21}$$

Here, Δp is the difference of the pitch with respect to the nearest harmonic situation, and Δf is the deviation of f_1 and f_2 from the nearest harmonic values. The constant η can be interpreted as the rank number of the harmonic that acts as the centre of gravity of the complex. However, it is no longer restricted to be an integral number. If the aural spectrum would consist of the two primary components only, η could only lie between n and n + 1 when n is the rank number of the lower component. The results of the pitch matches indicate that this does not hold true. In terms of Eq. (21), η should be lower than n to account for the data. From each pitch match, the equivalent value of η can be computed and the average values of η so obtained are included in Fig. 31. If we interpret η as the rank number associated with the centre of gravity of the aural spectrum, we see from the figure that η always lies *inside* the existence region of the combination tones. This lends considerable support to the assumption that the combination tones act as extensions to the spectrum and are involved in the determination of residue pitch.

Several additional arguments in favour of this hypothesis can be formulated. At the outset of Section G. 6, it was mentioned that there are essentially two categories of listeners. There proved to be a significant correlation with the existence region: the lower limit λ of the existence region for combination tones was lower for those subjects that judged pitch on the basis of the complex perceived as a whole, and it was higher for those subjects that judged pitch on the basis of individual part-tones. In further experiments, the effective width of the existence region for combination tones was experimentally manipulated. The existence region becomes smaller when the signals are weaker, and one may expect a larger value of λ as well as η. Similar effects are expected when the frequency region below the two primary tones is masked by a band-limited noise signal. Both expectations were confirmed by experiments.

From this work, we can conclude that we now have a fair understanding of the large pitch shifts found in experiments with inharmonic signals consisting of a very low number of components. If the primary components have low rank numbers, the effective rank number η will tend to refer to the fourth harmonic because of the principle of dominance. If the stimulus frequencies are all above the dominant region, the pitch will be determined by the lowest components perceived by the ear. In most cases, the effective rank number η then points to the region where combination tones have considerably enriched the spectrum.

For the highest complexes for which a pitch determination was reported to be possible, the effective value of η was limited to about 8. This suggests that generation of residue pitch by two-component signals requires at least one part-tone (combination tones included) that corresponds to a frequency not exceeding the eighth harmonic. In the series of experiments reported, this limiting situation occurred for $n = 14$. Note that in PLOMP'S experiment (1964) all harmonics up to the ninth were found to be at least partly analyzable (see Section F. 6). The experiments reported in this section cannot yield a decision between the two alternatives for the residue pitch mechanism, temporal versus spectral processing. The results do lend more support to the theory involving spectral pattern recognition. Combination tones behave like tones that are physically present in the ear (see the next section). Thus, a pitch mechanism based upon the spectrum may be simply extended to include combination tones. An appropriate non-symmetrical weighing of the components will lead to predicted pitch values that agree with experimental ones. At the point where the analyzability of components fails completely, about the ninth to tenth harmonic, the pitch mechanism fails as well.

A simple spectral pattern recognition theory cannot account for all experimental results. For instance, RITSMA reported that for a three-component signal, pitch matches could be made for signals with harmonic numbers up to 20. Even when combination tones are included, such signals have only components with harmonic numbers of 14 and higher. And even the lowest components of such a complex cannot be analyzed by the ear at all. Most likely the mechanism of pitch determination is a combination of two processes, one based on temporal and the other on spectral cues. Reference may be made to the work by WALLISER (1969b) in which it is suggested that a first estimate is based on temporal properties and that the final pitch value is selected as the submultiple of the frequency of the lowest component that is closest to the first estimate. In any event, the results of the experiments reported by SMOORENBURG suggest strongly that detection of the residue pitch requires at least spectral information and confirm that combination tones considerably enrich the aural spectrum.

8. Physiological Considerations

The experiments described above have contributed enormously to our knowledge of the properties of residue pitch. It has not yet been possible to decide upon the most probable underlying mechanism of pitch determination. In any event, a process of pattern recognition of partly resolved spectral components is almost certainly involved. If that is true, a search for the physiological counterpart of the residue pitch mechanism will be extremely difficult if not impossible at this moment. It does not seem likely that we know enough about the physiological processing of signals to be able to pinpoint any physiological mechanism as being the residue pitch processor. Furthermore, we must consider all physiological evidence with extreme caution. From time to time, there appear reports about neurons in the auditory pathway that have properties closely associated with the envelope of auditory waveforms. Although temporal processing may play a part, the ultimate determination of residue pitch seems more likely to be based on spectral cues. Hence, the temporal mechanisms discovered may be involved in residue processing but almost certainly they are *not* related to the ultimate pitch processor. This is one of the reasons why the present author has considered it necessary to describe residue theory at such a length and to discuss so many auxiliary phenomena so deeply. For a complicated, and possibly multi-stage process like the one under discussion, it seems unlikely that the neurophysiological counterpart will be discovered soon; neurophysiologists should be aware of this situation. The author hopes that neurophysiologists do find the required background in the present text and invites readers to communicate their opinions, criticisms and questions to him.

The situation is entirely different for auxiliary phenomena like the production of harmonics and combination tones. Distortion products in electro-physiological potentials have received considerable attention (*e. g.*, Wever and Lawrence, 1954; Dallos, 1969; Tonndorf, 1958). There is every reason to believe that the cause for the perceptual appearance of distortion products is located in the cochlea, and, hence, the physiological findings are highly relevant. Let us concentrate on the CDT which is the most interesting distortion component in the present context.

The CDT is also the component about which the largest amount of physiological knowledge has been gathered, but not conclusive evidence as we shall see presently. Distortion products are prominent in the cochlear microphonic potential (CM). Harmonics and difference tones as well as higher-order distortion products are readily detected with specialized electronic equipment. A systematic study of the properties of the CDT (Dallos, 1969) has yielded the result that the CDT in the CM follows the theoretical description rather accurately. No evidence about an essential nonlinearity was found. In this respect, the physiological findings are quite contrary to the psychophysical results as described above. This discrepancy has led Dallos to conclude that the microphonic potential, although undoubtedly due to important physiological events in the cochlea, is an epiphenomenon that does not reveal the presence of all audible components.

Dallos' results are somewhat remarkable since there are many reasons why the cause of CDT production is to be sought either in the mechanics of the cochlea or in the sensori-neural transduction process, and it seems strange that the distortion does not reflect back on the CM. Let us now consider cochlear mechanics and the generation of action potentials in single fibers of the auditory nerve in somewhat more detail. The most direct evidence that cochlear movements are nonlinear has been obtained by Rhode (1971). He measured the response of a particular spot on the basilar membrane as a function of frequency. The experiments were performed in squirrel monkeys and detection of the minute movements of the basilar membrane was possible by the use of an advanced measurement method borrowed from nuclear physics. The location chosen for observation showed a clear and pronounced resonance at a frequency near 7 kHz but only for weak sound stimuli. For sound signals of over 70 dBL, the resonance was observed to be damped, and at 90 dBL, the peak demonstrating the resonance had almost disappeared. Unfortunately, RHODE could not perform measurements at sound levels below 70 dB; hence, we have no direct evidence on mechanical nonlinearity at extremely low levels. The observations suggest, however, that the mechanical nonlinearity operates only at high levels, and that it dampens the resonance progressively when the stimulus level exceeds 70 dB.

The type of nonlinearity observed by RHODE is of the right type to yield a CDT, and, as such, it has been used as a basis for models of cochlear mechanics (KIM and PFEIFFER, 1973; HALL, 1974). The models constructed do show production of a CDT (and higher-order combination tones) but only for moderately loud to loud stimuli. Insofar as explanation of CDT behaviour at these higher levels is concerned, such models are quite powerful (*cf.* SCHROEDER, 1975). If our inference from RHODE'S data is correct, cochlear mechanics on the level of basilar-membrane movement does not contain essential nonlinearities. Hence, the origin for the observed CDT at low sound levels is probably not located in cochlear mechanics.

9. Single-Fiber Responses

Let us now turn to the responses of single fibers of the auditory nerve. Here, we find neurophysiological counterparts of several psychophysical phenomena, notably frequency selectivity and masking (see the authorative review by EVANS and WILSON, 1973). The sharpness of frequency selectivity is much higher than the sharpness reported for cochlear mechanics by VON BÉKÉSY (1960). Although it is well recognized now that VON BÉKÉSY'S experiments were performed at such high levels that nonlinear distortion must have swamped out almost all resonance, and that more modern methods have yielded considerably sharper resonance curves, there is still no agreement about the question whether mechanical selectivity is sufficient to explain neural selectivity. The most widely held opinion is that the selectivity displayed by responses of single fibers is substantially sharper than the selectivity of mechanical response curves. However, it could be that mechanical sharpness is greater than measurements have hitherto revealed it to be.

If we adhere to the most common opinion, we must allow frequency selectivity to be achieved in two stages, the first by way of hydrodynamical transformations in the cochlea and the second via the mechano-neural transduction mechanism. The two stages are referred to as the *first* and the *second filter*, respectively. Although the existence of a second filter is usually assumed, there is no evidence at all as to its nature (several theoretical possibilities are currently being worked out (see, *e. g.*, STEELE, 1973; HELLE, 1974; ZWICKER, 1974; DUIFHUIS, 1975). Furthermore, the part played by outer hair cells versus inner hair cells in the cochlea, insofar as achieving the ultimate frequency selectivity is concerned, is not clear at all (*cf.* RYAN and DALLOS, 1975).

Responses of auditory-nerve fibers reveal many nonlinearities (*cf.* PFEIFFER and KIM, 1973). For the purpose of the present review, it is most important to note that the firings of an auditory-nerve fiber may be linked to the CDT. This property has been studied by GOLDSTEIN and KIANG (1968). They found that responses of a nerve fiber, when the sound consists of two components with frequencies f_1 and f_2, can show synchrony to the compound waveform but to the CDT with frequency $2f_1 - f_2$ as well. Synchrony is to be interpreted in a probabilistic sense: firings do not occur in every cycle of the waveform, but the probability of firing goes up and down just like the waveform. Hence, if the firings are observed on a time scale that is synchronous with periods of the CDT, it is noted that the firing probability may have a tendency to show the same period as the CDT. About the proper interpretation of these results, some discussion arose, but the net result was that we may safely assume that auditory-nerve fibers respond to a CDT as if this distortion product is physically present in the cochlea (GOLDSTEIN, 1970). This property tallies with the conclusion from psychophysical experiments: aural combination tones act as if they were part of the sound stimulus.

In this way we are led back to the problem, where in the chain from cochlear movement pattern to excitation of a nerve fiber are combination tones generated by a mechanism displaying all the properties of an essential nonlinearity? Rather than answering this question directly, we turn once more to psychophysical evidence and ask whether we can distinguish several stages in the production of the CDT. The answer to this question is affirmative. The subjective strength of the CDT is strongly dependent upon the frequency ratio f_2/f_1 (see Section G. 5). This suggests that the stimulus undergoes a certain amount of filtering before the site of the nonlinearity is reached. Secondly, the CDT behaves as a normal spectral component; hence, it should undergo a filtering process before the locus of its detection (nerve fibers tuned to its frequency) is reached. There appear to be two filters involved in the processing of the CDT, and the nonlinearity can be viewed as sandwiched in between. How far these filters coincide with the two filters involved in straight frequency selectivity is not known. Moreover, it may well be that the two filters involved in this concept of CDT production are partially coincident.

It may be concluded that, from psychophysical as well as neurophysiological evidence, we have obtained a fair idea about the mechanisms involved in CDT production. As to the exact site of production, we have as yet no idea (see, however, the literature cited above). And, about the nature of the nonlinearity, we know little more than that it appears to be an essential nonlinearity. Chances are likely that the same nonlinearity is responsible for other manifestations of nonlinearity in the cochlea (SMOORENBURG, 1972b; DE BOER, 1975), notably two-tone suppression (SACHS and KIANG, 1968).

H. Return to Place (?) (Increasing Importance of Spectral Concepts)

1. Doubts About Relevance of Periodicity

Are we right in concluding that "residue pitch" is growing away from "periodicity pitch"? On the basis of the foregoing, it certainly would seem so. Originally, the residue was considered as intimately connected with periodicity, but, as research progressed, more and more emphasis was put on spectral resolution rather than on interference of unresolved components. Doubts about the necessity of purely temporal processing arose from various sides. WHITFIELD (1970) reported that extensive neurophysiological research had failed to uncover a mechanism capable of measuring relatively long (several milliseconds) time intervals to a 1% accuracy. This does not prove that such a mechanism does not exist, of course, but the argument is very suggestive. In a completely different context, SIEBERT (1970) showed that the representation of nervous activity in the auditory nerve is accurate enough to account for human frequency discrimination, and that temporal information is not needed by the central nervous system. Again, not a conclusive proof, but an important piece of knowledge.

Let us recall, at this moment, which psychophysical experiment was the most important one that served to stress spectral aspects relative to temporal phenomena. That experiment was the finding of the dominant spectral region for residue pitch. In the dominant region, the components of the signal are well resolved by the auditory system, and it appears unlikely that the dominant residue is mediated by the remaining weak interactions between adjacent components. It is easier to believe that, in this case, pitch is mediated by a mechanism that tries to detect the pseudo-fundamental, a mechanism that can be described as an extension of the place theory.

Let us look now at the evidence from the other side. One of the main arguments in favour of temporal processing has been the existence of phase effects (see Section E. 8). The timbre of a sound was found to depend on the phase relation between components only when the components were close to one another in frequency. Since, in that case, adjacent components interfere because of insufficient resolution, the idea of waveform processing is immediately strengthened. For sound complexes showing clear phase effects, it would be unreasonable to assume that the residue pitch is mediated by the frequencies of the only marginally resolved components.

What part have combination tones played in the problem of theories on temporal versus spectral processing? Since combination tones are audible only when the (relative) frequency spacing is small, they seem to have strengthened the position of the temporal theory. Combination tones might even be important for phase effects (just how much, will be seen later). May we conclude that we would have been much more inclined earlier to accept an extended place theory for pitch, if there were no combination tones? The experiments to be reported in the next sections suggest an affirmative answer to this question.

2. Houtsma and Goldstein's Experiments

An essential basis for practising the art of music is the human auditory system's ability to perceive melody. On listening to a sequence of musical sounds, one can generally retrieve the series of "notes" played regardless of the spectrum generated by the particular musical instrument. Melodies can be recognized even when the sound signals are passed through a bandpass filter which only passes a small number of spectral components. Hence melody appears invariant over a large class of spectral transformations. Houtsma and Goldstein (1971, 1972) reported a series of psychophysical experiments that were designed within a musical framework to explore how the auditory system retrieves pitch from a sequence of periodic sound signals. These experiments have had a profound influence on contemporary theories about residue pitch – so much, in fact, that an almost complete reversal of opinion has taken place.

The experiments differed in several respects from all earlier ones. First, instead of single sounds, a simple musical message consisting of a series of sounds was used as the material. Second, the experimental subjects had an extensive musical

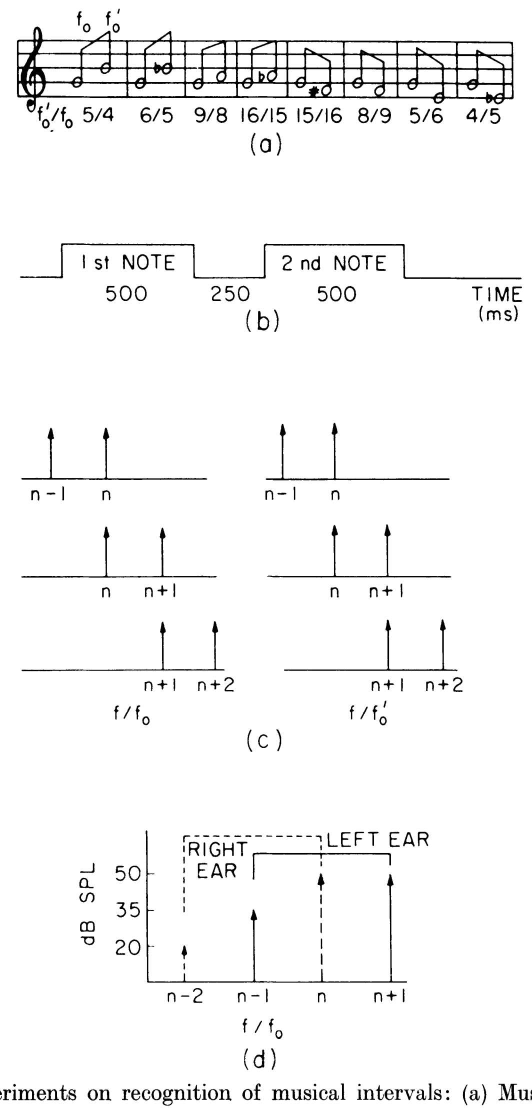

Fig. 32a—d. Experiments on recognition of musical intervals: (a) Musical intervals to be identified. (b) Timing of the stimulus tones. (c) The three possible two-component stimuli for each of the two tones; for each tone, a random choice was made among the three possible stimuli. (d) Stimulus presentation for the dichotic experiments, with the addition of simulated combination tones (see Section H. 3). After Houtsma and Goldstein (1972)

background. In order to test whether the experiments truly reflected the musical behaviour of melody perception, several control tests were carried out. In these control tests, the musical messages consisted of sequences of four "notes", each being a two-tone complex with two successive harmonics as components. The fundamental frequencies were in the range 200–400 Hz. The lower component was chosen randomly as the 3rd, 4th, or 5th harmonic. The subjects were required to identify the series of notes. Since all of the 10 subjects were familiar with musical dictation, an almost perfect pitch score could be expected and was, in fact, achieved. This shows that even for the unfamiliar two-tone stimulus, the subject's response reflects his natural musical skill and that very little or no special training was required for this type of task.

The main experiments were carried out with sequences of two signals. Each signal consisted of two components which were successive harmonics; for each signal, the number of the lower harmonic was chosen randomly from three successive integers, $n-1$, n, $n+1$. The parameter, n, was one of the independent parameters in the experiments. The fundamental frequencies f_0 and f'_0 of the two signals formed a musical interval: it was the subject's task to identify this interval.

The repertoire of intervals is depicted in Fig. 32a.

For reasons of convenience, all intervals are shown as starting with the same note. In the experiments, the fundamental frequency f_0 of the first signal was the second independent variable; in each series of experimental runs f_0 was the same, and, in successive runs, n was incremented by one. The intervals were presented in random order, of course. Nearly perfect scores of identification were usually achieved in the first run of a series. Increments were added to n until response dropped to chance level (12.5% correct, corresponding to one out of eight intervals guessed correctly). Section b of Fig. 32 gives details of the timing of the stimuli; Section c shows the three possibilities for the spectrum of each of the two signals. Section d will be referred to below.

Some representative results are shown by Fig. 33a. All these experiments were performed with monotic presentation[4]. The figure shows the contours for equal performance (P_c is the percentage of intervals correctly identified) in the $f_0 - \bar{n}$ plane. In Fig. 33a the ordinate $\bar{n}$ refers to an average of (integral) values of n. The results show clearly that the best performance is achieved with the lowest harmonic numbers. In other words, the larger the spacing between harmonics (on a logarithmic frequency scale), the better the performance. This is in general agreement with the principle of dominance, and it suggests once more that the two harmonics employed in this experiment are processed through separate channels of the auditory system to obtain successful identification.

3. Dichotic Pitch Recognition

The hypothesis about separate channels was tested in a second experiment employing probably the most extreme channel separation one can think of,

[4] Dichotic: Two-ear presentation; different signals are fed to the two ears. Monotic: One-ear presentation, or two-ear presentation with the same signal being fed to the two ears.

namely separate ears. The experimental paradigm was the same as in the monotic experiment, except that the components of each signal were presented dichotically, one to each ear. Figure 33b shows results for dichotic presentation. Comparison with Fig. 33a proves that each subject's performance is slightly below

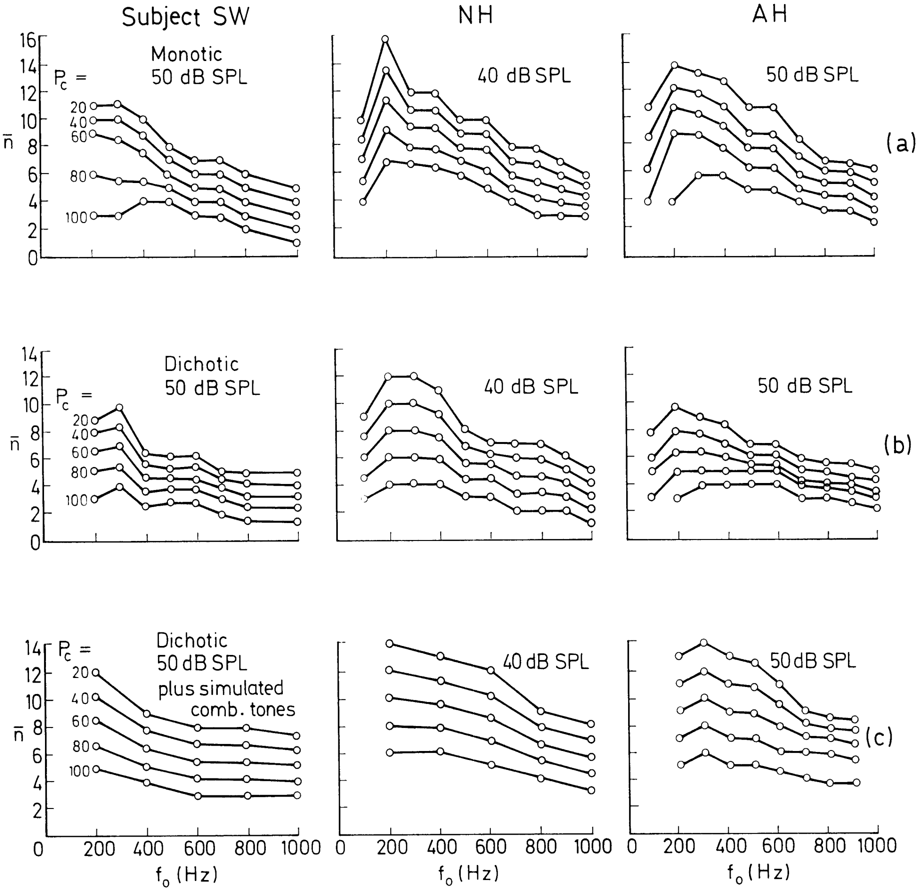

Fig. 33a—c. Equal-performance contours for identification of musical intervals. (a) Monotic presentation at 50 (40) dB SL. (b) Dichotic presentation (one component fed to each ear). (c) Dichotic presentation, with two simulated combination tones added. After Houtsma and Goldstein (1972)

that obtained under monotic presentation[4]. This finding considerably strengthens the hypothesis that a central mechanism integrates and processes information from both cochleas. Note that, in the dichotic case, the information from each cochlea involves only one component.

In an additional experiment, the influence of aural combination tones, which is undeniably present in the monotic case, was simulated for dichotic presentation. Two additional components were generated and distributed over the ears as is indicated by Fig. 32d. It may be expected that the total spectrum perceived in the dichotic case is now similar to that for the monotic case. As is shown by Fig. 33c, performance contours undergo an upward shift of n of approximately 2, compared to those of Fig. 33b, and there appears to be very little difference in performance between monotic and dichotic presentation of the signals. This is strong evidence that such differences in performance as occur between monotic and dichotic stimulus conditions can be attributed to aural combination tones.

Musical relevance of the experiments was tested once more by employing inharmonic stimuli. When the frequencies of the components were adjusted to produce the correct pitch despite inharmonicity (*cf.* Figs. 15 and 19) and care was taken to avoid pitch ambiguities, the subjects' performance was nearly perfect for $\bar{n}$ up to 6 ($\bar{n}$ is now the average of non-integral n-values). All experiments, then, indicate that what we have called residue pitch is identical with the concept of pitch that musical persons have acquired through their training. In the experiments described, the concept of "pitch" was approached from the side of musical intervals. Conversely, the concept of residue pitch may be extended safely toward musical intervals. The results of the experiments indicate that this holds true only for low values of n, at least for two-component signals. There are reasons to believe that a similar but probably higher limit is valid for signals with more than two components (*cf.* Section G. 7).

The dichotic experiments prove directly that *a fundamental period in the cochlear output is not required for the perception of residue pitch.* This finding points to a definite inadequacy of the period-detection theory. In the words of the authors (HOUTSMA and GOLDSTEIN, 1972): the findings "... suggest that fundamentals of complex-tone stimuli are retrieved by means of a central mechanism which operates on those stimulus tones or combination tones that can be resolved in the cochlea". That performance is bounded by harmonic number in the monotic case might be attributed to the limit of aural frequency resolution. That performance for dichotic stimulation is bounded in the same way cannot be accounted for by frequency resolution alone; the cause must be more central.

Little can be said about the neural mechanism that mediates retrieval of the fundamental. The central mechanism will operate on neural signals derived from components that are resolved by the peripheral organs. It is known that information about the frequency of components is preserved by the peripheral organ in two ways: "place" of activated nerve fibers and, at least for low frequencies, partial synchrony between firings of a nerve fiber and waveform of the stimulus. Either of these two cues could provide the required information for the central mechanism, and it will be quite difficult to decide between these alternatives on the basis of psychophysical experiments alone.

4. Conclusions — Wightman's Theory

The experiments by HOUTSMA and GOLDSTEIN have given considerable strength to the opinion that the final processor that determines pitch for a complex stimulus operates on the *spectrum* of the signal rather than on the *waveform.* There is only one aspect of the residue that seems not to agree with this opinion, namely, the existence of phase effects (see *e.g.* Section E. 14). However, it should be remembered that phase changes between components do indeed produce changes in the timbre of the sound but that the pitch remains unaffected. The arguments can be found in WIGHTMAN'S first 1973 paper (1973a). Hence, it is probable that the pitch extraction mechanism operates independently of the timbre mediating mechanism, and, if that is true, the pitch mechanism might well operate entirely and solely upon spectral cues. Accordingly, two theories have been formulated that describe such a mechanism in abstract terms. We shall not endeavour to describe the theories in detail. Readers in possession of the necessary mathematical background knowledge can consult the pertinent references (WIGHTMAN, 1973b; GOLDSTEIN, 1973). The following descriptions attempt to make the ideas clear to readers who are well oriented in physiology or psychology but who are not willing to follow every step in all its abstract or formalistic detail.

In this section and in the next, we shall describe the basic ideas contained in WIGHTMAN'S (1973b) pattern-transformation theory for residue pitch. We shall follow this author closely in the outline of some basic steps in a process of pattern recognition. Consider the following analogy between a visual and an auditory pattern-recognition problem. The letter "A", as it is printed here, has some particular property in common with the same letter printed elsewhere. That property allows us to identify it as an "A" despite the fact that the physical features of all A's are different. Recognition may now be described as a *sequence of transformations*; in each transformation, some of the most detailed information is lost while some specific information is preserved. After the final transformation, all information about the differences between various A's is lost and only the information relevant to the recognition of a letter remains. This is enough to enable the task of *recognition* to be performed.

In neurophysiological terms, the transformations involved are transformations of patterns of neural activity. For the pitch problem, the first transformation is carried out by the peripheral sensory system. The nervous system is assumed to carry out a transformation of the peripheral pattern of neural activity into another pattern in such a way that all stimuli with the same pitch have a similar representation. The pattern-transformation theory proposes specific forms for the transformations involved. In each of the stages, it is important to list what kind of information is preserved and what is lost. Let us take the first transformation, the transformation of the acoustical stimulus into what is termed the "Peripheral (neural) Activity Pattern" (PAP). Note that the latter pattern is not necessarily to be identified with the pattern of neural activity in the auditory nerve(s) since the experiments described in the previous sections strongly indicate that information coming from the two ears is processed jointly.

Fig. 34. The peripheral activity pattern (schematic) for a periodic pulse series. Abscissa: frequency — or location in the cochlea labelled by rank number of component. Ordinate: neural activity. Thin lines: contributions by individual components. Thick line: total activity

The first transformation (Stage I) is assumed to be a frequency analysis process with limited resolution. In many respects, it is similar to the signal transformation carried out by the cochlea, but, in the theory, only the major features are needed. The output of this stage, the PAP, is a pattern of neural activity that represents the spectral contents of the stimulus in a crude way. In factual terms, the output is a function ("amount of neural activity", for instance, the mean rate of firing) that depends on location (representing, in a way, the "place" in the cochlea). The analysis performed in this transformation is, as said, limited in resolving power just as the cochlea does not produce a single locus of neural activity for each sinusoidal component of the sound stimulus. The limit of resolution agrees in general terms with the critical band; the bandwidth is assumed to be proportional to frequency. Details about this limit are not important here, but the idea of explicitly including a limited power of resolution in this transformation is one of the main features of the theory.

As a consequence of this restriction, the representation of a multi-component stimulus is substantially smeared out; only the lower components will stand out as more or less isolated patches of neural activity (the locus of largest activity corresponding to frequency) and the higher components, *e. g.*, those with a rank number of 10 and higher, are fused. See Fig. 34 for a stylized representation of the PAP. In the transformation of Stage I, all information about phase is lost, and so is the information about the waveform formed by the fused components.

5. Wightman's Theory, Continued

The second transformation (Stage II) is intended to bring out the main features of the PAP in terms of its oscillatory properties. If the stimulus were a periodic pulse series, having a large number of equidistant components, the PAP will clearly indicate the first 10, or so, of these components. This occurs in the form of regular undulations as in Fig. 34. When the stimulus is a different signal, Stage II tries to bring out a similar feature. Hence, it is not surprising that Stage II is formally assumed to operate like a Fourier analyzer. If a more or less regular undulation is present in the PAP, the Stage II transformation brings out its "frequency". Actually, the latter parameter is a temporal parameter since a periodic repetition in the original signal is the feature that causes an oscillation in the PAP. Because of the two transformations involved, this parameter is not the same as the time t; hence, we shall designate it by a different symbol, namely τ. Summarizing, then, the output of the second stage is a function (again probably an amount of neural activity) dependent upon τ. Whenever a specific form of undulation or oscillation is present in the PAP, Stage II shows that activity is larger for a certain value of τ.

The third stage of the proposed model identifies the maximum in the transformed version of the PAP, and pitch is assumed to be directly associated with the value of τ for which the maximum occurs. As an extra feature, the magnitude of the maximum gives an indication of the subjective strength, or the "clearness", of pitch, a feature that no other theory possesses. In WIGHTMAN's paper, the predictions of the model are compared with the results from several experiments on residue pitch. In many cases, a satisfactory agreement is found, somewhat surprising in view of the extreme simplifications that have been built in to make the model manageable in computational terms.

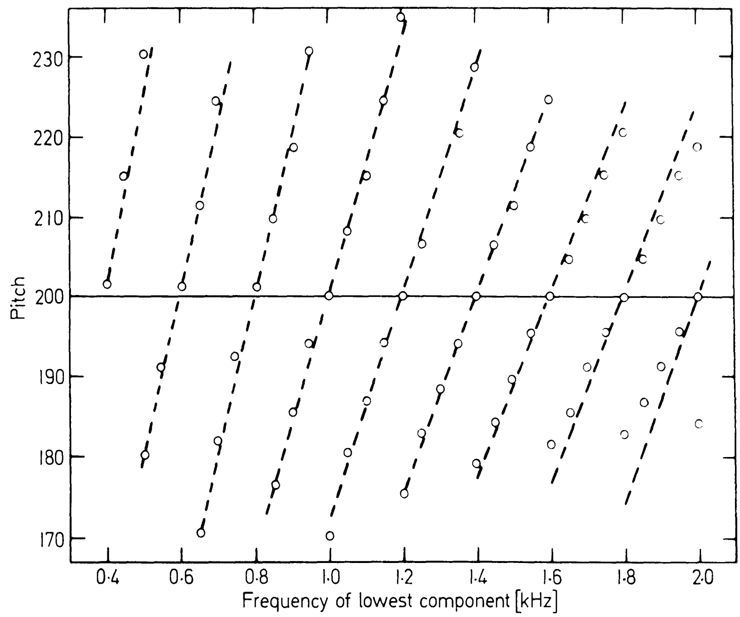

Fig. 35. Comparison between theory (circles) and experiment (dashed lines) on the pitch of inharmonic signals. After WIGHTMAN (1973b)

For example, the model rather accurately predicts the pitch shifts for inharmonic signals. Data are taken from PATTERSON's (1973) measurements, and the comparison between theory and experiment is shown by Fig. 35. Contrary to common usage, the circles in this figure are predictions from the theory and the dashed lines indicate the trend of the experimental results. Due to the fact that limited aural resolution is built in, and that, therefore, the lower components are more prominent in the PAP, the first as well as the second effect of pitch are represented accurately. An exception is formed by the region of higher frequencies where the theory predicts smaller pitch shifts than the measured ones. Two modifications appear necessary. The first is the addition of combination tones to the PAP. The second modification is that these combination tones should be represented with an intensity nearly that (within a few decibels) of the stimulus components. The latter modification implies some kind of amplitude-compression scheme. It should be stressed that these points constitute only refinements; the basic theory successfully predicts much of the experimental data.

WIGHTMAN's model predicts more than just pitch *value*; it also gives estimates of pitch *strength*. It appears that the theory is rather insensitive to changes of parameters in the Stage I transformation as regards pitch values. With the exception of the point discussed in the preceding paragraph, the parameters of the model can, thus, not easily be fixed by comparing theory and experiment for pitch values. The predictions for pitch strength, however, are strongly dependent upon the parameters inserted in the model. There are as yet too few experimental results on this issue, but the theory in its present form agrees with the tendencies reported in the literature.

One aspect of pitch strength is the principle of dominance. The theory predicts that pitch strength is a smoothly declining function of harmonic rank number. The upper limits are in good agreement with RITSMA's (1962) values for the existence region for the three-component residue signal. The model fails, however, to account for the lower limit of the existence region, and, hence, does not predict that the 3rd, 4th, and 5th harmonics are dominant with respect to pitch. It appears difficult to improve the theory on this point without making the model considerably more complicated. Apart from this deficiency, the theory ties together in a meaningful and elegant way a great number of properties of residue pitch.

6. Goldstein's Theory — Basic Constraints

A second theory on residue pitch has been published by GOLDSTEIN (1973). In one sense, it is related to WIGHTMAN's theory: it is assumed that *all* cues for pitch are obtained from the spectrum. In all other respects, it is quite dissimilar; in particular, the concepts involved are quite different. Furthermore, the concepts used in GOLDSTEIN's theory do not allow a simple interpretation in terms of neural activities. The stages involved are much more abstract; the beauty of the theory lies more in its capability of quantitatively synthesizing inferences from all available sets of relevant psychophysical data than on the possibilities for experimental verification of its essential features.

Basically, the central processor is viewed as a recognizer of spectral patterns that are supplied by the peripheral frequency analyzers. The *peripheral analyzer* extracts from a complex stimulus all the components that differ from their neighbours by more than some resolution limit. It then produces *only* information about the frequency of each component to the central processor. The following constraints are involved in the analysis stage:

a) only aurally *resolved* components contribute,

b) the *phase* relations are irrelevant,

c) only the *presence* of a component is reported – the amplitude is irrelevant (within limits),

d) the information about component frequency is basically *inaccurate*, a non-negligible variability is involved.

The *central processor* accomplishes recognition by selecting out of its repertoire of stored patterns the one that is best matched to the information received. The matching is done on the basis of maximum likelihood estimation. The stored patterns with which the matching is carried out have one constraint:

e) it is assumed that the information received corresponds to stimuli in which the components are *successive harmonics*.

The points listed under a) through e) above, are the fundamental restrictions under which the proposed mechanism is assumed to operate. GOLDSTEIN's paper describes these points in detail, providing all the relevant justification for each one of them. We suffice by mentioning a few basic relations. Assumptions a) and b) appear reasonable in view of the growing role of spectral concepts in residue theory that we have witnessed. Assumption c) was already necessary in WIGHTMAN's theory and can be traced back to SMOORENBURG's work. Assumption d) is based on the motivation that frequency information about resolved components is conveyed to the central nervous system by noisy neural channels.

The last Assumption (e) is the one involving successive harmonics. One argument is contained in DE BOER's observation that a signal consisting of odd harmonics of 100 Hz does not convey a residue pitch of 100 Hz but yields two possible pitches in the 200 Hz region (see Section E. 6 and Fig. 15). Hence, it is better to assume that the central processor operates on the principle that successive harmonics are present than that a periodic stimulus is present. It is to be noted that the central processor has no information about which rank number is to be associated with each component. The estimation of the best-fitting series of (successive) rank numbers is an essential part of the pitch recognition procedure, and it directly accounts for ambiguities in pitch as they are often observed (see Section E. 14).

The only free parameter in the theory is the one describing variability (fundamental inaccuracy) associated with frequency information. This parameter is assumed to be a function only of frequency. It proved possible to derive a functional form for this parameter (labelled σ) that is sufficient to produce good agreement between theory and experiment. The experimental data were taken from HOUTSMA and GOLDSTEIN's work on recognition of musical intervals (1971, 1972). The best estimates for the standard deviation $\sigma(f)$ that is associated with the measurement of a component frequency f are depicted in Fig. 36. With the appropriate parameter values inserted, the theory can predict performance in the musical interval experiment in great detail, as Fig. 37 shows. Experimental data are averaged over various stimulus conditions (compare Fig. 33a). Because of its strictly statistical structure, the theory is capable of predicting contours of equal performance, and the agreement with experimental results is observed to be quite good.

Unlike WIGHTMAN's theory, the present theory predicts several details of the existence region for residues as well as of the region of dominance. In this way, it accounts for RITSMA's (1967) as well as PLOMP's (1967) data, albeit that there are indications that the highest components should be given a smaller weight in some way or other. [Note: this is difficult in view of Assumption (c)]. In particular, the theory accounts for the fact that the lowest components – including the fundamental – do not belong to the dominant region. One reason for this effect is seen in Fig. 36; for low frequencies, the standard deviation σ increases when frequency decreases.

A separate check on the derived values of the measurement uncertainty σ is possible on the basis of measurements of frequency discrimination of residues. RITSMA (1963) tried to measure the fundamental accuracy of pitch matches by comparing two residues with one another. To avoid matches based on perception of part-tones, the two signals to be compared comprised disjoint frequency regions; the test stimulus consisted, *e. g.*, of harmonics 6–8, and the standard stimulus of harmonics 9–11. GOLDSTEIN shows that his theory is in excellent agreement with these data. Incidentally, the accuracy with which residue pitches are determined is decidedly inferior to that with which the pitch of pure tones can be determined. Hence, the variance $\sigma(f)$ describing the precision with which

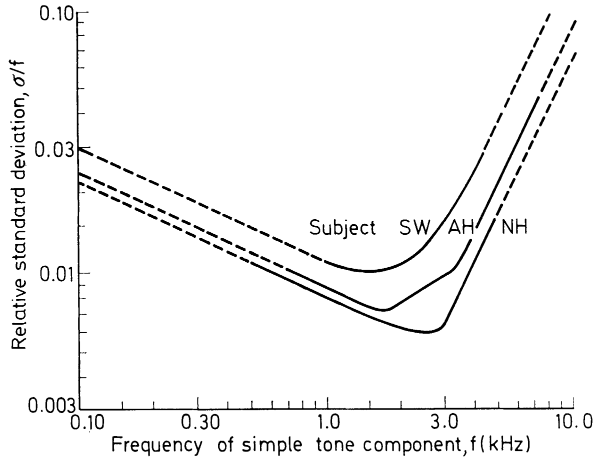

Fig. 36. Relative standard deviations characterizing the precision with which frequency in formation from aurally resolved components is conveyed to the central pitch processor. After GOLDSTEIN (1973)

frequency information is conveyed to the central processor, has a considerably larger value than the variance corresponding to pure-tone frequency discrimination. The conclusion that the accuracy of representation of components is inferior to that of isolated pure tones is also possible outside the framework of GOLDSTEIN's theory; as yet there seems to be no reasonable explanation for this property.

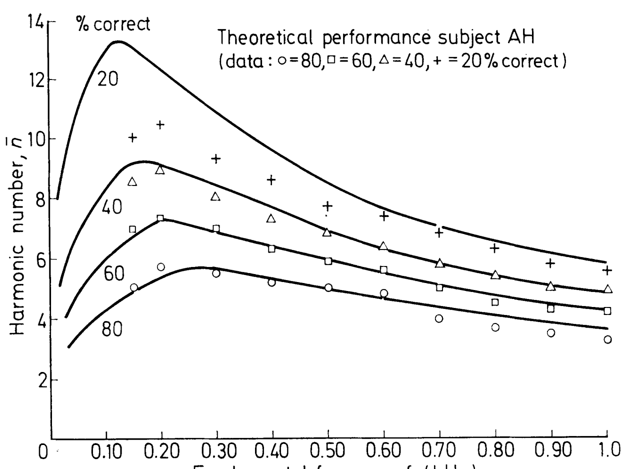

Fig. 37. Theoretical and measured performance in experiments on musical intelligibility. Data are averaged from data published by HOUTSMA and GOLDSTEIN (1972). After GOLDSTEIN (1973)

7. Goldstein's Theory — Verification

A major point concerns, of course, the pitch shift associated with inharmonicity of the sound stimulus. GOLDSTEIN'S theory predicts, in an approximate form, that the effective rank number η describing the pitch shifts is the average of the rank numbers of the lowest and the highest component:

$$\eta = \frac{n_{\text{low}} + n_{\text{high}}}{2}. \tag{22}$$

For the meaning of η, we refer to Eq. (21) and the discussion in Section G. 7. If we include combination tones, the predicted values of η in Fig. 31 would lie midway the curve for λ (the lower limit for combination tones) and the line representing f_2, the highest component (higher combination tones being ignored). Up to $n = 10$, the agreement is quite remarkable; it is recalled that n is the rank number of the lower one of the primary components.

Above $n = 10$ there are deviations; to account for these, it is necessary to add the rule that only components that are more than 10% remote from one another contribute information to the central processor. It is exactly at this point that the limit of peripheral aural resolution creeps in. In this connection, it is interesting to note that the frequency inaccuracy implied in σ reflects a different dependence on f than the limit of aural resolution as represented by, *e. g.*, the critical band. In particular, σ is observed to increase sharply above 3000 Hz, a fact which is not paralleled by critical-band behaviour.

The role of combination tones is unseparable from the concept of inharmonicity pitch shifts. A rather lengthy part of GOLDSTEIN's paper is devoted to the confounding role of the combination tones. During the history of the residue theory, there were several instances in which experimental evidence indicated primary involvement of the spectrum rather than the waveform. The first of these findings was DE BOER's description of the "wide residue" (Sections E. 9 and E. 10). The second was the discovery of the principle of dominance (see Sections F. 2 and F. 3). The third point to be mentioned in this connection is the frequent underestimation of pitch shifts due to inharmonicity. But it appears that the contributions of combination tones, associated with narrow spacing of components as they are, have over-emphasized the possible role of interaction between components. Once the confounding role of the combination tones was recognized, many properties of the residue appeared as less puzzling. One point stressed in GOLDSTEIN'S paper is that combination tones must be resolved by the auditory system in order to be effective in pitch determination. It is suggested that combination tones would also be responsible for at least some of the effects associated with narrow-band sound complexes. As stated earlier, we shall come back to this subject later (Section H. 11).

8. General Conclusions

GOLDSTEIN's theory describes a unifying logic that underlies the residue phenomenon and, thereby, defines a problem for physiology to solve. The frequency of the constituent harmonics is communicated to the central processor in a noisy way, and the question is left open whether the signals are coded by their place or their temporal course. Present physiological knowledge is insufficient to answer the question.

The theory starts from the assumption that the properties of residue pitch cannot be accounted for by peripheral events localized within a small region of frequencies; the pitch processor synthesizes information from wide frequency regions as well as from both ears. In this respect, the theory is much more than a consistent framework describing the formation of a pseudo-fundamental frequency. When aural frequency analysis fails because of insufficient resolution, residue pitch ceases to exist. However, more often residue pitch ceases to exist because of too much ambiguity. Hence, the peripheral analysis of the stimulus into its constituent components is a necessary but not a sufficient step in the creation of residue pitch. The emphasis of the pitch-determining mechanism is displaced from a peripheral to a central site. It is then logical to see the mechanism of pitch extraction equipped with properties typical for a neural network; despite the noisy way in which information is received, a remarkable consistency of the result of the processing is evident. Yet the pitch extraction mechanism has an innate tendency to arrive at ambiguous pitch values; a fact that is corroborated by many experimental findings. In GOLDSTEIN's theory this tendency can be traced to the problem of determining the proper value of n. WIGHTMAN's theory has the same tendency, but here it is more difficult to pinpoint the underlying cause. With these somewhat more general remarks, we close the discussion on unifying theories of residue pitch.

9. Dichotic Pitch Phenomena — Dichotic Repetition Pitch

Long before the period we are describing here, several dichotic pitch phenomena were known. To elicit these effects, two signals are presented to the two ears that are the same in many respects but differ in one. For the HUGGINS pitch, for example, two white-noise signals are used that differ only in the interaural phase relation over a narrow range of frequencies (CRAMER and HUGGINS, 1958). A faint pitch is heard corresponding to the frequency region for which the signal is truly dichotic, *i. e.*, for which the two signals at the two ears are different. Another manifestation of dichotic pitch perception is described by FOURCIN (1970). Repeated attempts were made to explain these phenomena but they remained puzzling. A relation with lateralization seemed evident and has even led to definite studies of this relation even for monaural residue phenomena (NORDMARK, 1963). More recently, this type of study has been taken up again (possibly prompted by the success of HOUTSMA and GOLDSTEIN's work on dichotic interval recognition) and it appears that the results now allow for a more consistent interpretation. We shall describe some experiments on "Dichotic Repetition Pitch" as reported by BILSEN and GOLDSTEIN (1974) and point out in due course the relation with, *e. g.*, the FOURCIN pitch.

We recall that "repetition pitch" (RP) is produced when to a noise signal $n(t)$ the same signal, but delayed in time, is added. Monaural repetition pitch (MRP) is described in Section F. 8, and Eq. (11) describes the stimulus used. Monaural repetition pitch can be heard when the time delay τ is between 1 and

10 msec. A much fainter but distinct pitch can be heard when the original noise signal $n(t)$ is led to one ear and the delayed version $n(t - \tau)$ of the same noise signal to the other. The pitch heard is called "dichotic repetition pitch" (DRP). Unlike MRP, DRP exists only when the time delay exceeds 3 msec. (Note that the time delays associated with lateralization are smaller than 1 msec.)

The pitch of DRP was measured by matching it with an MRP stimulus; this procedure proved possible over the range of τ indicated above. Furthermore, it proved possible to determine DRP with tests involving musical intervals. It is not surprising to find that DRP corresponds to the frequency $1/\tau$.

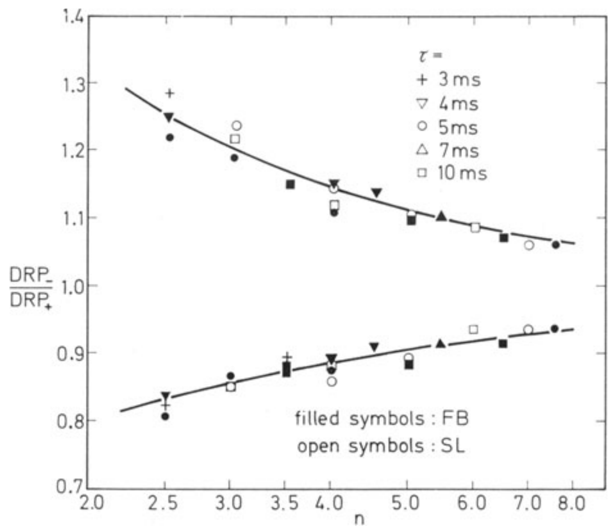

Fig. 38. Relation between DRP_- and DRP_+ for band-pass filtered signals. The thick lines are computed from $DRP_-/DRP_+ = \tau/(\tau \pm 1/2f_0)$, where f_0 is the centre frequency of the filter. Note that the abscissa is labelled as $n = f_0 \cdot \tau$. After BILSEN and GOLDSTEIN (1974)

Just as for MRP, further experiments were carried out on the effect produced by inverting the delayed noise signal $n(t - \tau)$ before it is presented to the ear. Let us designate dichotic repetition pitch by DRP_+ when the two signals are presented with the same polarity, and by DRP_- when the delayed version is inverted. It was found that DRP_- deviates significantly from DRP_+. In general, two DRP_- pitches can be perceived, one a little higher, the other a little lower than $1/\tau$. For wide-band signals, the best agreement was obtained when, for DRP_-, the actual time delay τ (in msec) was increased or decreased by 0.8 msec; the measured pitches corresponded well with $1/(\tau + 0.8)$ and $1/(\tau - 0.8)$. Note that this result holds for DRP only; for MRP it is different (see Section F. 8).

The wide-band signals employed covered the frequency region up to 2 kHz. Experiments with the same paradigm but executed with narrow-band signals (bandwidth: one third of an octave) yielded pitch deviations that were a distinct function of the central frequency f_0. In Fig. 38, the results are plotted. The ordinate shows the relation between DRP_- and DRP_+ values; the abscissa is labelled as n, the equivalent rank number associated with f_0 (actually, n is calculated from $n = f_0 \cdot \tau$). The two sets of data correspond to the two possible pitches. If we compare this figure with Fig. 26, we see a very close correspondence. For the narrow-band DRP_- pitches, the same formula is proposed as for MRP with negative p; the pitch corresponds to $1/(\tau \pm 1/2f_0)$. This relation is presented in the figure in the form of two solid lines, the experimental data points are observed to be described rather accurately by this formula.

Returning once more to wide-band signals, the results of the experiments indicate that there is a single frequency region of dominance; the value of 0.8 msec is equivalent to half a period of a centre frequency of 625 Hz. This dominant region appears to be effective for all values of τ employed. It is recalled that in the monaural case the dominant region appears to adjust itself to the value of τ used in the experiment (Section F. 8).

10. Relation with "BMLD" and the "EC" Theory

It is clear that the DRP phenomenon cannot be explained by simply assuming that the two spectral representations produced by the two peripheral hearing organs are added. There must be involved some subtle form of interaction. Binaural interactions are well-known phenomena in view of directional hearing, lateralization, etc. Regrettably, there is not one type of binaural interaction that is capable of tying together all experimental evidence of this nature. Perhaps the auditory phenomenon that is closest to DRP is the so-called Binaural Masking-Level Difference (BMLD). Space does not permit a full description, so only a few basic remarks will be presented, sufficient to convey the gist of the ideas involved. Suppose a tone is presented at such a level that it is just masked by a (wide-band) noise signal. If tone and masking signals are presented identically at the two ears, there will be no advantage over the monaural situation; in this case, the binaural masked threshold is the same as in the monaural condition. Now, let the polarity of the tone signal as it is presented to one of the ears be inverted. The tone at the other ear remains the same, and the noise is presented to the two ears with identical waveforms, as before. The surprising fact is that the masked threshold can be considerably lower in this situation.

This "release from masking" (as it is sometimes called) can be explained if we assume that somewhere in the nervous system an operation is performed which is equivalent to a subtraction of the two waveforms of the two stimuli. Even when such a subtraction would be imperfect, contaminated by neural noise, for instance, the tone would stand out in the resulting signal with respect to the noise. This idea has been worked out in great detail by DURLACH (1972). The subtraction stage was characterized as the process of "cancellation". To account for all experimental results on BMLD's, it was necessary to assume that the subtraction process did not involve the signals as they are presented at the ears, nor the signals as they might be processed by the peripheral organs. In DURLACH's theory, another idea is incorporated; before being subject to the subtraction process, the signals originating from the two ears are made alike as much as possible. That is, they are made to have the same average amplitude, and, furthermore, they are assumed to be shifted in time so as to make them as similar as possible. This part of the mechanism is referred to as the "equalization process".

The basic ideas involved in this EC-theory (Equalization and Cancellation) are taken over by BILSEN and GOLDSTEIN to the domain of dichotic repetition pitch. In order to achieve a resulting signal that displays the characteristics of an MRP signal, addition is postulated instead of subtraction. In this way, the properties of DRP are brought back to those of MRP. In line with recent developments, the explanation of the properties of pitch is sought entirely in the spectral domain. In other words, the proposed addition process is invoked only to obtain a resulting signal with the spectral properties required for an MRP.

The theory also offers a qualitative description for the FOURCIN pitch. In FOURCIN's experiments, two independent signal pairs, each of the type described above, were used. The first pair consists of the noise signal $n_1(t)$ and its delayed version $n_1(t - \tau_1)$. The second pair consists of another noise signal $n_2(t)$ and the same noise, delayed over a different time interval τ_2, to be written $n_2(t - \tau_2)$. The signals presented at the ears consist of various combinations of these four signals. It appears reasonable to assume that the "equalization stage" is presented with partly conflicting cues, and it thus sets a different standard for each condition. By the use of this expedient, many aspects of FOURCIN's work receive an explanation. The FOURCIN pitch then appears as an extension of the dichotic repetition pitch, and the latter appears as the more fundamental phenomenon.

A monaural version of the FOURCIN pitch experiments has been studied by ROSENBERG (1975). The sum of three signals was presented to the ear: a noise signal $n(t)$, a delayed version of it $n(t - \tau_1)$ and a second delayed version $n(t - \tau_2)$. Among the several pitches audible there was one corresponding to $\tau_1 - \tau_2$. If the resulting signal is inspected on its temporal properties, *e. g.*, in terms of its autocorrelation function (*cf.* Section E. 4 where an operational neural model is described capable of carrying out the analysis), this result is easy to understand. In terms of the recognition of the shape of the spectrum, the location of peaks and valleys, etc., the existence of this pitch is more difficult to visualize. We mention this experiment mainly because of its connection to the FOURCIN pitch phenomenon; the consequences of ROSENBERG's experiment have not yet been worked out in detail.

11. The Internal Spectrum

We now come to the latest types of evidence that are relevant to the problem of spectral versus temporal processing. One of the major arguments in favour of a theory of temporal processing of residue signals has always been based on the existence of *phase effects* (see, *e. g.*, Sections E. 8 and E. 14). Figure 16 shows several typical waveforms of three-component residue signals. The waveforms are characterized by prominent envelopes (indicated by the dotted lines). Any change of the phase relation of the components that leaves the waveshape of the envelope the same does not lead to an audible change of timbre. All phase changes obeying DE BOER's phase rule (1956a, 1961) belong to this class [see the description in Section E. 8 and Relation (10)]. Any *other* type of phase change may be perceptible, especially when the spacing of components is relatively narrow. Such phase changes affect the prominence of peaks in the waveform of the envelope; in the framework of a temporal residue theory, this would lead to changes in the synchronization of nerve firings and, ultimately, to changes in prominence of the residue. If, in line with the recent trends as described in the preceding sections, we are led to believe that all residue processing is based upon spectral amplitudes, the phase effects remain a puzzle. There are strong indications that phase effects are intimately connected with combination tones. Evidence for this notion is presented in the following paragraphs.

First, it should be made clear that the "spectrum", upon which the central residue processor is assumed to operate, contains the combination tones in exactly the same form as the acoustic components. In a series of tests, it was shown by BUUNEN *et al.* (1974) that a combination tone and an objective acoustic component of exactly the same frequency follow the normal rules of addition when they are combined. The second point demonstrated in BUUNEN'S paper is that the residue is the most prominent for that phase relation of the components for which the strength of the (third-order) combination tone is highest.

A third piece of what may be termed circumstantial evidence is the proof that a combination tone that is coincident in frequency with one acoustic component shows exactly the same phase change as that component for all phase changes obeying DE BOER'S phase rule. This proof is described in the same paper. We present the proof here in a simplified form. Consider a three-component signal, the upper two component frequencies being called f_1 and f_2. The third-order combination tone with frequency $2f_1 - f_2$ is coincident with the lowest acoustical component. If all acoustical components are given the same phase change φ_0 [*cf.* Relation (10) in Section E. 8], the combination tone acquires a phase change of $2\varphi_0 - \varphi_0$ and this is equal to φ_0, the phase change of the lowest acoustic component. A phase change proportional to frequency is equivalent to a shift in the origin of the time axis, and, therefore, cannot affect any combination tone in another way than an acoustic component. This completes the proof that the combination tone with frequency $2f_1 - f_2$ acquires exactly the same phase change as the objective component of the same frequency. It is thus seen that the phase rule not only describes situations leaving the envelope of the waveform invariant but also the situations leaving the relation between combination tones and objective components invariant.

From these studies, it can be concluded that the "internal spectrum" is composed of both objective components and combination tones without distinction between these two. For phase changes of the objective signal that conform to the linear phase rule (10), the internal spectrum remains the same no matter whether combination tones contribute or not. The residue is the most prominent when the internal spectrum shows the largest component amplitudes.

12. Phase Effects — Buunen's Work

We now turn to the phase changes that do not obey the linear-phase rule (10). The associated changes in timbre are most easily noticed when the components of a three-component signal are not locked to one another but are produced by free-running oscillators. We can represent that signal in the following way: assume first that the frequencies are exactly equidistant and replace one of the components by one with a slightly different frequency. BUUNEN and BILSEN (1974) describe several experiments executed with this type of signal of which we will present the one that is the most relevant at this stage. The original frequencies are labelled $f_c - \Delta f$, f_c, and $f_c + \Delta f$. (The symbol f_c is equivalent to the carrier frequency f of Section E. 6ff., F. 2, etc. and Δf is equivalent to g.) The components with frequencies $f_c - \Delta f$ and $f_c + \Delta f$ are presented directly to the ear, but f_c is replaced by f'_c deviating a few Hz from f_c. If $\Delta f/f_c$ is sufficiently small (*e. g.* < 0.15), beats are heard that are easily associated with the varying phase relation between the f'_c-component and the other components. Note that the phase of the central component differs from the average phase of the two outer components and that, consequently, the phase variations do not conform to (10).

A component is added to the signal with a frequency of $2f'_c - (f_c + \Delta f)$, exactly the same frequency as the aural cubic difference tone (CDT). Amplitude and phase of this component can be adjusted by the listener just as in a cancellation experiment (see Section G. 4). If the phase effects would result from variations of the signal's envelope, no setting of amplitude and phase of the added component would result in their disappearance. BUUNEN and BILSEN report that their listeners were able to adjust the cancellation tone so that the beats disappeared. Unfortunately, they give no details as to whether the setting corresponds to that in which the combination tone is fully cancelled. In an informal experiment, the present author was able to corroborate their finding. At a particular setting of amplitude and phase of the cancellation tone, the timbre beats of the residue disappear, only a very faint beat associated with the lowest parttone remains. The setting necessary to achieve this is very close to that for cancellation of the CDT.

The implication of these findings is clear: phase effects are strongly related to combination tones. We may extrapolate from these findings the following. Changes in the timbre of the residue are related to changes in the internal spectrum composed as it is of acoustical components and combination tones. Any phase change that leaves the internal spectrum the same does not affect the residue timbre. Any phase change that causes one or more of the internal components to change its amplitude may produce an audible change of the residue timbre. The prominence of residue pitch is a monotonic function of the strength of those components of the internal spectrum that are dominant with respect to pitch. It should be noted that these remarks are extrapolations which may not be justified in the most general situation, *e. g.*, the effect produced by changes in the phase of only one component amidst a great many components. It can be expected that such a situation is a most difficult and complicated one from the experimental point of view since several combination tones contribute to the internal spectrum.

13. Binaural Diplacusis and the Residue (van den Brink)

The importance of aural isolation of partials is confirmed by a series of experiments by VAN DEN BRINK (1970, 1975a, b). Binaural diplacusis is used as a tool to measure details about the way the pitch of the residue is derived from the pitches of the (resolved) partials. When a pure tone is presented alternatingly to the left and to the right ear, the pitch may not seem the same. The pitch difference "between the ears" can be measured by presenting tones with different frequencies alternatingly to the two ears and having the observer adjust the frequency of the tone at one ear so as to achieve equality of pitch. The pitch differences measured in this way show an irregular pattern when viewed as a function of frequency; this pattern may show day-to-day variations but over a period of several hours it is fairly stable. *Binaural diplacusis* is the term denoting the phenomenon that the same tone may have a different pitch at the two ears; the magnitude of this effect can be expressed as the (relative) shift of frequency that is necessary to compensate for the difference in pitch.

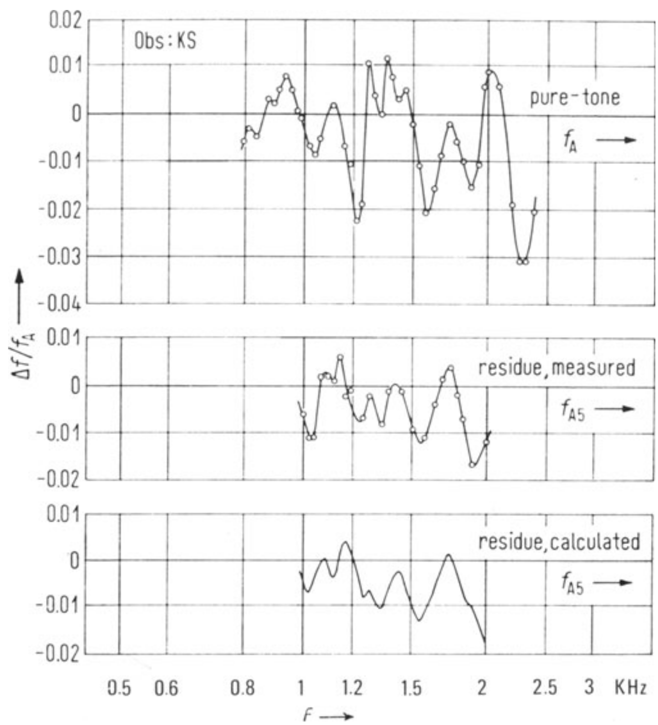

Fig. 39. Measured diplacusis for pure tones and residue signals (upper and centre panels). The lower panel shows the theoretical prediction (see text). The abscissa is the frequency f of the pure tone or the 5th harmonic. After VAN DEN BRINK (1975a)

What is described here for pure tones holds true also for residue signals; binaural diplacusis for residue pitch can be measured by essentially the same method. In all cases, a great measurement accuracy is needed, of the order of 0.1%. In VAN DEN BRINK's experiments, the residue signals consisted of the 4th, 5th, and 6th harmonics. Signal A (fundamental frequency f_A) presented to the left ear was alternated with Signal B (frequency f_B) presented to the right ear. The subjects were instructed to vary f_B until the pitches of the sounds at the two ears were the same. Care was exercised to ensure that the pitch matches were indeed based upon judgements of residue pitches and not on the comparison of aurally isolated partials (part-tones). The magnitude of diplacusis can be plotted as the ratio $(f_B - f_A)/f_A$ necessary to achieve equality of pitch; the central panel of Fig. 39 shows some representative results. The abscissa (f) is the frequency of the central component, in this case five times the fundamental f_A. The upper panel of the figure shows diplacusis values for pure tones; the abscissa is now the frequency of these tones. Evidently, there is some similarity between these two sets of data for one observer.

From the pure-tone diplacusis values, a prediction can be made about the diplacusis of the residue. For any of the residue signals employed, the diplacusis values can be determined for each of the components; averaging of these values yields the data plotted in the lower panel of the figure. It is surprising how well this prediction agrees with experimental results for residues (central panel). The pitch of the residue apparently correlates well with the (average of the) pitches that each of the components would have if it were presented in isolation. *It is the "pitch information" of the part-tones rather than the "frequency information" of the components that determines the pitch of a residue.* If we translate the pitch information contributed by each component into the "place" in the cochlea corresponding to maximal excitation, we arrive at an extended place theory, just as DE BOER implied in his pseudo-fundamental theory (Section E. 9) or as it is crystallized in WIGHTMAN'S and GOLDSTEIN'S theories. Compare also WALLISER'S results (1968). This conclusion holds true only under the conditions described: signals containing components near the region of spectral dominance. With higher rank numbers of the components, there will be a larger overlap of the excitation patterns in the cochlea and a non-negligible contribution of combination tones; indeed it was observed that the agreement between theory and experiment deteriorates drastically for rank numbers beyond 8. This point will not be pursued in the present review. The main results of the study were corroborated by several experiments in which the pitch-frequency relations of one of the ears were manipulated. The application of low-pass noise or the execution of the test where one ear is under the condition of auditory fatigue are examples of the methods used in additional experiments. In all cases, the basic findings could be confirmed.

One of VAN DEN BRINK'S experiments (1975b) is interesting also because of the pitch percepts involved. Signal A consisted of the 4th, 5th, and 6th harmonics (of frequency f_A), all presented to the left ear. Signal B consisted of the same harmonics (of a different frequency, f_B) but these were divided between the ears; the 4th and 6th harmonics were presented to the left ear while the 5th harmonic was presented to the right ear. Here, we have the peculiar situation that the components at the left ear have a fundamental one octave above f_B. Yet the residue pitch corresponds to f_B provided the sound is not too loud. Another surprising fact is that this residue is localized near the left ear. The pitch perceived on the left jumps by one octave when the right-ear signal is switched off. A pure-tone sensation corresponding to the single component at the right ear is perceived at that side. During the pitch matches it proved quite easy to concentrate upon the residues perceived on the left side and to disregard the pure-tone percept on the other side.

In the light of HOUTSMA and GOLDSTEIN'S work, these observations are not too surprising. They suggest that the information coming from the two ears interacts to determine residue pitch, and the same pitch is then associated with the side receiving the most complex information. The experiments described here were intended to serve quite another purpose, namely, to determine accurately how much the pitch information of one single component contributes to the pitch of the residue. It could be concluded that the weights of the three components are about equal in this process. This holds for signals consisting of the 4th, 5th, and 6th components. For other rank numbers, the results are not conclusive.

14. Spectral Aspects of Time Separation Pitch and Monaural Repetition Pitch

The main tendency of the studies reported in the preceding sections is that information about pitch is mainly conveyed by spectral information and not by temporal information. So we must have another look at those experiments that once seemed to confirm that pitch is strongly related to periodicity. Let us reconsider, therefore, the interpretations of time-separation pitch. (Section E. 12) and repetition pitch (Section F. 8). There is no doubt that the simplest explanation of the presence of pitch can be based on temporal aspects of the signals. The signals involved have the peculiarity that the waveforms are almost identical to time-shifted versions of themselves, and it is not too far-fetched to conclude that a neuron arrangement like the one of Fig. 12 would bring out this property easily. An interpretation in spectral terms is slightly more difficult. Since the spectral description of a noise signal is more involved than that of a deterministic signal, we shall consider the case of a double-pulse series like that depicted in Fig. 18. This signal leads to perception of time-separation pitch (TSP).

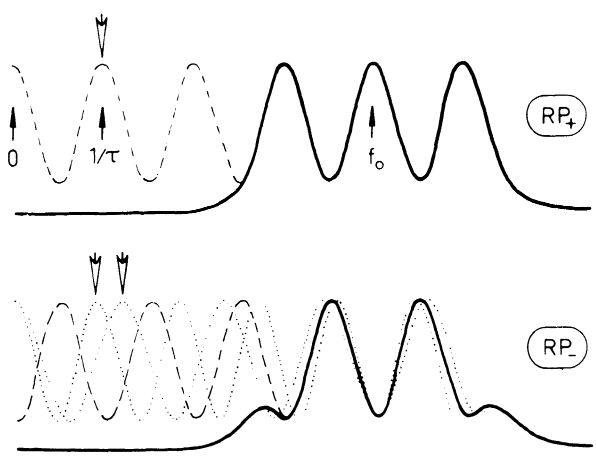

Fig. 40. Illustrating the matching procedure carried out by the central pitch processor. The thick lines indicate the spectrum of the (filtered) stimulus. The dashed line is the spectrum of the same stimulus but unfiltered. The dotted lines show the spectrum that the central processor tries to match to the actual spectrum. The two panels pertain to two possible repetition pitch (RP) conditions, RP_- and RP_+. In the latter case two different matches are possible arrow. After BILSEN and GOLDSTEIN (1974)

One periodic series of pulses has a spectrum consisting of many components (see, *e. g.* Fig. 4a). The presence of the "echo" causes undulations in the spectrum. In particular, the components with frequencies at or near odd multiples of $1/2\tau$ (τ is the delay time) will be reduced drastically. The spectrum will show peaks between these points, at integral multiples of $1/\tau$; *i. e.*, at even multiples of $1/2\tau$. When the polarity of the "echo" is inverted, the spectral magnitude will be reduced at even multiples of $1/2\tau$, including the frequency zero. (The same type of spectral modification occurs when an echo is added to a noise signal; note that in this case the spectrum must be interpreted in the statistical sense.) In terms of the newer residue theories, we must assume that the general shape of the spectrum is recognized by the central processor and that the successive peaks are interpreted as successive "harmonics" of a common "fundamental frequency". The latter frequency is the pitch of the corresponding residue.

Now, if the signal presented is band-limited, *i. e.*, restricted in its frequency range, the task for the central pitch processor is much more difficult. An idea of how the interpretation in terms of "successive harmonics" is achieved is presented in Fig. 40, taken from Bilsen and Goldstein's (1974) paper. The two cases illustrated pertain to "echo's" of positive and negative polarities, respectively. For reasons of clarity, the spectra are not shown as groups of lines (each line corresponding to an integral multiple of the fundamental frequency) but as continuous functions. In this respect, the same figure describes also the average spectrum of the signals used for repetition pitch (This explains why the two panels are labelled by RP_+ and RP_-.)

For the upper panel (positive polarity of the echo), a satisfactory match is achieved between the actual spectrum and a hypothetical one (dashed line) containing lobes at integral multiples of $1/\tau$. This makes it plausible that the pitch will correspond to $1/\tau$ despite the fact that the spectrum is incomplete. For the lower panel, the actual spectrum is a part of a spectrum that has zero amplitude at a frequency of zero and that has maxima at odd multiples of $1/2\tau$ (see the dashed line). The spectrum is interpreted by the central processor as a "regular" spectrum (one that has a maximum also at the frequency zero); this is indicated by the figure. Actually, there are two possible matches – in terms of other theories: there are two possible pseudo-fundamentals. Both matching spectra are shown as thin dotted lines.

The principal point illustrated by this figure is that the determination of a pitch value is achieved by a process of pattern recognition. Compare Wightman's theory in this respect; in the present case, the central processor tries to find the best index (pseudo-fundamental) to describe the observed undulations in the spectrum. If we assume that the accuracy of the matching procedure is determined by the critical band on the one hand, and the irrelevance of the lower spectral lobes on the other hand, we can easily understand why for a wide-band signal the region of spectral dominance seems to adjust itself in accordance with the value of τ (*cf.* Section F. 8). In all conditions, the third to fifth peaks in the spectrum will be the most important ones for the determination of pitch, and, as a result, the frequency region of these peaks acts as the dominant region.

I. "Time-Out" (General Reflections)

1. The Importance of "Set"

Must we say goodbye to the "residue"? We might be inclined to say: yes. In the course of this treatise, we witnessed the gradual change of meaning of the term "residue", a change that finally led to a new and less restrictive definition (see *e. g.* Section E. 10). At the same time the musical aspects of "residue pitch" grew more and more important. Was the residue originally thought to be one bearer of pitch, later it appeared that the residue formed by a few of the lower components is dominant even with respect to the fundamental. What has remained is the notion that the percept of the residue can only be brought about by a particular form of collaboration of components, and, in this form, the residue is dominant even with respect to the fundamental. Allusions to phenomena like the "pitchiness" of chopped noise belong to the past; a much more refined correlation between spectral components is necessary to invoke a pitch percept in the musical sense. Instead of chopped noise, we nowadays discuss phenomena like repetition pitch and combination tones.

What is also new is the ambiguity involved in residue pitch. This was recognized rather hesitantly at first (Section E. 6); later, it was considered an essential feature of the mechanism (Section E. 14). How essential it is did not become completely clear before the work on two-tone complexes and musical intervals (Section H. 2). In last instance, even the harmonic number n of the lower partial appeared to be essentially ambiguous (Section H. 6). The ambiguity involves one other concept that is, in the opinion of the present reviewer, of the greatest importance. Every psychophysical result is the product of the specific methodology involved and the "set" or "state of conditioning" of the subject. When we listen to a sound, we do not perceive this or that aspect of the sound all the time; what we perceive depends completely on our training. Let us recall, for instance, the introspection involved in the recognition of the residue in SCHOUTEN's terms (Section D. 9), the "classical" counterpart of the "instruction phase" involved in present-day psychophysics. In this connection, we recall also that there are two groups of listeners as regards their primary reaction to two-tone signals (Section G. 6). But we must appreciate as well that a particular listener involved in a particular task produces a result that is highly dependent upon his training, motivation, ability, etc. Hence, the fact that the result of a particular experiment turns out to be extremely clear-cut does not necessarily imply that every naive listener is able to perceive the signals in the same way. It is not too difficult to associate a (residue) pitch with a two-tone complex presented monaurally. But to do the same with a two-tone complex presented dichotically is quite another matter. A similar difference obtains for repetition pitch: every naive listener can immediately perceive monaural repetition pitch but the dichotic phenomenon (Section H. 9) is so much weaker that many listeners experience extreme difficulties with it. Yet, in both cases, we see that psychophysical tests (admittedly executed under ideal circumstances) indicate a performance that is nearly the same for dichotic and monotic conditions.

We may conclude that we must be extremely cautious in our conclusions. The "normal" mechanism for extracting pitch is not "connected" to the two ears as it appears to be functioning in HOUTSMA and GOLDSTEIN's dichotic experiments (Section H. 3). But by proper training and instruction, the mechanism can switch from the monaural to the binaural mode and it seems to operate with the same perfection and with the same limitations. It is then logical to endow the underlying mechanism with a set of constraints that operate in the same way under both types of conditions. However, there is no direct evidence that this is fully justified – it remains a theory.

We may carry this scrutiny over to the domain of dichotic repetition pitch, and, in that case, the conclusion may be slightly different. In order to explain dichotic repetition pitch (Section H. 4), it is necessary to postulate some interaural interaction that is partly based on waveform aspects. The idea is borrowed from work on binaural masking level differences (BMLD's), and it serves its purpose well. But we do not necessarily have to conclude that a similar mechanism is operative under "daily listening conditions". Moreover, there is no indication as to whether or not the subjective "strength" of dichotic repetition pitch is comparable to that of BMLD phenomena, and we have no inkling as to whether or not the same amount of additional training is needed to perceive it. The negative results of HOUTSMA and GOLDSTEIN's experiments (1971) on the perceptibility of dichotic phase effects only augment our doubts. This is one of the fundamental problems in the interpretation of advanced psychophysical experiments; comparison of different experiments must proceed with the greatest possible caution. The present reviewer hopes that the reader will interpret with the same caution all that has been written in this chapter on the residue phenomenon.

2. Guide to Related Literature

In collecting and selecting the material for this chapter, many subjects were left out of the discussion entirely. As is pointed out in the introduction, the selection process was a careful and deliberate one. Apologies are due to numerous authors that have contributed to residue research but do not find their work cited appropriately. No attempt will be made to fill that gap, or to be complete in that respect, but some indications are in order as to where to find other pertinent material in the literature. This may guide the cautious and interested reader to domains of pitch phenomena that have not been treated in the present text.

Residue pitch is an important aspect of speech perception. Yet it has proved extremely difficult to incorporate it in the analysis and synthesis of speech. Many different types of "pitch extractors" have been proposed but none of these was entirely satisfactory. We mention only two possibilities for a pitch extractor. The "cepstrum" technique (NOLL, 1964) is, in a way, analogous to WIGHTMAN's pattern-transformation theory (Sections H. 4 and H. 5) but without the limited frequency resolution of this theory. SCHROEDER (1968) proposed two pitch extractors that symbolize the two fundamental aspects upon which the present chapter is concentrated; one extractor operates on regularity in the temporal, the other on regularity in the spectral domain. The latter type of pitch extractor can be connected with GOLDSTEIN's theory (Section H. 6) and could be modified easily to mimic almost all aspects of our perception of intonation in speech. Many other pitch extractors work on the principle of reconstituting the fundamental by nonlinear distortion; as we have seen, these procedures have no connection with pitch perception.

Pitch in music has important uses in the formation of melodies and of harmonies (chords). Studies on the relation between musical scales and fundamental properties of the auditory system can be found in TERHARDT's work (1972a, b, 1974a). Musical intervals exist in two forms, successive and simultaneous. The concepts of consonance and dissonance are crucial for simultaneously presented tones. The relation between consonance and frequency analysis in hearing was studied by PLOMP and LEVELT (1965). Unfortunately, this work was not continued with residue signals; hence, the conclusions are partly the result of extrapolation and relate more to the physics of signals than to the logic of perception. More modern views on, *e. g.*, beats of mistuned consonances (PLOMP, 1967a) invoke the involvement of combination tones (HALL, 1972b). At present, much work is concentrated on dynamic aspects of pitch successions; the opinions seem not to be crystallized enough to cite one authorative review.

In its original conception, the term "residue" was conceived to encompass phenomena associated with the perception of unresolved components, hence, with the perception of temporal phenomena. This aspect of the residue has received less and less emphasis, at least in studies where especially the musical pitch aspects have been pursued. This does not imply, of course, that the perception of temporal properties of signals has been neglected. On the contrary, several schools of research have concentrated on this subject. The entire range of possibilities between tonal residues (distinguished by a clear pitch) and a-tonal residues (distinguished by a rattle-like quality) has been explored by RITSMA and HOEKSTRA (1974). The perception of roughness has been studied very thoroughly by TERHARDT (1968, 1974b), who was amongst the first authors to question the alleged relationship between pitch perception and unresolved components (1970, 1972).

A completely novel way of studying psychophysical phenomena has been introduced by HOUTGAST (1972, 1973). Normally, thresholds are determined as the limits for detecting the presence of a signal. In the pulsation threshold method, the signal under study alternates with another one (encompassing the same frequency region). Under appropriate conditions, an auditory illusion is created; the listener does not perceive the alternations but hears the test signal as a continuous signal. By varying the stimulus conditions the limit of continuity is determined; this is called the "pulsation threshold". This new method has opened up a vast area of applications; the first of these have already been quite successful, a definite confirmation of the fact that a phenomenon like lateral suppression exists also for auditory perception (HOUTGAST, 1972). The consequences of this type of work may reach very far, indeed; at present, they cannot even be assessed in a very approximate way.

3. Conferences

Many European scientists have contributed to residue research. They have had the opportunity to discuss their work in a truly international setting in three important conferences. The first two of these meetings took place in the Netherlands; the third, in Germany. It is interesting to note the shifts in emphasis from the published proceedings of these conferences (PLOMP and SMOORENBURG, 1970; LOPES CARDOZO, 1972; ZWICKER and TERHARDT, 1974). At the time of the first conference (in Driebergen, 1969), emphasis was still on the periodicity aspects of residue signals. This applies to the psychophysical as well as to the physiological research reported. Opinions that tended to disagree with the current view were ventured rather casually and with great reserve, yet they caused a great deal of discussion. The second conference was held in 1972 at the IPO in Eindhoven (SCHOUTEN's institute). At that time, the major part of the switch from periodicity aspects to spectral pattern recognition had already taken place. In the present author's recollection, several of the discussants seemed almost too eager in accepting the newer theories and denouncing the older ones. The balance between enthousiasm and sound criticism was reached somewhat later; the proceedings of the third conference (held in Tutzing, 1974) indicate a major shift of focus. It is no longer general pitch theory that holds the main part of attention. It was recognized that the neural mechanisms underlying the extraction of pitch information are too complex to allow for a fruitful study. Consequently, most of the papers describe research directed at more fundamental aspects of the mechanism of the cochlea, and very few papers have a direct bearing on pitch perception.

At present, there seem to be more puzzling problems concerning basic cochlear operation – problems about mechanics, sharpening, nonlinearity, to name but a few – than in pitch. However, *do we have to solve all these problems completely before we can tackle other questions*? Questions such as: in what form is information about the frequency of partials transferred to the central nervous system? Is it by way of "phase" or by "periodicity" (in the sense of partial synchrony between the stimulus waveform and the firings of auditory nerve fibers)? And: why is information about partials substantially less accurate than that for isolated pure tones?

4. Place and Period Coding in Nerve Fibers

Because of his personal interests, the present author cannot refrain from making some comments on these fundamental issues. The concept of "place" has received a different meaning nowadays. When a pure tone is presented to the ear, the activity (average firing rate) is maximal in a confined group of auditory nerve fibers. This group is characterized by rather sharp boundaries but not by a well localized or very pronounced maximum. Every nerve fiber carries information about the spectral content of the stimulus in a specific frequency region; this is the "principle of specific coding" (DE BOER, 1973). The information, when considered in terms of rate of firing, is not too precise, so that "place" information is not very specific. However, a nerve fiber also carries information in its pattern of firing. In fact, the firings are partly synchronous with a (band-pass) filtered form of the stimulus, the type of filtering being specific to the nerve fiber. If the ear is stimulated by a complex signal consisting of a number of components, many nerve fibers will be excited by more than one component. Still, for many fibers one component will be dominant over the others, and the nerve fiber's firings will be predominantly synchronous with that particular component (ROSE *et al.*, 1969; HIND *et al.*, 1967; DE BOER, 1973). This is what we mean by stating that information about components of a stimulus is also carried in terms of temporal aspects of nerve firings.

One of the major questions at issue during the 1969 Driebergen conference was whether the central nervous system possesses the means of extracting periodicity information from nervous activity patterns contributed by many nerve fibers. Recent research has suggested very strongly that this question is irrelevant insofar as the cues for residue pitch are concerned. Hence, a more modern question would be whether or not *the central nervous system utilizes temporal information from neural firing patterns for the isolation of partials.* This question cannot be fully answered now. But it is difficult to understand auditory analysis of highly specialized signals such as, *e. g.*, the ones used by DUIFHUIS (Section F. 7) on the basis of a "pure" place theory, with neural firing rate as the only variable.

It should be emphasized that it is not implied that the pitch of a partial is measured on the basis of the periodicity of nerve firings. The arguments presented above only concern the *isolation* of a partial, *i. e.*, the tying together of the information conveyed by the nerve fibers from a region in which one partial dominates. In the light of these elaborations, the author expresses the desire that, in the near future, research will be devoted to the processing of periodicity and regularity of firing patterns by the central nervous system. Such a study would be essential in a quest for revealing the mechanisms involved in the isolation of partials.

5. The One-Tone Residue

Returning once more to the subject of residue pitch, we have witnessed a continuous reduction of the number of components. Remarkably enough, the musical quality of the signals used has increased. It has been pointed out in Section I. 1 how cautious we must be in the interpretation of this statement; the listener must be brought into a certain "state" of competence and attention in order to perform the required task. The most extreme form of this effect is encountered in the case of the "one-component residue" (HOUTGAST, 1974). The experiment can only be carried out in a situation of low signal-to-noise ratio; *i. e.*, the signals are presented at levels slightly above the threshold (the threshold may be elevated by a continuously presented wide-band noise signal). The stimuli are two-tone complexes consisting of successive harmonics just as in SMOORENBURG'S and in HOUTSMA and GOLDSTEIN'S work. By way of a task of pitch comparisons, the subject is brought into a state in which he has to utilize his "central pitch processor". In the course of the experiment, one of the stimuli is sometimes replaced by a single component (one harmonic). The subject appears not to notice this, and his scores, when evaluated for those more or less isolated instances, are in line with his other scores. Apparently, the central pitch processor has substituted a subharmonic of the single harmonic just as it has tried to determine common subharmonic frequencies for all the two-component signals.

This experiment demonstrates in an extreme form how much a pattern-transformation process can be fooled before it ceases to operate. When one of the single-component stimuli is presented in isolation, there is no listener who would ascribe a pitch to it corresponding to a subharmonic, *i. e.*, an integral submultiple of the frequency. Yet, in the paradigm of HOUTGAST'S experiment, these instances are bound to occur. Should we conclude that the listener is subject to illusions ? Certainly not, or else we should conclude that all residue pitch perception phenomena are illusions. The experiment only demonstrates the distinction between *performance* in a task and *perception* in an introspective sense. Quite a few authors, when writing about the residue (in the classical sense, associated with periodicity pitch) failed to include the notions about introspection that were so clearly described and emphasized by SCHOUTEN. Modern authors should try to avoid a related pitfall; there is no one-to-one relation between perception and the outcome of a particular experiment.

HOUTGAST'S experiment serves another function; it points once more to the modern-day dilemma described in the previous section. Is the pitch of the subharmonic recognized on the basis of the temporal pattern of nervous activity? Or on the basis of "place", *i. e.*, the locus of maximal activity? These are the same fundamental questions as before but their meaning has become quite different.

Formerly, these questions had to do with the periodicity of the sound and its processing. In this sense the situation is clear:

Is there a "place" for "periodicity" ? The answer is: no.

Are we in a "period" of "place" ? Answer: it seems so.

At present, the same questions can be asked about recognition and isolation of part-tones:

Is there a "place" for "periodicity" ? The answer is: probably.

Are we in a "period" of "place" ? Answer: not necessarily.

"Place" seems necessary in the present "period". But it may not be sufficient; partials may be processed by temporal mechanisms. We do not know for certain.

Period.

References

BÉKÉSY, G. VON: Experiments in hearing. New York: McGraw-Hill Co. 1960.

BILSEN, F. A.: Repetition pitch: Monaural interaction of a sound with the repetition of the same, but phase shifted, sound. Acustica **17**, 295—300 (1966).

BILSEN, F. A., GOLDSTEIN, J. L.: Pitch of dichotically delayed noise and its possible spectral basis. J. acoust. Soc. Amer. **55**, 292—296 (1974).

BILSEN, F. A., RITSMA, R. J.: Repetition pitch and its implication for hearing theory. Acustica **22**, 63—73 (1969, 1970).

BOER, E. DE: On the "residue" in hearing. Academic thesis, Amsterdam 1956a.

BOER, E. DE: Pitch of inharmonic signals. Nature (Lond.) **178**, 535—536 (1956b).

BOER, E. DE: A note on phase distortion and hearing. Acustica **11**, 182—184 (1961).

BOER, E. DE: On the principle of specific coding. J. Dynamic Systems, Measurement, and Control (Trans ASME) **95**-G, 265—273 (1973).

BOER, E. DE: Spectral transformations by essential nonlinearities. Internal memorandum Bell Telephone Laboratories (1975).

BOER, E. DE, BOUWMEESTER, J.: Critical bands and sensorineural hearing loss. Audiology (Basel) **13**, 236—259 (1974).

BRINK, G. VAN DEN: Experiments on binaural diplacusis and tone perception. In: PLOMP, R., SMOORENBURG, G. F. (Eds.): Frequency analysis and periodicity detection in hearing. Leiden, Netherlands: Sijthoff 1970.

BRINK, G. VAN DEN: The relation between binaural diplacusis for pure tones and for complex sounds under normal conditions and with induced monaural pitch shift. Acustica **32**, 159—165 (1975a).

BRINK, G. VAN DEN: Monaural frequency-pitch relations as the origin of binaural diplacusis for pure tones and residue sounds. Acustica **32**, 166—174 (1975b).

BUUNEN, T. J. F., BILSEN, F. A.: Subjective phase effects and combination tones. In: ZWICKER, E., TERHARDT, E. (Eds.): Facts and models in hearing. Berlin-Heidelberg-New York: Springer 1974.

BUUNEN, T. J. F., FESTEN, J. M., BILSEN, F. A., BRINK, G. VAN DEN: Phase effects in a three-component signal. J. acoust. Soc. Amer. **55**, 297—303 (1974).

CARDOZO, B. L.: Proceedings IPO conference on hearing theory. Eindhoven, Netherland: IPO 1972.

CRAMER, E. M., HUGGINS, W. H.: Creation of pitch through binaural interaction. J. acoust. Soc. Amer. **30**, 413—417 (1958).

DE BOER, E.: See under B.

DALLOS, P.: Combination tone $2f_1$—f_2 in microphonic potentials. J. acoust. Soc. Amer. **46**, 1437—1444 (1969).

DUIFHUIS, H.: Audibility of high harmonics in a periodic pulse. J. acoust. Soc. Amer. **48**, 888—893 (1970).

DUIFHUIS, H.: Audibility of high harmonics in a periodic pulse. II. Time effect. J. acoust. Soc. Amer. **49**, 1155—1162 (1971).

DUIFHUIS, H.: Cochlear nonlinearity and second filter; possible mechanism and implications. Submitted for publication, J. acoust. Soc. Amer. (1975).

DURLACH, N. I.: Binaural signal detection: equalization and cancellation theory. In: TOBIAS, J. V. (Ed.): Foundations of modern auditory theory. New York-London: Academic Press 1972.

EVANS, E. F., WILSON, J. P.: The frequency selectivity of the cochlea. In: MØLLER, A. A. (Ed.): Basic mechanisms in hearing. New York-London: Academic Press 1973.

FISCHLER, H.: Model of the "secondary" residue effect in the perception of complex tones. J. acoust. Soc. Amer. **42**, 759—764 (1967).

FLANAGAN, J. L., GUTTMAN, N.: On the pitch of periodic pulses. J. acoust. Soc. Amer. **32**, 1308—1319 (1960a).

FLANAGAN, J. L., GUTTMAN, N.: Pitch of periodic pulses without fundamental component. J. acoust. Soc. Amer. **32**, 1319—1328 (1960b).

FLETCHER, H.: The physical criterion for determining the pitch of a musical tone. Phys. Rev. **23**, 427—437 (1924).

FLETCHER, H.: Speech and Hearing. London: McMillan 1929.

FLETCHER, H.: Auditory patterns. Rev. Mod. Phys. **3**, 258—278 (1931).

FLETCHER, H.: Speech and hearing in communication. New York: Van Nostrand 1953.

FOURCIN, A. J.: The pitch of noise with periodic spectral peaks. In: Proceedings 5th International Congress on Acoustics, Liège, Vol. Ia, B 52 (1965).

FOURCIN, A. J.: Central pitch and auditory lateralization. In: PLOMP, R., SMOORENBURG, G. F. (Eds.): Frequency analysis and periodicity detection in hearing. Leiden, Netherland: Sijthoff 1970.

GOLDSTEIN, J. L.: Auditory spectral filtering and monaural phase perception. J. acoust. Soc. Amer. **41**, 458—479 (1967a).

GOLDSTEIN, J. L.: Auditory nonlinearity. J. acoust. Soc. Amer. **41**, 676—689 (1967b).

GOLDSTEIN, J. L.: Aural combination tones. In: PLOMP, R., SMOORENBURG, G. F. (Eds.): Frequency analysis and periodicity detection in hearing. Leiden, Netherland: Sijthoff 1970.

GOLDSTEIN, J. L.: An optimum processor theory for the central formation of the pitch of complex tones. J. acoust. Soc. Amer. **54**, 1496—1516 (1973).

GOLDSTEIN, J. L., KIANG, N. Y.-S.: Neural correlates of the aural combination tone $2f_1$—f_2. Proc. IEEE **56**, 981—992 (1968).

GOLDSTEIN, M. H. JR.: Neurophysiological representation of complex auditory stimuli. MIT-RLE, Technical Report **323** (1957).

GUTTMAN, N., FLANAGAN, J. L.: Pitch of high-pass-filtered pulse trains. J. acoust. Soc. Amer. **36**, 757—765 (1964).

HALL, J. L.: Auditory distortion products $f_2—f_1$ and $2f_1—f_2$. J. acoust. Soc. Amer. **51**, 1863—1871 (1972a).

HALL, J. L.: Monaural phase effect: cancellation and reinforcement of distortion products $f_2—f_1$ and $2f_1—f_2$. J. acoust. Soc. Amer. **51**, 1872—1881 (1972b).

HALL, J. L.: Two-tone distortion products in a nonlinear model of the basilar membrane. J. acoust. Soc. Amer. **56**, 1818—1828 (1974).

HELLE, R.: Amplitude und Phase des im Gehör gebildeten Differenztones dritter Ordnung. Acustica **22**, 74—87 (1969, 1970).

HELLE, R.: Selektivitätssteigerung in einem hydromechanischen Innenohrmodell mit Basilar- und Deckmembran. Acustica **30**, 301—312 (1974).

HELMHOLTZ, H. VON: Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik. Braunschweig: Vieweg 1862. Fifth edition 1896. On the sensations of tone as a physiological basis for the theory of music. First English edition 1897. Paperback edition. New York: Dover 1954.

HIND, J. E., ANDERSON, D. J., BRUGGE, J. F., ROSE, J. E.: Coding of information pertaining to paired low-frequency tones in single auditory nerve fibers of the squirrel monkey. J. Neurophysiol. **30**, 794—816 (1967).

HOOGLAND, G. A.: The missing fundamental. A place theory of frequency analysis in hearing. Academic thesis, Utrecht (1953).

HOUTGAST, T.: Psychophysical evidence for lateral inhibition in hearing. J. acoust. Soc. Amer. **51**, 1885—1894 (1972).

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

HOUTGAST, T.: Personal demonstrations (1974).

HOUTSMA, A. J. M., GOLDSTEIN, J. L.: Perception of musical intervals: evidence for the central origin of the pitch of complex tones. MIT-RLE, Technical Report **484** (1971).

HOUTSMA, A. J. M., GOLDSTEIN, J. L.: The central origin of the pitch of complex tones: evidence from musical interval recognition. J. acoust. Soc. Amer. **51**, 520—529 (1972).

KIANG, N. Y.-S., WATANABE, T., THOMAS, E. D., CLARK, L. F.: Discharge patterns of single fibers in the cat's auditory nerve. Cambridge, Mass.: MIT-Press (1965).

KIM, D. O., MOLNAR, C. E., PFEIFFER, R. R.: A system of nonlinear differential equations modeling basilar-membrane motion. J. acoust. Soc. Amer. **54**, 1517—1529 (1973).

LICKLIDER, J. C. R.: A duplex theory of pitch perception. Experientia (Basel) **7/4**, 128—134 (1951).

LICKLIDER, J. C. R.: "Periodicity pitch" and "place pitch". J. acoust. Soc. Amer. **26**, 945 (A) (1954).

LICKLIDER, J. C. R.: Auditory frequency analysis. In: Proceedings Third London Symposium on Information Theory 1955.

LOPES CARDOZO, B.: See under C.

MATHES, R. C., MILLER, R. L.: Phase effects in monaural perception. J. acoust. Soc. Amer. **19**, 780—797 (1947).

MCCLELLAN, M. E., SMALL, A. M., Jr.: Pitch perception of pulse pairs with random repetition rate. J. acoust. Soc. Amer. **41**, 690—699 (1967).

MILLER, G. A., TAYLOR, W. G.: The perception of repeated bursts of noise. J. acoust. Soc. Amer. **20**, 171—180 (1948).

NOLL, A. M.: Short-time spectrum and "cepstrum" techniques for vocal-pitch detection. J. acoust. Soc. Amer. **36**, 296—302 (1964).

NORDMARK, J.: Some analogies between pitch and lateralization phenomena. J. acoust. Soc. Amer. **35**, 1544—1547 (1963).

OHM, G. S.: Über die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen. Ann. Phys. Chem. **59**, 513—565 (1843).

OHM, G. S.: Noch ein Paar Worte über die Definition des Tones. Ann. Phys. Chem. **62**, 1—18 (1844).

PATTERSON, R. D.: The effects of relative phase and number of components on residue pitch. J. acoust. Soc. Amer. **53**, 1565—1572 (1973).

PFEIFFER, R. R., KIM, D. O.: Considerations of nonlinear response properties of single cochlear nerve fibers. In: MØLLER, A. A. (Ed.): Basic mechanisms in hearing. New York-London: Academic Press 1973.
PLOMP, R.: The ear as a frequency analyzer. J. acoust. Soc. Amer. **36**, 1628—1636 (1964).
PLOMP, R.: Detectability threshold for combination tones. J. acoust. Soc. Amer. **37**, 1110—1123 (1965).
PLOMP, R.: Experiments on tone perception. Academic thesis, Utrecht (1966).
PLOMP, R.: Beats of mistuned consonances. J. acoust. Soc. Amer. **42**, 462—474 (1967a).
PLOMP, R.: Pitch of complex tones. J. acoust. Soc. Amer. **41**, 1526—1533 (1967b).
PLOMP, R., LEVELT, W. J. M.: Tonal consonance and critical bandwidth. J. acoust. Soc. Amer. **38**, 548—560 (1965).
PLOMP, R., SMOORENBURG, G. F.: Frequency analysis and periodicity detection in hearing. Leiden, Netherland: Sijthoff 1970.
RHODE, W. S.: Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique. J. acoust. Soc. Amer. **49**, 1218—1231 (1971).
RITSMA, R. J.: Existence region of the tonal residue. I. J. acoust. Soc. Amer. **34**, 1224—1229 (1962).
RITSMA, R. J.: On pitch discrimination of residue tones. Intern. Audiol. **2**, 34—37 (1963).
RITSMA, R. J.: Frequencies dominant in the perception of the pitch of complex sounds. J. acoust. Soc. Amer. **42**, 191—198 (1967).
RITSMA, R. J., HOEKSTRA, A.: Frequency selectivity and the tonal residue. In: ZWICKER, E., TERHARDT, E. (Eds.): Facts and models in hearing. Berlin-Heidelberg-New York: Springer 1974.
ROSE, J. E., BRUGGE, J. F., ANDERSON, D. J., HIND, J. E.: Some possible neural correlates of combination tones. J. Neurophysiol. **32**, 402—423 (1969).
ROSENBERG, A. E.: Personal communication (1975).
RYAN, A., DALLOS, P.: Effect of absence of cochlear outer hair cells on behavioural auditory threshold. Nature (Lond.) **253**, 44—46 (1975).
SACHS, M. B., KIANG, N. Y.-S.: Two-tone inhibition in auditory-nerve fibers. J. acoust. Soc. Amer. **43**, 1120—1128 (1968).
SCHOUTEN, J. F.: The perception of subjective tones. Proceedings Kon. Acad. Wetensch. (Neth.) **41**, 1086—1094 (1938).
SCHOUTEN, J. F.: Synthetic sound. Philips Tech. Rev. **4**, 153—180 (1939).
SCHOUTEN, J. F.: The residue, a new component in subjective sound analysis. Proc. Kon. Acad. Wetensch. (Neth.) **43**, 356—365 (1940a).
SCHOUTEN, J. F.: The residue and the mechanism of hearing. Proc. Kon. Acad. Wetensch. (Neth.) **43**, 991—999 (1940b).
SCHOUTEN, J. F.: De toonhoogtegewaarwording. Philips Technisch Tijdschr. **5**, 298—306 (1940c).
SCHOUTEN, J. F., RITSMA, R. J., CARDOZO, B. L.: Pitch of the residue. J. acoust. Soc. Amer. **34**, 1418—1424 (1962).
SCHROEDER, M. R.: "Period histogram and product spectrum", new methods for fundamental-frequency measurement. J. acoust. Soc. Amer. **43**, 829—834 (1968).
SCHROEDER, M. R.: Relation between critical bands in hearing and the phase characteristic of cubic difference tones. J. acoust. Soc. Amer. **46**, 1488—1491 (1969).
SCHROEDER, M. R.: The amplitude behavior of the cubic difference tone. J. acoust. Soc. Amer. **58**, 728—732 (1975).
SEEBECK, A.: Beobachtungen über einige Bedingungen der Entstehung von Tönen. Ann. Phys. Chem. **53**, 417—436 (1841).
SEEBECK, A.: Ueber die Sirene. Ann. Phys. Chem. **60**, 449—481 (1843).
SIEBERT, W. M.: Frequency discrimination in the auditory system: place or periodicity mechanisms? Proc. IEEE **58**, 723—730 (1970).
SMALL, A. M., Jr.: Some parameters influencing the pitch of amplitude modulated signals. J. acoust. Soc. Amer. **27**, 751—760 (1955).
SMALL, A. M., JR., MCCLELLAN, M. E.: Pitch associated with time delay between two pulse trains. J. acoust. Soc. Amer. **35**, 1246—1255 (1963).

SMOORENBURG, G. F.: Pitch perception of two-frequency stimuli. J. acoust. Soc. Amer. **48**, 924—942 (1970).

SMOORENBURG, G. F.: Audibility region of combination tones. J. acoust. Soc. Amer. **52**, 603—614 (1972a).

SMOORENBURG, G. F.: Combination tones and their origin. J. acoust. Soc. Amer. **52**, 615—632 (1972b).

STEELE, C. R.: A possibility for sub-tectorial membrane fluid motion. In: MØLLER, A. A. (Ed.): Basic mechanisms in hearing. New York-London: Academic Press 1973.

TERHARDT, E.: Über die durch amplitudenmodulierte Sinustöne hervorgerufene Hörempfindung. Acustica **20**, 210—214 (1968).

TERHARDT, E.: Frequency analysis and periodicity detection in the sensations of roughness and periodicity pitch. In: PLOMP, R., SMOORENBURG, G. F. (Eds.): Frequency analysis and periodicity detection in hearing. Leiden, Netherland: Sijthoff 1970.

TERHARDT, E.: Zur Tonhöhenwahrnehmung von Klängen. I. Psychoakustische Grundlagen. Acustica **26**, 173—186 (1972a).

TERHARDT, E.: Zur Tonhöhenwahrnehmung von Klängen. II. Ein Funktionsschema. Acustica **26**, 187—199 (1972b).

TERHARDT, E.: Pitch, consonance, and harmony. J. acoust. Soc. Amer. **55**, 1061—1069 (1974a).

TERHARDT, E.: On the perception of periodic sound fluctuations (roughness). Acustica **30**, 201—213 (1974b).

THURLOW, W. R., SMALL, A. M., JR.: Pitch perception for certain periodic auditory stimuli. J. acoust. Soc. Amer. **27**, 132—137 (1955).

TONNDORF, J.: Localization of aural harmonics along the basilar membrane of guinea pigs. J. acoust. Soc. Amer. **30**, 938—943 (1958).

VAN DEN BRINK, G.: See under B.

WALLISER, K.: Zusammenwirken von Hüllkurvenperiode und Tonheit bei der Bildung der Periodentonhöhe. Academic thesis, Stuttgart 1968.

WALLISER, K.: Zusammenhänge zwischen dem Schallreiz und der Periodenhöhe. Acustica **21**, 319—329 (1969a).

WALLISER, K.: Über ein Funktionsschema für die Bildung der Periodentonhöhe aus dem Schallreiz. Kybernetik **6**, 65—72 (1969b).

WEVER, E. G.: Theory of hearing. New York: Wiley 1949.

WEVER, E. G., LAWRENCE, M.: Physiological acoustics. Princeton: University Press 1954.

WHITFIELD, I. C.: Central nervous processing in relation to spatio-temporal discrimination of auditory patterns. In: PLOMP, R., SMOORENBURG, G. F. (Eds.): Frequency analysis and periodicity detection in hearing. Leiden, Netherland: Sijthoff 1970.

WIGHTMAN, F. L.: Pitch and stimulus fine structure. J. acoust. Soc. Amer. **54**, 397—406 (1973a).

WIGHTMAN, F. L.: The pattern-transformation model of pitch. J. acoust. Soc. Amer. **54**, 407—416 (1973b).

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

ZWICKER, E.: Der kubische Differenzton und die Erregung des Gehörs. Acustica **20**, 206—209 (1968).

ZWICKER, E.: Ein hydromechanisches Ausschnittmodell des Innenohres zur Erforschung des adequaten Reizes der Sinneszellen. Acustica **30**, 313—319 (1974).

ZWICKER, E., FLOTTORP, G., STEVENS, S. S.: Critical bandwidth in loudness summation. J. acoust. Soc. Amer. **29**, 548—557 (1957).

ZWICKER, E., TERHARDT, E.: Facts and models in hearing. Berlin-Heidelberg-New York: Springer 1974.