Pitch of the Residue

J. F. SCHOUTEN, R. J. RITSMA, AND B. LOPES CARDOZO

Instituut voor Perceptie Onderzoek, Eindhoven, Netherlands

(Received April 2, 1962)

The residue is defined as the joint perception of a number of Fourier components. Depending on circumstances outside the scope of this paper, it has a pronounced pitch. The consequences of this phenomenon for the theory of hearing are briefly reviewed in the light of past experiments. Special attention is then called to what are termed the first and second effects of pitch shift. The first effect is found when equidistantly shifting the entire Fourier spectrum. The second effect shows itself primarily in a slight drop in pitch when increasing the frequency spacing of the Fourier components. Presented are rather extensive measurements of these effects for a spectrum consisting of three components. Their inherent connection is shown along with their mathematical relationship. As an important experimental finding, the ambiguity of pitch is presented, measured in two more-or-less independent ways. All these phenomena strongly point towards a pitch-extracting mechanism different from and subsequent to the basilar membrane and operating in the time domain.

1. RESIDUE AS A PERCEPT

A SOUND consisting of a single objective frequency f is perceived as a pure tone with a subjective pitch p corresponding to that frequency. Whenever a sound consists of a small number of frequencies, spaced sufficiently widely, subjective sound analysis allows us to hear each Fourier component as a separate pure tone with corresponding pitch (Ohm's acoustical law). If, however, the frequencies are narrowly spaced, the ear fails partly or completely to analyze the Fourier components into corresponding pure tones and, instead, hears one single percept of sharp timbre. If, moreover, the frequencies are harmonics of one fundamental frequency, the percept may have a decidedly low pitch.

Such a percept is called a "residue." It was discovered when listening to a periodic pulse of repetition rate of 200 cps, the spectrum of which contained a score of Fourier components.[1,2] In subjective sound analysis, only the lower harmonics can be heard separately with a loudness decreasing with increasing order. The higher harmonics are not individually perceptible, but are heard as a single sharp note with a pitch equal to that of the fundamental tone. The loudness of this sharp note greatly exceeds that of the fundamental tone, provided the number of unresolved higher harmonics is sufficiently large.

The name "residue" was meant to account for the sharp note as the joint perception of those higher Fourier components which the ear fails to resolve. Indeed, its loudness depends on the number and intensity of the higher harmonics and is fairly independent of those of the lower ones. The sharpness of timbre increases with the average frequency of the spectral region involved. Beats of a residue and a pure tone of corresponding pitch do not occur. Similarly, mutual masking of a residue and a pure tone of equal pitch is not observed.[3,4] Both beats and masking do occur when the frequency of the pure tone is in or near the objective spectral region of the residue.

In view of the above facts, Ohm's acoustical law, stating that the components in subjective sound analysis correspond in a one-to-one relation with the objective Fourier components, breaks down when the spacing of these components is too narrow. It should be restated in the following manner:

1. The ear analyzes a complex sound into a number of separate percepts.
2. Some of these percepts correspond with the Fourier components present in the sound field within the inner

[1] J. F. Schouten, "The Perception of Subjective Tones," Proc. Koninkl. Ned. Akad. Wetenschap. 41, 1086–1093 (1938).
[2] J. F. Schouten, "The Residue, a New Component in Subjective Sound Analysis," Proc. Koninkl. Ned. Akad. Wetenschap. 43, 356–365 (1940).
[3] J. F. Schouten, "The Perception of Pitch," Philips Tech. Rev. 5, 286–294 (1940).
[4] J. C. R. Licklider, "'Periodicity' Pitch and 'Place' Pitch," J. Acoust. Soc. Am. 26, 945 (1954).

ear. These percepts have the timbre of a pure tone and the pitch of the corresponding Fourier component.

3. Moreover, there may exist one or more percepts (residues) not corresponding with any individual Fourier component. These do correspond with a group of Fourier components. Such a percept (residue) has an impure, sharp timbre and, if tonal, a pitch corresponding to the periodicity of the time pattern of the unresolved spectral components.

De Boer[5] distinguished between tonal and atonal residues. Indeed, we find that the tonal residue is particularly pronounced in a frequency region up to 3000 cps and for reciprocal periodicities between one-fifth and one-twentieth of the average of the constituent frequencies. Beyond this region, the residue is atonal, i.e., without definite pitch.

2. RESIDUE AS A CONCEPT

In trying to establish what is trivial and what is striking about the residue, it is necessary to distinguish clearly between two functions of the total hearing mechanism: First, the analyzing function, providing some form of spectral analysis. Second, the pitch-extracting function, enabling us to attribute the notion of pitch to the various distinguishable products of the analyzer.

It should be borne in mind that the normal place theory in fact[6] consists of two completely separate hypotheses: First, with respect to its analyzing function, that the ear performs a Fourier analysis of the sound (Ohm) in terms of a spatial distribution of the frequencies along the basilar membrane,[7] though with a limited resolving power.[8] Second, with respect to its pitch-extracting function, that pitch is determined by the area of stimulation on the basilar membrane.[7]

The first hypothesis, which may be called the place hypothesis of analysis, is by no means in contradiction with the experimental facts mentioned in Sec. 1. The explanation of the residue in terms of nonlinear behavior of the inner ear[9,10] has been disproved by a number of arguments, viz., the occurrence of the residue at moderate loudness and its behavior in phenomena of beating and masking.[3,4]

The second hypothesis, however, which may be called the place hypothesis of pitch, is refuted by the experimental evidence that a group of high harmonics is perceived in bulk as a single percept with a low pitch corresponding to that of the fundamental tone. The experimental evidence, assuming the place hypothesis of analysis to be correct, can now be stated as follows:

1. Stimulation of a particular area of the basilar membrane may give rise to widely different sensations of pitch, depending upon the pattern of stimulation.
2. One sound may consist of several percepts of identical or almost identical pitch.

Ever since Helmholtz, the place hypothesis of pitch has been largely taken for granted. Its seeming obviousness was due to theoretical simplicity rather than to inherent logic or to experimental evidence. The phenomenon of the residue forces us to discard it. Thus, we have to question anew what physiological mechanism performs the measurement of pitch whenever there is a stimulation of a particular area of the basilar membrane with a particular wave pattern.[11,12]

3. PITCH AND TIMBRE

Hitherto, the concepts "timbre" and "pitch" have been used rather vaguely. In fact, many a subject when first listening to a residue is at a loss whether to describe the sound as "high" or "low," since it seems to carry both aspects simultaneously.

Consider, for example, a periodic impulse with a repetition frequency of 200 cps. The Fourier components have frequencies of 200, 400, 600. . .cps. If a filter is applied, passing only the components of 1200, 1400, and 1600 cps, a residue is heard with a pitch equal to that of the fundamental. Again, if the filter passes the components 1800, 2000, and 2200 cps, a residue of equal pitch is heard. Both residues are easily distinguishable, the latter having a sharper timbre than the former. The operational definition of both pitch and timbre provided by the above example may be generalized to other sounds.

Many familiar periodic sounds, such as vowels in speech, yield a Fourier spectrum containing a large number of higher harmonics. In subjective sound analysis, the perception of the residue often predominates over that of the lower harmonics and determines the timbre (formant region) as well as the pitch of the sound. The same applies to the sounds of many string and wind instruments.

In view of this interpretation, it is scarcely surprising that both the timbre and the pitch are hardly dependent upon the presence of the fundamental tone. We realize that, normally, timbre involves more attributes of the sound than the mere spectral distribution of its Fourier components. In this article, however, dealing with stationary sounds, we use timbre in a limited meaning, mainly in order to disentangle the concepts of pitch and

[5] E. de Boer, "On the 'Residue' in Hearing," Ac. Thesis, Amsterdam, 1956.

[6] J. F. Schouten, "The Residue and the Mechanism of Hearing," Proc. Koninkl. Ned. Akad. Wetenschap. 43, 991–999 (1940).

[7] H. L. F. von Helmholtz, *Die Lehre von den Tonempfindungen als physiologie Grundlage für die Theorie der Musik*. (Vieweg Verlag, Braunschweig, Germany, 1862), Chap. 4.

[8] G. von Békésy, "Über die Resonanzkurve und die Abklingzeit der verschiedenen Stellen der Schneckentrennwand," Akust. Z. 8, 66–76 (1943).

[9] H. Fletcher, "A Space–Time Pattern Theory of Hearing," J. Acoust. Soc. Am. 1, 311–343 (1929).

[10] G. von Békésy, "Über die nichtlinearer Verzerrungen des Ohres," Ann. Physik 20, 809–827 (1934).

[11] J. C. R. Licklider, "A Duplex Theory of Pitch Perception," Experientia 7, 128 (1951).

[12] J. C. R. Licklider, "Auditory Frequency Analysis," in *Information Theory*, Proc. 3rd London Symposium on Information Theory, edited by Colin Cherry (Butterworths Scientific Publications, London, 1956), pp. 253–268.

timbre. It is felt that our operational definition is sufficient for this purpose, since subjects after some training proved to be able to make pitch judgments of the residue consistently without being hampered unduly by differences in timbre.[13]

Special mention should be made of the case of the pure sinusoid. Since a change in frequency involves both a shift along the basilar membrane and a change in periodicity, the perceptual change, according to our definition, should be interpreted as consisting of both a change in timbre and a change in pitch.

4. PHYSICAL CORRELATE OF PITCH

We assume that the analyzing mechanism provides us with a rough resolution into separate spectral regions leading to percepts distinguishable by virtue of their timbre and characterized by a more or less pronounced pitch.

We find that stimulation of a particular area of the basilar membrane may give rise to widely different sensations of pitch, ranging from the highest for stimulation with a pure tone down to about one-twentieth of that value.

Hence, the pitch extractor, operating upon the selected spectral region at this particular area, must derive its notion of pitch from certain physical parameters of the spectral region involved. Its location is unknown, as yet, except that it is posterior to the analyzing function of the basilar membrane and, thus, most probably of neural origin.

This pitch extractor might, in principle, operate in two entirely different ways. Either it operates in the frequency domain or in the time domain. When operating in the frequency domain, a pitch equal to that of the fundamental tone could be derived by some process measuring the spacing of the spectral components. When operating in the time domain, however, a pitch equal to that of the fundamental tone could be derived by a process measuring the spacing of certain marking points in the time function.

The former hypothesis has, to our knowledge, never been proposed. We would not consider it a very probable one in view of anatomical evidence, as well as in view of the experimental evidence described here later. The latter hypothesis gathered momentum during the last twenty or so years on account of both perceptual[11,12] and anatomical evidence.[14,15]

5. PHENOMENON OF PITCH SHIFT

The example given in Sec. 3 was a case of an harmonic complex, i.e., a complex the components of which are integer multiples of one frequency. If all components of the harmonic complex are shifted from $f_i=1800$, 2000, 2200 cps to, say, $f_i+\Delta f=1840$, 2040, 2240 cps, the complex is no longer harmonic. Yet, a residue is clearly heard characterized by an evident pitch which, however, is unequal to that of the original harmonic complex. It is observed to move upwards from $p=200$ to, roughly, 204 cps.[3] As an empirical rule, it was found that the pitch of such a residue moves up and down in linear proportion to the center frequency of the constituent components. We refer to this phenomenon as the first effect of pitch shift.

Stating the first effect in more-general terms, the constituent frequencies are given by

$$f_i=f_0-g_0, f_0, f_0+g_0, \tag{1}$$

with

$$f_0=ng_0$$
$$f_i+\Delta f=f_0-g_0+\Delta f, f_0+\Delta f, f_0+g_0+\Delta f. \tag{2}$$

The pitch of the harmonic residue is given by

$$p_0\equiv g_0, \tag{3}$$

and the first pitch-shift effect by

$$\Delta p=\Delta f/n. \tag{4}$$

More-elaborate measurements[5] confirmed the first effect and, in addition, showed up another effect which may be called the second effect of pitch shift. This effect consists of two different experimental findings:

1. The observed pitch shift was found to be slightly but consistently larger than given by (4).
2. When changing the spacing g with constant center frequency, the pitch was found to *decrease* slightly when *raising* the frequency spacing g.[16]

Since these two observations can be correlated, it seems preferable to subsume under one term the second effect of pitch shift.

Let us describe the three frequencies $f-g$, f, $f+g$ as resulting from amplitude modulation of the carrier frequency f with a modulating frequency g. The signal is then given by

$$\begin{aligned} s(t) &= (1+m\cos 2\pi gt)\sin 2\pi ft \\ &= \tfrac{1}{2}m\sin 2\pi(f-g)t+\sin 2\pi ft+\tfrac{1}{2}m\sin 2\pi(f+g)t. \end{aligned} \tag{5}$$

The harmonic situation is characterized by

$$f_0=ng_0, \tag{6}$$

in which n is an integer.

Let us assume that, starting from the harmonic situation, f and g are changed by Δf and Δg, respectively, and that, in first approximation, the pitch shift Δp is a linear function of Δf and Δg. Thus,

$$\Delta p=a\Delta f-b\Delta g. \tag{7}$$

If the two frequency shifts are such that the sound re-

[13] R. A. Jenkins, "Perception of Pitch, Timbre, and Loudness," J. Acoust. Soc. Am. 33, 1550–1557 (1961).

[14] I. C. Whitfield, "The Physiology of Hearing," Progr. in Biophys and Biophys. Chem. 8, 43 (1957).

[15] M. H. Goldstein, "Neurophysiological Representation of Complex Auditory Stimuli," Tech. Rept. 323, MIT Research Laboratory of Electronics (1957).

[16] The latter experiment was repeated by Goldstein[15] with inconclusive results.

mains harmonic,

$$\Delta f = n\Delta g, \tag{8}$$

we have, according to (3),

$$\Delta p = \Delta g. \tag{9}$$

Hence, according to (7),

$$1 = an - b, \tag{10}$$

and

$$\Delta p = [(1+b)/n]\Delta f - b\Delta g. \tag{11}$$

It is easily seen that indeed the harmonic change (8) fulfills the obligation (9). It is also seen that, for $b=0$, (11) reduces to the first effect of pitch shift (4). Thus, the constant b is a measure of the second effect of pitch shift. If b is positive, the pitch shift as a function of Δf must be larger than (4). Moreover, the pitch must decrease when increasing g. The value of b calculated from de Boer's measurements[5] at $f_0 = 1800$ cps, $g = 200$ cps as a function of either Δf or Δg, leads to $b = 0.35$ and $b = 0.40$, respectively, which is a fair agreement and an indirect verification of his accuracy of measurement.

The first and second effects of pitch shift seem to be of crucial importance for an understanding of the phenomenon of the residue and its underlying mechanism. Therefore, we repeated and tried to extend measurements of the two effects. In order to obtain an experimental situation that was both clear-cut and easy, we chose a sound complex consisting of three equally spaced Fourier components. This complex has the experimental advantage of being easily generated by modulation techniques. Moreover, it permits a simple mathematical description.

6. SIGNAL GENERATION

A test signal consisting of the three frequencies $f-g$, g, and $f+g$ was generated by amplitude modulation of a carrier frequency f (center frequency) with a modulation frequency g (frequency spacing). Thus,

$$s(t) = (1+m\cos 2\pi gt)\sin 2\pi ft$$
$$= \tfrac{1}{2}m\sin 2\pi(f-g)t + \sin 2\pi ft + \tfrac{1}{2}m\sin 2\pi(f+g)t. \tag{5}$$

The modulation depth m was normally taken 0.9 in the experiment. Leak of the modulation frequency g through the modulator was reduced to -60 dB.

By choosing f an integer multiple of g, we obtain the periodic signal

$$s(t) = \tfrac{1}{2}m\sin 2\pi(n-1)gt + \sin 2\pi ngt + \tfrac{1}{2}m\sin 2\pi(n+1)gt. \tag{12}$$

Shifting the carrier frequency by Δf results in an additive shift Δf of all three components:

$$s(t) = \tfrac{1}{2}m\sin 2\pi[(n-1)g+\Delta f]t + \sin 2\pi(ng+\Delta f)t + \tfrac{1}{2}m\sin 2\pi[(n+1)g+\Delta f]t. \tag{13}$$

As $|\Delta f|$ increases, the complex becomes increasingly anharmonic.

A matching signal, equally consisting of three frequencies, was generated by recording an harmonic test signal (5) on tape and played back at variable speed. The tape speed was controlled by an oscillator of adjustable frequency. The matching signal is thus given by

$$r(t) = \tfrac{1}{2}m\sin 2\pi(n-1)\gamma gt + \sin 2\pi n\gamma gt + \tfrac{1}{2}m\sin 2\pi(n+1)\gamma gt, \tag{14}$$

γ representing the ratio of playback over recording speed.

Deviations from $\gamma = 1$ result in a multiplicative shift of all three components by a factor γ. The matching complex remains harmonic.

All frequencies were measured with a crystal-controlled frequency counter. Listening was performed binaurally with PDR 10 headphones at 30- to 40-dB sensation level.

7. METHOD OF MEASUREMENT

When listening to the subjective effect of a frequency shift Δf, a change of pitch is clearly heard up or down to about a major third.

Using the subjective scale of musical pitch, settings can be made instructing the subject either to judge the

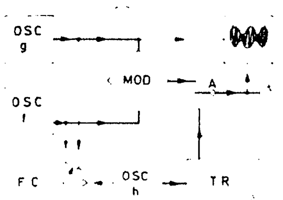

FIG. 1. Block diagram of apparatus. A test complex is generated with the aid of a modulator (MOD) and two sine-wave oscillators (OSC.f and OSC.g). As a matching signal, an harmonic complex from a tape recorder is used (TR). The pitch of this matching signal can be changed by varying the tape speed (OSC.h). A crystal-controlled frequency counter (FC) is used for measuring the frequencies of the oscillators. The switch A consists of two buttons.

pitch difference as a function of Δf or to adjust the Δf so as to obtain one of the discrete musical intervals ranging from half a tone in steps of half a tone up to about the major third.

Although the general behavior of pitch shift can be easily checked and measured in this way, the accuracy is rather poor. A better accuracy can be obtained with a matching method in which the subject is instructed to adjust the parameter γ of a matching signal so as to match the pitch of the test signal. In this method, contrary to the former ones, the ear is used as a zero instrument with respect to pitch differences.

We found, in accordance with others,[17–19] that pitch matchings between a residue and a pure tone are difficult. This difficulty is due to the widely different tim-

[17] A. M. Small, Jr., "Some Parameters Influencing the Pitch of Amplitude Modulated Signals," J. Acoust. Soc. Am. 27, 751–760 (1955).

[18] H. Davis, S. R. Silverman, and D. R. McAuliffe, "Some Observations on Pitch and Frequency," J. Acoust. Soc. Am. 23, 40–42 (1951).

[19] W. A. Rosenblith, Harvard Univ., Psycho-Acoustic Laboratory Progr. Rept. 2 (PNM-6, 1947), as quoted in reference 12.

bres and leads to poor consistency of settings. Therefore, the anharmonic residue (13) was matched with an harmonic residue (14) of closest-resembling timbre. Consistent settings were obtained in this way.

By pressing one of two buttons, the subject could listen to either the test signal or the matching signal. He was free to press the buttons in any order without time limitation. Simultaneous pressing of both buttons was not allowed. In Fig. 1, a block diagram is given of the experimental setup.

8. PROCEDURE

The procedure for measuring pitch shifts is explained by examining the way in which a typical set of matchings was brought about.

Starting from, say, $f_0=2000$ cps and $g_0=200$ cps ($n=10$), f was varied in steps of $\Delta f=50$ cps upwards and downwards, keeping g constant. The γ of the matching signal (with $n=10$) was adjusted until its pitch matched that of the test signal. For larger values of Δf, the sensation of pitch along this particular line of settings gets lost gradually and sets a limit to reproducible measurements. Similarly, g was varied in steps of $\Delta g=5$ cps upwards and downwards, keeping f constant.

It is essential in this procedure that the attention of the subject is directed towards changes along a line of constant n.

Half a dozen well-trained subjects participated in the trial experiments in which f_0 ranged from 1000–3000 cps, g_0 from 100–300 cps. Three of these participated in the final experiments are reported in this paper.

9. RESULTS

The results are shown in Figs. 2 and 3. In Fig. 2, the abcissa is the center frequency f and the ordinate is the pitch p defined as the modulating frequency γg of the harmonic-matching complex. The results of three subjects are shown separately by circles, triangles, and dots. Each point represents the mean of three series of four separate measurements. Solid lines are drawn according to best fit of the experimental points. Dashed lines represent the first effect of pitch shift given by (4).

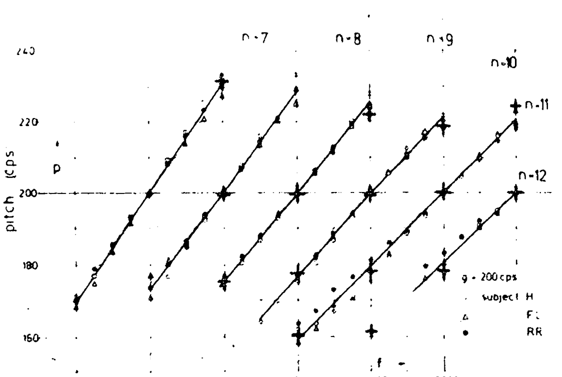

FIG. 2. Pitch as a function of the center frequency for a three-component complex. The frequency spacing g is 200 cps. The circles, triangles, and dots represent the means of twelve matchings made by three subjects, respectively. Solid lines are drawn according to best fit of the experimental points. Dashed lines represent the first effect of pitch shift given by (4). The crosses are the means of clusters like those in Fig. 6.

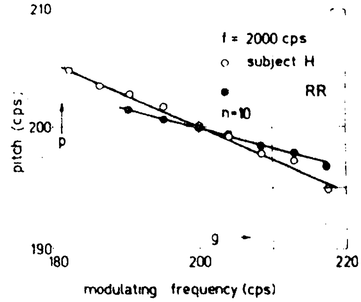

FIG. 3. Second effect of pitch shift. Pitch as a function of the modulating frequency for a three-component complex. The center frequency is 2000 cps. The results of two subjects are shown by circles and dots. Solid lines are drawn according to best fit of the experimental points. These lines fulfill Eq. (11) with $b=0.27$ or 0.16, respectively.

In Fig. 3, the abcissa and ordinate are the modulating frequency g and the pitch p, respectively. The results are shown for two subjects at center frequency $f=2000$ cps. The standard deviation in an experimental point is 0.5–1.0 cps.

10. CONTROL MEASUREMENTS

In Sec. 7, reasons were given for preferring the method of zero settings of pitch difference to settings in which a setting according to a subjective estimation of musical intervals is involved.

Yet, one has to be aware of the dangers inherent in such a matching method. In both the test signal and the matching signal, the low-pitched residue is by far the most prominent percept. Even so, the constituent frequencies can be heard faintly as pure tones with corresponding high pitch. This applies particularly to complexes with relatively wide spacing. Thus, the subject might mistakenly match the pitch of two pure tones rather than that of the residues.

Such settings occurred indeed during early trial experiments. Instructions were corrected accordingly and satisfactorily.

The phenomena of pitch shift, however, cannot possibly be explained in terms of such settings of pure tones. This is already evident from the judgments of musical intervals.

A corroboration within the framework of the matching method was obtained by choosing a test signal consisting of the frequencies near 1800, 2000, and 2200 cps ($n=10$) and a matching signal consisting of the frequencies near 1200, 1400, and 1600 cps ($n=7$). For the rest, the measurements were made according to the procedure described under Sec. 8. Now, the matching signal contained components all of which were unequal

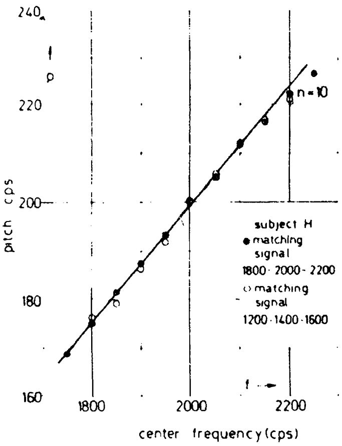

FIG. 4. Control experiment. Plot of subjective pitch against center frequency for different matching signals. Dots show the result obtained by using a recorded complex with frequencies 1800 2000 2200 cps ($n=10$). The circles show the result obtained by using a recorded complex with frequencies 1200 1400 1600 cps ($n=7$). The frequency spacing was 200 cps.

to those of the test signal. Moreover, the timbres of both signals were different. Matching a pair of components or matching the timbres would have given a completely different result from that given in Fig. 2. The results, however, were the same within the experimental errors, as can be seen in Fig. 4. Other control measurements were done with test signals with center frequencies $f_0=1400$ ($n=7$), 1800 ($n=9$), and 2200 cps ($n=11$) and a matching signal with $f_0=2000$ cps ($n=10$). All these combinations gave the same result within the experimental error.

11. DISCUSSION

It can be seen from Fig. 2 that the pitch of the anharmonic residue can be followed and measured along a line of constant n up to the adjoining harmonic situation for positive Δf and, at times, even further for negative Δf. This is a considerable extension of the results of de Boer[5] where the lines ended half-way between the harmonic situations and, thus, presented a sawtooth-like pattern. In fact, when traveling along a line of constant n, the sensation of pitch fades away gradually rather than jumping abruptly towards a line of different n. The first effect of pitch shift is thus amply confirmed and extended.

The second effect of pitch shift is corroborated as shown in Fig. 2 where the solid lines can be seen to deviate from the dashed ones. The second effect of pitch shift was also corroborated by measuring the pitch p as a function of g, as shown in Fig. 3.

In order to check the correlation of the two aspects of the second effect on pitch shift, the constant $b=0.27$, calculated from Fig. 4 for subject H, was used to predict the behavior of pitch as a function of f (Fig. 5). The observed values fit in very well with the calculated ones.

It should be mentioned that b is not an invariant. Among other things, it varies with f and g, as well as from one subject to another.

The f range in which measurements have been made goes from $f=1200$ to $f=2400$ cps. Trial measurements were done for lower f values. In this region, however, subjects experienced increasing difficulties in concentrating on the residue pitch, since the separate components became obtrusively audible.

The region of $f>2400$ cps is characterized by a gradual waning of the phenomenon of the residue as such. Trial measurements with $g=100$ and $g=300$ cps gave the same general tendency.

12. AMBIGUITY OF PITCH

We already drew attention to the fact that, in Fig. 2, the lines of pitch settings for various values of n extend up to and beyond the adjoining harmonic situations.

It was noted also during trial experiments that subjects at times agreed to widely different pitch settings. These stray settings fell pronouncedly outside of the normal error interval. Upon inspection, it turned out later that these stray settings corresponded to settings belonging to lines of different n.

The procedure described in Sec. 8 partly forces the subject to stick to a particular value of n, but evidently his attention goes astray and picks up a pitch belonging to a different value of n.

This means that even in a purely harmonic complex a number of discrete pitches are acceptable to the subject.[6]

In order to obtain a direct experimental access to this phenomenon, subjects were instructed to ascertain whether in an harmonic residue more than one pitch was observable and whether these pitches could be matched with the matching signal. A histogram of such settings for an harmonic complex with $f=1990$ and $g=199$ cps is given as an example in Fig. 6. The histogram does not intend to show the relative predominance of the alternative pitches, since subjects were instructed

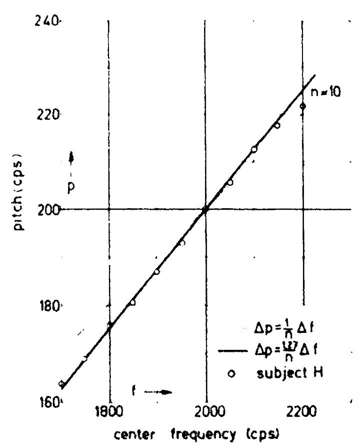

FIG. 5. Check on Eq. (11). The pitch as a function of the center frequency f for a three-component complex. The frequency spacing g is 200 cps. Direct measurements are plotted as circles. The solid line is calculated according to (11) with $b=0.27$ (compare Fig. 3).

to concentrate rather on the secondary pitches than on the most predominant primary pitch. The significance of Fig. 6 is to show the discrete character of the distribution of these settings. Similar measurements were carried out with center frequencies $f_0 = 1600$, 1800, 2200, and 2400 cps. The centers of the clusters found in this way have been plotted as crosses in Fig. 2. They show good agreement with the measurements along lines of constant n.

13. SOME PROPERTIES OF THE PITCH EXTRACTOR

The phenomena of pitch shift described above conclusively rule out an operation of the pitch extractor in the frequency domain (see Sec. 4). They strongly point towards an operation in the time domain measuring time intervals in one way or another. *They also rule out, at least for the tonal residue, the hypothesis that pitch is determined by the period of the envelope of the waveform.* It is evident from (5) that the envelope is affected by g only and independent of any change in f. Thus, the pitch extractor must take the fine structure of the waveform into account.

De Boer[5] pointed out that a fair description of the first effect of pitch shift is given when assuming that the distance of f peaks closest to, but not necessarily coinciding with, g peaks is measured (see Fig. 7). This assumption also elegantly describes the phenomena of ambiguity of pitch as also the fact that the experimental lines of settings (Fig. 2) extend up to and beyond the adjoining harmonic situations.

Of course, the pitch extractor need not necessarily operate on f peaks within g peaks, but might just as well operate on, e.g., zero crossings of the carrier signal at a particular phase of the envelope. As yet, no satisfactory explanation of the second effect of pitch shift can be brought forward.

The outstanding problem really is whether the second effect has anything to do with the first. It is possible that the second effect will give us the very clue for understanding the mechanism of pitch extraction. It is also possible, however, that the second effect is due to some additional mechanism not inherently bearing upon the explanation of the first.

It may be remarked in passing that a pitch extractor operating by measuring time intervals between peaks or zeros would be prone to similar ambiguities when measuring the pitch of pure tones. This effect might be at the root of the well-known difficulties of pitch judgment, such as between two tones an octave apart.

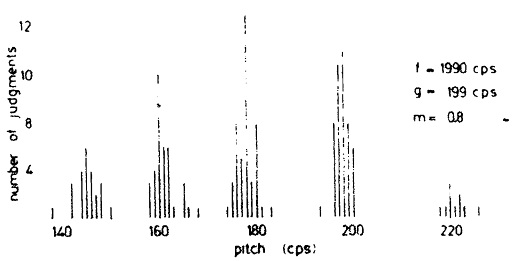

FIG. 6. Histogram of pitch measurements for an harmonic complex consisting of three components. The center frequency $f = 1990$ cps; the frequency spacing $g = 199$ cps.

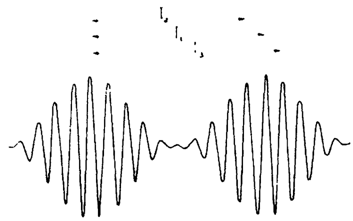

FIG. 7. Behavior of the hypothetical pitch extractor. Waveform of an amplitude modulated sinusoid. A pitch extractor measuring time intervals between main peaks in the waveform will measure a time interval equal to an integral multiple of the interval between successive peaks. In this figure, the main time intervals I_1, I_2, and I_3 correspond to $n = 10$, 9, and 11, respectively. This effect is assumed to account for the observed ambiguity of pitch.

An interesting feature of the tonal residue is that no pitches were found lower than about one-twentieth of the spectral region involved. In view of these considerations, a further exploration of the domain of existence of the tonal residue in terms of f and g would be highly valuable.

14. CONCLUSIONS

1. The first effect of pitch shift[3] is corroborated for complex sounds consisting of three components in a wide range of frequencies (cf. Fig. 2).

2. The second effect of pitch shift[5] is also corroborated for complex sounds consisting of three components in a rather wide range of frequencies (cf. Figs. 2 and 3).

3. The second effect of pitch shift shows itself in two ways, viz. as a small correction to the first effect (4) and also as a drop in pitch with increasing frequency spacing (cf. Fig. 3). The relation between both aspects is described mathematically (11) and was verified experimentally (cf. Fig. 4).

4. The ambiguity of the pitch of an harmonic residue, which was conjectured earlier by one of the authors,[6] has been established experimentally.

5. The phenomenon of the residue necessarily leads to a hypothetical pitch extractor different from and subsequent to the analyzer. As a consequence of the pitch shifts, the operation of the pitch extraction in the frequency domain is highly improbable. Therefore, the hypothetical pitch extractor probably operates in the time domain (e.g., with delay-line techniques). This conclusion can be upheld even without the converging physiological evidence.

6. The inevitable conclusion from the pitch shifts is that the hypothetical pitch extractor does not simply operate on the envelope of the signal. It is highly probable that it takes the fine structure of the signal into account.[5]

ACKNOWLEDGMENTS

The authors wish to express their appreciation to J. E. Evetts (Pembroke College, Cambridge, England) and to J.'t Hart and F. F. Leopold of this Laboratory who participated in the experiments.