HALL,III, JOSEPH WALTON. On the Encoding of Pitch and the Quantitative Measurement of Pitch Strength. (1974) Directed by: Dr. David R. Soderquist. pp.49

The purposes of this study were (a) to examine the level, peripheral or central, at which complex stimuli are encoded or analyzed, and (b) to obtain quantitative measures of the strength of complex stimuli. Theories of complex pitch perception (especially Wightman's 1973b. The pattern transformation model of pitch. J. Acous. Soc. Amer., _54_, 407-416) bearing on the physiological level of pitch encoding and the strength of complex stimuli were reviewed, and a number of complex stimulus presentation methods that would test the predictions of these theories were described.

Peripheral vs. central pitch analysis was examined in binaural conditions in which complex tones having different fundamentals were presented to separate ears. The stimuli were selected such that taken _together_, they had the same fundamental. If two pitches (each corresponding to the fundamental of the complex at each ear) were heard, evidence for peripheral coding would be supplied. If one pitch (corresponding to the fundamental of both complexes taken together) were heard, evidence for central coding would be supplied.

Pitch strength was measured by three dependent variables: amount of noise required to mask a complex pitch;

intensity of a comparison tone required to correspond to the intensity of a complex tone; and the variability of successive pitch value matches of a complex tone. Pitch strength was examined in terms of (a) spectral region of component stimuli (1200-Hz vs. 2000 Hz), (b) number of components comprising a complex tone (three vs. five), and (c) spacing between complex components (200 Hz vs. 400 Hz).

Data were collected and analyzed in a repeated measures design. Two pitches were reported in the binaural conditions, so a peripheral encoding hypothesis was supported. The noise intensity measure of pitch strength indicated that the 1200-Hz spectral region yielded stronger pitches than the 2000-Hz region; complexes with five components were stronger than those with three components; and complex tones with 400 Hz-spacing were stronger than those with 200 Hz-spacing. The other measures of pitch strength did not yield significance but revealed trends that were in the same direction as that for the noise variable.

It was concluded that while pitch analysis may be a central mechanism, information from each ear may be kept separate and be analyzed independently. It was further concluded that pitch strength can be quantified, and seems to depend upon spectral region, number of components in a complex, and spacing between complex components, as theorized by Wightman (1973b).

ON THE ENCODING OF PITCH AND THE QUANTITATIVE MEASUREMENT OF PITCH STRENGTH

by

Joseph W. Hall, III

A Thesis Submitted to
the Faculty of the Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Arts

Greensboro
April, 1974

Approved by

D. R. Soderquist
Thesis Advisor

APPROVAL SHEET

This thesis has been approved by the following committee of the Faculty of the Graduate School at the University of North Carolina at Greensboro.

Thesis Advisor D. R. Soderquist

Oral Examination Committee Members Kendon Smith

D. R. Soderquist

Walter Salinger

April 15, 1974
Date of Examination

ACKNOWLEDGMENTS

The author wishes to thank Dr. Kendon Smith, Dr. Walter Salinger, and Dr. David Soderquist for serving on the Thesis committee.

Special gratitude is due Dr. David Soderquist for his guidance, criticsm, and willingness to devote his time to this undertaking.

Thanks is also extended to Tali Hughes, Don Hall, and Mary Hall, all of whom served as subjects.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

A longstanding task in the area of sensory psychology has been the development of an adequate theory of pitch perception. If the adequacy of a theory of pitch perception is judged in terms of degree of experimental data "explained" by the theory, and degree of consonance between the theory and auditory physiology , it is clear that no theory so far advanced can be termed completely adequate. Although no adequate theory of pitch perception has been established, it is nevertheless apparent that several theories have remained popular over a period of many years. Obviously some theories do account for or explain much of the data, and supporters of such theories usually consider them largely correct. Furthermore, theories that seem to account for large portions of the data are reluctantly abandoned when it is possible that they may be modified to accord with data that are initially discrepant with theory, or if it is possible to show that discrepant data are due to some kind of experimental error.

A theory that has retained some degree of popularity for over 100 years has been one form or another of the place theory proposed by Helmholtz in 1863. Basically, Helmholtz held that (a) the inner ear analyzed sound by means of a system of resonators differentially sensitive to frequency, with low and high resonators occupying places on the basilar membrane which were apical and basal, respectively, and (b) the part of the basilar membrane stimulated determined what pitch was heard. Stimulation of basal resonators resulted in the perception of high pitch, and stimulation of apical resonators resulted in the perception of low pitch. Tenet (a), which characterizes the basilar membrane as a kind of Fourier analyzer of acoustical stimuli, has received qualified support. That is, it is believed that the cochlea can "pick out" various frequencies which are presented together in a complex sound wave. The resolution of a particular frequency by a particular section of the basilar membrane seems to be far less refined than was originally conceptualized by Helmholtz, however; viz., a broad portion of the basilar membrane is placed into movement by a simple sinusoid, rather than the selective movement of a small number of finely tuned reonators. Tenet (b) has received much less general support. This tenet is an extension of Mueller's law of specific energies; that is, each place on the basilar membrane is assumed to be neurally connected to a central (psychological) pitch extractor in such a way that a certain place of basilar vibration results in the perception of a certain pitch. The most serious evidence against tenet (b) stems from research dealing with complex tones. The complex tones of interest have been those composed of several harmonically related sinusoids, the frequency difference between any sinusoid and its closest neighbors being the same for all components of the complex. It happens that the pitch of such a complex tone is nearly always the same as the frequency of the fundamental (lowest possible component of the complex), even when the fundamental is not physically present in the complex waveform. With the fundamental absent or missing, it is difficult to imagine, by the place theory, how one could perceive the pitch of the fundamental when, in fact, the area of the cochlea corresponding to the fundamental frequency is presumably unstimulated. Place theorists have tried to defend their position by positing that the perception of the "missing" fundamental may be caused by the reintroduction of the fundamental on the basilar membrane by means of nonlinear distortion within the cochlea. The fact that the fundamental is still heard even when that portion of the basilar membrane which corresponds to the fundamental frequency is saturated with white noise (thus preventing the perception of the "reintroduced" fundamental through masking) places tenet (b) of the place theory in serious doubt (Licklider, 1956). While it is possible that place principles may determine pitch for simple tones and some complex tones, it has been made clear that a theory based solely on place cannot account for all pitch perception data.

The most popular candidate to supplement or supplant place theory has been one version or another of a theory based on the periodicity of the acoustic stimulus. The earliest periodicity theories were founded upon the premise that acoustic stimuli were represented by peripheral neural firing, and that the periodic rate of neural firing determined pitch (Rutherford, 1893; Wundt, 1893). In its simplest form, periodicity theory holds that the cycles per second of an acoustic stimulus are represented by a one-to-one neural firing rate. The major line of resistance against this theory has been based on the fact that the rates of neural firing required for the perception of higher pitches (e.g., 800 to 20,000 Hz) are physiologically incapable of being generated. Estimates of ceiling rates for individual neuronal firing have ranged from 50 impulses per second (Fletcher, 1923) to 2000 per second (Wever, 1949). Wever (1949) suggested a principle which made a theory based on periodicity more tenable. He supposed that pitch may be coded by rate of firing of single neurons, up to frequencies of about 500 Hz, whereas frequencies from about 500 Hz to about 4000 Hz could be coded by the synchronized firing of coordinated groups of neurons responding in volleys. The volley principle can be exemplified by considering two groups of neurons each firing at a rate of 400 impulses per second. If each group fires when the other group is exactly midway between impulses, it can be seen that taken together the two groups fire at a rate of 800 impulses per second. Wever suggested that for frequencies above 4000 Hz, encoding by the volley principle was not physiologically possible and suggested that frequencies above this value were encoded by the place principle.

In the past 30 years there has been a shift in interest from pitch perception resulting from stimulation by simple sinusoids to pitch perception resulting from controlled but more complex stimuli. The interest in complex stimuli stems largely from the fact that the perception of the "missing" fundamental with such stimuli is at once damaging to the place theory and challenging for alternative theories. Recent experimentation involving the pitch of complex tones has led many theorists to believe that the periodicity of such tones may substantially, if not wholly, underlie the pitch perception of complex stimuli. The present investigation is devoted to the concept of periodicity pitch and to the neural location of pitch encoding and extracting processes. Some of the more recent developments in periodicity theory will now be reviewed.

Periodicity Pitch

Schouten (1940) performed experiments on the perception of harmonically related complex stimuli and noted (a) the lower harmonics which are presented together can be perceived individually and have almost the same pitch as when sounded separately, and (b) higher harmonics which are presented together are perceived collectively as one pitch called the "residue". Further, the residue has a pitch equivalent to the fundamental of the complex stimulus. In order to account for his observations, Schouten proposed that the complex stimulus is first Fourier analysed by the basilar membrane, as in tenet (a) of the place theory, with the lower harmonics resolved well and perceived separately. However, because the cochlea is limited in its ability to resolve higher harmonics, these components are coded in terms of their collective temporal activity and are perceived as a residue. The pitch of the residue, then, was thought to be peripherally coded as the collective, periodic firing of the first order neurons excited by upper harmonic components. Further, it was proposed that the residue was determined by the periodicity of the collective waveform of the physical stimulus.

Schouten's pioneering work indicated that the pitch of the residue corresponded to the frequency of the modulation of the waveform envelope (see Figure 1). Thus a complex

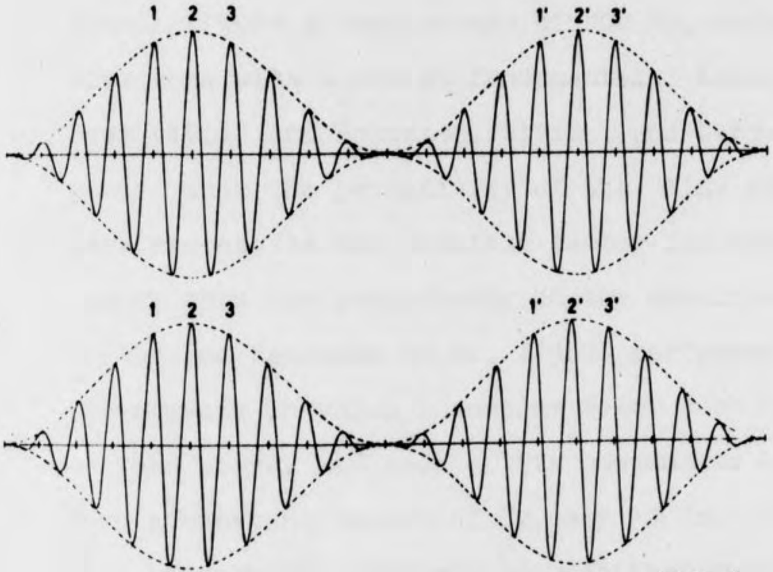

Figure 1. Envelopes and fine structures of shifted and unshifted waveforms. The dashed lines represent the envelope of a complex waveform. Top waveform is a complex made up of 1800, 2000, and 2200 Hz components. Bottom waveform is made up of the same components, but with each shifted up 30 Hz. Fine structure is represented by the wave patterns inside the envelopes. Numbers over the fine structure peaks indicate peaks likely to be picked or used in pitch analysis (adapted from Plomp, 1968).

tone with components of 1800, 2000, and 2200 Hz would result in the perception of a pitch of 200 Hz, since positive peaks occur in the waveform envelope just as if the stimulus were a simple tone of 200 Hz, rather than a complex tone with a 200 Hz fundamental. Later, however, De Boer (1956) and Schouten, Ritsma, and Cardozo (1962) suggested that the periodicity of the "fine structure" of the waveform may be the critical factor for pitch encoding, rather than the periodicity of the envelope. De Boer (1956) and Schouten et al. (1962) performed pitch-shift experiments in which a complex tone, such as the one described above, had each of its components shifted up or down a constant number of Hz, say 30 Hz. The only difference between the shifted and unshifted complex waveforms is the fine structures of the two waveforms. The envelopes of the two stimuli are identical. Thus, for both shifted and unshifted complexes, the envelope of the waveform has positive peaks every 5 msec. The fine structures within the envelopes of shifted and unshifted complexes differ, however; the "temporal distances" between the fine structure peaks are relatively shorter for the complex that has been shifted up 30 Hz. A fine structure theory of pitch perception maintains that pitch is determined by the temporal distance between two fine structure peaks in adjacent waveform envelopes. Further, the peaks most likely to be "picked" or used by a psychological pitch analyser are the "tallest" peaks, or those with the greatest amplitude (see Figure 1). Results of the pitch shift experiments showed that complexes which are shifted up in frequency are also shifted up in perceived pitch, thus supporting a fine structure theory. As the fine structure theory holds that pitch is determined by the temporal distance between peaks near the crests of adjacent waveform envelopes, some pitch ambiguity would be predicted. This is because a psychological pitch analyser may choose any of the fine structure peaks near the crest of an envelope for analysis (in Figure 1 these peaks are identified as 1, 2, 3, and 1', 2', and 3'). Therefore, different peak-to-peak time differences may be used for analysis. Pitch ambiguity for complex waveforms such as that represented in Figure 1 has been found experimentally, thus further supporting a fine structure theory of pitch (Schouten et al., 1962).

Although Schouten's residue-fine structure theory has enjoyed wide and growing support since it was postulated, recent objections have been lodged against it by Wightman (1973a) and Houtsma and Goldstein (1972). Wightman contended that the weakest point of residue theory is the supposition that the pitch of complex stimuli is determined by the <u>interacting</u> <u>unresolved</u> components. By this supposition, Wightman maintains, complex tones composed of very high (and therefore unresolved) components should result in a relatively strong residue pitch. Conversely, complex tones composed of lower component frequencies (and therefore better resolved) should result in a relatively weaker residue pitch (pitch strength can be thought of as the clarity or perceptibility of a pitch). Reports from listeners, however, indicate that a stronger residue pitch is heard for complexes with lower harmonic components; further, if component frequencies exceed 5000 Hz, no residue pitch is heard at all (Ritsma, 1962). So, although residue theory states that residue pitch is determined by the interaction of unresolved components, residue pitch is strongest for complexes whose components are relatively well resolved.

A second objection which Wightman raised against the residue theory (and all other theories based upon the temporal fine structure of complex waveforms) concerns the fact that such theories are generally phase sensitive. A change in the phase relations of components will result in a change in temporal fine structure which, in turn, should result in a change in perceived pitch. Recent research by Patterson (1973), Bilsen (1973) and Wightman (1973a) indicates that the value of perceived pitch is actually phase insensitive, however.

Another line of evidence against the residue theory comes from an experiment by Houtsma and Goldstein (1972) in which components of complex stimuli were presented to Ss binaurally. All stimuli were composed of only two components. As an example, a 1000-Hz sinusoid was presented to the right ear, and a 800-Hz sinusoid was simultaneously presented to the left ear. A pitch perceived was that of the fundamental (200 Hz), just as though a complex stimulus with components of 1000 and 800 Hz had been presented monaurally. Thus it was argued that the residue pitch was not coded and determined by the interaction of unresolved components at the level of the cochlea, as the residue theory suggested. Houtsma and Goldstein argued that, instead, pitch was determined more centrally. They held that components are resolved at the peripheral level, but that this peripheral resolution and encoding does not immediately determine pitch. Pitch determination or extraction is the end result of a central analysis on all afferent information (that is, information from both cochleae). Thus Houtsma and Goldstein attack the peripheral encoding foundation of residue theory.

Wightman (1973b) has recently proposed a mathematical model of pitch perception which holds that pitch is independent of the temporal fine structure of complex stimuli, and which is compatible with central mediation of pitch.

The model and some of its predictions will be very briefly outlined. Essentially, Wightman invokes both place and periodicity principles. There are three stages in his model. At Stage 1 the power spectrum of the acoustic waveform is coded as a Peripheral Activity Pattern (PAP). Here frequency is coded as place in the periphery of the auditory system, and power is coded as the amount of activity at each place. Taking into account the limited resolving power of the cochlea, the model proposes that the Stage 1 coding is coarse, especially for high-frequency representation. In Stage 2 the PAP of Stage 1 is autocorrelated (a Fourier cosine transform of the power spectrum is performed). After this transformation, place of neural activity corresponds to the temporal dimension of autocorrelation , and the amount of neural activity at a particular place corresponds to the value or magnitude of the autocorrelation. According to Wightman, all stimuli which result in a relatively large amount of neural activity at the same Stage 2 place will result in the same pitch perception. Therefore the temporal dimension of Stage 2 autocorrelation determines pitch and is coded as neural place. The autocorrelation value at each neural place determines the strength of pitch. If two or more Stage 2 places are neurally active an explanation for pitch ambiguity becomes apparent. Places having the most neural activity are thought to have strong pitch values, and therefore are perceived most clearly. If the absolute difference of amount of neural activity between two places is not great, however, more than one pitch may be perceived. The final stage of Wightman's model can be conceived of as the decoding of Stage 2 place activity into perceived pitch. The pitch is taken as the inverse of the autocorrelation time value associated with the place of greatest neural activity.

Wightman's model makes several predictions about the strength of complex stimuli. In the literature previous to Wightman's (1973b) paper there are only qualitative reports on the strength of complex stimuli. Typically a complex stimulus has been reported as strong or weak in relation to another complex stimulus on the basis of subjective report of a listener. Wightman's mathematical model generates predictions which invite quantitative measures of pitch strength as a test of his theory. Some of these predictions are summarized in Figures 2 and 3. Figure 2 shows that as the number of components in a complex stimulus increases, strength also increases. Figure 3 shows that as spacing between components increases , strength also increases. Both Figures 2 and 3 indicate that pitch strength decreases as the frequency of the lowest component of a complex increases. It should be noted that the decrease in pitch strength predicted for higher-frequency components

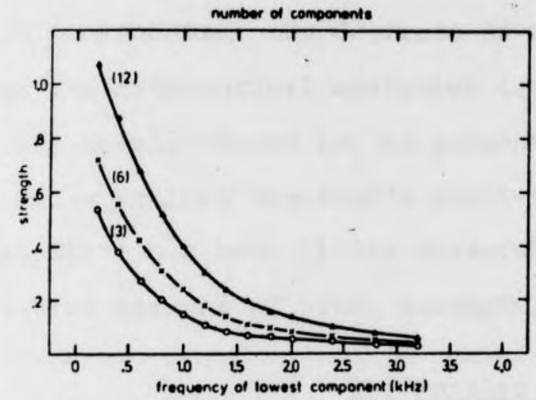

Figure 2. Predicted variation of pitch strength with regard to number of components per complex, and spectral region. Number in parentheses represents number of components per complex (adapted from Wightman, 1973b).

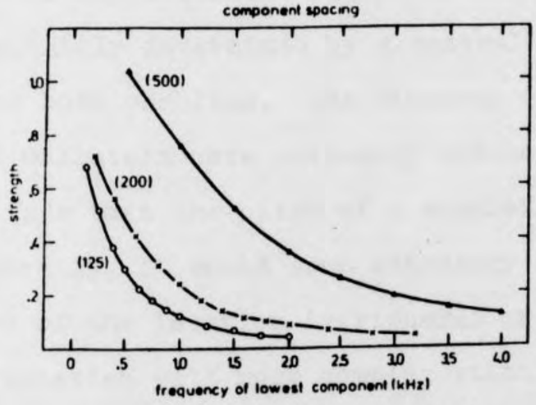

Figure 3. Predicted variation of pitch strength with regard to spacing of components in complex, and spectral region. Number in parentheses represents spacing (in Hz) between components (adapted from Wightman, 1973b).

is not merely a function of relatively poor resolution of high components. The decrease is directly predicted from the mathematical operation (autocorrelation) proposed by the model. There is, at present, little support for or evidence against Wightman's predictions due to the fact that there has been little research directed at the quantitative measure of pitch strength.

The Problem

The goals of the present experiment were twofold. The first goal concerned the neural level at which encoding of acoustical stimuli determines perceived pitch. The question under investigation was whether the neural activity at peripheral levels (the cochlea or eigth cranial nerve) may determine pitch, or whether coding of pitch is ultimately determined by a central mediation of information from both cochleae. The binaural stimuli used by Houtsma and Goldstein were extremely simple. Before the statement is made that the pitch of a complex stimulus is determined centrally, it would seem necessary to investigate the question of the location (peripheral or central) of pitch determination with more complex stimuli. While the Houtsma and Goldstein experiment strongly suggested that the pitch of complex tones may be mediated centrally, it is argued here that the stimuli used did not actually make a fair test of whether a complex stimulus may also be coded for pitch at a more peripheral level. This argument rests on the fact that although complex stimuli were presented binaurally, the stimuli that were presented to one ear or the other were single sinusoids. If the pitch of a complex stimulus can somehow be encoded and determined at a peripheral level, the Houtsma and Goldstein stimuli did not provide an adequate test for such encoding. A better test of the peripheral encoding hypothesis would involve the presentation of a strong complex stimulus (at least three components) to each ear simultaneously. The complex stimuli should have different fundamentals, but should be chosen in such a way that when "mixed together", or presented as a single complex stimulus, components should be harmonically spaced so that there is only one fundamental. Complexes of 1800, 2000, and 2200 Hz (fundamental of 200 Hz), and 1600, 2000, and 2400 Hz (fundamental of 400 Hz) would fill these requirements. The monaural presentation of each three-component complex stimulus should result in a different pitch perception. The question of interest here is what perception will occur when one complex is presented to the right ear, and the other complex is presented to the left ear. A purely central theory would predict that all of the components would be represented centrally and that a pitch of 200 Hz would be heard because the components are 1600, 1800, 2000, 2200, and 2400 Hz (fundamental of 200

Hz). A peripheral encoding theory would predict that there are two separate complex stimuli encoded at the two peripheral levels, one of which corresponds to a pitch of 200 Hz, and the other of which corresponds to a pitch of 400 Hz. Thus if the pitch of a complex stimulus can be peripherally coded and if encoding at each cochlea can be kept "separate", two pitches should be heard.

The second second goal of the experiment involved the quantitative measurement of pitch strength with regard to the predictions made by Wightman (Figures 2 and 3). To this end, pitch strength was measured in terms of report of loudness of pitch and variability of successive matches of pitch value. These measures were chosen because pitch strength is usually characterized as clarity or the quality that distinguishes pitch value from background noise. Thus Ss may be thought to report that the loudness of a weak stimulus is relatively low, while the loudness of a strong pitch is relatively high. Pitch strength was measured with respect to (a) spectral region of components, (b) number of components per complex stimulus, and (c) spacing between the components of a complex stimulus. Pitch strength results will be discussed largely in terms of their consonance with the theoretical predictions made by Wightman (1973b).

METHOD

Design

A 2X5 repeated measures design was used to investigate the influence of factors (a) spectral region and (b) composition and presentation method, on the perception of complex stimuli. One of the spectral regions investigated was centered at 2000 Hz (components in this region being 1600, 1800, 2000, 2200, and 2400 Hz); the other spectral region was centered at 1200 Hz (components in this region being 800, 1000, 1200, 1400,and 1600 Hz). At each of these spectral regions Ss were subjected to five listening conditions (see Table 1). In each condition the stimulus to which a S was to attend was presented to his "best" ear. The "best" ear was the more sensitive ear of a subject. The conditions were: (a) a monaural condition in which three components were presented, one of which was the central component of a spectral region, and the other two of which were spaced 200 Hz above and below the central component (condition M-3 200); (b) a monaural condition in which three components were presented, one of which was the central component of a spectral region, and the other two of which were spaced 400 Hz above and below the central component (M-3 400); (c) a monaural condition in which all of

TABLE 1

REPRESENTATION OF EXPERIMENTAL DESIGN

Spectral Region	1200 Hz					2000 Hz				
Presentation Condition	M-3 200	M-3 400	M-5	B-3 200	B-3 400	M-3 200	M-3 400	M-5	B-3 200	B-3 400
$\underline{S}_1$										
$\underline{S}_2$										
$\underline{S}_3$										

the components of a spectral region were presented (M-5); (d) a binaural condition in which the 200-Hz-spaced components around the central frequency were presented to the "best" ear, and the 400-Hz-spaced components around the central frequency were presented to the other ear (B-3 200); and (e) a binaural condition in which the 400-Hz-spaced components around the central frequency were presented to the "best" ear, and the 200- Hz-spaced components were presented to the other ear (B-3 400). The dependent variables were perceived pitch value (in Hz), and pitch strength. There were three measures of pitch strength. One was the intensity to which S adjusted a comparison tone to match the perceived intensity of a complex tone. The second measure was the intensity of white noise required to mask the pitch of a complex tone. The third measure was the standard deviation of successive pitch value matches of the same complex tone. It was thought that standard deviations would be relatively smaller for strong pitches, and that noise and comparison intensities would be larger for strong pitches.

Complex Stimuli

All complex stimuli consisted of 100% amplitude modulated sinusoids recorded on tape. The following stimuli were used and recorded separately: 2000 Hz modulated by 200 Hz (components of 1800, 2000, and 2200 Hz); 2000 Hz modulated by 400 Hz (components of 1600, 2000, and 2400 Hz); 1200 Hz modulated by 200 Hz (components of 1000, 1200, and 1400 Hz); and 1200 Hz modulated by 400 Hz (components of 800, 1200, and 1600 Hz). Stimuli for the M-5 conditions were generated by amplitude modulating the 2000 Hz carrier with 200 and 400 Hz simultaneously (components of 1600, 1800, 2000, 2200, and 2400 Hz), and modulating the 1200 Hz carrier with 200 and 400 Hz simultaneously (components of 800, 1000, 1200, 1400, and 1600 Hz).

Procedure

Three adult Ss, two female, one male, were used. The the amount of musical training to which each S had been exposed differed: subject D.H. had eleven years of musical training with the trumpet; T.H. had five years of musical training with the piano; M.H. had no musical training. The Ss sat in a sound-attenuated chamber and matched the pitch of one of the complexes formerly defined, to the pitch of a square wave from a function generator. A square wave was used because the timbre of such a wave is much like that of the complex stimuli used. If two stimuli are to be matched for pitch, matching is facilitated if the stimuli are similar in timbre. All stimuli were presented through a pair of calibrated earphones. The comparison square wave was presented monaurally, to the best ear, in all conditions, as was the white noise when masking was performed. The Ss had at their disposal (a) a switch that could select either the complex stimulus or the comparison stimulus, (b) knobs that controlled the intensity and frequency of the comparison tone, and (c) an attenuator knob that controlled the intensity of noise.. The knobs that controlled the intensity and freuqency of the comparison were visually hidden from Ss by a partition. The partition had an aperature large enough so that S could put his hand through and adjust knobs. The purpose of the partition was to avoid errors of habituation due to visual cues. The knob of the attenuator that controlled noise intensity was in full view of S, but errors of habituation were controlled for since E randomly changed the voltage at S's attenuator before each trial.

At the beginning of each trial, the noise was turned off and S was told to switch to the complex stimulus and report the number of pitches heard. If two pitches were heard in the binaural conditions, S was instructed to match just one throughout the entire session. At the start of a trial, E instructed S to turn the intensity of the comparison tone to zero and the frequency dial fully clockwise or counterclockwise, alternating direction on successive trials. The alternation of dial position over trials was intended to reduce errors of habituation due to kinesthetic cues. The S's task was to listen to the complex and comparison tones alternately and adjust the intensity and frequency of the latter tone to match the loudness and pitch of the complex. The S was allowed to switch freely from the comparison to the complex and to vary both the intensity and frequency of the comparison tone continually, until satisfied with his pitch and loudness matches. When the matches had been made, E recorded the frequency of the comparison tone (in Hz) from a frequency counter, and the intensity of the comparison (in volts) from an RMS voltmeter. The masking noise was then turned on, and S was instructed to mask the pitch of the complex signal. If S heard two pitches he was to attend to and mask only the pitch which he had just matched, and to "ignore" the other pitch. The S was also told to "bracket" or vary the noise intensity above and below the point at which the pitch seemed to be masked until he obtained the minimum noise intensity required to just mask the pitch. When the noise masking was completed, the intensity of the noise (in volts) was recorded from an RMS voltmeter. This completed one trial.

The 10 experimental conditions were randomized over Ss with each session comprising one of the 10 conditions. Each session consisted of 20 trials with a five minute rest midway through the session.

Training

Each S was trained in two one-hour sessions. During the first session Ss performed monaural matches with a complex stimulus having a carrier of 1200 Hz, and aquainted themselves with the operation of the function generator and noise attenuator. In the second session, Ss performed monaural matches to a complex stimulus having a 2000-Hz carrier, and binaural matches at both the 1200- and 2000-Hz spectral regions. During training Ss were given guidance in avoiding octave errors when pitch matching.

Apparatus

A diagram of the apparatus is presented in Figure 4. The masking noise in this experiment was generated by a Grason-Stadler white noise generator. The noise band width, 10-6000 Hz, was determined by a 3100 Krohn-Hite band pass filter. Noise intensity was controlled by two Hewlett-Packard 350-D attenuators. The square wave comparison tone was generated by a Phillips PM 5168 function generator. The noise and comparison tone were mixed by a solid state Coulbourn Audio Mixer-Amplifier (model 582-24), and the complex tone and the output from the Mixer-Amplefier were matched to TDH39 earphones (circumaural muffs) by a Grason-Stadler E10589A impedance matching transformer. In the binaural conditions, the complex stimulus presented to the best ear was mixed with the noise and comparison tone prior

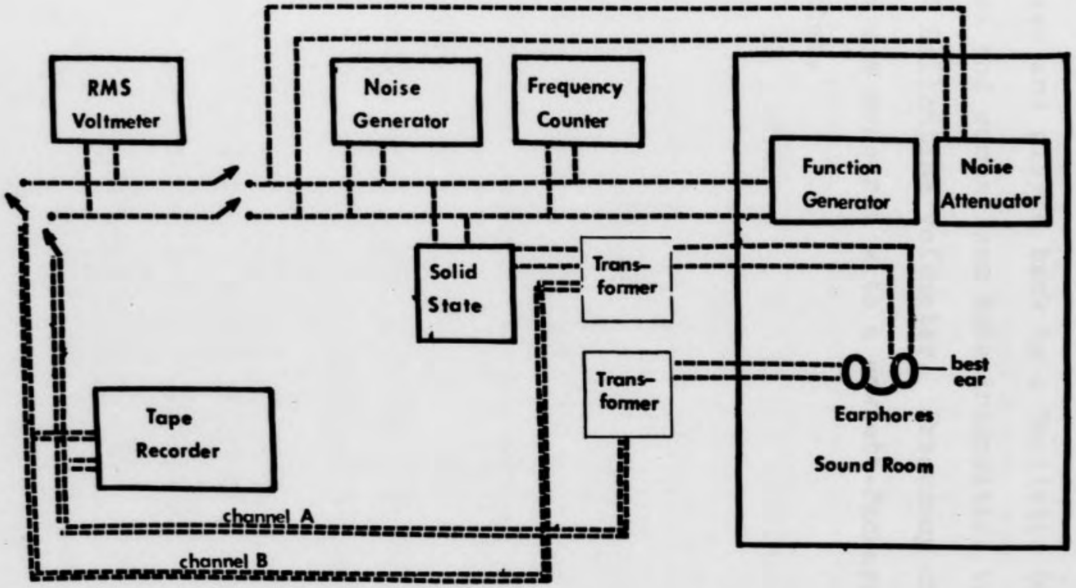

Figure 4. A schematic diagram of experimental apparatus.

to the transformer. The other complex went to another transformer and the other ear. The complex tones were recorded and played back by a Realistic 909A tape recorder. Noise and comparison tone intensities were measured with a 320A Ballantine voltmeter. Frequency of the comparison tone was measured with a Hewlett-Packard 5221B frequency counter.

RESULTS

The most striking and unambiguous result was that Ss heard two pitches (200 Hz and 400 Hz) in the binaural conditions. The histograms in Figures 5, 6, and 7 show the pitches matched in the experimental conditions. The only S who failed to hear two pitches in all binaural conditions was M.H. This S did not hear a 200-Hz pitch in condition 2000 B-3 200. As expected, all Ss perceived a pitch of 200 Hz in the M-3 200 and M-5 conditions and a pitch of 400 Hz in the M-3 400 conditions.

Results concerning pitch strength are less clear cut. Analyses of variance were performed on each of the dependent measures of pitch strength (comparison intensity (C), noise intensity (N), and standard deviation of pitch match (S)) presented in Table 2. On each dependent measure three analyses of variance were performed. The first was done on the monaural data (six conditions) with all three Ss. The second was done on the binaural data with two Ss (D.H. and T.H.). Subject M.H. was not included in this analysis because, as noted above, she heard no pitch of 200 Hz in condition 2000 B-3 200. The third analysis was done on data from Ss D.H. and T.H. and included all 10 conditions. There were two reasons for performing three separate analyses on the data. First, the binaural conditions seemed to

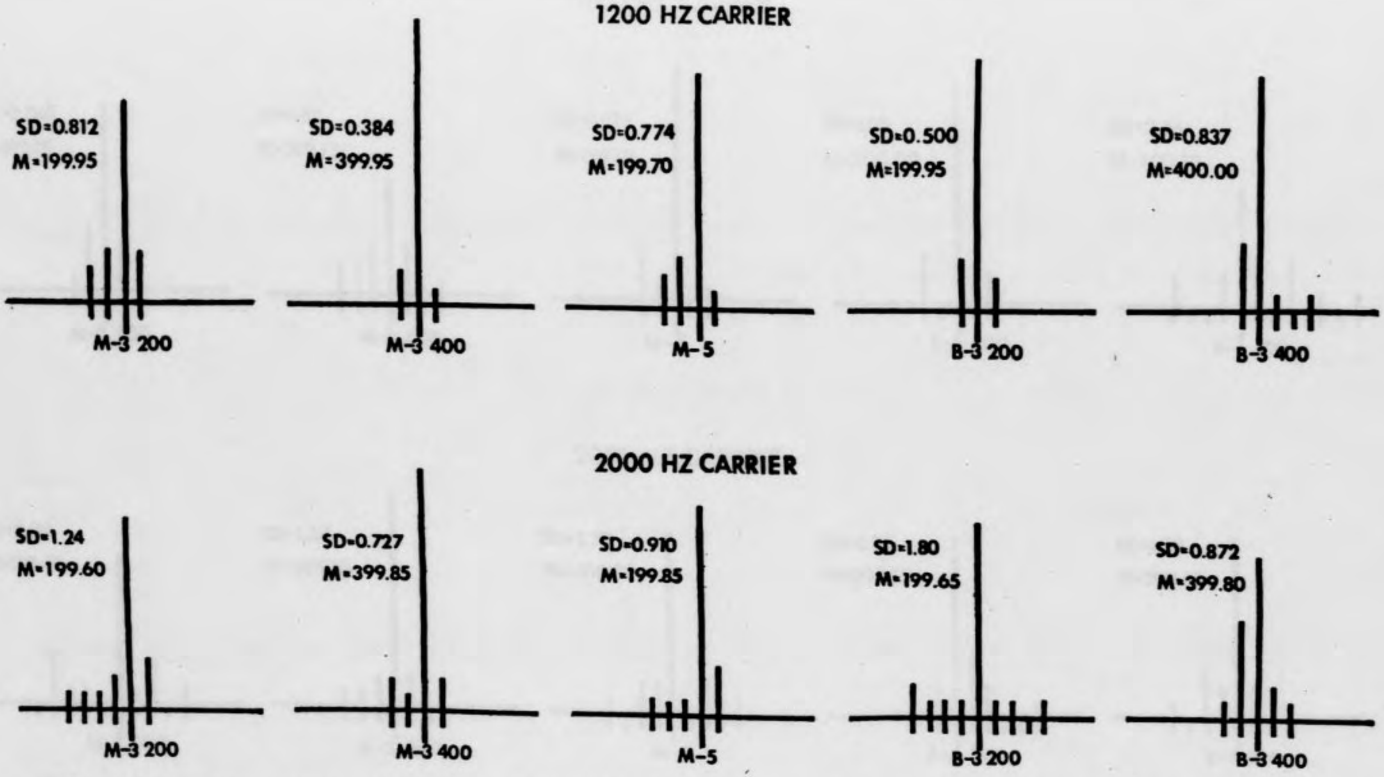

Figure 5. Histograms for S D.H. for the 10 experimental conditions. "SD" refers to standard deviation, and "M" refers to the mean of the 20 pitches matched in a condition. The longest line below the horizontal axis corressponds to the fundamental frequency. Vertical lines to the left and right of this line descend and ascend in one Hz steps.

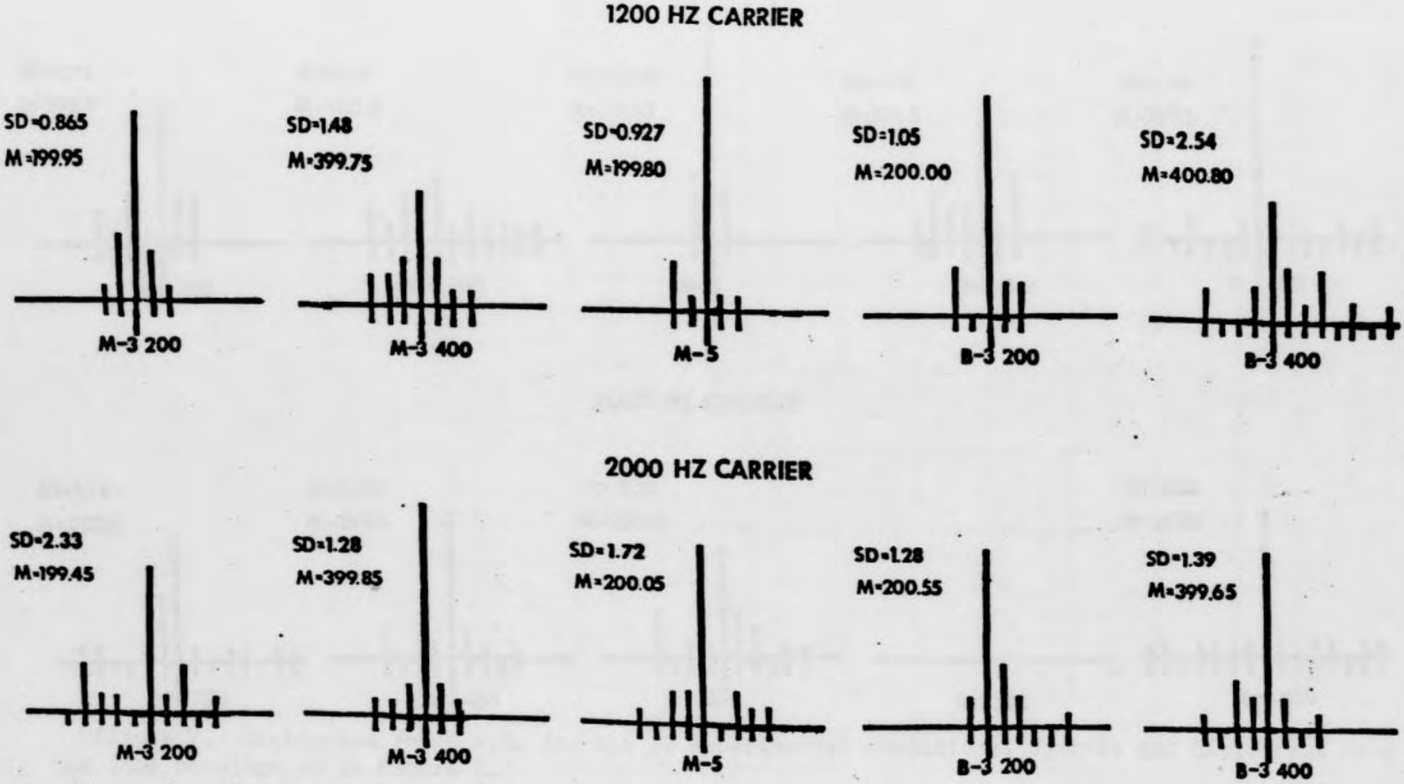

Figure 6. Histograms for S T.H. for the 10 experimental conditions. Symbols and histograms have the same meanings as in Figure 5.

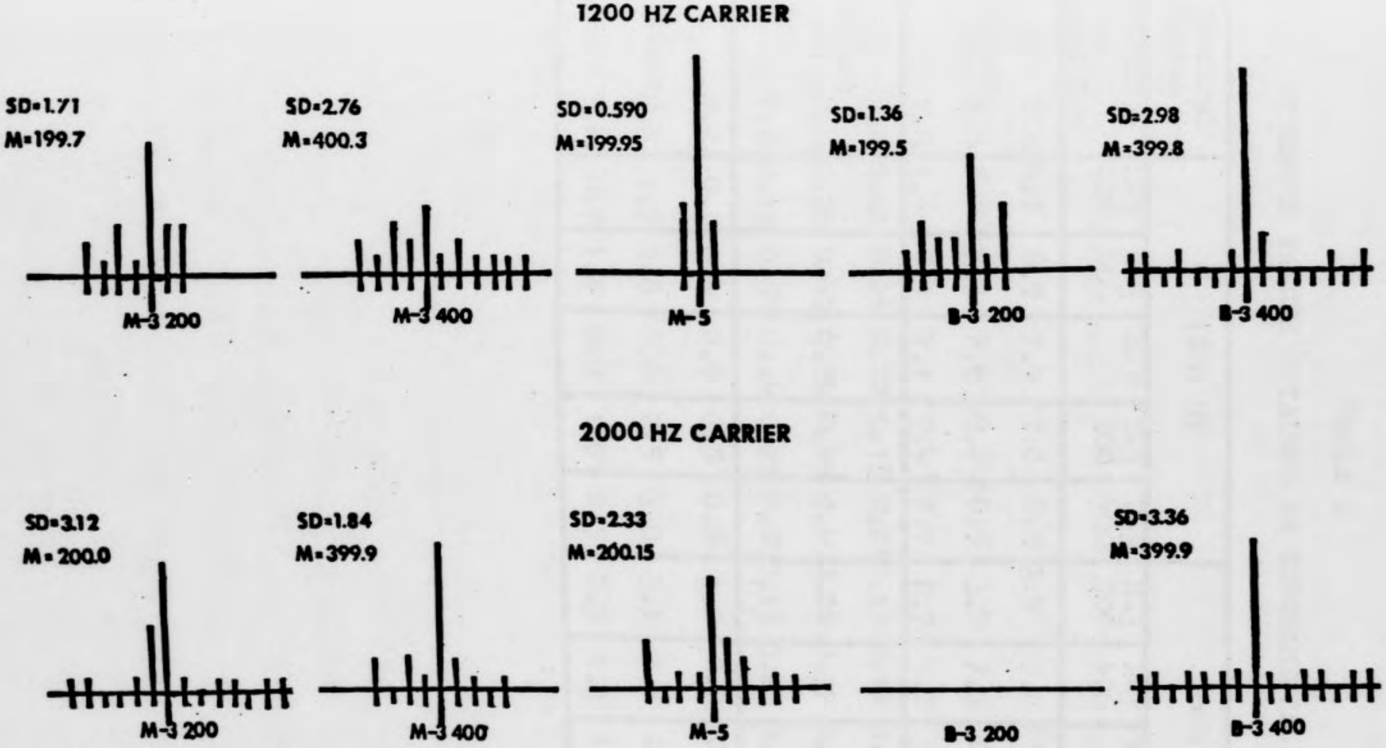

Figure 7. Histograms for S M.H. for the 10 experimental conditions. Symbols and histograms have the same meanings as in Figure 5.

TABLE 2

STRENGTH VALUES OBTAINED IN EXPERIMENTAL CONDITIONS

Spectral Region		1200 Hz					2000 Hz				
Presentation Condition		M-3 200	M-3 400	M-5	B-3 200	B-3 400	M-3 200	M-3 400	M-5	B-3 200	B-3 400
N Values (mv)	D.H.	4.1	6.9	7.4	7.0	8.0	3.9	5.2	5.7	2.6	5.6
	M.H.	5.8	7.4	7.6	9.2	10.5	3.7	7.3	6.2	--	11.2
	T.H.	3.6	5.1	7.1	7.4	7.7	0.7	4.2	5.4	6.5	7.0
C Values (v)	D.H.	0.21	0.24	0.22	0.15	0.21	0.17	0.12	0.21	0.07	0.18
	M.H.	0.36	0.42	0.54	0.49	0.47	0.28	0.37	0.42	--	0.39
	T.H.	0.15	0.54	0.34	0.27	0.29	0.13	0.27	0.18	0.07	0.22
S Values	D.H.	0.8	0.4	0.8	0.5	0.8	1.2	0.7	0.9	1.8	0.9
	M.H.	1.7	2.8	0.6	2.5	3.0	3.1	1.8	2.3	--	3.4
	T.H.	0.9	1.5	0.9	1.0	2.5	2.3	1.3	1.7	1.3	1.4

be inherently more difficult for pitch matching, loudness matching, and noise masking than the monaural conditions. The relative difficulty of the binaural conditions was assumed because Ss reported that they were not always, if ever, sure that they were adjusting the comparison and noise intensities only for the pitch to which they were instructed to attend. For instance, if a S heard two pitches, he was not sure that he could adjust the comparison or noise intensity to match or mask just one pitch, totally ignoring the other. Data from the binaural conditions might therefore be expected to be less valid than monaural data as a measure of pitch strength. The second reason for performing several analyses was that M.H. had no data for condition 2000 M-3 200. While this made it impossible to do an analysis for all three Ss over all conditions, an analysis on the six monaural conditions using all three Ss could be performed.

Of the nine analyses performed three showed significance. The 2X3 analysis of variance with noise masking as the dependent variable yielded significant main effects for the two spectral regions (1200 Hz and 2000 Hz) and the three monaural presentation conditions ($F_{.95}(1,2)=18.5$ and $F_{.99}(2,4)=18.0$, respectively). Noise values were larger for the 1200-Hz region (see Figure 8). A Tukey-b post hoc test showed that noise values for conditions M-3 400 and M-5

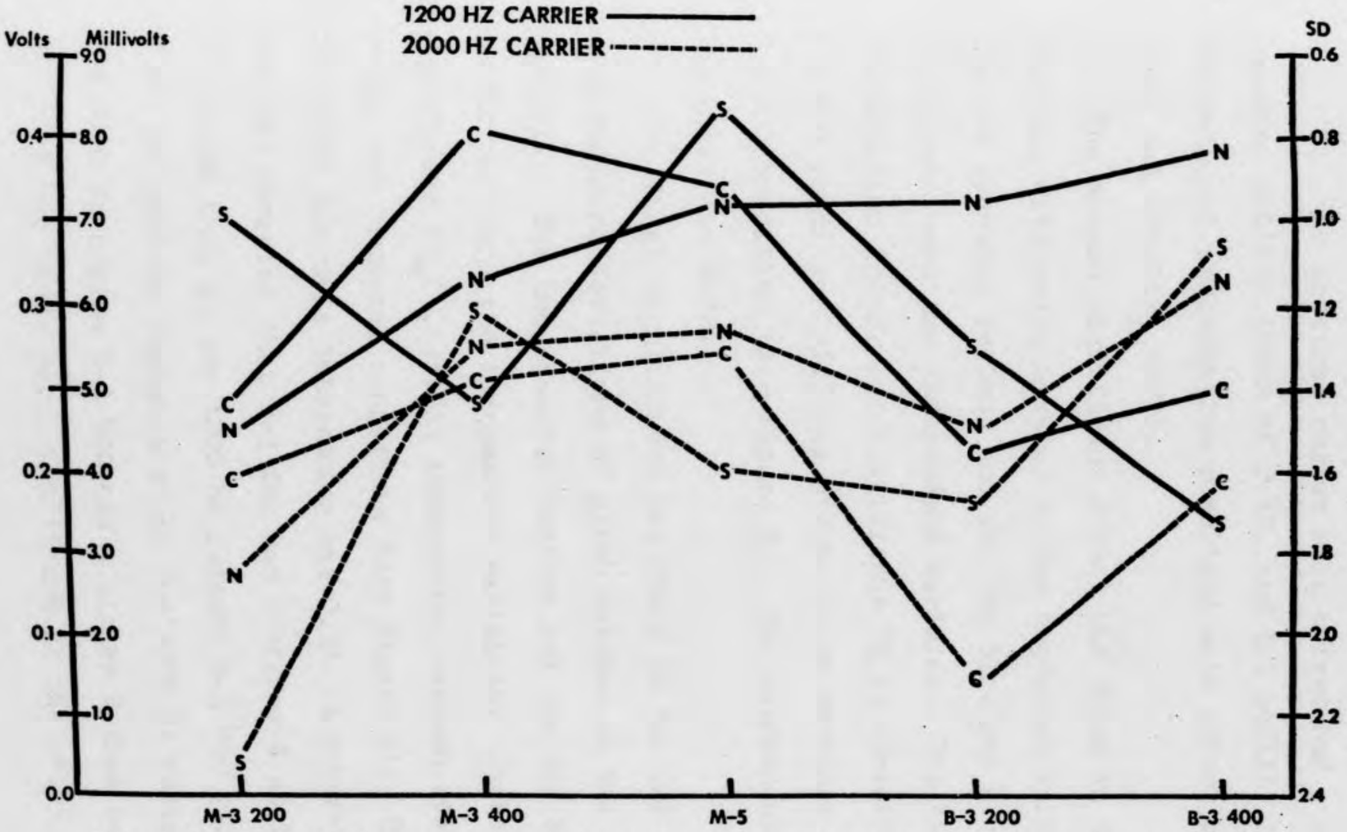

Figure 8. Representation of pitch strength at the 1200 Hz and 2000 Hz spectral regions. Strength is in terms of noise intensity, N (in millivolts), comparison intensity, C (in volts), and standard deviations of pitch matches, S. The solid lines corresponds tothe 1200 Hz spectral region, and the dotted lines correspond to the 2000 Hz region.

were significantly larger ($p<.05$) than those for condition M-3 200. The spectral region main effect had a corresponding utility index of 0.12, and the utility index for the monaural presentation condition main effect was 0.38 (Dodd and Schultz, 1973).

The second significant effect was found in the 2X2 analysis with noise masking as the dependent variable. The two spectral regions and the two binaural presentation conditions were the independent variales. The binaural presentation factor was significant ($F_{.95}(1,1)=161$) with the B-3 400 condition yielding higher noise settings than the B-3 200 condition (see Figure 8). The corresponding utility index was 0.22.

The final significance was found in the 2X2 analysis with standard deviations of pitch matches as the dependent variable. The two spectral regions and the two binaural conditions were the independent variables. There was a significant ($F_{.95}(1,1)=161$) interaction between spectral region and binaural conditions (see Figure 9). The utility index for this interaction was 0.21. A Tukey-b post hoc test revealed that B-3 200 had a higher S at the 2000 Hz region than at the 1200 Hz region; B-3 400 had a higher S at the 1200 Hz region than at the 2000 Hz region; and at the 2000 Hz region B-3 400 had a higher S than B-3 200. All post hoc tests were significant at the $p<.05$ level.

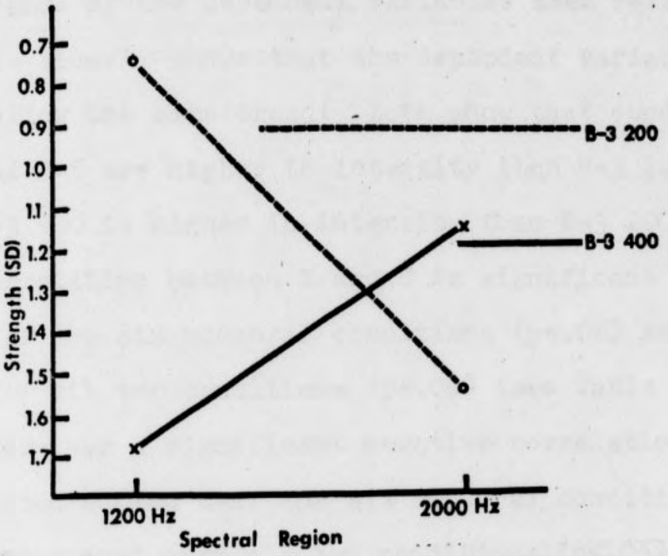

Figure 9. Representation of the interaction between spectral region and binaural conditions. Standard deviation is the dependent measure.

Although C and S did not show significance for either of the two factors, it can be seen in Figure 8 that the trends of the dependent variables seem related. This figure clearly shows that the dependent variables N and C follow the same trend: -both show that conditions M-3 400 and M-5 are higher in intensity than M-3 200 and that B-3 400 is higher in intensity than B-3 200. The positive correlation between N and C is significant when summed over the six monaural conditions ($p<.01$) and when summed over all ten conditions ($p<.06$) (see Table 3). Further, there was a significant negative correlation between N and S when summed over the six monaural conditions ($p<.05$) and when summed over all ten conditions ($p<.05$) (see Table 3). Thus while neither C nor S showed significance for main effects, they correlate highly with the dependent variable that did show significance. The correlations between these variables support the idea that they were all indeed measures of the same thing (pitch strength).

Finally, an analysis of variance was performed to determine whether there was any difference among <u>S</u>s in terms of variability of overall pitch matching. The analysis showed significance ($F_{.99}(2,27)=5.48$) and a Tukey-b post hoc test revealed that D.H. had significantly lower S values than M.H. and T.H., and that T.H. had significantly lower S values than M.H. (see Table 4).

TABLE 3

CORRELATIONS FOR N, C, AND S

	Monaural Conditions			All Conditions		
	N	C	S	N	C	S
N	▬	+.86**	-.75*	▬	+.51●	-.61
C	▬	▬	-.50	▬	▬	-.26
S	▬	▬	▬	▬	▬	▬

** p<.01
* p<.05
● p<.06

TABLE 4

SUBJECT DIFFERENCES IN PITCH MATCHING VARIABILITY

		D.H.	T.H.	M.H.
		0.89	1.49	2.38
D.H.	0.89	▬	0.60*	1.49**
T.H.	1.49	▬	▬	0.89**
M.H.	2.38	▬	▬	▬

** p<.01
* p<.05

DISCUSSION

Central vs. Peripheral Pitch Encoding

The results indicate that the attack by Houtsma and Goldstein (1972) and Wightman (1973a) on the peripheral pitch encoding hypothesis of residue theory was illfounded. Their arguments against peripheral encoding of complex stimuli seem to rest on the assumption that the final "decision" on what pitch is to be perceived is made by a central analyzer. For harmonically related stimuli, this analyzer is sensitive to component input from both ears but is <u>not</u> sensitive to the complex imput to one ear opposed to the input to the other ear. The results of the experiments reported here do not support such a simple central encoding hypothesis. That is, it appears that the neural mechanism which makes decisions on pitch perception is indeed sensitive to peripherally encoded information, and the integrity of peripheral encoding at each ear seems to be preserved antecedent to a neural pitch decision. In each binaural condition <u>S</u>s reported hearing two pitches (each corresponding to the fundamental of a complex presented to one or the other ear). Only one <u>S</u> reported hearing just one pitch in the 2000 B-3 400 condition, and the pitch perceived (400 Hz) was not consonant with a predicted pitch of 200 Hz by the central encoding hypothesis.

The results of Houtsma and Goldstein have been replicated by Bilsen (1973). Thus, the finding that a pitch equal to the fundamental of two binaurally presented sinusoids (one sinusoid to each ear) can be perceived seems to be established as fact. What is not established is the leap from this empirical finding to the theory of a central pitch extractor outlined previously.

The results of Houtsma and Goldstein and the results of the experiment reported here are not thought of as contradictory. Indeed, the two sets of results may be thought of as unrelated in the sense that in one experiment simple tones were presented to each ear, and in the other experiment complex tones were presented to each ear. From the Houtsma and Goldstein experiment the conclusion can be drawn that a pitch decision mechanism seems to be sensitive to the harmonic relation between complex components at *each* ear.

One question which it was hoped would be answered here was that of whether the kind of central "mixing" of peripheral components proposed by Houtsma and Goldstein might be taking place, to some degree, even if evidence for maintenance of peripheral integrity were found. The key to answering this question was thought to lie in pitch strength measures. It was thought that a comparison of the 200-Hz pitch strength from a binaural condition at one spectral region to the M-3 200 condition at the same region would test the central "mixing" hypothesis. If the B-3 200 condition were found to yield higher strength than the M-3 200 condition the central "mixing" hypothesis would be supported. Unfortunately, as previously noted, the binaural conditions entailed inherent perceptual distractions which cast doubt on the validity of all binaural pitch strength measures. In general, S̲s found it difficult to attend to the pitch and intensity of only one complex stimulus when two pitches were perceived. Certain results of the experiment indicate that the inherent problem is not insurmountable, however.

A solution to the problem may lie in proper selection of S̲s and of stimulus intensities. It was found, for instance, that the S̲ with the most musical training, D.H., had the least difficulty in "picking out" two separate pitches in the binaural conditions and that the S̲ with no musical training, M.H., had the most difficulty. Subject M.H. did not hear two pitches in one binaural condition and T.H. heard two pitches in all binaural conditions. Subject D.H. could not only hear two pitches in all the binaural conditions but could also localize the pitches--for instance, D.H. could correctly report that he heard the lower pitch in his "best" ear and the higher pitch in his other ear in a B-3 200 condition. Neither of the other two Ss could report such localization. Further, D.H. had lower S values than either T.H. or M.H. and T.H. had lower S values than M.H. (see Table 4). Here the trend seems to be one of better pitch matching precision with more musical training. Although innate musical abilities could account for the differences, also, the nature-nurture distinction is not critical to the present argument. It seems logical to conclude that Ss with musical training may be better equipped to attend selectively to one pitch when two pitches are present. It is therefore proposed that Ss with extensive musical training could be used to test the central "mixing" hypothesis, especially if they were given some special training in attending to the type of complex stimuli that are generated in the laboratory. To further aid Ss in attending to the 200 Hz pitch in the B-3 200 conditions, it might be helpful if the intensity of the 400 Hz complex stimulus were attenuated; each S reported that the higher pitch in the B-3 200 condition was "much louder" than the lower pitch. The 400 Hz complex intensity could be lowered to make its corresponding pitch less distracting to the 200 Hz match. Therefore a test of the central "mixing" hypothesis may be possible by proper selection of Ss and component intensities.

Pitch Strength and Theoretical Considerations

The significant differences in pitch strength found with N as the dependent measure, along with the significant correlations between N and both C and S, establish that pitch strength can be reported quantitatively as well as qualitatively. An obvious observation is that noise masking seems to be the best measure of pitch strength in that it seems most likely to show significant differences where the other measures show only trend. In summary, the noise masking measures of pitch strength indicated that conditions M-3 400 and M-5 yielded stronger pitches than M-3 200 at both spectral regions; and the 1200 Hz spectral region yielded stronger pitches than the 2000 Hz region.

Referring back to Figures 2 and 3, it can be seen that, in some respects, the results conform to the predictions of Wightman's autocorrelation model for pitch strength: stimuli with relatively more components are predicted to be stronger than those with fewer components (M-5 vs. M-3 200); stimuli with greater spacing between components are predicted to be stronger than those with lesser spacing (M-3 400 vs. M-3 200); and stimuli with components in lower spectral regions are predicted to be stronger than those at higher spectral regions (1200 Hz carrier vs. 2000 Hz carrier). These predictions are met by the data. Note however, that in Figures 2 and 3 pitch strength differences, both as a function of number of components and component spacing, are predicted to become less as frequency of lowest component increases. As shown in Figures 10 and 11, the results of the present experiment actually indicate a slight trend in the opposite direction. That is, differences in pitch strength were slightly enhanced at higher frequencies. Obviously more data points on Figures 10 and 11 must be filled in before this trend is substantiated but the results depicted in these figures may be viewed as damaging to Wightman's theory.

Recently Ritsma (1967) and Bilsen (1973) have advanced empirical evidence that there is a "dominant" spectral region for the perceptibility of complex stimuli. This dominant region consists of the 3rd, 4th, and 5th harmonics of the fundamental tone. Dominance is not a theoretical concept but simply an empirical phenomenon. Complexes containing dominant components are found to be relatively strong. Pitch strength predictions based upon the concept of dominance and those based on Wightman's theory overlap considerably for the stimuli used in this experiment. There are, however, conditions in which dominance and Wightman's model predict different strength values. As can be seen in Table 5, condition M-3 200 has one dominant component at the 1200 Hz spectral region while M-3 400 and M-5 each have two dominant components. At the 2000 Hz

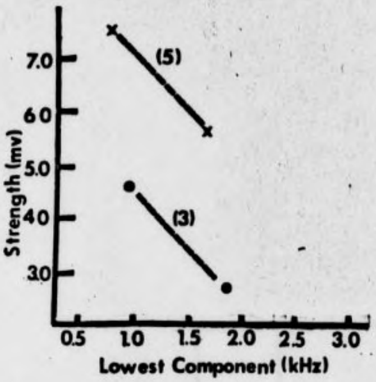

Figure 10. Variation of pitch strength (N as dependent measure) with regard to number of components per complex, and spectral region. Number in parentheses is number of components per complex.

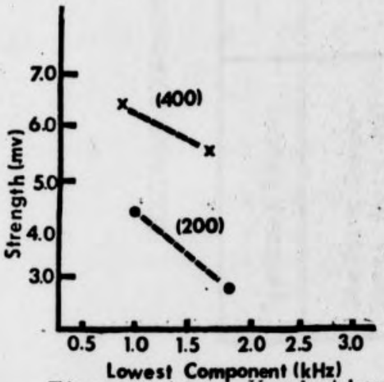

Figure 11. Variation of pitch strength (N as dependent measure) with regard to component spacing, and spectral region. Number in parentheses represents spacing (in Hz) between components.

TABLE 5

DOMINANT COMPONENTS IN THE SPECTRAL REGIONS

Spectral Region	1200 Hz	2000 Hz
Dominant Components for a 400 Hz Fundamental	1200 Hz 1600 Hz	1600 Hz 2000 Hz
Dominant Components for a 200 Hz Fundamental	800 Hz 1000 Hz	--- ---

spectral region condition M-3 200 has no dominant components while M-3 400 and M-5 have two and zero dominant components, respectively. Dominance therefore predicts the higher strength of M-3 400 and M-5, relative to M-3 200, at 1200 Hz; dominance also predicts the higher strength of M-3 400 relative to M-3 200 at 2000 Hz. What dominance does not predict is the higher strength of M-5 relative to M-3 200, at 2000 Hz, or the higher strength of M-3 400 at 1200 Hz relative to M-3 400 at 1200 Hz. On the other hand, Wightman's theory does account for these data. In general, the data here seem to be explained more adequately by Wightman's autocorrelation model than by the concept of spectral dominance.

When the results of this experiment are viewed in relation to the issue of peripheral vs. central encoding of complex stimuli and to Wightman's model of pitch perception, it can be seen that one more stumbling block has been laid in the path of the pitch theoretician. From the data of Houtsma and Goldstein, it appears that a Stage 2 process may autocorrelate the PAP's from both ears to yield one complex pitch value. From the data of the present experiment, however, it would seem that the Stage 2 process autocorrelates the PAP's from <u>each</u> ear to yield two pitch values. Any theory of pitch perception must account for this new finding that pitch perception resulting from binaural stimulation seems to be dependent upon the complexity of the stimuli at each ear. A model based on autocorrelation must be able to explain why PAP analysis sometimes seems to be performed on the input from both ears, as if all components were part of one complex tone, and other times as if the complex input at each ear corresponded to one of two different complex tones.

The last result that warrants discussion is the significant interaction found between binaural condition and spectral region (see Figure 9). This interaction was not expected, and is not predicted by any theory. The dependent measure used in obtaining this interaction was S; neither N nor C yielded anything resembling this interaction. At the present time no explanation can be offered for this result.

BIBLIOGRAPHY

Bilsen, F., 1973. On the influence of the number of phase harmonics on the perceptibility of the pitch of complex signals. Acustica, 28, 60-65.

De Boer, E., 1956. On the residue in hearing. Doctoral Dissertation, University of Amsterdam.

Dodd, H., and Schultz, R.F., 1973. Computational procedures for estimating magnitude of effect for some analysis of variance designs. Psych. Bull., 79, 391-393.

Fletcher, H., 1923. Physical measurements of audition and their bearing on the theory of hearing. Bell Syst. Techn. J., 2, 145-180.

Helmholtz, H.L.F. Sensations of Tone, A.J. Ellis (Trans.), Longmans, Green, New York, 1930.

Houtsma, A., and Goldstein, J., 1972. The central origin of pitch of complex tones: Evidence from musical interval recognition. J. Acous. Soc. Amer., 51, 520-529.

Licklider, J.C., 1956. Auditory frequency analysis. In: C. Cherry, ed. Information Theory. London, Butterworth, 1956, 253-268.

Patterson, R., 1973. Physical variables determining the residue pitch. J. Acous. Soc. Amer., 53, 1565-1572.

Plomp, R., 1968. Pitch, timbre, and hearing theory. Internat. Audiol., 7, 322-344.

Ritsma, R., 1962. Existence region of the tonal residue. J. Acous. Soc. Amer., 34, 1224-1229.

Ritsma, R., 1967. Frequencies dominant in the perception of the pitch of complex sounds. J. Acous. Soc. Amer., 42, 191-198.

Schouten, J.F., 1940. The residue, a new concept in subjective sound analysis. In: J. Harris, ed. Forty Germinal Papers in Human Hearing, J. Audit. Res., 1969, 107-117.

Schouten, J.F., Ritsma, R., and Cardozo, B., 1962. Pitch of the residue. J. Acous. Soc. Amer., 34, 1418-1424.

Wever, E.G., 1949. Theory of Hearing, John Wiley and Sons, Inc., New York.

Wightman F., 1973a. Pitch and stimulus fine structure. J. Acous. Soc. Amer., 54, 417-426.

Wightman, F., 1973. The pattern transformation model of pitch. J. Acous. Soc. Amer., 54, 407-416.