

# On the origin of the enlarged melodic octave

William Morris Hartmann

*Department of Physics, Michigan State University, East Lansing, Michigan 48824*

(Received 17 September 1992; accepted for publication 4 March 1993)

The perceptual octave is larger than the physical octave, i.e., most listeners perceive that two tones are an octave apart when their frequency ratio is greater than 2. This result is known as the octave enlargement effect. There are two theories for the effect, one of them a central template theory, the other a peripheral timing theory. In principle, it is possible to determine which theory is better by discovering whether or not octave enlargement occurs for centrally generated dichotic pitches such as the Huggins pitch. Experiments show that octave enlargement does indeed occur for Huggins pitch. This is the result predicted by the central template theory, but it can be argued that the result does not entirely eliminate the timing theory. A detailed examination of the two theories shows that each requires revision in order to make the octave enlargement prediction follow logically from its premises. The central template theory requires the auditory system to differentiate excitation caused by different harmonics of a complex tone on some basis other than place of excitation. Neural synchrony is suggested as a basis. The timing theory, originally formulated in terms of a neural interspike interval timing, can be made internally consistent by replacing the neural interspike-interval circuit by a neural autocorrelator.

PACS numbers: 43.66.Ba, 43.66.Hg, 43.66.Rq [WDW]

## INTRODUCTION

The musical interval known as the "octave" corresponds to a frequency ratio of two to one. It is a special interval, so special that in the western musical tradition, tones that are separated by an octave are given the same note name. For example, the tone with a frequency of 440 Hz is called "A;" a tone with a frequency of 880 Hz is an octave higher, and it is also called "A." For some musical purposes, tones that differ by an octave are essentially equivalent (Krumhansl and Shepard, 1979). Ethnomusicological studies suggest that the interval of the octave is prevalent in the performed music of most, if not all, human cultures (Burns and Ward, 1982). The octave even appears to be a special interval in early infancy (Demany and Armand, 1984).

It is not difficult to find explanations for the special role of the octave. All theories of musical consonance from Pythagorean mysticism to modern neurophysiologically based psychoacoustical models suggest that a frequency ratio of 2 ought to lead to a particularly consonant interval. Therefore, it is surprising to find that in terms of human perception, the octave does not correspond to a precise factor of two. Instead, there is a systematic frequency enlargement in the perceptual octave.

The enlarged octave is most easily seen in a melodic context, where two tones, approximately an octave apart, are heard sequentially with no overlap in time. In modern times, the effect was first seen in an adjustment experiment (Ward, 1954). Here the listener has the task of adjusting the frequency of a tone so that it sounds one octave higher (or lower) than a nonsimultaneous standard tone. The results show that the average matching frequency is significantly greater than a factor of two above (or below) the standard. The experiment has been repeated many times,

with variations, but with no change in conclusions. All studies agree: The octave is enlarged. This is true whether the matching is up an octave or down an octave (Ohgushi, 1983), whether the tones are pure or complex (Terhardt, 1971b; Sundberg and Lindqvist, 1973; Rakowski and Miskiewicz, 1985), or are residue pitches (van den Brink, 1977), or whether listeners are musically trained or not (Dobbins and Cuddy, 1982). (See, however, Hartmann, 1993.) While most of the experiments have used the method of adjustment (pitch matching), Dobbins and Cuddy made an important methodological change and used constant stimuli; the conclusions however, were unchanged.

The amount of enlargement depends upon the frequency range of the experiment. If the lower frequency  $f_1$  of the interval is between 100 and 500 Hz, the effect is usually small, between 0.5% and 1%. For higher values of  $f_1$ , 1000 to 2500 kHz, the enlargement increases to 2% or 3% (Ward, 1954; Walliser, 1969; Ohgushi, 1983). The enlargement becomes as large as 2% also for low frequencies, where  $f_1=50$  or 75 Hz (Terhardt, 1971b; Sundberg and Lindqvist, 1973). The amount of enlargement also depends upon the listener. Most experimenters report considerable individual variation; the percentages given above are averages over listeners.

## I. PITCH MODELS—INTRODUCED

There are essentially two explanations of the octave enlargement effect, corresponding to the two principal types of pitch theory, an excitation pattern (or place) theory and a timing theory.

## A. Excitation-pattern-learning theory

In the early 1970s Terhardt (1969/70, 1971b, 1974) developed a place theory of pitch perception that is a synthesis of two ideas. The first idea is a model of auditory neural excitation patterns and interaction amongst these patterns, following the tonotopic orientation of the physiological observations of von Békésy (1960) and the psychoacoustical masking studies by Zwicker and others (Zwicker and Fastl, 1990). The second idea is a matrix model of associative learning (Kohonen, 1978) whereby the pitches corresponding to the harmonics of periodic complex tones are stored together with the pitches of the fundamentals.

The derivation of octave enlargement in this theory proceeded from the plausible assumption that the memory trace of the octave is established by repeated hearing of complex tones that have second harmonics. Most notable of such tones are human vowel sounds. The second harmonic is an octave above the fundamental. However, because of the interaction among neural excitation patterns, the excitation associated with the second harmonic is partially masked by the excitation of the fundamental. This masking from below causes the center of gravity of the excitation to be shifted along the tonotopic axis, in the direction of a higher frequency place. Therefore, the tonotopic memory trace of a complex tone has an enlarged spacing between successive harmonics.

In an octave matching experiment, with lower and upper sine tones that are not simultaneous, no partial masking occurs, of course. Therefore, as the listener attempts to adjust an interval to match his memory of the separation between a fundamental and a second harmonic, he adjusts an enlarged interval. In the end, the listener tunes sharp octaves in a sequential matching task because he is using a measuring template that has been distorted by exposure to periodic complex tones.

Comments on the excitation-pattern-learning theory are given in Sec. V. What is most important for present purposes is that this theory attributes octave enlargement to a central process, the memory of the enlarged spacing of harmonic excitation patterns. This memory trace may have been originally distorted by the interaction of excitation patterns at the periphery. That does not change the fact that when the listener is given a new pair of sequential sine tones to judge, no peripheral interaction is involved, only the central memory trace of an enlarged octave plays a role.

## B. Timing theory

In 1983 Ohgushi proposed that octave enlargement is caused by anomalies in the timing of neural spikes. His argument began by supposing that the sensation of pitch is determined by a neural representation which is essentially an interspike interval (ISI) histogram. In this respect the theory resembles the pitch acuity model of Sruulovicz and Goldstein (1983), which relates the frequency difference

limen to variability in the ISI histogram. Major contributions to the pitch sensation come from an average over the first several peaks in the ISI histogram.

Ohgushi (1978) observed that the interspike intervals in the peripheral auditory nervous system are not precisely locked to the stimulus wave form. Instead, because of neural refractory effects, both the first and (to a lesser extent) the second peak of the ISI histogram are shifted to longer time intervals. Because refractory effects are more important for higher frequencies, the shift is greater the higher the tone frequency.

The model next assumed that the octave sensation corresponds to a precise factor of two in the average ISI timing. However, because the shift to longer interspike intervals is larger for the upper tone of the octave than for the lower, the upper tone frequency must be more than a factor of two greater than the lower tone frequency in order to achieve the factor of two in average interspike intervals. In essence, the theory says that the octave enlargement in frequency compensates for a refractory effect that increases with tone frequency.

Comments on this timing theory appear in Sec. V. What is most important about the theory for present purposes is that it attributes octave enlargement to a peripheral process, the refractory delay in peripheral neurons. For any pair of sequential tones presented to the listener, the refractory factor applies, so long as timing information is established at the periphery.

## II. CRITICAL EXPERIMENT—HUGGINS PITCH

To try to determine whether the octave enlargement phenomenon is a central effect, as suggested by the excitation-pattern-learning theory, or is a peripheral effect, as suggested by the timing theory, we performed octave judgment experiments using a dichotic-noise-pitch stimulus, the Huggins pitch (Cramer and Huggins, 1958; Durlach, 1962), where the sensation of pitch originates in the central auditory system. The definition of the central auditory system is vague here, but the logic of the present argument only requires that the word "central" refer to a site higher than the eighth-nerve where Ohgushi found the refractory effects.

The Huggins pitch effect uses wideband noise. The phases of the spectral components are random variables, as for any noise. The wideband noise is sent to one of the listener's ears. An identical noise is sent to the other ear, except that over a narrow frequency region there is a relative phase difference. This interaural phase difference ranges from 0 to 360 deg as the component frequency increases across the phase-difference region. Therefore, the low-frequency components in the two ears are identical (relative phase of zero), and the high-frequency components are identical too (relative phase of 360 deg, equivalent to zero), but in the center of the phase-difference region the relative phase is 180 deg, as shown in Fig. 1.

This dichotic stimulus sounds like broadband noise with a sine tone embedded in it. The pitch of the tone corresponds approximately to the frequency where the interaural phase difference is 180 deg (Guttman, 1962;

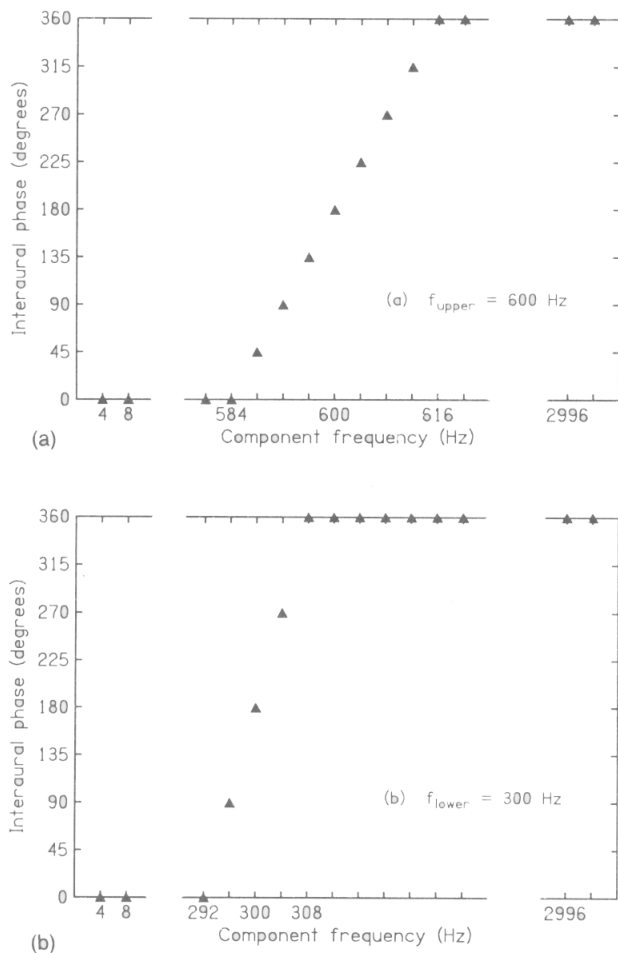


FIG. 1. A digital version of the Huggins pitch stimulus is made with 750 spectral components, equally spaced by 4 Hz. Part (a) shows the interaural phase difference as a function of component frequency, used to make a high-frequency pitch near  $f_u=600$  Hz. Part (b) shows the interaural phase difference used to make a low pitch near  $f_l=300$  Hz.

Bilsen, 1977). What is fascinating about the Huggins pitch is that it is centrally generated. There is no frequency-specific information in the noise sent to either ear. Given a single channel of the stimulus it is impossible to identify a pitch by any means whatsoever. In order to obtain a pitch sensation, both channels must be available, with phase information intact, to a central processor.

Because of its central character, the Huggins pitch resembles the visual effects that are known as “cyclopean perception” (Julesz, 1971). Here, the image seen by one eye is a random-dot pattern, analogous to the random noise used for dichotic pitch. An identical image is seen by the other eye except for local relative deviations in dot position. When both left eye and right eye patterns are presented to a central processor, the relative deviations are interpreted as the outlines of a shape.

Huggins pitch and cyclopean perception are similar in that the signal to each peripheral sensory organ is, by itself, amorphous. The final auditory or visual image is encoded, not so much in the noise patterns, as in the relationship between the noise patterns. The final image is available only to the central nervous system that can fuse separate inputs from the two peripheral sensory organs. This is the

central cyclopean eye or the single central ear.<sup>1</sup> Julesz emphasized the special research role that cyclopean perception can play in locating the site of function within the visual system. If a particular visual effect or optical illusion is observable by the cyclopean eye, then the origin of the effect in question is in the central nervous system; it is not in the retina. In principle, a dichotic pitch can be used in the same way, and this is the basic idea of the experiments of the present article. However, as will be shown in Sec. V below, the analogy is not perfect, and the Huggins pitch experiment is not necessarily decisive. What can be said definitively is that if octave enlargement is observed with sine tones presented either monaurally or diotically, and is not observed with Huggins pitch then octave enlargement must be a peripheral effect; it cannot be central.

### III. METHODS

As noted above, the octave enlargement effect is largest for frequencies above  $f_l=1000$  Hz, and perhaps for frequencies below  $f_l=100$  Hz. These are rather awkward frequency ranges for an experimental method that uses Huggins pitch. The Huggins pitch, like all binaural effects, is strongest in the vicinity of 500 or 600 Hz. It vanishes as frequencies increase beyond 1500 Hz and is relatively weak below 200 Hz. Even at its strongest the Huggins pitch is not strong. Adding to the problem is the obvious fact that the pitches in the experiment have to span an octave. This difficult experimental situation was confronted by doing experiments in a frequency range where the Huggins pitch is strong, and accepting a dauntingly small octave enlargement. A pilot experiment on matching the Huggins pitch, reported in Appendix A, supported the notion that the Huggins pitch might be stable enough to get meaningful data on small octave enlargements.

The experimental method was constant-stimuli, two-alternative forced-choice. The listener heard two dichotic noises in succession with phase-difference center frequencies separated by a musical interval of about an octave. The listener’s task was to decide whether the pitch was larger or smaller than an octave, and to indicate the decision by pushbuttons on a response box. In addition to the main experiment using Huggins pitch, there were two control experiments, sine tone and sine in noise. It is easiest to describe the sine-tone control experiment first, then the dichotic (Huggins) experiment, and then the sine-in-noise experiment.

#### A. Sine-tone experiment

In the sine-tone control experiment, tones were presented to both ears by Sennheiser HD480 headphones while the listener was seated in a quiet sound-attenuating room. The experiment was done in three frequency ranges, specified by the nominal lower frequency of the pair,  $f_l = 300, 350,$  and  $400$  Hz. Upper frequencies of the pair  $f_u$  were approximately a factor of two higher. On successive trials the nominal frequency of the lower tone was kept fixed, and the frequency of the upper tone took on values that were larger or smaller than a physical octave by  $0\%, \pm \frac{2}{3}\%, \pm \frac{4}{3}\%,$  or  $\pm 2\%$ . On a scale of cents these corre-

spond to octave mistunings of 0,  $\pm 11$ ,  $\pm 23$ , or  $\pm 34$  cents. Therefore, there were seven possible upper tones. For example, in the 300-Hz range, the lower tone was nominally 300 Hz and the upper tone took on the nominal values: 588, 592, 596, 600, 604, 608, or 612 Hz.

On a given trial, the listener heard a series of lower-tone upper-tone pairs. Throughout a trial the upper-tone frequency was fixed at one of the seven nominal values. The actual frequencies of the tones differed from the nominal values of  $f_l$  and  $f_u$  by a factor between 0.95 and 1.05, chosen randomly from a rectangular distribution for each trial. In the  $f_l=300$ -Hz range, for example, the actual lower-tone frequency was between 285 and 315 Hz. In the description of results below, actual frequencies are rescaled to their nominal values.

Tones were presented at a level of 54 dB SPL. Each tone was 300-ms long and was turned on and off with a 10-ms raised-cosine envelope. There was a gap of 400 ms between two tones of a pair and a 1200-ms gap between successive pairs. The series for one trial consisted of ten pairs, after which the listener had to make a response. However, the listener could make a response at any time after the first pair, thereby ending the trial immediately. There was no feedback.

In an experimental run, trials with different nominal upper-tone frequencies appeared in random order. Each upper-tone nominal frequency appeared in four trials, for a total of 28 trials in a run. It required between 3 and 6 min to do a run; then the listener could come out of the quiet room and rest. Runs were of two kinds, up or down. In an up-run the lower tone was the first of the pair and the listener judged an ascending octave. In a down-run, the upper tone was the first of the pair, and the listener judged a descending octave. A listener did 5 up-runs and 5 down-runs, in quasirandom order, in each of the three frequency ranges for a total of 30 runs with sine-tone stimuli.

## B. Stimulus generation and control

Stimuli were created by cycling a 2048-word digital buffer containing the tone and converting to analog with 16-bit digital-to-analog converters (DACs), at a nominal rate of 8192 samples per second. The intertone intervals were long enough to permit the controlling computer to reload the buffer prior to each tone. The three different frequency ranges, as well as the plus and minus five percent randomization were accomplished by changing the sample rate.<sup>2</sup>

Tones were low-pass filtered with a cutoff frequency of 3000 Hz and a slope of  $-115$  dB/octave. Tones were shaped by computer-controlled amplifiers. The experiment was controlled by a PDP 11/73 computer, which also collected and stored the response data.

## C. Dichotic noise experiment

The dichotic noise experiment was identical to the sine-tone experiment except that sine tones were replaced by the noise bands that created the Huggins pitch. One digital buffer included a noise file that was common to all

the stimuli. This noise was sent to the right ear. A second digital buffer contained the complementary file, with a phase-difference region centered on one of the frequencies given above, the lower-tone frequency or one of seven nominal upper-tone frequencies. This buffer was reloaded prior to each experimental interval. This noise was sent to the left ear. The DACs for left and right ear noise signals were synchronized by a common clock that initiated each conversion.

The bands of noise were computed by adding 1000 sine components of equal amplitude, equally spaced in frequency by a nominal spacing of 4 Hz. The nominal component frequencies therefore ranged from 4 to 4000 Hz, but the 3000-Hz low-pass cutoff reduced the bandwidth of the noise actually presented to listeners. The phases of the components of the common noise were random, rectangularly distributed from 0 to 360 deg. The phases of a complementary file were identical to the common file except in the phase-difference region, with a total width of four components for lower-frequency centers and eight components for upper-frequency centers. At the center of each phase-difference region was a component with an interaural difference of 180 deg. Figure 1 shows the interaural phase spectra.

Dichotic noises were presented at a level of 70 dB SPL (35-dB spectrum level). All other aspects of the experiment, including timing and envelopes, were identical to the sine-tone experiment. The trial and run structure were identical too, except that for some listeners in some frequency ranges, an eighth and ninth upper-frequency stimulus were added to the set, with shifts of  $\pm 2 \frac{2}{3}\%$  ( $\pm 46$  cents) from the physical octave. Runs of this kind included 36 trials.

## D. Sine-in-noise experiment

The experiment of major interest in this work is the dichotic noise experiment because the main question is whether octave enlargement is seen when pitch is created centrally. The sine-tone experiment is a control. However, it is not clear that it is the *right* control. The Huggins pitch stimulus sounds like a sine tone in noise. In fact, if it were not for a characteristic lateralization, the Huggins pitch might easily be confused with a diotic sine-in-noise stimulus. It is well known that the pitch of a sine tone is changed by embedding it in noise (Egan and Meyer, 1950; Terhardt and Fastl, 1971; Houstma, 1981). If the change should happen to be different for tones that are about an octave apart, the octave seen with the Huggins pitch effect might be susceptible to shifts that do not occur with the sine-tone control. Therefore, it seemed possible that the most appropriate diotic control for the dichotic experiment would be one with a sine tone embedded in noise.

The sine-in-noise experiment was essentially a rerun of the sine-tone experiment, with the same sine signals and the same procedure. However, to each sine signal was added a noise which was simply the common stimulus from the dichotic noise experiment. The noise was enveloped on and off with the sine tone. The sine to noise ratio was chosen by the listener to match the apparent strength

of the Huggins pitch in the noise, a match that listeners found natural to make (see the Appendix). As a result, the sine-in-noise experiment sounded rather like the Huggins experiment. On the average, the matched sine level was 16 dB below the noise level, i.e., a sine level of 54 dB in a 70-dB noise band.

### E. Procedure

The three experiments, sine, dichotic, and sine-in-noise were actually done contemporaneously in a quasirandom order. During a two-hour session, a listener generally did runs from all three experiments. The experimental sessions were blocked by frequency range. A listener began in the  $f_l=350$ -Hz range because the Huggins pitch seemed to be easiest in that range. After completing all three experiments in that range, a listener began experiments in another range.

### F. Listeners

There were six listeners, A, B, C, J, M, and P. Listeners A and M were females; A, C, and P were singers with performance experience in amateur productions. J was a professional recording engineer, and M was an amateur pianist. Listener B was the author. Except for B and J, the listeners were students with an average age of 20.0. The ages of J and B were 28 and 51, respectively. Listener M did not do runs in the lowest frequency range  $f_l=300$ . Listeners C and J were able to do additional runs in higher frequency ranges, as high as  $f_l=550$  Hz, before the Huggins pitch became unusable.

## IV. RESULTS

The results of the forced-choice experiments were plotted as psychometric functions, combining up-runs and down-runs. An example is given in Fig. 2, for listener B in the frequency range with nominal lower frequency  $f_l$  equal to 350 Hz. The function shows the percentage of presented intervals that were judged to be larger than an octave. An exact factor of two (physical octave) corresponds to a nominal upper frequency  $f_u$  equal to 700 Hz. The psychometric functions for all three signal conditions (sine, sine in noise, and Huggins pitch) show an octave enlargement because the crossing at 50% occurs at a frequency higher than the physical octave.

Data from a psychometric function, such as those in Fig. 2, were fitted with a cumulative normal, by converting the percentages to  $z$  scores using the inverse error function and then fitting the best straight line, minimizing the squared error. Zeros and 100% data points were extrapolated on the  $z$  coordinate. The frequency at which the best straight line crossed the 50% line  $f_{50}$  determined the octave enlargement. The percentages of enlargement are defined as  $(f_{50}-2f_l)/2f_l$ , consistent with the definition of Terhardt and others.

Octave enlargement percentages determined from this procedure are shown in Fig. 3(a) and (b). The values of octave enlargement show individual differences. However,

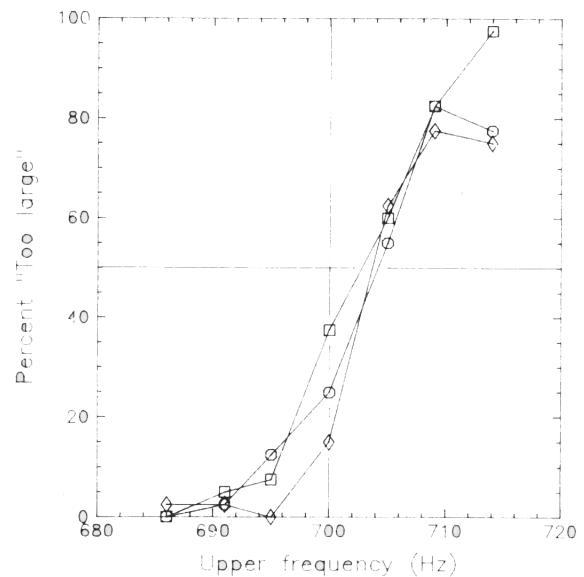


FIG. 2. The psychometric function shows the percentage of the trials in which the listener said that the melodic interval was larger than an octave. The lower frequency of the interval was 350 Hz, the higher frequency is given on the horizontal axis. Different symbols indicate different stimuli: Circles for sine tones, squares for sine tones in noise, and diamonds for Huggins pitch. The crossings at near 50% and 704 Hz indicate an octave enlargement of approximately  $704/700-1=0.57\%$ .

most of the values lie between zero and 1%, which is typical of other studies in this frequency range.

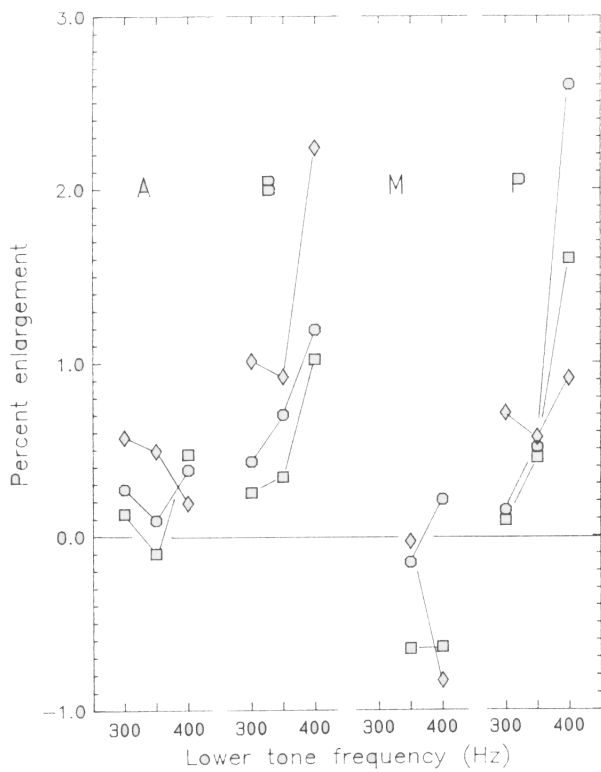
The diamond symbols in Fig. 3 leave little doubt that there is octave enlargement for the Huggins pitch. In fact, there is a tendency for the Huggins pitch to show the greatest enlargement at high frequencies. For listeners, C and J, who were pushed to the limit, the enlargement for the Huggins pitch became unmeasurably large. For other frequencies or other listeners, the enlargement for the Huggins pitch was similar to the enlargement for sine tone and sine tone in noise. In a number of instances the detailed behavior of the enlargement in the case of Huggins pitch paralleled that for sine or for sine in noise, or for both. One listener, listener M, showed an octave contraction; it is significant that the octave contraction occurred for Huggins pitch as well as for sine in noise. The results can be summarized by saying that octave enlargement seen with tonal stimuli does not disappear when pitches are created centrally.

## V. PITCH MODELS—RECONSIDERED

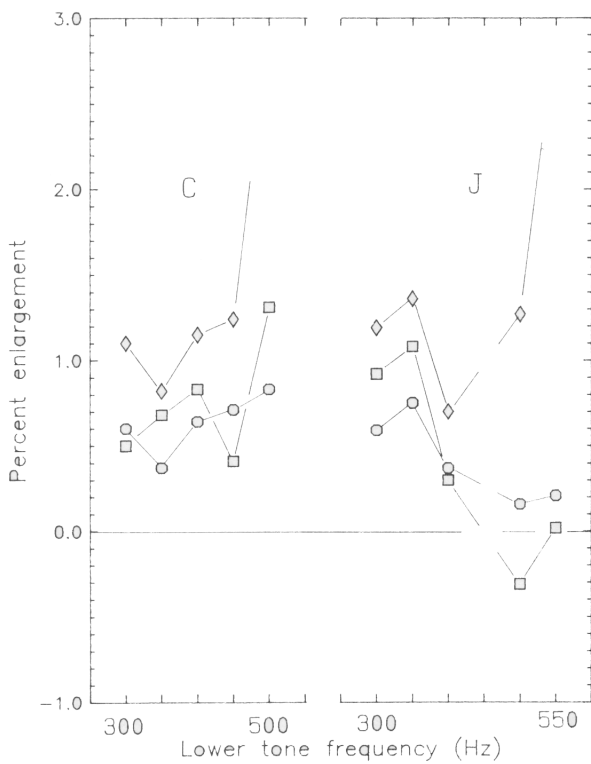
The results of the dichotic experiment are consistent with the excitation-pattern-learning model proposed by Terhardt. There is reason, therefore, to examine this model in more detail.

### A. Central excitation-pattern-learning theory

The excitation-pattern-learning model of octave enlargement requires that the spectral pitch of the second harmonic of a complex tone be shifted upward by the presence of the fundamental, and/or that the spectral pitch of the fundamental be shifted downward by the presence of



(a)



(b)

FIG. 3. (a) Octave enlargement percentages for listeners A, B, M, and P, calculated from the 50% points on psychometric functions. Different symbols indicate different stimuli as in Fig. 2: Circles for sine tones, squares for sine tones in noise, and diamonds for Huggins pitch. (b) The same for listeners C and J.

the second harmonic. It is these shifts that give rise to the distorted template in memory. There is evidence that such shifts occur in Terhardt's data (1969/70, 1971a,b) but there is also contrary evidence, indicating no shift, in the data of Peters *et al.* (1983).

### 1. Octave enlargement and partial masking

If it is granted that spectral pitch shifts do exist, there remains a question about the origin of such shifts. The excitation pattern approach to the effect invokes partial masking. For definiteness we consider the second harmonic in a 200-Hz complex tone. All by itself, this second harmonic is a 400-Hz sine tone with a certain loudness and pitch. The loudness is established by the neural firing rate, and the pitch by the place on a tonotopic coordinate where the firing rate is high. When this tone is embedded in the rest of the spectrum of the complex tone with an intense 200-Hz fundamental component, the loudness of the 400-Hz tone is reduced and its pitch is shifted, presumably upward. These effects are explained by the model as the result of excitation pattern interaction. Some neurons which were responding to the 400-Hz tone in the absence of masking by the 200-Hz fundamental are expropriated by the fundamental component. Therefore, the loudness of the 400-Hz tone is reduced. The expropriated neurons are mainly those with characteristic frequency (CF) at the low end of the 400-Hz range. This leaves only higher-CF neurons responding to 400 Hz, and this mechanism provides a place-theory explanation for the upward pitch shift.

A logical problem with this explanation of pitch shift is that it assumes that a particular place on the tonotopic axis can first be regarded as part of a 400-Hz excitation and can then be regarded as part of a 200-Hz pattern. A system that registers only gross firing rate and its place on a tonotopic axis cannot do that. A system which *can* do that is one which recognizes the difference between 400-Hz excitation and 200-Hz excitation on the basis of neural timing. But if neural timing allows the system to demarcate regions of the tonotopic axis that are excited by different frequencies then neural timing is contributing directly to frequency discrimination and at least indirectly, if not directly, to pitch perception. Thus neural timing is brought into a pitch model which began as a pure place theory.<sup>3</sup>

### 2. Simultaneous octave enlargement

A second difficulty for the excitation-pattern-learning model comes from the fact that octave enlargement appears in experiments with simultaneous sine tones (harmonic octave) as well as for successive tones (melodic octave) (Ward, 1954; Demany and Semal, 1990). The problem is that in the case of simultaneous presentation, the excitation pattern interactions that led to the stretched memory template are present also in the stimulus that the listener needs to adjust or judge. One might therefore predict no octave enlargement for simultaneous sine tones, contrary to experiment.

This difficulty is solved, in principle anyhow, by the level dependence of the pitch shifts. A reasonable assump-

tion is that excitation patterns are learned from environmental spectra in which the second harmonic has a lower level than the fundamental. However, octave matching experiments for simultaneous sine tones are done with both tones of the same level. Whereas partial masking of the second harmonic may be present in both cases, one expects greater masking for the environmental spectrum. Thus the shift in the template is greater than the shift in the test tones. As a result, octave enlargement may be observed in the tuning of simultaneous octaves.

This idea can be made quantitative with a calculation using the algorithm of Terhardt, Stoll, and Seewann (1982): In a complex tone with a 200-Hz, 60-dB SPL fundamental and 10 harmonics with level decreasing at a rate of 6 dB per octave, the interval between the second harmonic and the fundamental is enlarged by 2.6%. This estimates the enlargement stored in the memorized template. The calculation for a complex tone with two components, 200 and 400 Hz each at 60 dB, leads to an enlargement of 2.0%. The difference of 0.6% is the predicted value for the enlargement in tuning seen experimentally, and this value is actually in reasonable agreement with experiment. However, similar calculations for higher frequencies lead to vanishing values of octave enlargement (for example 0.06% for  $f_1=1500$  Hz and a  $-12$ -dB/octave template) which disagrees with experiment. That this idea fails for higher frequencies is unfortunate because it is at higher frequencies where octave enlargement is the greatest (Ward, 1954). It seems possible, however, that this idea is fundamentally correct, but that the mathematical algorithm is not accurate enough to permit calculations that involve differences as this one does.

The most important point made by this section on the excitation-pattern-learning model is that in order to develop an octave template that is enlarged, the auditory system must separate excitation caused by the second harmonic from excitation caused by the fundamental. Undifferentiated firing rate is insufficient information. In fact, a template derived from undifferentiated firing rate would be compressed and would lead to a contracted octave rather than an enlarged octave, because the centroids of two excitation peaks or the loci of their maxima are brought closer together when the excitations are superimposed.

## B. Peripheral timing model

The results of the dichotic experiment above show that octave enlargement is central. This result appears to disagree with the model proposed by Ohgushi (1978,1983) wherein octave enlargement results from the refractory nature of peripheral neurons. However, a more careful analysis, requires an examination of the Huggins pitch effect itself. The important question is whether refractory delays in peripheral neurons can affect a centrally created pitch.

As it happens, existing models of the Huggins pitch (Durlach, 1962; Raatgever and Bilsen, 1986) are essentially tonotopic. The role of neural timing in these models is only to create lateralization anomalies for a particular tonotopic region which then codes for pitch. The pitch itself is, therefore, placed based and is unaffected by refrac-

tory delays. Such models are inconsistent with Ohgushi's pitch theory from the outset and cannot rescue the theory from the problem posed by the octave enlargement seen with Huggins pitch.

Existing models notwithstanding, one can imagine a model of the Huggins effect in which the pitch is encoded in a central timing pattern. In this case, refractory delays in the tuned channels of the periphery are simply passed along to the central timing pattern where they could indeed result in octave enlargement. Because such timing models are not ruled out by our experimental results, Ohgushi's timing approach is next discussed in some detail.

## C. Timing theory

Historically, timing models of pitch perception have been vulnerable to the complaint that neural timing is too rigidly tied to the period of the stimulus to allow for the many pitch shift effects that are known to occur. For example, von Békésy argued that the existence of tonal diplacusis makes timing theories implausible (Green, 1977). By contrast, diplacusis could easily be accommodated within an excitation pattern model (Ward, 1963; see however, Turner *et al.*, 1983).

It was, therefore, a matter of some importance when Ohgushi's 1978 paper liberated timing models from bondage to the stimulus. Ohgushi showed that the locations of the peaks of an ISI histogram can be flexible. He further argued that physiological observations of eighth-nerve neurons appeared to show that there actually were appreciable displacements of the peak modal time intervals from the intervals expected from the stimulus frequency.<sup>4</sup> If pitch is derived from a representation like the ISI histogram then, as noted by Burns (1982), pitch can be similarly flexible within a timing model. For example, an unpublished work by Jones *et al.* (1983) shows how Stevens' rules for the effect of intensity on pitch can be explained in terms of refractory effects on the ISI histogram.

Examination of the ISI concept shows that there are indeed only a few constraints on the ISI histogram. The first is simply a conservation of the number of spikes,

$$\sum_{i=1}^{N_b} n_i = N, \quad (1)$$

where  $n_i$  is the number of spike counts in bin  $i$  with interspike interval  $\tau_i$ .  $N_b$  is the number of bins in the histogram and  $N$  is the total number of spikes counted.

The second constraint conserves the duration of the counting experiment  $T_D$ ,

$$\sum_{i=1}^{N_b} n_i \tau_i = T_D. \quad (2)$$

In the case that the peaks of the ISI histogram are well separated, as they are for intense low-frequency tones, an alternative form of the time conservation equation is

$$\sum_{p=1}^{N_p} n_p \bar{\tau}_p = T_D, \quad (3)$$

where  $n_p$  is the number of counts in histogram peak  $p$  with average ISI of  $\bar{\tau}_p$  and  $N_p$  is the number of peaks in the analysis. These two constraints are weak and offer enormous latitude for pitch shift effects.

Ohgushi's explanation of the octave enlargement attributes the effect to the refractory nature of neural firings. Because a neuron is less likely to fire if it has recently fired, the first peak of the ISI histogram occurs at a time that is longer than the period of the stimulus. Similarly, but to a lesser extent, the second peak in the ISI histogram occurs at a time that is larger than twice the period. The higher the frequency of the stimulus the greater the delays in peak positions.

The explanation continues by arguing that the octave is perceived when the major peaks of the lower tone occur at times that are exactly a factor of two greater than the major peaks of the upper tone. Because the peaks of the upper tone are more delayed by refractoriness than the peaks of the lower tone, the period of the upper tone must be correspondingly reduced in order to obtain the factor of two. Therefore, the frequency of the upper tone must be made more than twice the frequency of the lower tone, and that is the reason for the octave enlargement. Ohgushi's model, like Terhardt's, has the desirable property that it predicts greater octave enlargement for higher frequency ranges, in agreement with experiment.

Despite its attractive features, this model suffers from a problem of internal consistency. The model begins with a physiological explanation for pitch in terms of interspike intervals, but in the end it resorts to a nonphysiological origin for the factor of 2 that characterizes the perceived octave.

What makes this leap out of the physiology seem so unsatisfying is that the physiological ISI analysis itself offers a compelling approach to the special nature of the octave: As has been frequently observed, to obtain a physiological octave one merely needs to line up consecutive interspike intervals in the low-frequency tone with the even-numbered intervals of the high-frequency tone. If the frequencies of two tones are chosen so that the interspike intervals align in this way, then a neural circuit that is sensitive to an ISI equal to an ISI of the lower tone is excited by both tones. Such a neuron automatically recognizes musical intervals, particularly the octave. For example, an ISI circuit tuned to the first ISI of the lower tone also responds to spikes in the second ISI of the upper tone.

The behavior of such a neural circuit given refractory neurons depends upon the nature of the refractoriness. If the refractory recovery is a function only of the time since the most recent spike then the model predicts no octave enlargement. Successive ISIs of the lower tone and the corresponding even ISIs of the upper tone are equally delayed. Alignment takes place when the upper tone frequency is simply twice the lower tone frequency. If the refractory recovery time of a neuron scales inversely with best frequency (an ecologically plausible assumption) then refractoriness produces a larger delay in the first and second peaks of the lower tone than in the second and fourth peaks of the upper tone. As a result, the model predicts an

octave contraction, contrary to experiment. There does not seem to be any way to get an octave enlargement with an internally consistent ISI circuit.

A solution to the dilemma, that maintains the spirit of Ohgushi's model, is to replace the ISI neural circuit with a neural autocorrelator. The difference between the ISI circuit and the autocorrelator is a simple one. Mathematically, the ISI circuit is an autocorrelator with an extra factor in the correlation integrand that eliminates the counting of spike pairs that are separated by an intermediate spike. This small difference in models makes a big difference in the predictions for octave matching. For example, an autocorrelator with lag equal to the first ISI of the lower tone not only responds to spikes in the second ISI of the upper tone, but can also respond to spikes in the first ISI of the upper tone, so long as the frequency of the upper tone is increased to compensate for the refractory delay. This delay is actually doubled because this case requires two spikes in succession.

One can go further along this line of reasoning. If the neural response to the upper tone exhibits refractoriness in spikes in the first ISI, but negligible refractoriness when spikes synchronize with larger multiples of the tone period then a neural autocorrelator with a lag equal to the period of the lower tone responds to two different upper tone frequencies, one higher than twice the lower tone frequency, the other equal to twice the lower tone frequency. Thus, the autocorrelation function model predicts that there are two perceptual octaves, one enlarged and the other exact. Such a prediction is in qualitative agreement with data from all the listeners in this study, including M. Additional evidence for a bimodal octave was found in a survey experiment (Hartmann, 1993).

## VI. CONCLUSION

The octave enlargement effect for successive sine tones has been explained by two different models, one based upon a place theory of pitch perception and an excitation-pattern-learning model of memory, the other based upon a timing theory of pitch perception and the refractory nature of neural firing.

The first of these models clearly attributes octave enlargement to a process of the central nervous system. The enlargement effect is said to occur because the sine tone stimuli are judged by a central template that has been bent out of shape by exposure to complex tones. The second model regards the effect as peripheral in that the timing information of the two sine tones themselves has been distorted by the refractory period, as evident at the level of the eighth nerve.

In principle, an experiment that studies octave matching for pitches that are generated centrally is capable of determining which of these two models is better. If it should happen that octave enlargement is observed for sine tones but not observed for Huggins pitch, the natural conclusion would be that octave enlargement is not the result of a distorted central template. The reason is that if the template matching process is central then the operation of the template would be the same regardless of how the



pitches themselves are generated. By contrast, the timing model would survive the experiment unscathed, it being supposed that timing information generated centrally would be immune to peripheral refractory delays.

Sections II–IV described octave judging experiments in which the Huggins effect was used to generate pitches centrally. The results of the experiments were unambiguous: Octave enlargement does occur for centrally generated pitches, just as for sine tones. Unfortunately, this result is not as easy to interpret as the opposite result would have been. The result actually observed is clearly consistent with the central template matching model. The difficulty comes in deciding whether it is consistent with the timing model. It is possible to argue that there is no inconsistency. Conceivably refractory delays occur at the periphery for the components of white noise in the same way as they occur for sine tones. Perhaps these delays are passed through to the point where the central system uses the binaural difference to create a pitch sensation and then exert their effect there. Alternatively, there may be refractory delays in central neurons following the central creation of pitch.

The experiments are, therefore, not conclusive. One might argue that the results favor the central template model of octave enlargement because the results are understandable there in a very straightforward way. On the other hand, one might argue that the results favor the timing model because there is a systematic (though not universal) tendency for the octave enlargement to be somewhat greater for Huggins pitches than for sine tones, and for this difference between Huggins and sine-tone results to increase with increasing frequency range. That binaural neurons might be subjected to greater refractory delays than monaural, especially at high frequency, is plausible. By contrast, one expects a central template matching process to be indifferent to the way that the pitches are generated.

Although it was found that both the central template model and the timing model can survive the experimental result, it is possible to make improvements on the original formulations of both. Section V showed that in the creation of a central template from the excitation patterns of complex tones, the auditory system must differentiate by some means between excitations caused by different spectral components. If the system does not differentiate then the central template matching process leads to an octave contraction and not to an enlargement. A plausible basis for differentiation is neural timing, whereupon the excitation pattern model is no longer entirely a place theory.

The timing model can be improved too. The original formulation suggested that the octave enlargement could be understood if the auditory system derived pitch from the ISI histogram. A suggested revision is to replace the ISI histogram by an autocorrelation function. Although these two functions are so similar that they are sometimes regarded as equivalent, they give quite different results in the calculation of octave matching. An ISI neural circuit predicts no octave enlargement or an octave contraction. The neural autocorrelator, by contrast, is quite capable of predicting an octave enlargement in agreement with exper-

iment. It seems possible that the octave enlargement phenomenon is the only efficient discriminator between these two models.

## ACKNOWLEDGMENTS

This article benefited from helpful discussions with Dr. Edward Burns, Dr. Alain de Cheveigne, Dr. Laurent Demany, Ms. Mitra Hartmann, and Dr. W. Dixon Ward. Dr. Kengo Ohgushi and Dr. Ernst Terhardt made useful comments on an initial version.

## APPENDIX: MATCHING THE HUGGINS PITCH

The success of the main experiment in this paper depends upon the ability of listeners to make precise judgments of the Huggins pitch. This ability was tested in a pitch matching experiment which required a listener to adjust the frequency of a sine tone to match Huggins pitches. The center frequencies of the phase-difference region were similar to those of the main experiment,  $f_c = 300, 400, 500, 600, \text{ and } 700 \text{ Hz}$ .

Stimuli were presented in a four-interval sequence: (1) Dichotic noise with the phase difference and Huggins pitch; (2) diotic noise; (3) diotic noise plus matching sine tone; (4) diotic noise. Each interval was 300-ms long; there were no gaps, and the sequence was repeated indefinitely until the listener was satisfied with the match. The listener could adjust both the frequency and the level of the matching tone. An experimental run consisted of a single match for each of the five center frequencies. Each listener did seven runs.

The matching sine tones adjusted by the listener were invariably close to the phase-difference center frequencies in the dichotic noise. Standard deviations over the seven matches are given in Table AI in units of cents.

The table shows that matching reproducibility was quite good. The average standard deviation of 8.6 cents corresponds to 0.5%. The largest of all the standard deviations, 19.6 cents, is 1.1%, smaller than the second step in the constant stimuli experiment,  $608/600 - 1 = 1.3\%$ . The conclusion of the matching experiment is that one can expect meaningful data from the main experiment.

TABLE AI. Standard deviations (units of cents) for matching the Huggins pitch. For five different phase-difference center frequencies  $f_c$ , six listeners made seven matches of a sine tone to the Huggins pitch. The standard deviations of the matches ( $N-1=6$  weight) are given in the table. (Note: To convert cents to percent, multiply by 0.0578.)

$f_c$ (Hz)	Listener						Mean
	A	B	C	J	M	P	
300	3.7	8.4	11.3	6.2	13.3	...	8.6
400	6.1	3.5	4.1	5.4	6.0	2.9	4.7
500	7.6	4.2	5.1	13.2	16.8	10.0	9.5
600	6.6	5.3	6.3	10.3	19.6	8.0	9.4
700	5.5	17.7	7.9	9.9	17.5	8.6	11.1
Mean	5.9	7.8	6.9	9.0	14.6	7.4	8.6

<sup>1</sup>Huggins pitch and cyclopean perception both have spatial attributes. The tonal sensation of Huggins pitch is lateralized to one side of the head because of the interaural phase difference. The cyclopean images are seen in depth, either behind or in front of the random-dot background.

<sup>2</sup>The lowest sampling rate occurred for the 300-Hz range and a random scale factor of 0.95, namely  $0.95 \times 8192 = 7782$  Hz. The highest sampling rate occurred for the 550-Hz range and a random scale factor of 1.05, namely  $550/300 \times 1.05 \times 8192 = 15767$  Hz.

<sup>3</sup>The central template theory introduced by Terhardt's work is actually not specific concerning pitch mechanisms. It is rather a set of formulas that generalize relationships between different pitch-related psychoacoustical measurements. The place-theory interpretation given in the present paper is the physiological model that is consistent with the formulas and is popular among auditory theorists.

<sup>4</sup>If auditory scientists are going to make further progress on timing models of perception then we need to know much more about the timing in eighth-nerve fibers, measured by ISI histograms and autocorrelation functions if not higher-order correlation. We need precise measurements of these functions with attention paid to count distributions within the individual peaks and the dependencies of these distributions on stimulus frequency, stimulus intensity, and the presence of other spectral components, also dependencies on fiber best frequency (*CF*), spontaneous rate, and threshold. We need to know the variability in these distributions among fibers of similar characteristics.

Békésy, Georg von (1960). *Experiments in Hearing* (McGraw-Hill, New York), reprinted by the Acoustical Society of America.

Bilsen, F. A. (1977). "Pitch of noise signals: Evidence for a 'central spectrum,'" *J. Acoust. Soc. Am.* **61**, 150–161.

Brink, G. van den (1977). "Octave and fifth settings for pure tones and residue sounds," in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 373–379.

Burns, E. M. (1982). "Pure tone pitch anomalies. I. Pitch-intensity effects and diplacusis in normal ears," *J. Acoust. Soc. Am.* **72**, 1394–1402.

Burns, E. M., and Ward, W. D. (1982). "Intervals, scales and tuning," in *The Psychology of Music*, edited by D. Deutsch (Academic, New York), pp. 241–269.

Cramer, E. M., and Huggins, W. H. (1958). "Creation of pitch through binaural interaction," *J. Acoust. Soc. Am.* **30**, 413–417.

Demany, L., and Armand, F. (1984). "The perceptual reality of tone chroma in early infancy," *J. Acoust. Soc. Am.* **76**, 57–66.

Demany, L., and Semal, C. (1990). "Harmonic and melodic octave templates," *J. Acoust. Soc. Am.* **88**, 2126–2135.

Dobbins, P. A., and Cuddy, L. L. (1982). "Octave discrimination: An experimental confirmation of the 'stretched' subjective octave," *J. Acoust. Soc. Am.* **72**, 411–415.

Durlach, N. I. (1962). "Note on the creation of pitch through binaural interaction," *J. Acoust. Soc. Am.* **34**, 1096–1099.

Egan, J. P., and Meyer, D. R. (1950). "Changes in pitch of tones of low frequency as a function of the pattern of excitation produced by a band of noise," *J. Acoust. Soc. Am.* **22**, 827–833.

Green, D. M. (1977). (Private communication).

Guttman, N. (1962). "Pitch and loudness of a binaural subjective tone," *J. Acoust. Soc. Am.* **34**, 1996(A).

Hartmann, W. M. (1993). "Auditory demonstrations on compact disc for large *N*," *J. Acoust. Soc. Am.* **93**, 1–17.

Houtsma, A. J. M. (1981). "Noise induced shifts in the pitch of pure and complex tones," *J. Acoust. Soc. Am.* **70**, 1661–1668.

Jones, K., Tubis, A., and Burns, E. M. (1983). "Temporal neural response correlates of pitch-intensity effects and diplacusis," *J. Acoust. Soc. Am.* **74**, S8.

Julesz, B. (1971). *Foundations of Cyclopean Perception* (Univ. of Chicago, Chicago).

Kohonen, T. (1978). *Associative Memory: A System-theoretical Approach* (Springer-Verlag, Berlin).

Krumhansl, C., and Shepard, R. N. (1979). "Quantification of the hierarchy of tonal functions within a diatonic context," *J. Exp. Psychol., Human Percept. Perform.* **5**, 579–594.

Ohgushi, K. (1978). "On the role of spatial and temporal cues in the perception of the pitch of complex sounds," *J. Acoust. Soc. Am.* **64**, 764–771.

Ohgushi, K. (1983). "The origin of tonality and a possible explanation of the octave enlargement phenomenon," *J. Acoust. Soc. Am.* **73**, 1694–1700.

Peters, R. W., Moore, B. C. J., and Glasberg, B. R. (1983). "Pitch of components of a complex tone," *J. Acoust. Soc. Am.* **73**, 924–929.

Raatgever, J., and Bilsen, F. A. (1986). "A central spectrum theory of binaural processing. Evidence from dichotic pitch," *J. Acoust. Soc. Am.* **80**, 429–441.

Rakowski, A., and Miskiewicz, A. (1985). "Deviations from equal temperament in tuning isolated musical intervals," *Arch. Acoust.* **10**, 95–104.

Srulovicz P., and Goldstein, J. L. (1983). "A central spectrum model: A synthesis of auditory-nerve timing and place cues in monoaural communication of frequency spectrum," *J. Acoust. Soc. Am.* **73**, 1266–1276.

Sundberg, J. E. F., and Lindqvist, J. (1973). "Musical octaves and pitch," *J. Acoust. Soc. Am.* **54**, 922–929.

Terhardt, E. (1969/70). "Oktavspreizung und Tonhöhenverschiebung bei Sinustönen," *Acustica* **22**, 345–351.

Terhardt, E. (1971a). "Die Tonhöhe Harmonischer Klänge und das Oktavintervall," *Acustica* **24**, 126–136.

Terhardt, E. (1971b). "Pitch shifts of harmonics, an explanation of the octave enlargement phenomenon," *Proceedings of the Seventh International Congress on Acoustics, Akademiai Kiado, Budapest*, pp. 621–624.

Terhardt, E. (1974). "Pitch, consonance, and harmony," *J. Acoust. Soc. Am.* **55**, 1062–1069.

Terhardt, E., and Fastl, H. (1971). "Zum Einfluss von Störtönen und Störgeräuschen auf die Tonhöhe von Sinustönen," *Acustica* **25**, 53–61.

Terhardt, E., Stoll, G., and Seewann, M. (1982). "Algorithm for extraction of pitch and pitch salience from complex tonal signals," *J. Acoust. Soc. Am.* **71**, 679–688.

Turner, C., Burns, E. M., and Nelson, D. A. (1983). "Pure tone pitch perception and low-frequency hearing loss," *J. Acoust. Soc. Am.* **73**, 966–975.

Walliser, K. (1969). "Über die Spreizung von empfundenen gegenüber mathematisch harmonischen Intervallen bei Sinustönen," *Frequenz* **23**, 139–143.

Ward, W. D. (1954). "Subjective musical pitch," *J. Acoust. Soc. Am.* **26**, 369–380.

Ward, W. D. (1963). "Diplacusis and auditory theory," *J. Acoust. Soc. Am.* **35**, 1746–1747.

Zwicker, E., and Fastl, H. (1990). *Psychoacoustics* (Springer-Verlag, Berlin).