Frederic L. Wightman
David M. Green

The Perception of Pitch

The pitch of a sound wave is closely related to its frequency or periodicity—but the exact nature of that relation remains a mystery

Along with loudness and timbre, pitch is one of the most obvious of the psychological attributes of sounds. But what is it about a sound that determines its pitch? Despite the apparent simplicity of this question, there is still no completely satisfactory answer. In the sixth century B.C., Pythagoras noted that if one string is half the length of another, then the pitch produced by plucking the shorter string is one octave higher than that produced by the longer string. The shorter string, of course, vibrated twice as fast as the longer string. Galileo, about 1640, wrote about vibrating bodies and also suggested that pitch is related to the number of vibrations per unit time. There is no denying the force of that argument. Pitch definitely is related to the frequency or period of the sound wave, but the relation is not simple, as we will demonstrate.

Frederic L. Wightman is Assistant Professor of Audiology in the Auditory Research Laboratories at Northwestern University. After receiving a Ph.D in experimental psychology at the University of California, San Diego, in 1970, he spent two years at the Institute for Perception-TNO, Soesterburg, the Netherlands, as an NSF Postdoctoral Fellow. There his interest in pitch perception was first aroused. His current research is in psychoacoustics, primarily in the areas of pitch perception and masking.
David M. Green is Professor of Psychology in the Department of Psychology and Social Relations at Harvard University and, for 1973–74, a Guggenheim Fellow at Cambridge University. His many contributions to the field of psychoacoustics reflect a wide range of research interests.
Both authors wish to acknowledge the support of the National Institutes of Health, Public Health Service, U.S. Department of Health, Education, and Welfare, in the preparation of this article. *Address: Auditory Research Laboratories, Northwestern University, 2299 Sheridan Road, Evanston, IL 60201.*

The problem of devising a theory of pitch perception can be appreciated if one considers musicians in an orchestra tuning their instruments to the same "pitch." An oboe clearly sounds different from a piano or a violin, and certainly the acoustic waveforms they produce are very different. Despite these differences, the pitches produced by the different instruments when they all produce the same note are the same. Any theory of pitch perception must explain this phenomenon. For simple sounds there is no great problem. In general, waveforms that have the same period have the same pitch. If the periods are equal, then their frequencies of repetition or fundamental frequencies are the same (see Fig. 1). However, it is not difficult to produce waveforms that appear to have the same pitch and yet have unequal periods.

Lord Rayleigh, in the introduction to his classic work, *The Theory of Sound* (1877), realized the mistake of associating pitch simply with period and wrote: "In saying that pitch depends upon period, there lurks an ambiguity, which deserves attentive consideration." He pursued the problem by considering a siren, which can be constructed by piercing holes along the perimeter of a disc (see Fig. 2). A windpipe is fixed perpendicular to the disc with its open end opposite the holes in the disc. Air is forced through the windpipe and, as the disc turns, a succession of air puffs emerges from the holes in the disc. These air puffs produce a sound with a definite pitch. Said Rayleigh:

> In the Siren experiment, suppose that in one of the circles of holes containing an even number, every alternate hole is displaced along the arc of the circle by the same amount. The displacement may be made so small that no change can be detected in the resulting note, but the periodic time on which the pitch depends has been doubled. And secondly, it is evident from the nature of the periodicity, that the superposition on a vibration of period τ, of others having periods $\frac{1}{2}\tau$, $\frac{1}{3}\tau$, etc., does not disturb the period τ, while yet it cannot be supposed that the addition of the new elements has left the quality of the sound unchanged. Moreover, since the pitch is not affected by their presence, how do we know the elements of the shorter periods were not there from the beginning?

From the preceding it should be clear that the fundamental problem in formulating a theory of pitch perception is that of invariance. Sounds that are vastly different in their physical properties produce the same pitch. What is it about the physical stimulus and its processing by the auditory system that allows this many-to-one transformation? How can we describe the operation performed by the auditory system in extracting the quality we call pitch?

Research in this area is difficult because pitch is a purely subjective attribute of sound. Certain physical features of sound, like frequency or intensity, are easy to measure directly. With our laboratory instruments, we can easily determine the frequency or intensity of a sound with an accuracy of better than one part in a thousand. But like loudness and timbre, pitch cannot be directly measured. There is no meter for pitch, for it exists only in the head of the listener. The only thing most listeners can tell us about pitch is whether the pitch two sounds are equal, or whether the pitch of one is higher or lower than the other.

Thus we are forced to use matching techniques as our indirect measure of pitch. In a matching procedure, the pitch of the sound in question is compared to the pitch of some reference sound. It is usual to choose a pure tone (such as that produced by a tuning fork or a laboratory sine-wave generator) as a reference sound, because pitch can then be defined simply in terms of the frequency of the tone. If a given sound has a pitch of 200 Hz, this means its pitch has been judged equal to that of a 200 Hz sinusoid.

This procedure has some disadvantages, however. Many interesting experimental sounds have a timbre or quality that is quite different from that of a pure tone, making it difficult for some listeners to match the sounds accurately. In these cases, a secondary reference is carefully selected to ensure that pitch matching between the secondary reference and a pure tone are straightforward. Throughout this paper we will adopt the convention of defining the pitch of a given stimulus as a certain frequency in Hz, implying that the pitch of a pure tone at that frequency is equal to the pitch of whatever reference sound was used.

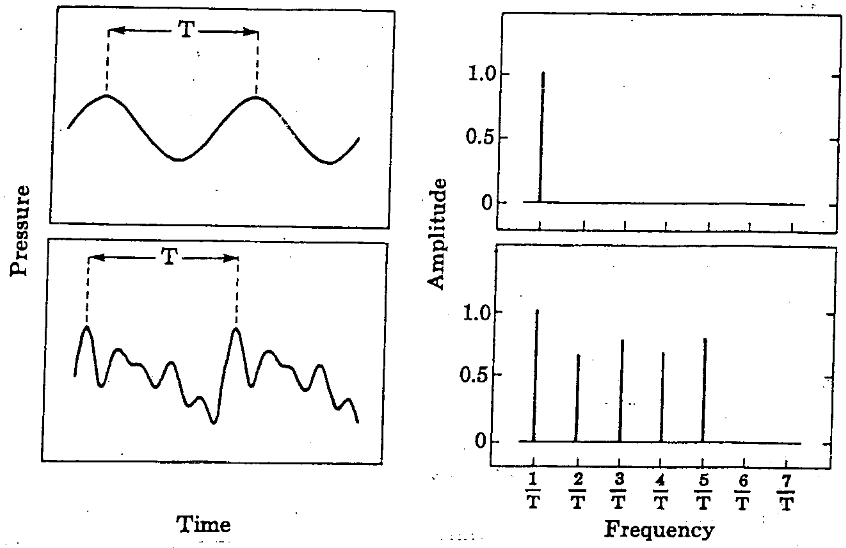


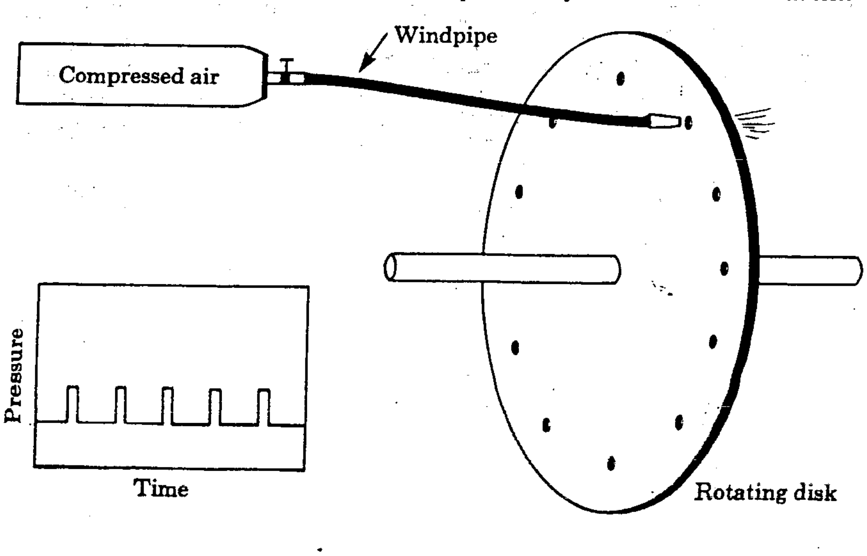
Figure 1. Examples of acoustic waveforms showing the relation between period of repetition and fundamental frequency. In the top case the waveform *(left)* is a simple sinusoid; its period is T. When this waveform is decomposed into its component frequencies, the result is a plot such as that shown at the right. In this plot the single vertical line at the frequency 1/T indicates that all the power in the sinusoidal waveform falls at that frequency. For any waveform with period T, the frequency 1/T is called the fundamental frequency of the waveform. A more complex waveform is shown in the lower part of the figure. The period of this waveform is also T. Because the waveform is complex, its frequency decomposition reveals that there is power in the waveform at several frequencies. Since the period of the waveform is T, the frequencies are at integer multiples of the fundamental frequency 1/T.

Early experiments

Systematic investigation of the mystery of pitch perception probably began with Seebeck's experiments in 1841. Hampered by the lack of modern electronic means for precise stimulus control, Seebeck utilized the siren described earlier to produce and control his stimuli. Small holes were punched equidistantly along a circle on the disc. The rate at which the disc was rotated and the spacing between the holes determined the frequency of the air puffs and, thus, the pitch of the resultant sound. A diagram of the sound wave that was produced by Seebeck's first siren is given in Figure 3A. Seebeck observed that this sound wave produced a very strong pitch which corresponded exactly to the reciprocal of the time between air puffs, or to the fundamental frequency of the sound wave. Moreover, when he doubled the number of holes in the disc (maintaining equidistant separation) and produced the sound diagramed in Figure 3B, the same relation held. In the second case the pitch was an octave higher than in the first, since the time between air puffs was reduced by a factor of 2, and thus the frequency of the puffs was doubled. Seebeck concluded that pitch was determined either by the periodicity of the sound wave or by its fundamental frequency.

Next, Seebeck used his siren to study the pitch produced by discs in which the holes were not equidistantly spaced. He constructed a disc for which the time between air puffs would be alternately t_1, t_2, t_1, t_2, etc. This stimulus is diagramed in Figure 3C. Again, a strong pitch was heard; the pitch was the same as that of a siren which produced equidistant pulses with a time spacing ($T = t_1 + t_2$). Since the pitch of the wave was equal to its periodicity and there was relatively little power at the fundamental frequency, Seebeck believed periodici-

Figure 2. Schematic diagram of a simple siren. The disk is rotated about the axle and, as air is forced through the holes in the disk, a sound is produced. An example of the type of acoustic waveform that might be produced by this siren is shown at the left.



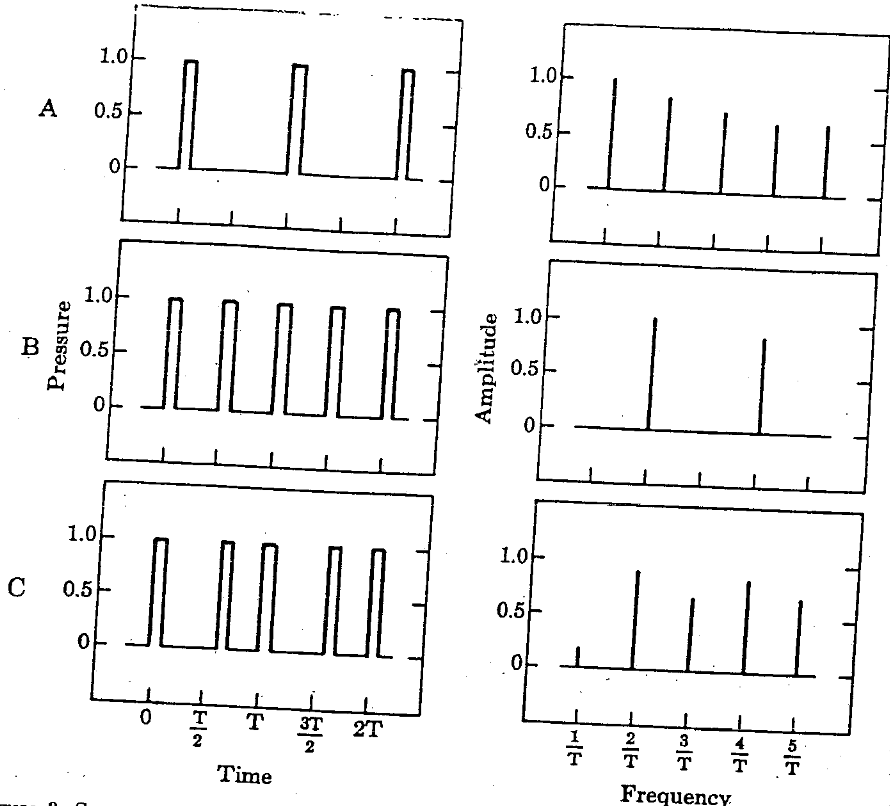


Figure 3. Some of the stimuli used by Seebeck in his experiments with the acoustic siren. At the left of each part of the figure is shown the acoustic waveform. At the right is a partial frequency decomposition of the waveform.

ty rather than fundamental frequency played the dominant role in determining the pitch judgment.

Two years later, in 1843, Ohm severely criticized Seebeck's interpretation. Ohm believed a pitch of a certain frequency could be heard only if the acoustic wave contained power at that frequency. This is the principal assertion of Ohm's famous "acoustical law." Ohm invoked Fourier's theorem on the frequency decomposition of complex waveforms, and showed that indeed the power spectra of Seebeck's waveforms did contain the necessary component, as Figure 3 shows. In this way, he tried to reconcile his law with Seebeck's results. However, Seebeck (1843) replied that the pitches he heard from his siren were much stronger than could be expected on the basis of Ohm's Law, especially in the case of the waveform shown in Figure 3C. In the spectrum of this waveform, there was very little power at the frequency 1/T ($T = t_1 + t_2$) but, nevertheless, the pitch of the waveform corresponded exactly to that frequency. Ohm (1844) finally suggested that this phenomenon was due to an "acoustical illusion."

It was not until nearly twenty years later that a possible resolution of the controversy was offered. In 1862 Helmholtz published his monumental work, *On the Sensations of Tone as a Physiological Basis for the Theory of Music.* Helmholtz strongly supported Ohm's position; in fact, he provided a possible physiological basis for the Fourier analysis of sound waves that Ohm's Law requires. Helmholtz suggested that the basilar membrane in the cochlea of the ear is composed of a sequence of transversely stretched fibers, much like the strings of a harp. The lengths and tensions of these fibers were presumed to make each fiber resonate at a slightly different frequency. A complex acoustical wave vibrating the membrane would be decomposed into its component sinusoids, because only those fibers would resonate that were tuned to the frequencies in the original wave. Thus, the basilar membrane was viewed as a simple Fourier analyzer. This alone, however, does not explain what Ohm called Seebeck's "illusion": that the strength of the pitch sensation in certain cases far exceeded what might be expected, given the physical intensity of the component at the fundamental frequency. Helmholtz's hypothesis of nonlinear distortion in the middle ear was believed to resolve this issue.

Helmholtz supposed that the transduction of sound from the eardrum to the cochlea was a nonlinear process. Nonlinearities of the type envisaged by Helmholtz would distort the incoming wave and, in general, introduce spurious spectral components, or distortion products. These distortion components, behaving as if they were part of the original input, would be analyzed by the proper resonant parts of the cochlea and heard as simple tones, following the tenets of Ohm's Law. For pure-tone inputs, the nonlinearities would generate distortion products at harmonics, or overtones of the input frequency. With complex sounds, such as those produced by Seebeck's sirens, additional distortion products would appear at frequencies given by the frequency differences between the spectral components. In Seebeck's first experiment (Fig. 3A), since the frequency spacing of the components is 1/T, a distortion product at that frequency would be introduced; its power could be assumed to add to that already present at the frequency 1/T, and thus to produce the very strong pitch Seebeck reported for that sound.

The same reasoning would predict a concentration of energy at 2/T for Seebeck's second stimulus (Fig. 3B). The controversy between Seebeck and Ohm is resolved when Helmholtz's distortion hypothesis is applied to Seebeck's third stimulus (Fig. 3C). Note that in this case the sound itself contains very little power at the frequency 1/T, despite Seebeck's report that the pitch of the stimulus corresponded to that frequency. However, since the frequency difference between most of the spectral components of the wave is 1/T, a strong distortion component would appear at that difference frequency. In other words, nonlinear distortion greatly increases the power present at the frequency 1/T. Thus, Helmholtz's nonlinear distortion hypothesis could quite adequately explain Seebeck's observations, within the context of Ohm's Law.

Helmholtz's position remained vir-

...lly unchallenged for nearly three-quarters of a century. Even the advent of precise electronic stimulus-generation equipment did not immediately bring its downfall. For example, in 1924 Harvey Fletcher, an acoustic scientist at the Bell Telephone Laboratories, used electronic equipment to generate sound waves similar to those originally studied by Seebeck. Fletcher found, in support of Seebeck, that even if he filtered out several of the lower harmonics of a complex pulselike waveform, the pitch remained the same. The pitch corresponded exactly to the fundamental, or difference, frequency, even though that frequency was missing from the acoustic waveform. Fletcher also invoked Helmholtz's nonlinear distortion hypothesis to explain this result, now commonly called the problem of the missing fundamental.

Georg von Békésy, although not addressing the issue of nonlinear distortion directly, provided support for other aspects of Helmholtz's theory. In a series of ingenious experiments on human cadavers, carried out in 1928, Békésy directly observed the waves created in the basilar membrane by sound stimulation. His main finding was that stimulation caused membrane vibration that was systematically related to the frequency of the sound. Just as Helmholtz had suggested, the point of maximal vibration of the membrane moved in an orderly way as the frequency of the sound wave was changed. Although the mechanical details of the movement were quite different from those proposed by Helmholtz, von Békésy had uncovered the spectral analyzer necessary for Helmholtz's theory. Von Békésy was later awarded the Nobel Prize for his work. It was not until a decade later that Helmholtz's distortion account of the missing fundamental was disproved.

Schouten's residue theory

In the late 1930s J. F. Schouten and his colleagues in the Netherlands began a long series of experiments on the problem of the missing fundamental. The results of these experiments proved the inadequacies of Helmholtz's hypothesis and laid the foundation for an entirely new theory of the pitch of complex tones. The first experiments were elegant in their simplicity. Schouten reasoned that if the pitch of a complex wave form were the result of nonlinear distortion, then the distortion product should behave just like a simple tone of that frequency. He produced a pulselike stimulus in which the repetition rate of the pulses was 200 Hz but in which all the energy at the fundamental frequency (200 Hz) was canceled out. The pitch of this stimulus, of course, corresponded to the repetition rate of the pulses, 200 Hz. Then Schouten added to this stimulus a pure tone of 206 Hz. If a nonlinear distortion product were responsible for the 200 Hz pitch, the addition of the 206 Hz tone would be expected to produce audible beats (a waxing and waning of the pitch sensation at a 6 Hz rate). No beats were heard, and the pitch of the complex was unaffected. This alone is rather compelling evidence against the distortion hypothesis.

However, Schouten carried his investigations one step further. Using amplitude-modulation techniques he produced complex waveforms in which the frequencies of the individual components could be shifted without disturbing the frequency spacing of the components. In all cases the components were evenly spaced, say with a 200 Hz frequency difference between adjacent components. Consider, for example, a waveform with component frequencies of 1000 Hz, 1200 Hz, 1400 Hz, etc. This waveform had a clear pitch of 200 Hz, corresponding, of course, to the missing fundamental. Now, recall that, according to the distortion hypothesis, the pitch of the complex should correspond to the difference frequency regardless of the individual component frequencies. Schouten showed quite convincingly that this was not always the case. When each of the components was shifted slightly upward in frequency, say to produce a waveform with components at 1040 Hz, 1240 Hz, 1440 Hz, etc., the pitch also shifted, to about 205 Hz. Since the difference frequency is still 200 Hz, this clearly contradicts the distortion hypothesis.

The years following Schouten's pioneering experiments brought more and more conclusive proof of the inadequacy of the distortion hypothesis. The pitch shift experiment described above was repeated many times. Thorough parametric studies were made by de Boer (1956) in Amsterdam and later by Schouten and his collaborators (1962). Ritsma (1962, 1963) mapped out the entire range of conditions for which the pitch shift could be observed. Clearly, it was not a second-order phenomenon of minor theoretical importance.

But despite Schouten's early work, not until much later did the scientific world become convinced that Helmholtz's account of the pitch of the missing fundamental was invalid. In 1954, at the national meeting of the Acoustical Society of America, J. C. R. Licklider conducted a very convincing demonstration. First he produced a stimulus consisting of a sequence of high harmonics of some missing fundamental. Then he showed that the pitch sensation produced by this stimulus was quite sufficient to carry a simple melody. (The pitch was changed to create the melody simply by changing the fundamental frequency.) Next, Licklider added low-frequency noise to his stimulus. The noise was intense enough to mask any distortion component at the fundamental frequency. The dramatic result was that the pitch of the complex was completely unaffected; the melody was heard in spite of the intense low-frequency noise. Licklider's experiment has been refined and repeated several times since 1954, most notably by Thurlow and Small in 1955 and Patterson in 1969. It is now absolutely clear that the pitch of the missing fundamental is not the result of nonlinear distortion in the ear.

With the introduction of his so-called residue theory, Schouten provided the first reasonable alternative to the distortion hypothesis. In fact, the principle on which Schouten built his theory became the basis for several of the modern theories of pitch. In this class of theories, which we will call "fine-structure" theories, pitch is assumed to be derived by some sort of neural operation on the internal representation of the microstructure (or fine-structure) of the in-

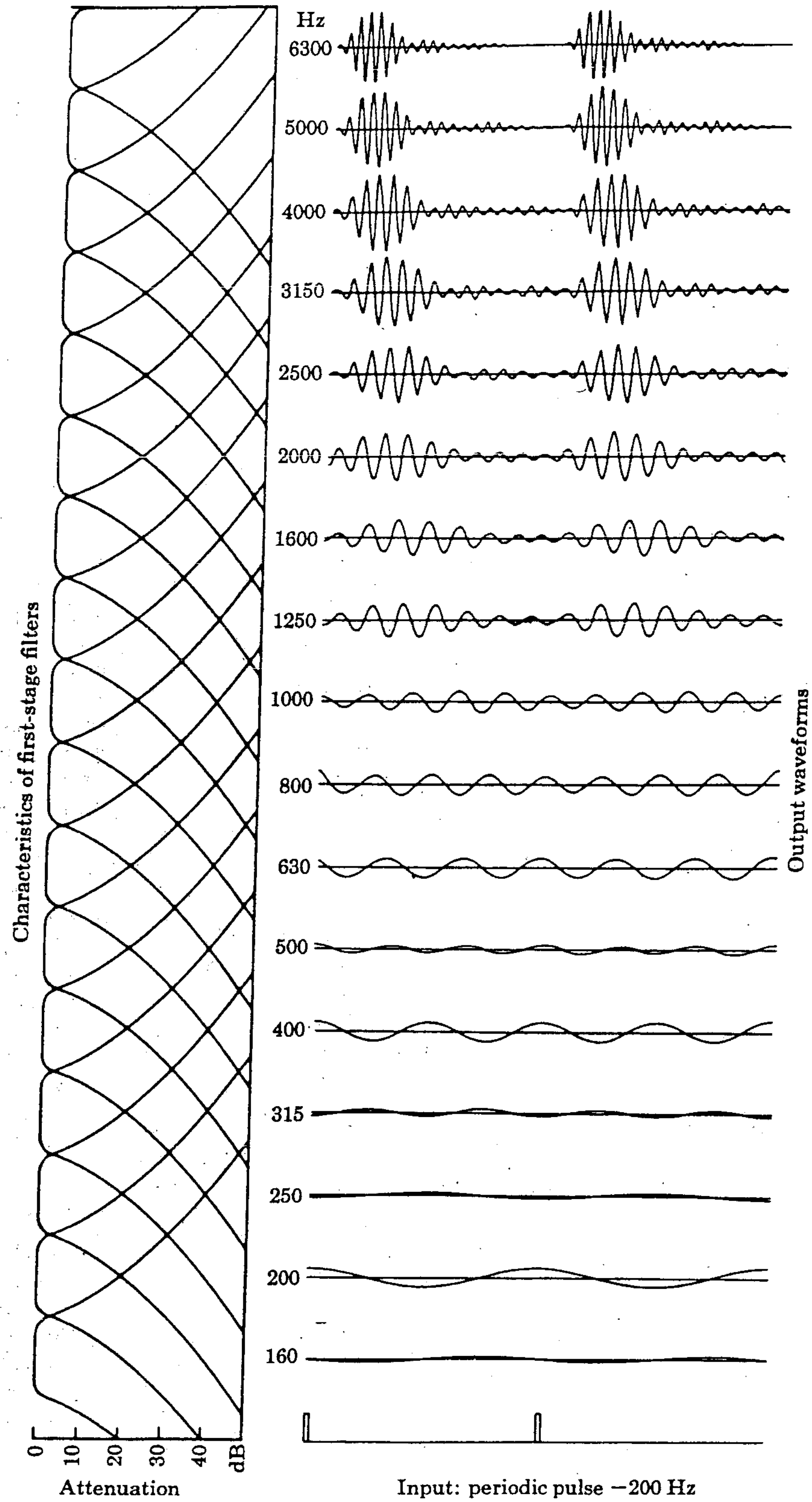


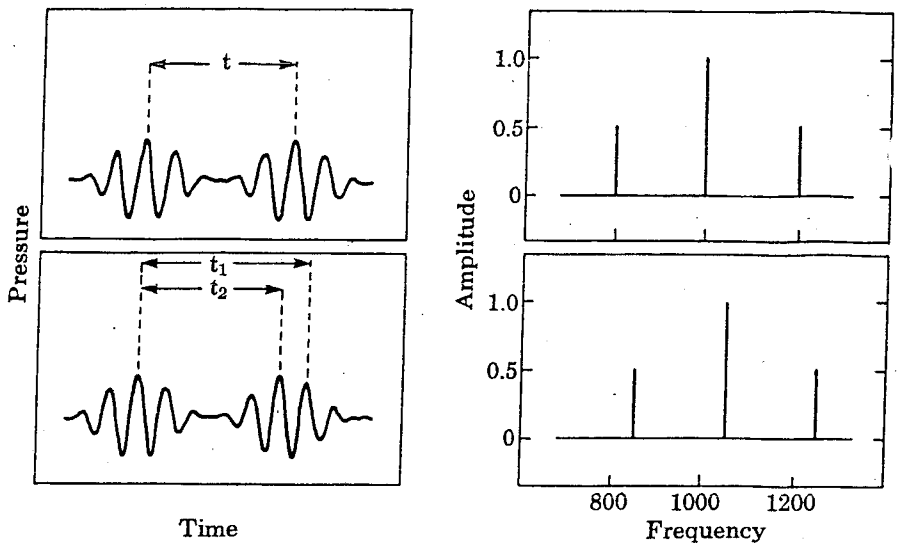
Figure 4. Schematic diagram of the bank of bandpass filters that constituted the first stage of Schouten's model of pitch perception. The high-frequency filters are represented at the top of the figure and the low-frequency filters at the bottom. The center-frequency of each filter is given by the number at the right of the filter characteristic. For this example, the input to the model is assumed to be a pulse-train consisting of all the harmonics of 200 Hz. Note that the outputs of the low-frequency filters are essentially sinusoids, indicating that only one harmonic is passed by each filter. On the other hand, since many harmonics are passed by the high-frequency filters, these outputs are complex. (Figure from Plomp, 1966, *Experiments in Tone Perception*, Institute for Perception RVO-TNO, reproduced by permission.)

coming sound wave. In constructing his theory, Schouten drew heavily on the work of von Békésy, who had observed that pure tone stimuli vibrated the basilar membrane in such a way that the point of maximal vibration depended in an orderly way on the frequency of the stimulus, high frequencies maximally stimulating one end of the membrane and low frequencies the other, so that the basilar membrane functions as a crude spectral analyzer. Schouten suggested that this analyzing property of the basilar membrane could be modeled with an electrical analogue, a bank of bandpass filters. This was the first stage of Schouten's model (see Fig. 4). Because von Békésy had observed that the frequency resolution on the membrane was much poorer at high frequencies than at low frequencies, Schouten assumed that the bandwidths of the first-stage filters were much larger at high frequencies than at low frequencies. In fact, Schouten proposed that the bandwidths of these filters were proportional to their center frequency.

The second stage of Schouten's model consisted of what he called a neural "transmitting mechanism." This device operated in such a way that the temporal fine-structure of the waveform at the output of each of the first-stage filters would be preserved in the temporal patterns of nerve firings and thus be "transmitted" to higher centers. It was particularly important that the positions of the peaks in the fine-structure be coded, since Schouten proposed that pitch was determined by the time-distance between these peaks.

Schouten's theory predicts that listeners hear the low-frequency components of a complex sound as separate simple tones. This is because the bandwidths of the analyzer's filters are narrow enough so that at low frequencies the individual components are "resolved" (i.e. each passed by a different filter). Thus, simple sinusoids would appear at the appropriate low-frequency filter outputs (Fig. 4). On the other hand, the high-frequency components are not separately resolved. The bandwidths of the filters in this region are wide, so that several components of the input waveform

Figure 5. Examples of how Schouten's residue theory can be applied to explain pitch data. In the top part of the figure is shown a complex waveform (and its corresponding frequency decomposition) consisting of 3 harmonics, the 4th, 5th, and 6th, of the missing fundamental, 200 Hz. The pitch of this waveform corresponds to the reciprocal of the time, t, between the major peaks in the waveform, just as residue theory would suggest. This pitch is 200 Hz, the frequency of the missing fundamental. To create the complex waveform shown in the bottom part of the figure, the frequencies of the 3 components have been shifted slightly upward. The pitch of this waveform corresponds closely to the reciprocal of the time t_2. A second pitch that is also occasionally reported for this waveform corresponds to the reciprocal of the time t_1.



interact in each filter, producing a complex output. Schouten suggested then that the high-frequency components are heard together, as a separate percept which he called the "residue." The residue is assumed to have a pitch which corresponds to the periodicity in the waveforms produced by the interaction of the unresolved components. More precisely, the pitch of the residue is assumed to be given by the reciprocal of the time between the major peaks in the fine-structure of the waveform at the high-frequency filter outputs.

Schouten's residue theory had a tremendous impact on all subsequent theories of pitch perception. In the thirty years or so since the theory was first spelled out, scores of experiments on the pitch of complex tones have been reported. With few exceptions, the results of these experiments have been explained in terms at least reminiscent of Schouten's original theory. The theory clearly accounts for both the problem of the missing fundamental and, more important, the pitch shift effect. Figure 5 shows how the theory can be applied to explain these phenomena.

While the general concept embodied in residue theory has seen wide acceptance in the last few decades, the new data contain some rather compelling evidence that the theory is inadequate. One weak point is the assertion that pitch (of complex tones) is derived from the unresolved high-frequency components of the stimulus. A logical extension of this argument would be that as more components interact (i.e. as more components are passed by each high-frequency channel) the pitch sensation would be better or "stronger," since, in these circumstances, the peaks in the resultant waveform typically become more prominent (see Fig. 4). Therefore, a pitch extractor searching for these peaks would have an easier job finding them, and they would be located with more precision.

In terms of residue theory, more components can be made to interact either by moving the components closer together or by moving them to higher frequencies. The unfortunate fact is that the pitch of a complex is weaker under these conditions. For example, Ritsma (1962, 1963) shows that, given a 3-component complex, there are definite upper limits on the component frequencies (about 3500 Hz for a component spacing of 200 Hz) beyond which no residue pitch can be perceived. Moreover, in complex tones that can be assumed to contain both resolved and unresolved components, it appears to be the partially resolved, *low-frequency* components that determine the pitch of the complex.

Perhaps the most convincing evidence on this point comes from Ritsma's (1967) studies of the so-called spectral dominance phenomenon. Roughly speaking, spectral dominance refers to the fact that the pitch of a multicomponent complex appears to depend primarily on the behavior of the components that fall within a spectral region bounded by frequencies about 3 and 5 times the pitch value. For example, if a pitch shift in one direction is created by shifting the components within the dominant region, and an opposite shift by changing the components outside the dominant region, listeners tend to agree that the overall pitch is shifted in the same direction as the components within the dominant region. In other words, the 3rd, 4th, and 5th harmonics are "dominant" with respect to pitch. What is important here is that in any given complex tone, the "dominant" components are clearly among those *best* resolved. All this evidence points to the fact that the pitch extractor cannot work in exactly the way Schouten's residue theory suggests.

The problem of phase

Schouten's residue theory was the first example of a class of theories of pitch perception which we call fine-structure theories. Recall that in residue theory, pitch was derived from a fine-structure analysis of the high-frequency components of the complex tone. Pitch was determined by the time distance between major peaks in the waveform produced by the interaction (in the ear) of unresolved stimulus components. Residue theory was shown to be inadequate, primarily because of the assertion that the high-frequency components are so analyzed.

It is certainly possible to propose a similar fine-structure theory which analyzes the waveform in other regions of the spectrum. For example, Ritsma and others suggest that pitch is derived by a fine-structure analysis in the so-called dominant region. However, a general feature of all such fine-structure theories is that pitch is assumed to be directly related to details of the stimulus *waveform*, or some filtered version of it. Thus, these theories are said to be *phase sensitive*. That is, since the relative starting phases of the components of a complex tone determine the waveform fine-structure, changes in these phase relations might be expected to cause changes in the waveform pitch.

It is easy to show how changes in the relative phases of the components of a waveform can cause dramatic changes in waveform fine-structure. Figure 6 shows two pitch-producing waveforms consisting of the 5th, 6th, 7th, 8th, 9th, and 10th harmonics of 200 Hz. (Thus, the stimulus has components at 1000 Hz, 1200 Hz, 1400 Hz, 1600 Hz, 1800 Hz, and 2000 Hz.) In the left part of the figure, the components are added in cosine phase, resulting in the complex waveform shown at the bottom of the figure. Cosine addition results in a waveform with pronounced peaks. In the right part of the figure, the components are added in random phase (the starting phase of each component is determined randomly), with the result that the peaks of the waveform are much less pronounced. The basic issue is whether or not these waveforms produce the same pitch sensation.

While the results of experiments of this sort are not completely unanimous, the weight of evidence suggests that as far as *pitch* is concerned, the relative phases of the components does not matter. For example, Patterson (1973) reported several experiments in which listeners matched the pitches of complex-tone stimuli. The stimuli were made up of evenly spaced components (6 or 12 in all) added together either in cosine phase or in random phase. Patterson found no differences in the pitch matches to the cosine- and random-phase stimuli, despite the fact that the temporal fine-structures of the two types of waveforms were dramatically different. The waveforms did sound different: there were slight differences in the roughness of the sound depending on the phase. Those with less extreme peaks sounded smoother. However, what is important for our purposes is that they were *identical* in pitch.

There is a simple demonstration of a phase effect which may help the reader to understand why we feel that simple fine-structure theories of pitch are inadequate. Consider the two waveforms shown in Figure 7. The amplitudes and temporal relations among the peaks are quite different in Figure 7A and Figure 7B. The fine-structure models assume a sensitivity to these details. A simple discrimination experiment conducted in our laboratory has convinced us that not only do the two stimuli have the same pitch but they are for all practical purposes indistinguishable. This result clearly conflicts with the expectations of simple fine-structure theories.

A review of the evidence on the phase–fine-structure problem would not be complete without the mention of one experiment that does appear to support a fine-structure theory of pitch. In 1964 Ritsma and Engel described an experiment in which listeners made pitch matches to several 3-component stimuli. The phase of the center components had been shifted 90° with respect to the phases of the two side components (which were in cosine phase). While the variability in the data was rather large, the distribution of pitch matches tended to follow the predictions of their "peak picker" fine-structure theory.

However, in a repetition of the Ritsma and Engel study, Wightman (1973a) obtained quite different results. Using exactly the same stimuli, he found that some of his data agreed with theirs, but others

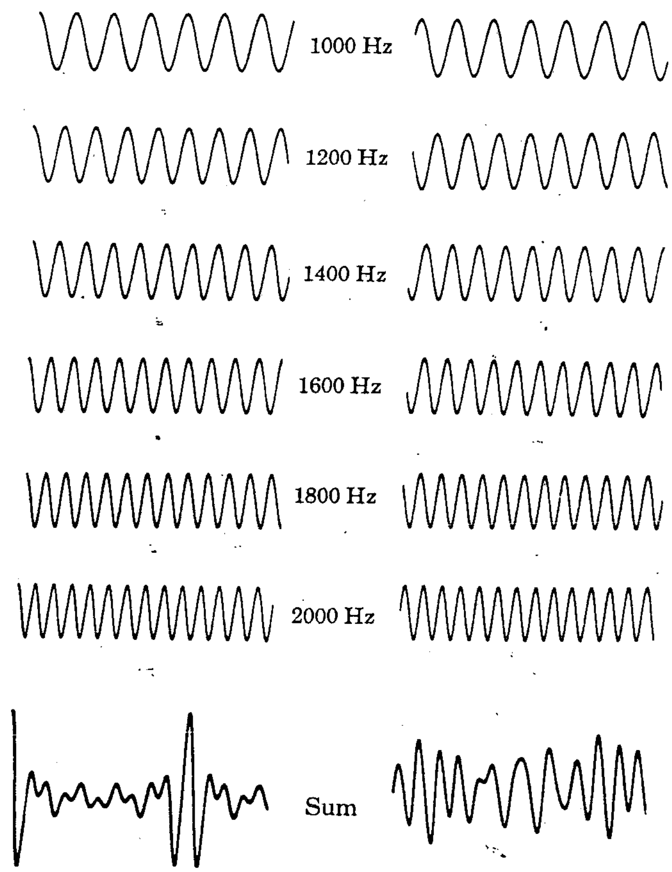


Figure 6. A demonstration of how the relative starting phases of the components of a complex waveform can radically affect the temporal fine-structure of the waveform. In the left-hand column, each component is started in cosine phase. The result of adding these components together is the waveform shown at the bottom of the column, a waveform with pronounced peaks. In the right-hand column, with the starting phase of each component determined randomly, the result is a waveform in which the peaks are not nearly so pronounced. In fact, the waveform looks almost like random noise.

did not. More important, in a separate experiment conducted under the same conditions, Wightman observed that when the center component was shifted back into phase with the side components, the distribution of pitch matches was virtually unaffected. Ritsma and Engel's fine-structure model would not make this prediction.

Wightman also attempted to replicate a small part of the Patterson experiment, by presenting several complex-tone stimuli in different phase configurations and observing the effect on pitch of the change in component phase. The results of this experiment were in complete agreement with Patterson's finding. The pitches of these stimuli were the same regardless of component phase. In another interesting experiment, Houtsma and Goldstein (1971) reported that if the center component of a 3-component complex is presented to one ear and the two side components to the other ear, subjects are completely unable to discriminate phase changes in the central component. Moreover, the pitch of these waveforms is the same as when all components are presented to the same ear.

In spite of the contradictory results from the Ritsma and Engel study, the evidence available now definitely supports the view that pitch is insensitive to details of stimulus fine-structure. It seems clear that all stimuli with the same spectral components have virtually the same *pitch* regardless of the relative phases of the components. Therefore, we are inclined to reject fine-structure or peak-picker models of the pitch-extraction process.

Alternative theories

At this point, two things should be evident: first, the operations by which the auditory system extracts pitch from an acoustic stimulus are anything but simple; and second, we still do not know what those operations are. Pitch perception is still very much a mystery. But we have learned a great deal in our investigation of this mystery. We know, for example, that pitch is not simply related to waveform periodicity, as Seebeck thought. We also know that pitch is not mediated solely by the presence of the corresponding spectral component, as Ohm, Helmholtz, and many others would have us believe. Finally, we can be nearly certain that pitch is not derived from a phase-sensitive operation such as a simple peak-picking analysis of the stimulus fine-structure.

We need to look at pitch perception in a different way—to formulate an entirely new approach to the problem. Wightman (1973b) has made one attempt to do this with what he calls the "Pattern Transformation Model." This promising model is appropriately phase-insensitive and can account for much of the data gathered from pitch-matching experiments. However, it remains to be seen whether the model bears any substantive relation to how pitch is actually extracted by the auditory system. Probably it is only one of a large number of possible models. Much more work on the problem of pitch perception is needed before we will be able to say the mystery has been solved.

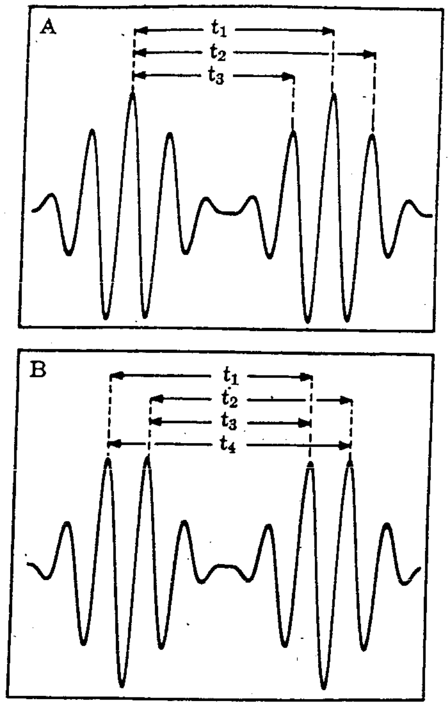


Figure 7. Two pitch-producing waveforms. Waveform B is simply an inverted version of waveform A. The time differences between the peaks of the waveform are similar in the two cases. For example, t_1 in A is the same as t_1 and t_2 in B. Despite these similarities, simple fine-structure theories of pitch would almost certainly predict a discriminable difference between the two waveforms. In fact they are indiscriminable.

References

Békésy, G. von. 1928. Zur Theorie des Hörens: Die Schwingungsform der Basilarmembran. *Phys. Z.* 29:793–810. Trans. G. von Békésy, *Experiments in Hearing* (New York: McGraw-Hill, 1960), pp. 404–29.

de Boer, E. 1956. On the residue in hearing. Ph.D. diss., University of Amsterdam.

Fletcher, H. 1924. The physical criterion for determining the pitch of a musical tone. *Phys. Rev.* 23:427–37.

Helmholtz, H. 1862. *On the Sensations of Tone as a Physiological Basis for the Theory of Music.* English ed., New York: Dover, 1954.

Houtsma, A. J. M., and J. Goldstein. 1971. Perception of musical intervals: Evidence for the central origin of the pitch of complex tones. *MIT Res. Lab. Elec. Technical Report 484.*

Licklider, J. C. R. 1954. Periodicity pitch and place pitch. *J. Acoust. Soc. Amer.* 26:945(A).

Ohm, G. S. 1843. Über die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen. *Ann. Phys. Chem.* 59:513–65.

———. 1844. Noch ein Paar Worte über die Definition des Tones. *Ann. Phys. Chem.* 62:1–18.

Patterson, R. 1969. Noise masking of a change in residue pitch. *J. Acoust. Soc. Amer.* 45:1520–24.

———. 1973. The effects of relative phase and the number of components on residue pitch. *J. Acoust. Soc. Amer.* 53:1565–72.

Rayleigh, Lord. 1877. *The Theory of Sound.* English ed., New York: Dover, 1945.

Ritsma, R. 1962. Existence region of the tonal residue, I. *J. Acoust. Soc. Amer.* 34:1224–29.

———. 1963. Existence region of the tonal residue, II. *J. Acoust. Soc. Amer.* 35:1241–45.

———. 1967. Frequencies dominant in the perception of the pitch of complex sounds. *J. Acoust. Soc. Amer.* 42:191–98.

Ritsma, R., and F. Engel. 1964. Pitch of frequency-modulated signals. *J. Acoust. Soc. Amer.* 36:1637–44.

Schouten, J. F. 1940. *Five Articles on the Perception of Sound (1938–1940).* Eindhoven: Institute for Perception.

Schouten, J. F., R. Ritsma, and B. Cardozo. 1962. Pitch of the residue. *J. Acoust. Soc. Amer.* 34:1418–24.

Seebeck, A. 1841. Beobachtungen über einige Bedingungen der Entstehung von Tönen. *Ann. Phys. Chem.* 53:417–36.

———. 1843. Über die Sirene. *Ann. Phys. Chem.* 60:449–81.

Thurlow, W. R., and A. Small. 1955. Pitch perception for certain periodic auditory stimuli. *J. Acoust. Soc. Amer.* 27:132–37.

Wightman, F. L. 1973a. Pitch and stimulus fine-structure. *J. Acoust Soc. Amer.* 54:397–407.

———. 1973b. Pattern-transformation model of pitch. *J. Acoust. Soc. Amer.* 54:407–17.