This is to certify that the

dissertation entitled

Two Studies of Pitch Perception

presented by

Mark Allen Klein

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Psychology

[signature]
Major professor

Date 14 Sept 1981

TWO STUDIES OF PITCH PERCEPTION

By

Mark Allen Klein

A DISSERTATION

Submitted to
Michigan State University

in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Psychology

1981

ABSTRACT

TWO STUDIES OF PITCH PERCEPTION

By

Mark Allen Klein

The first of these two studies describes a new dichotic noise band pitch effect. A pitch is heard when all low frequency components of a digital noise are sent in phase to the two ears and all high frequency components are sent at pi phase to the two ears. Each noise signal alone provides no pitch cues. We call this effect the Binaural Edge Pitch. In a method of adjustment task subjects matched a sine tone in diotic noise to the perceived pitch in the dichotic noise. Subjects reliably adjusted the matching tone to frequencies either just above the phase transition frequency or just below it. The Equalization-Cancellation model of binaural interaction provides a plausible explanation for these results.

The second study investigates the effect of intensity on pitch through data collection and computer modelling studies. Von Bekesy tracking was used to make detailed measurements of the microstructure in subjects' threshold of hearing curves. Measurements of shifts in perceived pitch with changes in intensity were made using a method of adjustment pitch matching task. The threshold microstructure information was used to calculate spatial excitation patterns. Four models of pitch extraction were evaluated. Predictions of performance on the pitch-intensity task were made. These predictions were compared to the data collected. None of the models yielded satisfactory predictions. The data collection procedures used were shown to be reasonable and demonstrated the existence and stability of threshold microstructure.

ACKNOWLEDGEMENTS

I would like to express my appreciation to Dr. Raymond Frankmann, Dr. Lester Hyman, Dr. Mark Rilling, and Dr. James Zacks for their help in serving as members of my dissertation committee. I am also very grateful to Dr. William Hartmann, my major advisor, whose guidance, assistance, and inspiration made this work possible. My special thanks to my wife, Diane, for her help in preparing this manuscript and understanding support throughout the entire duration of this project and my whole graduate education.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

This paper discribes two studies linked by a common objective: to discover more about how the human auditory system processes acoustic stimuli to generate a sensation of pitch. The first study is an investigation of a new dichotic noise pitch effect. The stimuli involved are complex, having over 250 sine wave components. The second is an investigation of an old phenomenon, the effect of intensity on the perceived pitch of a single sine tone.

An obvious difference between these two studies is the complexity of the stimuli used. The most important difference, however, is the portion of the auditory system under investigation. The latter study investigates a pitch effect that is based primarily on peripheral processes and characteristics. It explores the possibility that when perceived pitch is changed by altering only the intensity of the signal,the direction and magnitude of that change may be predicted by models of peripheral auditory function and the threshold curves for the particular ear being stimulated.

The first study describes a pitch effect that allows the peripheral auditory mechanisms to be bypassed and more central processes to be investigated directly. The information used by the subjects to generate the pitch sensations is not present in either of the separate signals presented to the individual ears. It is the relationship between the two signals that must be used. Phenomena similar to this are discribed by Julesz (1971) for vision. Random-dot-stereograms may be used to reinvestigate many classical visual perception experiments and to determine whether or not the normally observed results are dependent upon peripheral or central mechanisms.

Although Julesz seems to have taken much of his early motivation for his pioneering work in the field he labels "cyclopean perception" from musical and auditory demonstrations, his investigations are completely visual. The development of cyclopean audition has been virtually ignored. This new pitch effect might allow later researchers to attempt to "catch up" with at least some of Julesz' visual investigations. At this time, however, it is imperative to simply demonstrate the reality of the pitch effect and leave its applications to others.

Portions of this dichotic noise pitch investigation were published in Klein and Hartmann (1981).

CHAPTER I

I. Introduction

A. Huggins Pitch

The Huggins Pitch (Cramer & Huggins, 1958) is a binaural pitch effect created when a special correlation exists between the frequency components of broad band noise signals sent separately to the listeners' two ears. The noises to the two ears are identical except for a rapidly increasing phase discrepancy between the corresponding frequency components. Across a narrow frequency region the interaural phase relationship progressively changes from 0 to π to 2π. This shift is produced by passing a broadband noise through a narrowly tuned all-pass filter. When this phase-shifted noise is sent to one ear and the original unfiltered noise is sent to the other ear, the listener reports a sensation of pitch. This stimulus situation is shown in Figure 1A,B.

This stimulus is interesting in that, because the filter does not alter the amplitudes of the components in any way, the signal to each ear is still broad band noise with no frequency specific information content that could produce a sensation of pitch. When heard alone, either noise signal is devoid of any pitch or pitch-like characteristics. It is only when the two signals are combined within the auditory system that a pitch sensation occurs.

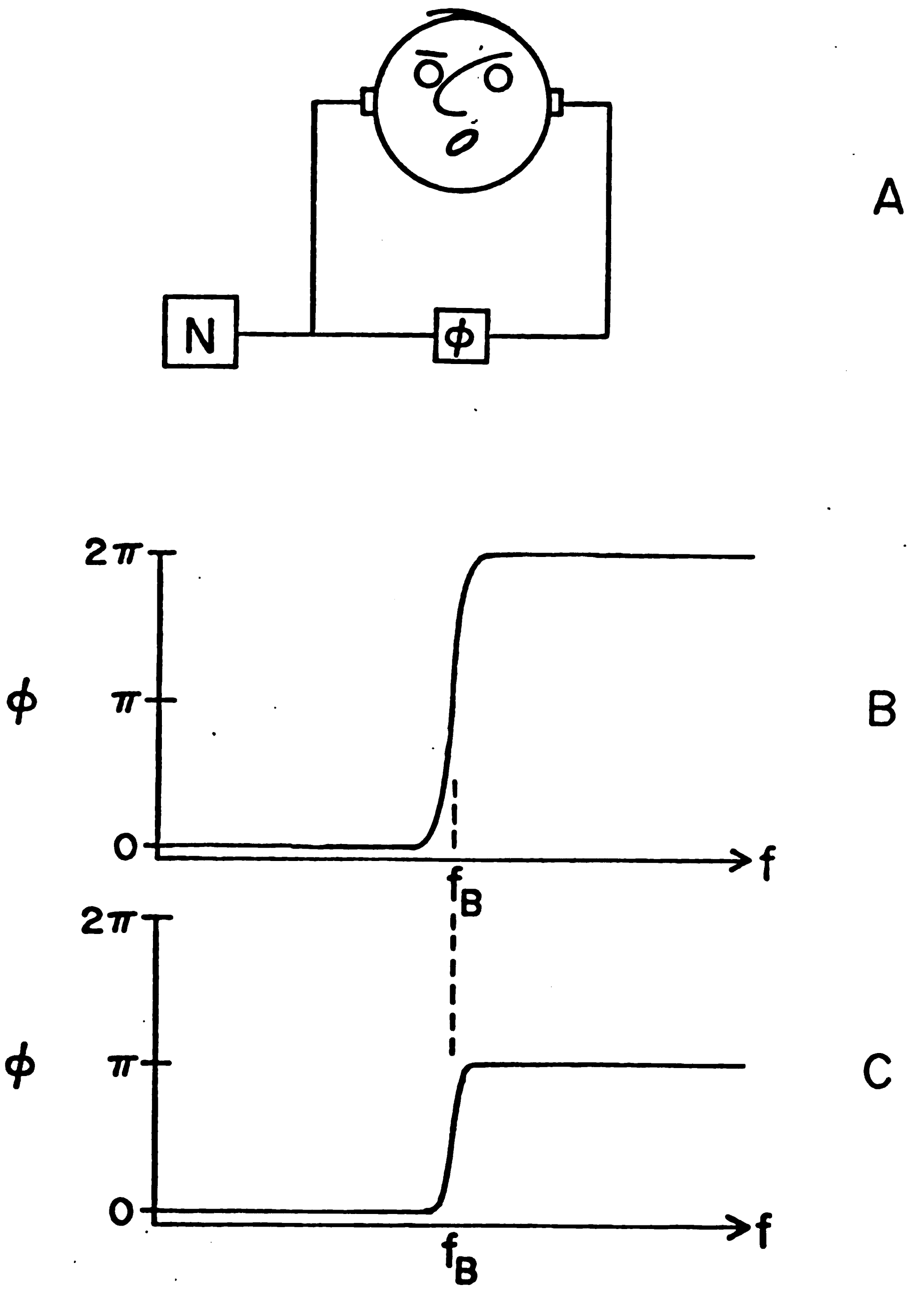

Figure 1. Stimulus configuration and phase characteristics

The pitch sensation is not random, but corresponds well with the center of the phase shift region (where the interaural phase shift is π). It is most easily heard when the center of the phase shift region is between 200 Hz and 1000 Hz and the $\pi/2$ and $3\pi/2$ points of the region are within 5% of that center. Cramer & Huggins found that under these optimum conditions, subjects could reliably identify the pitch with a standard error of about 3%.

Guttman (1962) was able to show that this pitch effect can also be obtained when the interaural phase shift goes from $-\pi$ to $+\pi$. This effect can be obtained simply by passing the inverse of the original noise signal through the all-pass filter. Wightman, Grantham, & Fowler (1977) were able to produce the effect digitally, by inverting only one of the frequency components of a broad band noise made up of many sine waves spaced 10 Hz apart. As for the analog Huggins effect, the original noise was sent to one ear and the noise with the phase alteration (this time only a single component) was sent to the other ear. By varying which component was phase shifted, they were able to play tunes that a subject could recognize.

Another interaural phase effect is the Masking Level Difference (MLD). This effect is characterized by an improvement in detection performance due to interaural phase discrepancies. In the classic experiment (Licklider, 1948; Hirsh, 1948) a signal is to be detected in noise. Performance on the detection task improves when the signal to both ears is out of phase across the head while the noise remains interaurally monophasic. Durlach (1963,1972) proposed a model to explain this MLD effect. According to his Equalization-Cancellation model (E-C), the binaural auditory system balances the inputs to the two ears and then finds the remaining differences between them in an attempt to maximize the signal-to-noise ratio in some single representation of the signal that is presented to a central processor. The Equalization part of this process has two operations: a frequency-independent amplitude adjustment and a frequency-independent phase shift. These uniform changes are applied to the inputs in such a way that as many differences between the two channels as possible are eliminated. Then, in the Cancellation stage, the equalized inputs are subtracted. The central processor then chooses between either of the separate inputs or the difference channel, continuing processing with the channel having the best signal-to-noise ratio.

The two Equalization parameters are frequency-independent in that the same adjustments are made at all frequencies. The binaural system optimizes those parameters in accordance with the goal of increasing signal-to-noise ratio. The process is not perfect and there is noise in both the Equalization and Cancellation stages making the results less than absolute, but this may be ignored for the present discussion.

The Huggins Pitch is explained within the E-C, model (Durlach, 1962)by having no phase or amplitude changes at all. Then, during the Cancellation stage, all frequencies having no amplitude or phase differences between the two channels are cancelled completely. The frequencies within the range of the filter do have phase discrepancies remaining and so during the cancellation stage a narrow noise band centered where the interaural phase shift is π remains. When the all-pass filter region is narrow, the resulting internal noise band (in the difference channel) may be narrow enough to produce a definite pitch sensation. The Guttman version of the stimulus is processed the same way except that the equalization consists of a uniform phase shift of π at all frequencies. This results in the same pair of signals being presented to the Cancellation stage as for the original Huggins Pitch.

The Huggins Pitch is, by no means, the only example of a broad noise band producing a pitch sensation. Bilsen (1977) reviewed a large collection of monaural and binaural stimuli that produce pitch sensations. Repetition pitch is the name given to the pitch heard when a noise and a time delayed version of the same noise are sent to the same ear (Bilsen, 1966). Sending the original noise to one ear and the time delayed version to the other ear (Bilsen, 1966) also produces a sensation of pitch. Foursin (1962, 1970) started with two uncorrelated noise sources. One original noise plus a time delayed version of the other source was sent to one ear. The second original noise plus the time delay of the first was sent to the other ear. Subjects heard a single pitch.

Fastl (1971) explored the pitch sensation produced by a single narrow noise band. When the noise band was very narrow a single distinct pitch was heard, much like a sine tone. As the band widened the sensation of pitch remained correlated with the center of the noise band but becomes weaker. When the noise band was approximately one fifth of an actave wide the sensation of pitch disappeared completely.

Bilsen (1977) proposed that the output from the E-C mechanism to the more central processes is analogous to the signal from a single ear. The spectrum of that internal difference channel is processed in the same way that the spectrum from a single ear is processed. He went so far as to say that there is only one "central spectrum" and both the single, right and left, channels input to it, and the products of the binaural system input to it.

Combining Bilsen's central spectrum with the findings of Fastl we have an explanation for why the Huggins Pitch disappears when the phase transition region is too wide. The narrow noise band remaining in the central spectrum after the cancellation stage (for Huggins Pitch) is processed exactly like Fastl's widening noise band. When the phase transition region becomes too wide the band in the central spectrum is also too wide. Like the noise band input through a single channel, it must remain below a certain width to continue to produce a distinct pitch. Because the two stimuli (the Huggins stimulus and the Fastl stimulus) are processed by the same mechanism, they will have exactly the same limit. The Huggins Pitch will disappear when the noise band in the central spectrum produced via the E-C mechanism exceeds the same threshold width determined by Fastl, roughly 1/5 of an octave.

B. Noise Band Edge Pitch

Fastl (1971) also showed that subjects can reliably match a tone to the pitch sensation resulting from low-pass or high-pass noise with a very sharp cut off. The frequency found to match the noise band pitches does not, however, correspond exactly to the cut off frequency of the spectral edges. That frequency is shifted <u>into</u> the noise, above the cutoff of high-pass noise and below the cutoff for low-pass noise.

This phenomenon is also observed for band pass noise. When the band width is too large to produce a single pitch and the cutoff slopes are very sharp, each cutoff acts as a separate noise signal. The lower cutoff produces a pitch like a high-pass noise and the upper cutoff produces a pitch like a low-pass noise. The single band pass noise may be matched by either of two frequencies, one just above the lower cutoff and one just below the upper cutoff.

C. Binaural Edge Pitch

The two preceding phenomena may be united in the following way. From Huggins Pitch we learned that the binaural system may modify the input on the way to producing a central spectrum by cancelling portions of the signals input to the two ears. From Noise Band Edge Pitch we learned that it only requires one edge in a signal to generate a sensation of pitch, not a narrow peak. Is it possible to create a sensation of pitch by using dichotic noise to produce a noise band edge in the central spectrum while presenting only broad band noise to each individual auditory channel (ear)?

We may try to generate this phenomena by presenting the two ears with broad band noise having an interaural phase shift of 0 to π across some transition region, or phase boundary, instead of from 0 to 2π as for the Huggins Pitch. All low frequencies are exactly out of phase. This configuration is shown in Figure 1A,C. With this type of interaural relationship it becomes impossible for an E-C model allowing only a single phase shift for all frequencies to cancel all components. The equalization mechanism has two choices: 1) It can introduce no phase shift. Then through the cancellation process the low frequency components will be suppressed and the high frequency components remain to generate a high-pass noise in the difference channel. 2) It can set up cancellation of the high frequency components by introducing a phase shift of π. The high frequency components, now forced out of phase, will result in a low-pass noise in the difference channel.

The experiments below are an investigation of this central edge pitch phenomenon. We performed experiments using dichotic broad band noise having an interaural phase shift varying from 0 to π. The results indicate that the expected pitch sensation is produced. We call this pitch effect the Binaural Edge Pitch (BEP).

This is a new pitch phenomenon and as such the first experimental priority is the establishment of its existence. This is carried out in Experiment I. Experiment II is a study repeating some of Fastl's work attempting to show (using our laboratory techniques and experimental paradigm) that the results from Experiment I using the dichotic stimuli are extremely similar to the results obtained using diotic noise band edges. Experiment III investigates the strength of this new phenomenon relative to the Huggins Pitch. In section V a number of auxiliary experiments are described. These experiments provide additional evidence of the existence of the BEP. They also show the behavior of the binaural system under various other interaural phase configurations; they shed light on how the implicit ambiguity of the stimulus may be dealt with. Does the E-C mechanism choose not to phase shift the stimuli and produce high-pass noise internally or does the system induce a phase shift (even though that forces previously equalized frequency components to be out of phase) to produce low-pass noise?

II. Experiment I

A. Stimuli

The noise stimuli used to establish the existence of the Binaural Edge Pitch (BEP) were generated digitally. (An analog all-pass filter cannot produce the required overall π phase shift within a narrow frequency region.) Two noise signals were generated each with 251 equally spaced equal-amplitude sine components at random initial phase angles. For the stimulus sent to the left ear all sine components below the phase boundary frequency were the same phase as those components for the right ear. Above the phase boundary all components to the left ear were at π phase relative to the corresponding components for the right ear. The noise to the right and left ears is described by the following formulas:

$$S_R(t) = \sum_{n=1}^{251} A \sin(n2\pi f_o t + \emptyset_n)$$

$$S_L(t) = \sum_{n=1}^{n_B} A \sin(n2\pi f_o t + \emptyset_n) - \sum_{n=n_B+1}^{251} A \sin(n2\pi f_o t = \emptyset n)$$

Here ϕ_n is a random variable. The uniform spacing of the spectral components is f_0. The phase boundary frequency is defined as $f_B=(n_B+1/2)f_0$, halfway between the highest 0 phase component and the lowest π phase component.

The two noise signals were created by two 12-bit digital to analog converters. (The difference between the two signals, when taken by analog electronics, showed a discontinuous 30 db drop at the phase boundary.) The subjects listened to the signals through Beyer DT-48 headphones at 60 dB SPL while seated in a sound proof booth.

The phase boundary frequency was varied from 126 Hz to 2438 Hz. We performed the experiments with phase boundaries in 3 overlapping frequency ranges, low, middle, and high. Within each range the waveforms were all identical and characterized by the boundary number n_B. For all ranges the phase boundary frequency was varied from trial to trial by changing f_0. This was done by changing the sampling rate. This meant that the spacing between the components and the frequency of the 251st component were changing on every trial along with the phase boundary. Table 1. shows these details. It is possible, then, that these other cues may be contributing to the perceived pitch sensation. In the results section G. below and in the auxiliary experiments we show, however, that these artifacts are unlikely to have been major sources of the BEP.

Table 1. Parameters of noise stimulus

	RANGE		
	LOW	MIDDLE	HIGH
n_B	40	100	180
$n_B/251$	0.16	0.40	0.72
Min f_B(Hz)	126	315	567
Max f_B(Hz)	420	856	2438
Min f_o(Hz)	3.15	3.15	3.15
Max f_o(Hz)	10.5	8.56	13.5

B. Procedure

In a method of adjustment procedure subjects matched a sine tone to the pitch that they perceived in the noise. Subjects controlled the frequency of the sine tone by rotating a knob that determined the control voltage sent to a voltage controlled function generator.

The dichotic noise stimulus was part of a repeating four-segment presentation structure lasting 1.6 seconds. As shown in Figure 2. the first segment contained the dichotic noise stimulus. In the second, third, and fourth segments the noise signal sent to the right ear during the first interval was sent in phase to both ears. During the third segment the sine tone matching signal was added to the diotic noise. This four-segment sequence repeated indefinitely until the subject signalled that he was satisfied with the match between the sine tone and the BEP.

The BEP is, in general, difficult to hear. Maintaining the diotic noise during the matching interval (third) made the task easier. Subjects were also able to vary the intensity of the matching sine tone in order to increase the similarity between the matching interval and the BEP and make the task easier. An additional switch was available to the subjects that allowed them to completely eliminate the matching tone without changing the frequency or amplitude settings.

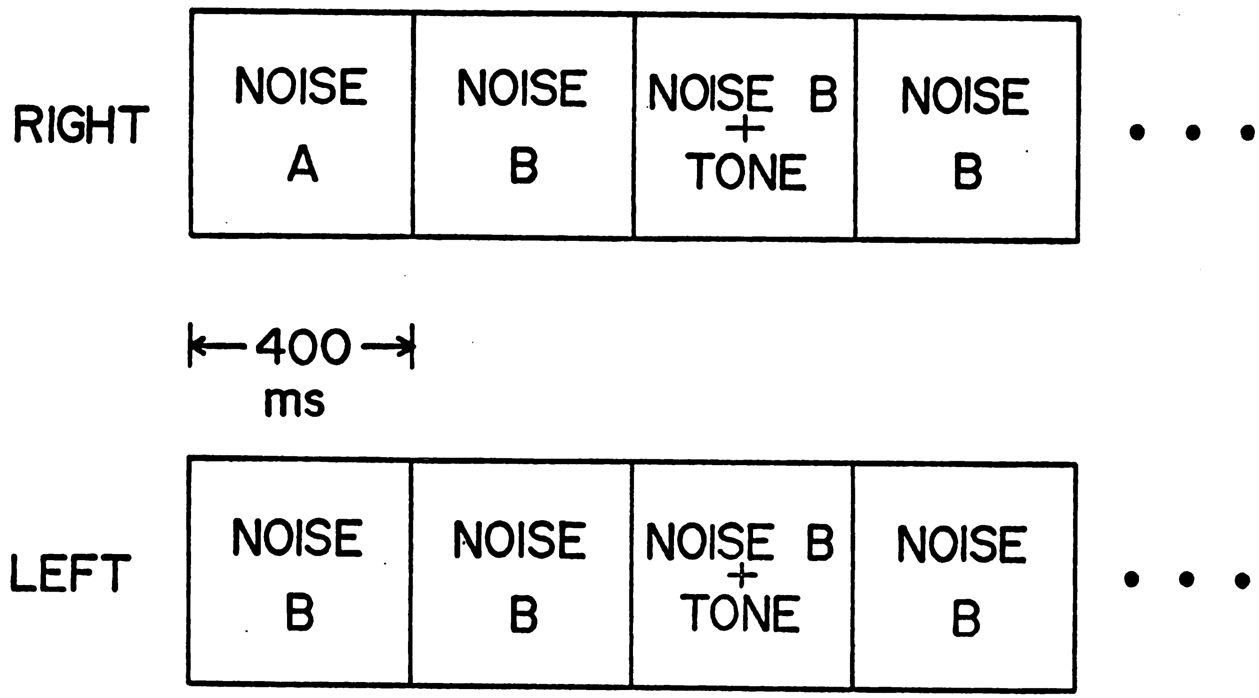

Figure 2. Stimulus presentation sequence

C. Middle Range Experiment

All subjects first performed the task with the phase boundaries of the middle frequency range. In a single experimental run subjects matched pitches to dichotic noise with 12 different phase boundary frequencies, presented in random order. Five subjects participated in this experiment. Subjects G, M, and W were experienced listeners. Subject M is the author. Subjects D and R had never previously participated in psychoacoustic experiments. Some of the subjects (G, D, and R) were given practice in matching pitches to the Huggins stimulus before participating in this experiment.

D. Middle Range Results

Four of the Five subjects produced reliable pitch matches to the binaural edge stimulus. The matches for these subjects, based on four or more runs, were remarkably consistent. The pitch associated with a phase boundary of frequency f_B was typically matched by a sine tone, with frequency f_M, about a quarter tone away from the boundary frequency. This result was independent of the phase boundary frequency. Subjects D, G, and M showed a bimodal distribution of responses making it reasonable to average the data for the 12 phase boundary frequencies and separated according to whether the matching tone was above or below f_B. Table II. gives these averaged results. The table shows the percent deviation of the matching frequency from the boundary frequency, $(f_m/f_b)-1$. . When divided this way the means of the two groups of data were significantly different at the $p<0.001$ level for both subjects G and M. The data for subject D showed higher errors, probably due to inexperience, but also appeared bimodal. The data for subject W was clearly unimodal with a mean at 3.5% above f_B, nearly coincident with the upper peaks for subjects G and M. The relative number of responses, above or below f_B is shown as a percentage in the P columns of the table.

Table 2. Distribution of pitch matches in the middle range of Experiment I for four subjects

Subject	Lower Peak (f_m/f_B-1) %	Lower Peak σ %	Lower Peak P %	Upper Peak (f_m/f_B-1) %	Upper Peak σ %	Upper Peak P %
W	-	-	0	3.5	0.9	100
M	-4.1	3.9	33	3.6	1.5	67
G	-5.2	1.4	40	3.1	1.1	60
D	-4.5	4.8	16	5.9	6.9	84

Figures 3 through 6 show the actual data for the four subjects plotted as cirlces. f_A/f_B is plotted as a function of f_B. The data have been divided into two groups, as described above, for each phase boundary frequency. The size of the circle represents the relative fraction of data in each of the two distributions. Error bars show plus and minus one standard deviation when shown. All other points had errors within the diameter of the circle at that boundary frequency. The error bars shown to the left of the circles show the median error for this middle frequency range.

The inexperienced subject R never learned to perform the task consistently despite over 20 attempts at matching the set of 12 phase boundary frequencies.

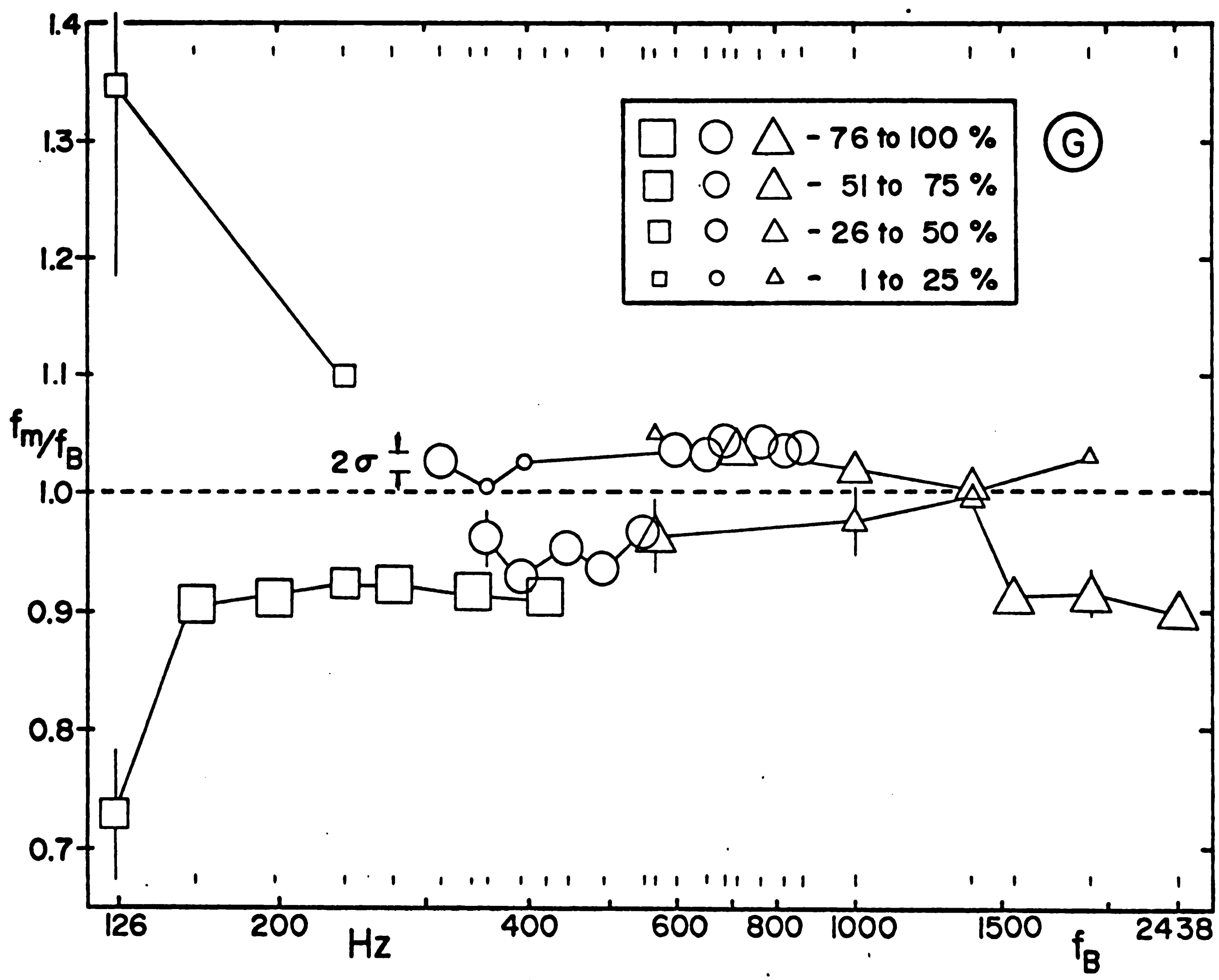

Figure 3. Data for subject G in Experiment I

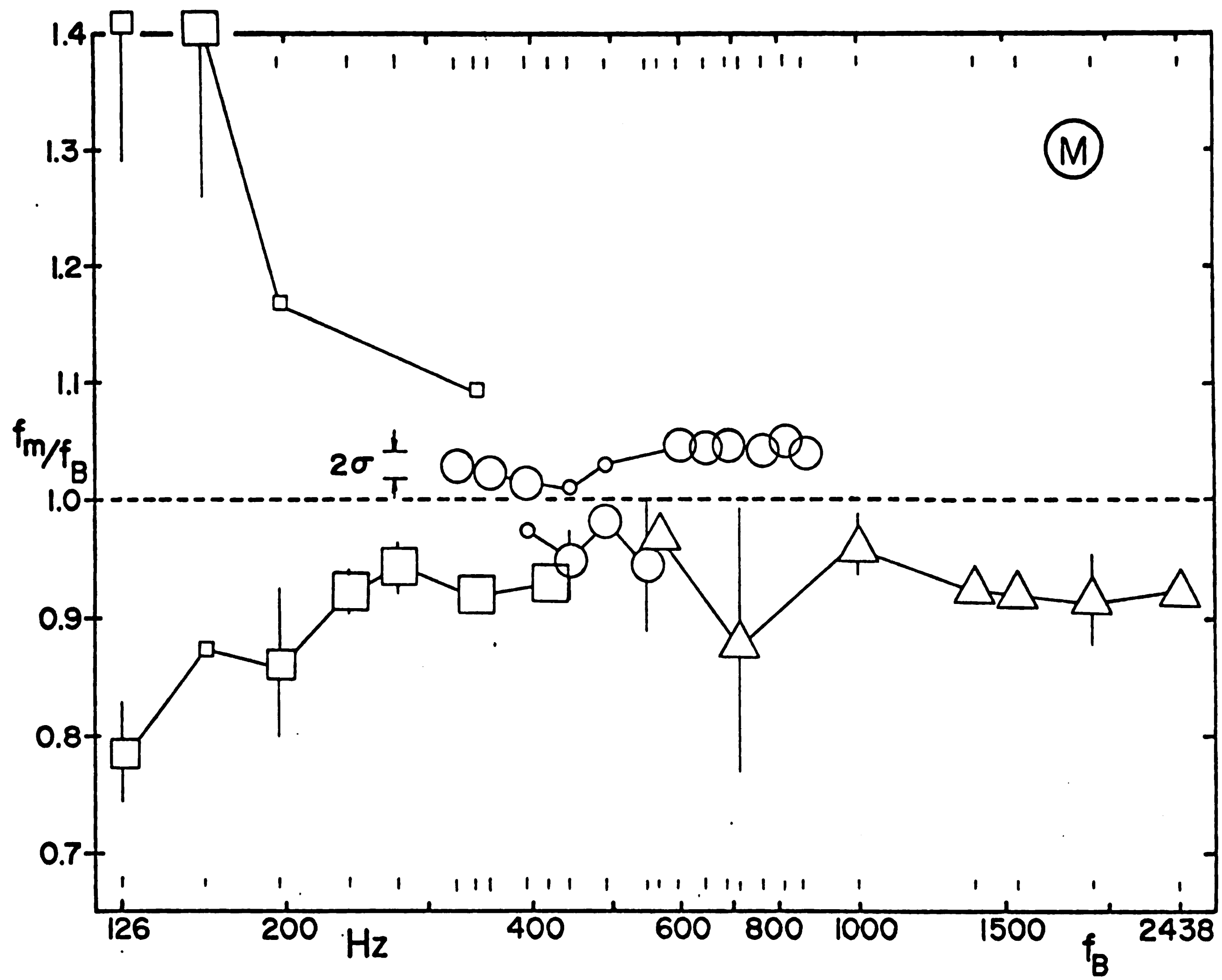

Figure 4. Data for subject M in Experiment I

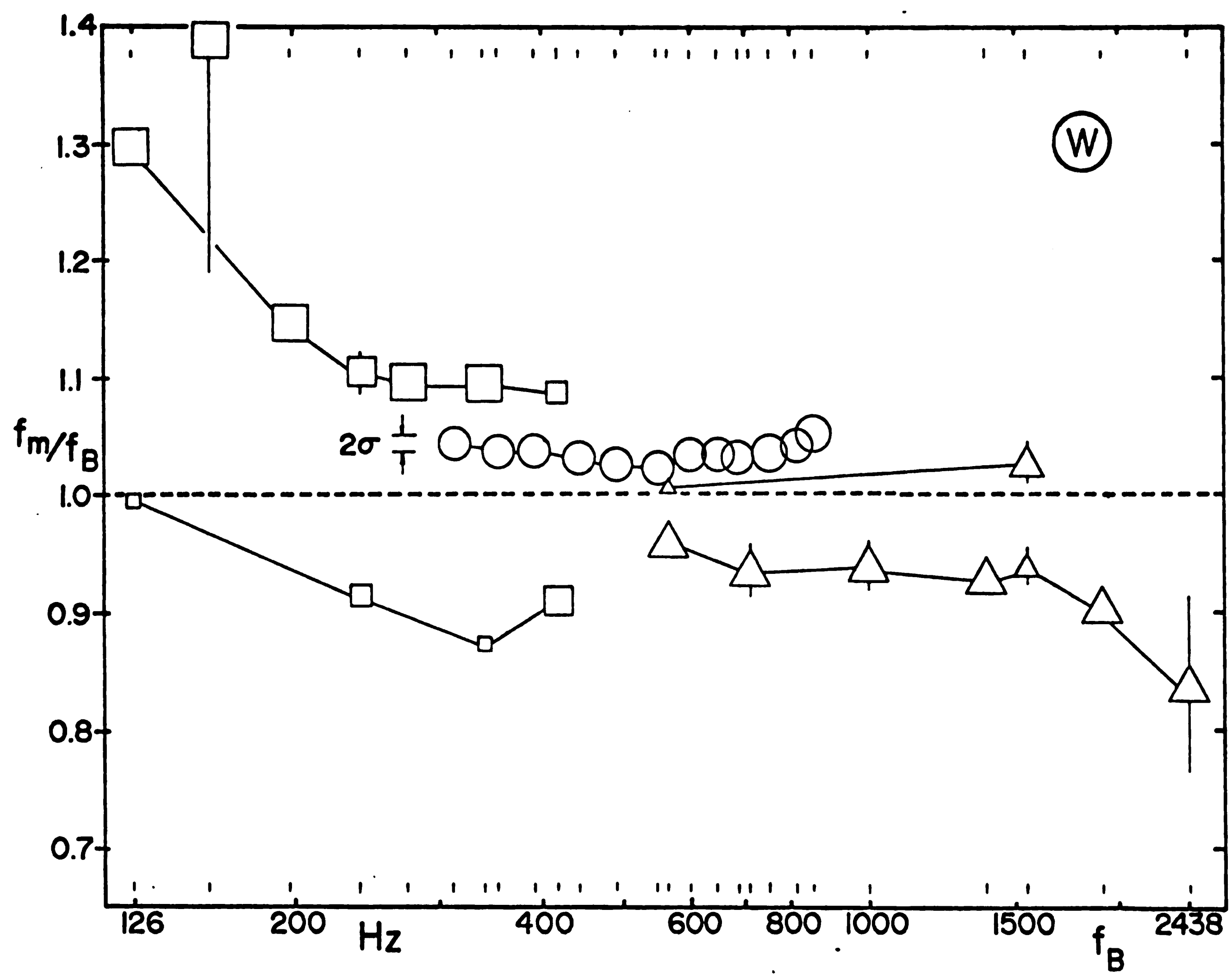

Figure 5. Data for subject W in Experiment I

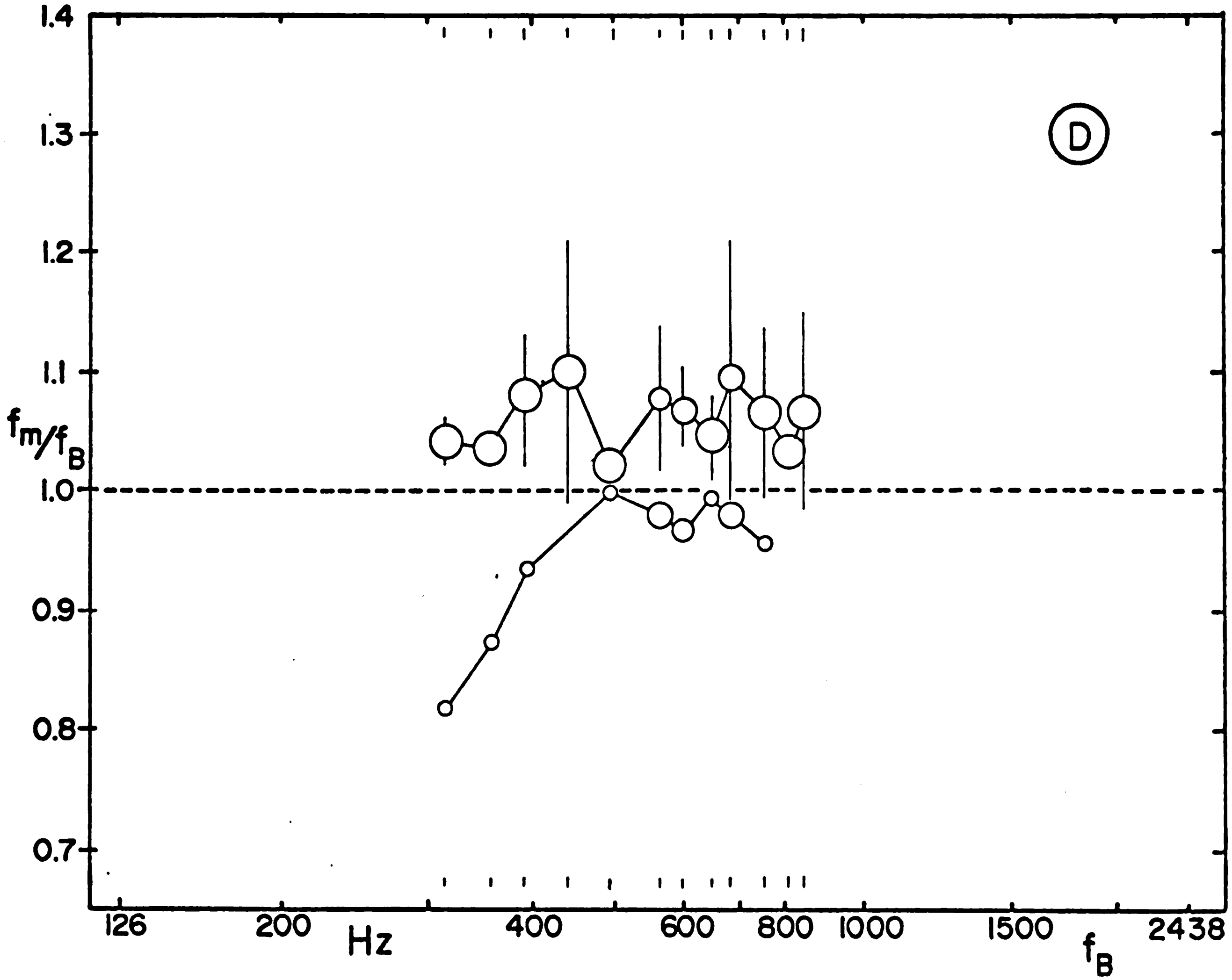

Figure 6. Data for subject D in Experiment I. Only middle frequency range tested.

E. Low and High Range Experiments

Subjects G, M, and W also matched pitches to dichotic noise with phase boundaries in the low and high ranges, respectively 126-420 Hz and 567-2438 Hz. Each of these frequency ranges included only 7 different phase boundary frequencies. In all other respects the experiments were identical to the middle range experiment. These frequency ranges are of interest in testing the range of existence of the BEP. It was expected that 126 Hz be below the range of existence for the BEP. 2438 Hz was thought to be above the existence range of the BEP because it is above the frequency range others have indicated is the limit of such binaural effects reviewed by Bilsen (1977). According to Cramer & Huggins (1958), the Huggins Pitch exists in the range of 200 Hz to 1600 Hz. Bilsen (1977) limits binaural pitch effects to 2000 Hz.

F. Low and High Range Results

The squares and triangles in Figures 3-6 show the results of the low and high range experiments respectively. Again the sizes of the symbols represent the relative fraction of responses above and below the phase boundary frequency. The following observations might be made about the low frequency range data:

1. The deviation of the matching tone from the phase boundary frequency increases at lower frequencies.
2. The error also tends to increase at lower frequencies.
3. All 3 subjects had responses both above and below the phase boundary frequency. Subjects G and M showed a preference for matching on the low side, while subject W continued to prefer the high side.
4. It is not unreasonable to say that the higher phase boundary frequencies in this range yielded responses that approached the middle frequency range responses where these ranges overlapped.

The following observations might be made regarding the high frequency range data:

1. The highest phase boundaries produced increased deviations of the matching tone from the phase boundary, but not to the same degree as the low frequency range.
2. The errors, again, tend to be larger in this extreme frequency range.
3. Only two of the 3 subjects responded with matches both higher and lower in frequency than the phase boundary frequency.

Subjects G and W had bimodal response distributions with W showing a strong preference for matching on the low side. This is completely opposite to W's preference for matching on the high side in both the low and middle frequency ranges. Subject M had a unimodal response distribution, always matching on the low side. This is similar to his preference for the low frequency range, but contrary to his preference in the middle range.

4. Again the matches at the low end of this frequency range coincided well with the higher phase boundary frequency matches of the middle range.

G. General Observations

The results from all 3 frequency ranges make it clear that the responses distribute well into two groups, one at higher frequencies than the phase boundary and one at lower frequencies. The sizes of the error bars show that the separation of these two distributions is quite complete. Very few of the data points or error bars lie on or across the phase boundary frequency. This is quite different from the Huggins Pitch where the perceived pitch is supposed to be exactly at the center of the phase transition region.

The results also show that deviation of the matching frequency from the phase boundary increases at both frequency extremes, but less so at the highest than the lowest phase boundaries. The errors at these extremes tend to become large, but not large enough to say the perception of a definite pitch has disappeared. If the BEP effect had completely disappeared at either extreme we would expect the errors to have been even larger showing a random distribution of matches.

In the frequency intervals where the ranges overlap the relative agreement between the matches from the different ranges provides evidence that the percieved pitch is due primarily to the phase boundary. For example, in the data for subject G at 687 Hz (middle range) the fundamental frequency is 6.87 Hz, the components are all separated by 6.87 Hz, the highest component is at 1725.4 Hz, and the boundary frequency is at 40% of that maximum frequency. The mean response is 4.2% above the phase boundary. At 712 Hz (high range) the fundamental frequency is 3.95 Hz, the highest frequency component is at 992.2Hz, and the boundary is at 72% of that highest frequency. Despite these major changes in the other characteristics of the stimulus the pitch matches still average 3.7% above the phase boundary frequency. Data for all 3 subjects show the same pattern. Major variations in the fundamental frequency, component separation, maximum frequency component, and placement of phase boundary within the noise band do not greatly alter the deviation of the matched frequency from the phase boundary frequency. We can conclude that the stimulus artifacts created by changing the placement of the phase boundary within the noise band and our stimulus production technique are not the major determinant of the BEP effect.

H. Qualitative Results

The matching task was initially difficult for the subjects. For an unpracticed observer the initial 12 stimulus run required about 45 minutes. With succeding runs this time was reduced to about 15 minutes. On an informal basis 6 other listeners attempted the task. Only 2 produced pitch matches consistently near the phase boundary. The others responded nearly randomly, with, at most, only 4 matches near the boundary. The experience of subject R indicates that some subjects may never be able to learn to perform the task.

Those subjects that could reliably match pitches to the BEP reported that the perceived pitch sensation sounded like a very narrow band noise added to the wide band noise. For these reliable subjects the pitch sensation was similar in strength and character to the Huggins Pitch. The pitch sensation tended to be localized toward one ear, but diffuse relative to the diotic matching tone on the third interval. This asymmetry occurs because only one channel changes from the first to the second interval. Reversing the head phones reverses the direction of the asymmetry.

The process involved in matching the sine tone to the BEP varied in difficulty and character from trial to trial. Sometimes the pitch sensation seemed stronger. When stronger, the matching task was much easier. Sometimes the pitch sensation was perceived immediately at the beginning of the trial while at other times random "searching" with the matching tone was required. As the matching tone approached the vicinity of the boundary frequency the BEP suddenly "popped out" and then was quite easily matched. At times the edge pitch was quite "ellusive" as the final adjustments of the matching tone were made. As the subject slowly increased the frequency of a matching tone that sounded flat, the matching tone suddenly was quite sharp. This evasion occurred as the matching tone approached the BEP from either direction and continued until the subject finally gave up and settled for a less precise tuning. There was no apparent correlation between these effects and the boundary frequency or range of the stimulus.

I. Conclusion

The results of Experiment I establish the existence of the BEP. Subjects can reliably match a sine tone to the pitch of the dichotic noise stimulus. The pitch matches obtained from the subjects were not exactly at the phase boundary. The matches were consistently higher or lower than the phase boundary by about 4%, in the middle range. At the extreme values of the phase boundary frequency this deviation was even greater.

Reliable pitch matches were made to all phase boundary frequencies. The increased errors indicated the weakening of the effect, but it had not yet disappeared, as expected, at the extreme frequencies, 126 Hz and 2438 Hz. Some break down in the ability of the binaural system to resolve these phase differences is expected at high frequencies when the real differences in the signals fall below the size of the errors in the system. 2438 Hz, however, is above previous estimates of that limit (Bilsen, 1977). One possible reason for our success might be the favorable conditions under which our subject perform the matching task.

The shift of the matching tone away from the boundary frequency is consistent with the hypothesis that the Equalization-Cancellation process creates a high-pass or low-pass central spectrum with a cutoff at the phase boundary.

III. Experiment II

The Equalization-Cancellation model explains the BEP by proposing the production of high-pass or low-pass noise in some internal channel. Bilsen (1977) indicates that this internally produced signal is processed by the same mechanism that processes externally introduced high-pass and low-pass noise. Fastl (1971) reports data from subjects matching a sine tone to the pitch from a high-pass or low-pass noise. The pitch associated with these noise signals was shifted into the noise, relative to the cutoff frequency. This leads to the expectation that the BEP should be shifted as well.

Fastl tested noise pitches using cutoff frequencies extending from 200 Hz to 4000 Hz, but he only showed data for 6 frequencies in between. High-pass noise was only presented at two of those middle frequencies. The detailed behavior of the BEP pitch matches has enough reliable structure to warrant reinvestigation of high- and low-pass noise pitches.

Experiment II is intended to determine the pitches elicited by high- and low-pass noise under conditions as similar as possible to those used for Experiment I. Both types of noise are tested with cutoff frequencies corresponding to each of the 26 phase boundary frequencies used in Experiment I. It is expected that the details of the obtained results will correlate well with the results of the BEP experiment.

A. Method

The procedure for this experiment was identical to the procedure for Experiment I. The stimulus was changed only in that the dichotic stimulus in the first interval of the sequence was replaced by either high-pass or low-pass diotic noise. The low-pass noise was produced by electronically adding the two original noise signals. The high-pass noise was produced by electronically subtracting the two original noise signals. The resulting noise spectra showed a discontinuous 30 dB drop in intensity at the noise band edge. This combined signal was sent in phase to both ears on the first interval. The other 3 segments of the sequence were unchanged from Experiment I. Subjects again matched a sine tone in noise on the third interval to the pitch of the edge stimulus in the first interval. The experiment was carried out with cutoff frequencies at the same 26 frequencies in 3 overlapping ranges as for Experiment I. Subjects G, M, and W participated.

B. Results

Figures 7, 8, and 9 show the results for the 3 subjects. Open symbols represent the pitch matches relative to the cutoff frequency f for high-pass noise. Filled symbols represent pitch matches relative to f for low-pass noise. As can be seen, the frequency of the matching sine tone always deviated from the noise band edge and shifted <u>into</u> the noise signal. For low-pass noise the matched pitch was below the edge frequency and for high-pass noise the matching tone was above the edge frequency. Subjects G and W commented that the pitch of the high-pass noise was easier to match that the pitch of the low-pass noise. Subject M had no preference.

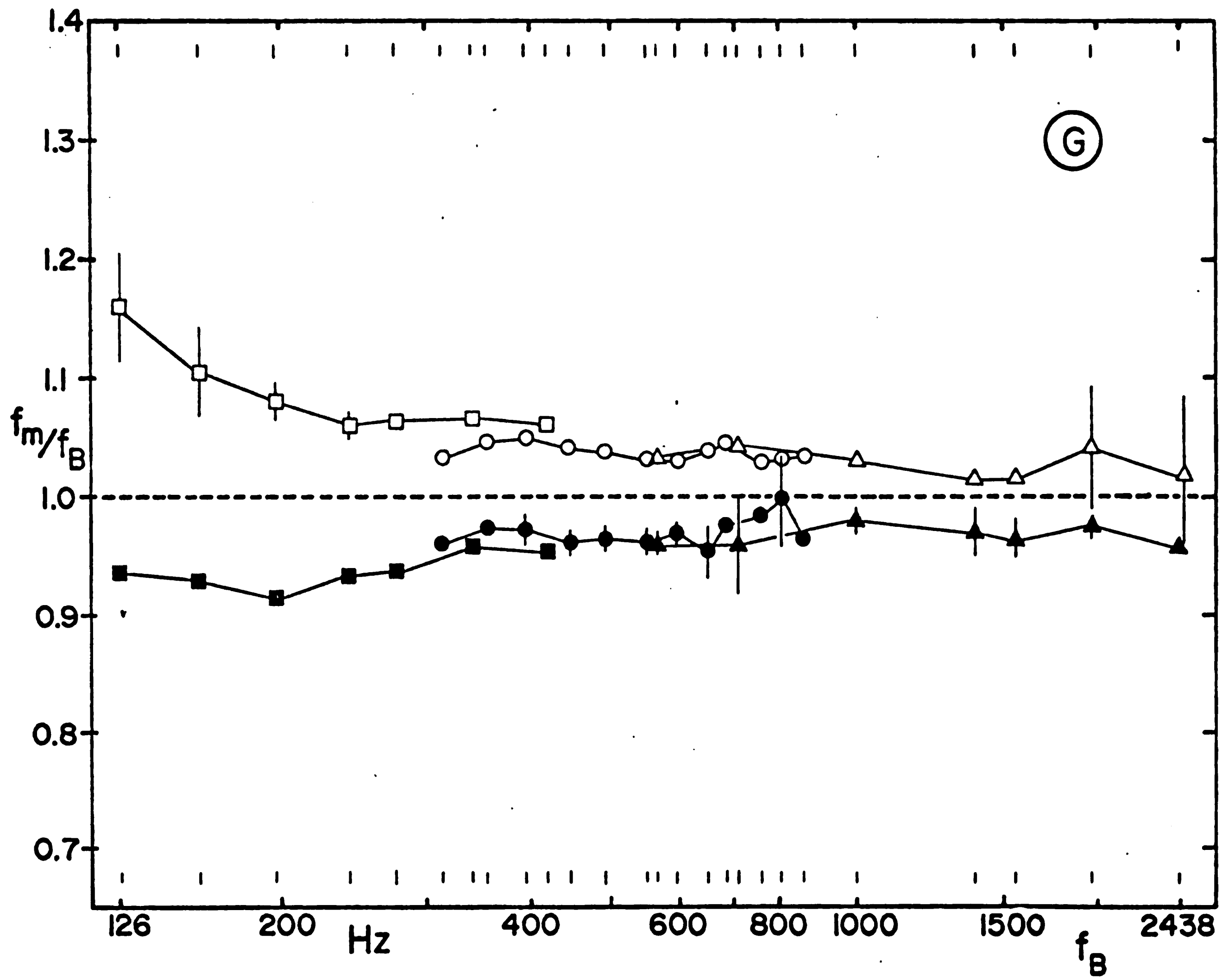

Figure 7. Data for subject G in Experiment II

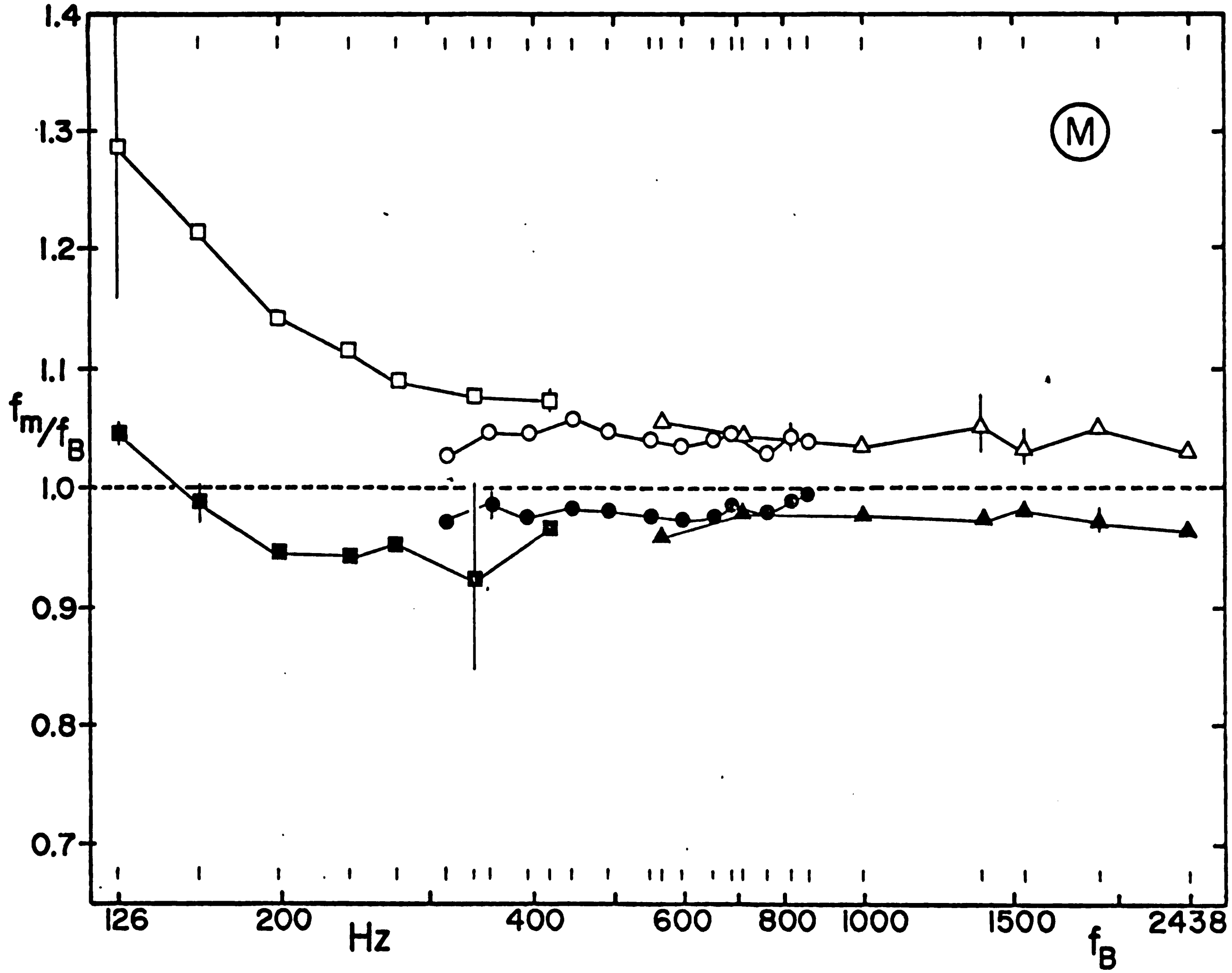

Figure 8. Data for subject M in Experiment II

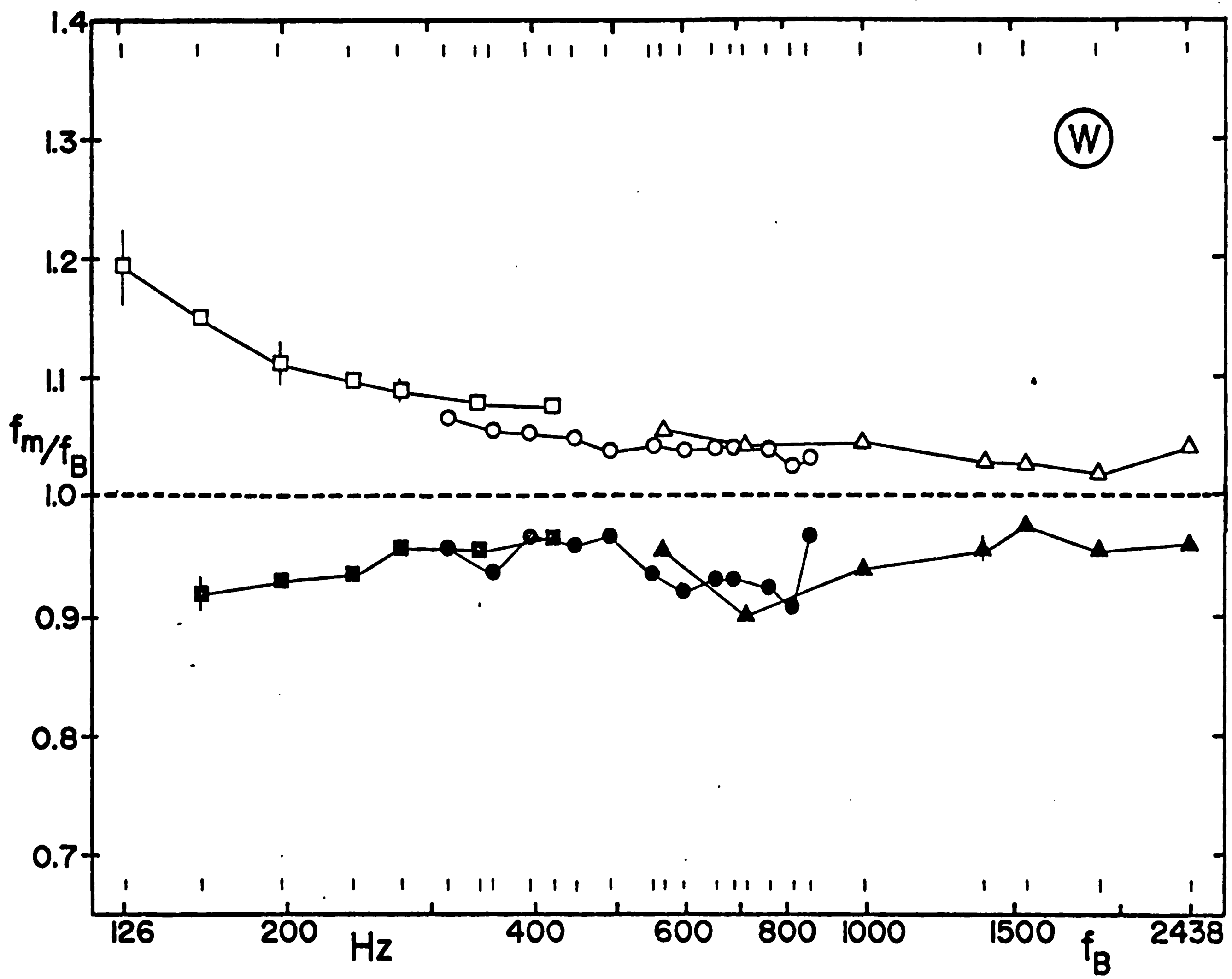

Figure 9. Data for subject W in Experiment II

C. Discussion

The similarity between the results of this experiment and the results of Experiment I are remarkable. The degree of deviation away from the boundary frequency is, within the limits of the error, nearly identical in the middle range. In the low frequency range the large increase in shift upward is quite similar. The downward shift for the low-pass noise is not quite as evident, however. At the overlap between the low and middle frequency ranges the same discontinuities are evident for all 3 subjects. Those discontinuities are small as for the BEP, but in the proper direction. In the high frequency range, however, none of the 3 subjects shows the increasing downward deviation below the boundary frequency shown by G and W in Experiment I. At the overlap between the middle and high frequency ranges the lack of discontinuity in the BEP data is also evident in the data for high- and low-pass noise.

The Equalization-Cancellation model provides a very compelling explanation for these data. The close correspondence between the results of Experiments I and II tends to justify the name, Binaural Edge Pitch, for the effect.

IV. Experiment III

In 1962 Guttman investigated the strength of the Huggins Pitch. His subjects adjusted the frequency and amplitude of a sine tone to match the pitch and pitch strength of a Huggins stimulus. Subjects found the Huggins Pitch to be 4.6 dB above masked threshold for the sine tone at the matching frequency.

Experiment III was designed to determine the strength of the Binaural Edge Pitch relative to masked threshold and relative to the Huggins Pitch. In Experiment I we showed that the BEP existed from at least 126 Hz to 2438 Hz. This contrasts with the claim by Cramer & Huggins (1958) that the Huggins Pitch only exists up to 1600 Hz. In this experiment particular attention is paid to the upper phase boundary frequencies. The stimulus set from Experiment I only included 1 stimulus with a phase boundary below 200 Hz, but 2 above 1600 Hz. By measuring the strength of the BEP in the middle and high frequency ranges more detailed information regarding the upper existence limit of both the Huggins Pitch and the BEP may be obtained.

A. Method

To make the BEP and Huggins stimuli as similar as possible for this experiment, both were generated digitally. The binaural edge generated for Experiment I involves a discontinuous transition and so we chose to use the Wightman-Grantham-Fowler (1977) version of the Huggins stimulus as our comparison. The sensations produced by these two stimuli are quite similar and so comparison is reasonable.

The BEP stimulus was produced as for Experiment I. The same spectral components were used to produce the Huggins stimulus. The one component chosen to be out of phase corresponded exactly to the phase boundary of the BEP stimulus. Each of these stimuli were presented in the same four segment stimulus structure used in the two previous experiments. The strength of the BEP and the Huggins Pitch were measured independently by having the subject adjust the sine tone in the third interval to match the pitch sensation on the first interval for both pitch and loudness. The measured intensity of the matching sine tone was recorded. The comparison of these sine tone levels for BEP and Huggins stimuli having the same phase transition frequency provides the measure of relative strength.

Masked threshold for the sine tones was determined by a method of adjustment task run in between the trials of the BEP and Huggins matching experments. After the match was made to the dichotic stimulus the first segment was removed from the sequence and the subject then adjusted the sine tone so that it was just barely audible.

Subjects M and W participated in this experment. The 19 stimuli from the middle and high frequency ranges were used. Two runs were performed by each subject in each frequency range for both the Huggins and BEP conditions.

B. Results

The results of Experiment 3 are shown in Figure 10. The loudness matches and thresholds are plotted against a vertical dB scale. The strength of the pitch sensations from the Huggins effect (filled symbols) and the BEP (open symbols) in the standard (circles) and high (triangles) frequency ranges are shown. The dashed line shows masked thresholds for sine tones of the corresponding frequencies in the two ranges. Because the spacing between frequency components of the noise varied from trial to trial with the phase transition frequency, these intensity values are all relative to the power density of the noise. This adjustment was made by subtracting the quantity ($60-10\log\Delta f$) from each of the measured values. Δf is the noise band width in Hz.

For subject W, the BEP is approximately the same strength as the Huggins Pitch up to about 800 Hz. At all higher frequencies the Huggins Pitch has a perceived strength very close to threshold. Accurate pitch adjustments could still be made, however, even though the pitch sensation was weak. The BEP remained relatively strong (6-8 dB above masked threshold) until the phase boundary reached approximately 1600 Hz. At the two highest boundary frequencies (1888 Hz and 2438 Hz) the BEP decreased in strength. It is expected that at slightly higher frequencies the BEP disappears completely.

For subject M, the Huggins Pitch is slightly weaker than the BEP from 300 Hz to 500 Hz. At all higher transition frequencies they are approximately equal in strength. At the highest frequencies, both pitch effects begin to decrease in strength relative to masked threshold. As for subject W it is expected that at slightly higher frequencies these dichotic pitch effects would decrease in strength and disappear for M as well.

The data from the two subjects show a rough equivalence for the strengths of the two dichotic pitch sensations. At lower frequencies of the phase boundary the BEP is slightly stronger than the Huggins Pitch. At higher frequencies, however, individual differences are significant. Both pitch effects for both subjects are expected to disappear by 3000 Hz.

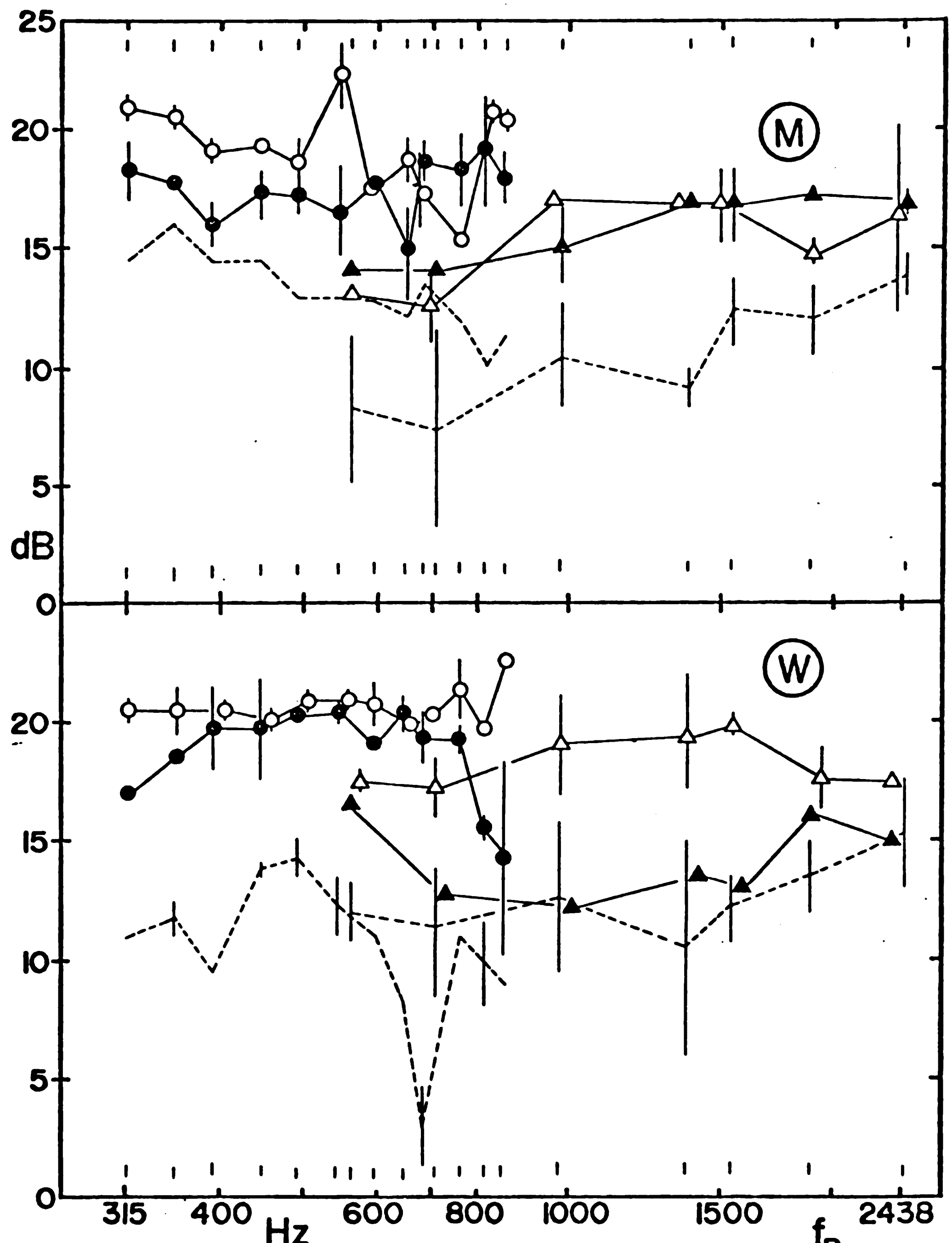

Figure 10. Data for Experiment III

C. Discussion

This experiment sets a new upper limit for this variety of dichotic noise pitch. Both the Huggins Pitch and the BEP may be matched by sine tones at phase transition frequencies up to and including 2438 Hz.

The reasons for this high existence limit seem to be related to the experimental procedure. In our experiment subjects are given repeated exposure to the stimuli. Cramer & Huggins (1958) used a forced choice task wherein the subject selected the Huggins stimulus with the higher pitch. One of the tasks used by Wightman, Grantham & Fowler (1977) was a forced choice task wherein the subject chose the dichotic stimulus over a diotic noise stimulus. Both these reports specify the limit of the effect to be below 2000 Hz, (C & H, 1600 Hz; W, G, & F, 1000Hz).

A number of other features of our experimental paradigm also contribute to the ability of subjects to hear these noise pitches at high frequencies.

1. The matching tone could serve as a probe device to focus attention on specific spectral regions.
2. Adjustment of the frequency and intensity of the matching tone by the subject allowed optimal listening conditions to be chosen.
3. The dichotic stimulus was immediately preceded and followed by diotic noise with a power spectrum identical to the power spectra of the two components of the dichotic stimulus. The subtle changes resulting from the dichotic stimulus are

emphasized by this procedure. In particular, the subject could hear the pitch in the dichotic stimulus turn on and off.

4. The diotic noise was continued through the matching interval which increased the similarity of the matching interval and the dichotic interval.

These procedural details allow easier discrimination of very weak phenomena.

V. Auxiliary Experiments

In the examination of a new phenomenon there are always a great number of variations on the original experiment to test. Below, some of the more obvious variations are described. All of these experiments were carried out, but on a less formal basis than Experiments I, II, and III. The remarkable result was that all of these experiments gave the same basic results. Subjects continued to match the BEP at approximately plus or minus 4% deviation from the phase boundary frequency.

Only subjects M and W participated in these experiments. The phase boundary frequencies were restricted to the middle frequency range, 315-856Hz. A minimum of two experimental runs were performed by each subject for each condition.

1. Reversed Discontinuous Binaural Edge: For this experiment the interaural phase differences, above and below the phase boundary, were reversed. All frequency components below the boundary were out of phase and all frequency components above the boundary were in phase relative to the opposite channel. Subject M continued to produce a bimodal distribution of pitch matches. Subject W again produced a unimodal response distribution at frequencies above the phase boundary. All deviations were approximately 4% away from the phase boundary as for Experiment I. This result demonstrates the ability of the E-C mechanism to equalize both with and without the π phase change in all components. The data for subject W indicate behavior equivalent to matching the pitch from high-pass noise for both the original and this reverse experiment. In the original

experiment a high-pass central spectrum was generated by not imposing a phase shift and in this experiment a high-pass central spectrum was generated by imposing a π phase shift internally.

2. <u>Quadrature Discontinuous Binaural Edge</u>:

For this experiment all components below the phase boundary were at $\pi/2$ phase relative to one another. All components above the phase boundary were at $3\pi/2$ phase relative to one another. The phase boundary remained as a discontinuous phase transition region of π, but all other components were now out of phase at least by $\pi/2$. This condition is unique in that taking the sum or difference with no phase alteration first results in a flat noise spectrum. No indication remains of the location of the phase boundary. Qualitatively, however, the stimulus sounds the same as the stimulus for Experiment I. The BEP can be heard and was matched by both subjects with matching tones about 4% above the phase boundary. This means that the Equalization stage of the E-C model is not restricted to 0 or π phase shifts. The fact that both subjects produced unimodal response distributions <u>above</u> the phase boundary, however, might indicate that those phase equalizations are restricted to the range of $-\pi$ to 0. An Equalization-stage phase shift of $-\pi/2$ produces a high-pass central spectrum and $+\pi/2$ produces a low-pass central spectrum. Based upon the previous data of subject M, a bimodal response distribution was expected. This possible restriction should be explored further.

3. Discontinuous Binaural Edge Superimposed on a Gradual Phase Shift: In this experiment a gradual phase shift was added to one of the stimulus channels. This was accomplished by passing the right channel through an All-pass filter. The gradual phase shift varied from π to $-\pi$. The phase shift was $\pi/2$ at 200 Hz, 0 at 500 Hz, and $-\pi/2$ at 1400 Hz. These characteristics remained constant for all 12 values of the discontinuous phase boundary. All tested phase boundaries occurred where the gradual shift was in the range of about $\pi/3$ to $-\pi/2$. This stimulus configuration approximates a constant interaural time delay plus a discontinuity. This contrasts with the previous experiments that consist of a constant (or zero) interaural phase change plus a discontinuity. The results of pitch matching to the BEP were unaffected.

In the original Equalization-Cancellation model only a single frequency-independent interval phase compensation operation is allowed. This means that noise with a constant interaural time delay could not be cancelled. The resulting central spectrum resembles comb-filtered noise. Even Durlach himself (1972) complained that the frequency independent phase shift was unrealistic. For the present experiment the binaural system must be able to equalize for both the time delay and the phase discontinuity. Two solutions are available. 1. Either the binaural system can equalize in 2 operations, one for time and one for phase, as well as for amplitude. Or 2. the binaural system is preceded by a frequency analysis allowing equalization and cancellation independently within narrow frequency regions. Within small frequency bands a time delay is equivalent to a phase shift. This latter hypothesis is consistent with current views of auditory processing (Colburn & Durlach, 1978).

4. Diffuse Binaural Edge: One of Cramer & Huggins (1958) manipulations was the width of their phase transition region. This was carried out using the BEP by allowing the interaural phase shift to vary linearly from 0 to π or π to 0 (reversed edge) over a frequency range of either 10% or 20% of the phase boundary frequency. The mean values of the matching frequencies did not change ($\pm$ 4%) nor did the errors increase significantly. This lack of change in the errors was not expected. Fastl (1971) found that errors generally increased when he decreased the filter slopes used for his high-pass and low-pass noise signals. A 30 dB drop in intensity over a 20% change in frequency is still equal to about 120 dB per octave, the steepest slope Fastl used, however. More diffuse edges should be tested.

5. Discontinuous Binaural Edge at Reduced Intensity: Both subjects performed the experiment with the signal levels at 40 dB SPL and 30 dB SPL. At 20 dB SPL the experiment was impossible to perform. For both intermediate levels the mean pitch matching results were unchanged. The errors increased. At 40 dB the error was twice that of the 60 dB stimuli. Reducing the level to 30 dB increased the errors by a factor of 2 over the error at the 40 dB level. The BEP, like the Huggins Pitch, requires moderate listening levels for best results.

6. Discontinuous Binaural Edge with Matching Tone in Quiet:

For Experiments I, II, and III the matching tone was presented with a noise background to increase the similarity between the matching interval and the dichotic interval. Egan and Meyer (1950) showed that the pitch of a sine tone was raised by a noise background. For experienced subjects, M and W, it was possible to run the experiment and eliminate the background diotic noise during the third segment. As expected, the frequencies of the matching tones were elevated. The change in mean deviation from the phase boundary frequency was up by 1% for matching tones above the phase boundary (plus 5%). For matching tones below the phase boundary, the mean deviation decreased to about 3%. The error increased as well.

7. Discontinuous Binaural Edge with Restricted Noise Band:

One of the peculiarities of the standard binaural edge stimulus was the placement of the phase boundary within the noise band. This phase boundary placement was different in each of the three frequency ranges, but the total number of components remained the same, 251. In this experiment the phase boundary remained constant between the 100th and 101st components while the upper frequency range of the noise band was reduced. Two different techniques were used. In the first, the standard stimuli were passed through an analog low-pass filter with a cutoff (6 dB down point) at 1500 Hz and a 48 dB/octatve slope. In the second technique the stimuli were recomputed to only include 151 components. The observed pitch matches were virtually unchanged indicating that the BEP is not particularly sensitive to the details of the noise spectrum.

8. Binaural Coherance Edge Pitch: Durlach (Personal Communication, 1980) suggested an experiment in which all frequency components below the phase boundary remained at 0 interaural phase while all components above the phase boundary were at random interaural phase. The E-C model, again, predicts that a edge is created in the central spectrum. This time the shape of the central spectrum above the phase boundary frequency is less well defined. In general, both subjects M and W performed as for the BEP. Subject M reported no perceivalbe difference between the Binaural Coherance Edge Pitch and the BEP. All data from M was consistent with data from the BEP. Subject W produced pitch matches consistent with the BEP data up to 763 Hz. With the phase boundary at 807 Hz, W only reported hearing a pitch on 3 of 5 presentations of the Binaural Coherance Edge Pitch. At the highest frequency phase boundary, 856 Hz, W could not hear a pitch on any presentation. Subject W reported that the pitch sensation heard at lower phase boundaries seemed to be correlated with the perceived lateralization of the pitch "image", (see discussion of Raatgever & Bilsen, 1977, below).

VI. Discussion

The existence of the Binaural Edge Pitch and its tentative explanation based upon the Equalization-Cancellation model draw our attention to two separate stages of processing. The first stage involves the binaural auditory processes that form a single output based upon a synthesis of the two input signals. The second is the pitch extraction mechanism that interprets that synthesized internal representation as having a pitch.

A. Binaural Synthesis

A comparison of the pitch matching data from Experiments I and II show that the internal representation or central spectrum resulting from the BEP is functionally equivalent to the central spectrum resulting from diotic high-pass or low-pass noise regarding pitch matching behavior. Are there other models, besides the E-C model, that predict high-pass or low-pass noise in the central spectrum given the BEP stimulus? The E-C nodel is a place-theory model. Are there models based upon neural timing that would generate the desired central spectrum?

Colburn & Durlach (1978) reviewed a great number of models of binaural processing, including the E-C model. Their review dwells upon the abilities of the models to accurately explain lateralization, MLD phenomena and binaural discrimination. They intentionally avoid the binaural-creation-of-pitch phenomena because little has been done to apply the models, except the E-C model, specifically to this problem. Presumeably any model that fully represents the binaural analysis processes of the auditory system should explain the lateralization, MLD, and discrimination results as well as the binaural-creation-of-pitch results. Drawing from the Colburn & Durlach review, there is no pure neural-timing model that can explain the first 3 phenomena. All plausible models require as a basic assumption that any binaural timing analysis carried out be specific to fibers with nearly identical characteristic frequencies. In other words, at this stage of processing the spectral components must already have been filtered into separate pathways logically ordered along some tonotopic axis, (ie. frequency-place mapping). Further binaural analysis will only occur using contralateral frequency specific pairs of nerve fibers.

According to Colburn & Durlach the E-C model is a special case of a more general model. In fact, almost all of the models reviewed can be classed as special cases of a single general model. This general model includes band-pass filtering at the peripheral level to seperate the spectral components, limited time equalization, cross-connected delay lines to generate a <u>cross-correlation function</u>, some binaural display to "interpret" the cross-correlations, and a decision process. The cross-correlation function is generated by a coincidence network where neural pulses from both ears must arrive simultaneously at a single synapse to generate an action potential in the post-synaptic neuron. Internal noise is added at a variety of locations depending upon how conveniently it may be incorporated in the particular model under consideration. The Equalization-Cancellation model fits this general model if it is based upon energy. In this case a central square law mechanism must operate after the coincidence network. The individual theories differ primarily in the binaural display stage, whose input is the cross-correlation function, and in the decision mechanism. The location of differences between pitch perception processing and porcessing for lateralization, MLD, and binaural discrimination will most likely occur at these final stages. All of the models generate roughly the same internal cross-correlation function. We will assume that the binaural display stage and the decision process incorporate the central spectrum and pitch extraction mechanisms that form the second part of the BEP processing. We now concentrate on how a pitch extraction mechanism generates the same pitch from the crosscorrelation function as from the high-pass or low-pass spectral edge.

B. Pitch of Noise Bands

Fastl (1971) and Experiment II showed that high- and low-pass noise bands have a pitch asssociated with the cutoff frequency of the band but shifted into the noise. It is conceivable that some neural-timing mechanism is capable of generating a pitch from a noise band. This is reasonable based on the fact that a discontinuity in the spectrum of a signal is associated with oscillations in the waveform at a frequency near that of the spectral discontinuity. We investigated these oscillations as the possible source of the noise band pitch by making plots of our high-pass and low-pass noise stimuli waveforms for the 3 frequency ranges. We located these oscillations by eye and determined the periodicity of each "cycle". We calculated a mean period T and standard deviation. If the subjects matched the noise bands with a pitch at frequency $f_m=1/T$ then the quantity $R=(f_B T)^{-1}$ is a reasonable estimate of the predicted pitch matches from a neural timing mechanism corresponding to the value f_M/f_B plotted in the figures.

The predictions from this model for low-pass noise were as follows: for n =40 (low frequency range) R=0.86 (23%). For n =100 (middle frequency range) R=0.78 (28%). For n_B=180 (high frequency range) R=0.61 (38%). The numbers in parentheses are the standard deviations as a percentage of the mean. The values of R are much smaller than the observed values of f_M/f_B. Also, the monotonic dependence of R upon n does not match the relatively constant nature of f_M/f_B for low-pass noise pitches. These problems, coupled with the extremely large standard deviations, provide very difficult conditions for a neural-timing model to cope with.

Visual inspection of the high-pass noise stimuli resulted in no reasonable periodicity estimates. The few oscillations that could be recognized produced values of R much larger that 1.0. This difference in the difficulty of judging the periodicities of the two wavefrm types must be considered in light of the lack of difference in the perceived strength of the pitch sensations generated by these waveforms. While it is possible that temporal fine structure and waveform envelopes within the high-pass and low-pass noise stimuli might be coded for pitch, there are also some very serious deficiencies in this information that must be dealt with. A neural-timing based pitch perception process seems to be an unlikely candidate for explaining noise band edge pitch.

An alternative temporal model for pitch perception is based upon the autocorrelation function of the stimulus. In general these auto-correlators have the temporal waveform as their input. Neural autocorrelator models have been popular mechanisms for pitch extraction (ct. Licklider, 1959 or Wightman, 1973). Klein & Hartmann (1981) have derived an expression for the autocorrelation function for digital noise. They show that the resulting function oscillates with the frequency of the spectral edge. Clearly, a mechanism that predicts the pitch of a noise band to be exactly at the cutoff frequency is not desireable. An ideal autocorrelator, such as this, cannot explain the observed shifts in pitch away from the boundary frequency. A neural autocorrelation process may be vulnerable to the same intensity-dependence mechanisms required by the models of Stern & Colburn (1978) to explain the time-intensity trading relationshiips found in localization. The mechanism Stern & Colburn use in their model to handle intensity effects is designed to process the power spectrum of the input signal. In this neural-timing based pitch extraction scheme this spectral information would serve only as a bias in the process interpretting the autocorrelation function. These intensity processes operate outside of the cross-correlation network on the separate channels. The results of this intensity operation are reinjected back into the binaural display after the cross-correlation stage. The intensity effects, possibly incorporated at this binaural display stage, affect the autocrrelation in the same way as the cross-correlation pitch is affected. If these intensity mechanisms can affect the pitch of a diotic noise band edge as generated by an autocorrelation process then a suitable dichotic stimulus passing through a cross-correlation process will be affected in the same way. Through this type of process a common mechanism may explain the bidirectional pitch shifts from both the noise band edge experiments and the Binaural Edge Pitch experiments.

Wightman's (1973) "pattern-transformation" model of pitch provides an equivalent alternative to the autocorrelation described above. Instead of autocorrelating the input waveform, however, the pattern-transformation model performs the autocorrelation on the power spectrum of the input. This allows the autocorrelator to be located after the band pass filters that divide the signal into its frequency components.

It is clear that noise band edge pitch may be heard when the stimulus is monaural or diotic. The question remains: to what extent do the monaural channels share pathways and processes with the binaural channels? Is there an autocorrelator that works on 3-dimensional activity patterns (frequency, power, time) in the same way that it works on 2-dimension activity patterns (frequency, power)?

C. Binaural Edge Pitch

The BEP data provide evidence against models of binaural pitch perception based upon pitch extraction from an ideal cross-correlation function. The noise bands used in the diotic experiment are created by electronically adding or subtracting the two channels of the BEP stimulus. By expanding the product in the autocorelation integrand derived by Klein & Hartmann (1981) for the diotic noise bands oscillations at the phase boundary frequency can be seen. These oscillations are exactly the same oscillations observed in the cross-correlation function for the BEP stimulus. The ideal cross-correlation model cannot account for the shifts in the perceived pitch of the BEP in the same way that the ideal autocorrelation model cannot account for the shifts in the perceived pitch of a noise band edge into the noise.

An ideal cross-correlation mechanism, as outlined by Jeffress (1948) in a neural coincidence network, was incorporated into Licklider's (1951) duplex theory of pitch perception to create the triplex theory of pitch perception (Licklider, 1956). This addition was designed specifically to deal with the Huggins Pitch. The Huggins Pitch is explained equally well by both the original E-C model and Licklider's ideal cross-correlation mechanism. Because of the symmetric nature of the phase shift region and the correspondence of the Huggins Pitch with the exact center of that region an ideal cross-correlation works as well as the E-C model. This is not true of the BEP. The asymmetric nature of the stimulus and the measured pitch shifts to either side of the phase boundary make the E-C model a plausible explanation of binaural pitch effects and an ideal cross-correlator implausible.

The binaural mechanism proposed by Bilsen (1977) and by Raatgever & Bilsen (1977) follows the same general model outlined by Colburn & Durlach (1978). Both the cross-correlation coincidence network of Licklider's mechanism and the equalization mechanism of the E-C model are included. The following two additions are made: 1) the intensity information from the input signals is maintained through the cross-correlation network resulting in a 3-dimensional activity pattern. The 3 dimensions include the spectral frequency along one axis, the cross-correlation or time-delay coincidence along another axis, and spectral power along the third axis. 2) The last modification is the specification of a pattern recognition process. Operating on the 3-dimensional activity pattern, this pattern recognition process feeds the decision mechanism with information required for simple detection, localization, discrimination, and pitch extraction.

This activity pattern is the central spectrum that Bilsen (1977) refers to. Pattern recognition for pitch finds and reports evidence of significant spectral properties at specific spectral locations. This process works identically on diotic noise band spectra and the cross-correlation function of the BEP dichotic stimulus. An extra property is that the cross-correlation information present in the BEP stimulus will produce a perception of lateralization in the dichotic stimulus not sensed in the diotic stimulus. For the BEP, that lateralization is correlated with the boundary frequency.

This pattern recognition process is capable of bidirectional pitch shifts with no added complexity. For the BEP the direction of that shift depends upon the time equalization added prior to the cross-correlation network.

VII. Conclusion and Summary

We have found a new dichotic noise pitch effect. A sensation of pitch is created by dichotic noise with a special interaural phase correlation. Components of the noise are at 0 interaural phase throughout one frequency region and at π interaural phase throughout an adjoining region. The change from one region to the next occurs over a narrow frequency region, called the phase boundary. The pitch perceived is not exactly at this boundary, but shifted by approximately 4% of the phase boundary frequency in either direction. We called this pitch effect the Binaural Edge Pitch (BEP).

BEP is strongest (4-9 dB above masked threshold) for phase boundary frequencies between 300 Hz and 800 Hz. The favorable listening conditions used in these experiments showed that the effect can be heard at frequencies as low as 125 Hz and as high as 2438 Hz. At these extremes the sensation of pitch is weaker, and pitch matching errors are larger. Characteristic pitch matching errors show the spread of matching frequencies to be 1-2% of the phase boundary frequency. The BEP does not depend upon the direction of the phase boundary. BEP is present for noise at 60 and 30 dB and for phase boundary widths that are 1/2%, 10%, or 20% of the phase boundary frequency. Qualitatively, the BEP is reported by experienced listeners, to be similar in nature and strength to the Huggins Pitch (overall phase transition from 0 to 2π across the phase boundary region), like a narrow band of noise added to wide band noise.

The BEP, like the Huggins Pitch, is easily explained by the Equalization-Cancellation model of binaural processing. Within the

E-C model the binaural system processes the dichotic noise by manipulating the interaural phase so as to produce a sharp edge in a central spectrum derived from the difference between the left and right channels. The central spectrum created by the binaural system may be equivalent to that produced by either high-pass or low-pass noise. The strongest evidence in favor of this explanation is that the bimodal distribution of BEP pitch matches above and below the phase boundary frequency corresponds well with the pitch sensations produced by high-pass and low-pass noise bands with sharp edges in the physical spectrum.

This type of explanation allows questions to be asked about two separate stages in the processing of the BEP stimulus. The first stage consists of processes specific to the binaural auditory system. Is there a pattern in the distribution of pitches above and below the phase boundary frequency? Under what conditions does the binaural system output a central spectrum equivalent to high-pass noise or to low-pass noise? Our results showed no consistent trends. Subjects often switch from one mode to another, though one subject was biased toward a high-pass central spectrum for all phase boundaries in the middle frequency range. All subjects preferred the low-pass central spectrum for the highest phase boundary frequencies. It is clear that the choice for which way to adjust the phase in the equalization stage does not depend upon an algorithm intended to minimize the noise power. At low phase boundary frequencies individual differences mask any trends at all.

The second stage of processing involves the extraction of pitch from the central spectrum. What is the representation of the dichotic stimulus and the representation of the diotic stimulus that are necessary and/or sufficient for pitch matching behavior on the two stimuli to be so similar? No clear answer was found. The BEP stimulus, however, with the bidirectional shift away from the phase boundary, provides a unique starting point for further investigations of central processes.

CHAPTER II

I. Introduction

Pitch perception is a multistage process. The physical stimulus is coded by the peripheral elements of the auditory system into a pattern of neural activity. That peripheral excitation pattern representation of the stimulus is then recoded into a central neural excitation pattern. It is from this central representation of the stimulus that the sensation of pitch is generated.

At each stage in the pitch perception sequence the information of the input signal is transformed and information is lost. The recoding of the incoming signal is affected by the general characteristics of the process of recoding and by the idiosyncratic characteristics of the individual listener. Coding of the physical stimulus into a neural excitation pattern in the periphery is guided by the general architecture of the auditory system and subject to the limitations of neurons. The pattern is further modified by the resonance properties of the particular individual's outer,middle, and inner ear and the pattern of normal versus damaged or destroyed hair cells along the basilar membrane. Thus, the excitation pattern at the peripheral level is already a caricature of the input stimulus. The transformation from the peripheral excitation pattern to the central representation of the input involves still more alterations and idiosyncratic variations. More information from the signal is lost; the sensitivities of the individual neurons add more structure to the final central excitation pattern. The central representation of the signal, from which pitch is "extracted," has many characteristics that are not directly attributable to the original stimulus.

Frequency is the physical variable describing the original stimulus that is most often associated with the sensation of pitch. Because of the many transformation points in the pitch perception process there are many circumstances that can modify the correlation between stimulus frequency and perceived pitch. The situation surrounding the particular perceptual event interacts with the processes modifying the recoding of the stimulus to change the final central excitation pattern. For two stimuli with the same frequency the central representation of those stimuli can be different enough that a single pitch extraction process will generate two noticeably different pitch sensations. Diplacusis refers to the phenomenon where a sine tone with constant frequency generates different pitches depending upon which ear is stimulated (van den Brink,1971). Thurlow (1943) has shown that yawning changes the perceived pitch of a sine tone. Clenching of the jaw also affects the sensation of pitch from a constant frequency sine tone (Corey,1950). Hartmann (1978) has shown that the pitch of a sine tone changes with the amplitude envelope of the tone.

There has been a continuing controversy regarding the form of the neural representation of sounds (cf. Nordmark, 1970). Is the neural activity pattern from which pitch is extracted a pattern in time or in space? Is the relevant information for pitch a function of the time of neural impulses or a function of which neuron is firing? If the relevant information is represented by the timing properties of neural spikes then that information is closely linked to the frequency of the stimulus. Ward (1963) pointed out that, given circumstances like those listed above, the timing characteristics of the excitation remain fixed with the frequency of the stimulus. The pitch is expected to remain constant. Because the pitch does change, those effects are evidence in favor of neural excitation patterns with the relevant information for pitch coded across space.

The coding of information for pitch across space is called the place theory of pitch. The frequency components of the input stimulus are each detected and coded by different locations along the basilar membrane. That spatial mapping of frequency is preserved in the transmission of the neural excitation patterns from the peripheral to the central mechanisms. This spatial array is called the tonotopic axis. The central neural excitation pattern along the tonotopic axis is the coding of the input stimulus from which the sensation of pitch is extracted.

The goal of pitch perception research is to discover the mechanism by which the sensation of a specific pitch is extracted from the central neural excitation pattern. The strategy of this study involves the utilization of assumed modifications to the excitation pattern as it is transmitted from the peripheral to the central levels of processing. By modelling the changes in the excitation pattern a representation of the final central excitation pattern is produced. Candidate models of the pitch extraction process are used to predict changes in the perceived pitch. A good model of this process will extract the correct pitch from the altered excitation pattern. The correct pitch is defined by measurements of pitch perception behavior for human listeners.

We begin with the assumption that the central excitation pattern for an acoustic stimulus is modified by the sensitivity of the listener to the individual frequency components of the stimulus. By measuring these sensitivities we learn about specific changes made in the overall shape of the excitation pattern. These changes affect the percieved pitch of the stimulus. We take the point of view that sensitivity is reflected in the curve of absolute threshold for hearing. We measure sensitivity by measuring threshold.

The next step involves the phenomenon of changes in perceived pitch due to changes in stimulus intensity (Fletcher, 1934). Stimulus intensity is assumed to make predictable alterations in the excitation pattern. These changes in the pattern result in significant changes in the pitch extracted by the mechanism we are attempting to learn about.

These two elements are combined in the following way. Changes in the shape of the central excitation pattern are calculated based upon the curve of sensitivity for an individual listener. The excitation patterns for high and low intensity tones are calculated. The two calculated excitation patterns are processed by the candidate pitch extraction models to predict a change in the perceived pitch of a tone across the two intensity levels. These predicted pitch changes are compared to measured pitch shifts with intensity changes for the same listener whose sensitivity curve was used in the earlier calculations.

In the sections that follow the components of this study are further explained. The next two sections review the literature relevant to the phenomena of pitch changes with intensity and threshold microstructure. Subsequent sections describe the two data collection procedures, threshold measurement and pitch shifts with intensity, and the models of pitch extraction to be evaluated. The final section summarizes the results of this study and the direction it gives to further research.

II. Chronology of Pitch-Intensity Studies

This section reviews the literature examining the effects of intensity on the perceived pitch of a tone. This phenomenon is basic to pitch perception and is crucial in the present experiment as a test of pitch extraction models.

A1. Early pure tone research-pre-1960

As early as 1828, Weber (Wever, 1949/1970) noted that the pitch of a tuning fork rose as its sound died away. Helmholtz (1863/1954), however, did not acknowledge of this phenomenon stating flattly, "Pitch depends solely on the length of time in which each single vibration is executed, or, . . . on the number of vibrations completed in a given time." (p.11a). Helmholtz missed a chance for a test of his resonance theory by denying this effect. Pitch changes with intensity were not missed by Helmholtz' contemporaries, though, as Thurlow (1943) referred to 7 different investigators attempting to explain a lowering of pitch with increasing intensity, from 1863 to 1899.

Zurmuhl (1930, cited in Wever, 1949) was the first investigator of the electronic age to investigate the phenomenon in detail. Working with frequencies from 128 Hz to 3072 Hz he verified the effect for low frequencies, but found little or no effect of intensity on pitch for the high frequencies in this range.

Fletcher (1934) proposed a standard for measuring pitch. Assuming the existence of an effect of intensity on pitch, he proposed that all pitch comparisons be made to tones at intensity levels set by the Fletcher & Munson (1933) equiloudness contour for 40 dB. Using this standard the pitch of any tone of arbitrary frequency and intensity is assigned a pitch value by adjusting the frequency of a pure tone at the specified loudness level (40 dB) so that its pitch matches the pitch of the target tone. The pitch value is the measured frequency of the standard intensity tone. Fletcher verified the findings of Zurmuhl, using his standard, for loudness levels of 60, 70,80, and 100 dB at frequencies from 50 to 10,000 Hz.

Not all later investigators followed the lead of Fletcher by using his proposed standard, but many were mindful of his distinction between loudness level and simple intensity in regards to their work.

Stevens (1935) used a method in which the subject adjusted the amplitude of one tone until the perceived pitch of the tone matched the pitch of a second tone at a fixed frequency difference. Using this method and showing the data from only one subject he reported much larger pitch changes with intensity than most studies before or since his work was published. Ward (1970) criticized this method pointing out that given the understanding that the effects of intensity on pitch are, in general, quite small then large variations in intensity will yield pitches that still acceptably match the target tone. Also, large frequency differences may have exceeded the limits of the stimulus generation system to allow an appropriate match, if indeed a match could have been made at all. The subject may simply have set the variable intensity to its largest difference and claimed an adequate match.

Despite its flaws, Stevens (1935) is often cited and his results have served as a standard against which many later studies compared their results. For this reason Stevens' general results will be referred to throughout the remainder of this paper as Stevens Rule. Stevens Rule may be sumarized as follows:

1) The pitch of a high frequency tone rises with increasing intensity.

2) The pitch of a low frequency tone goes down with increasing intensity.

3) The pitch of a middle frequency tone (800Hz to 2000 Hz) is relatively unaffected by intensity changes.

Snow (1936) followed Fletcher's (1934) recommendations for measuring pitch and found very large differences between observers and some very large shifts in pitch at low frequency. For one observer, an octave shift downward for 240 Hz at 120 dB loudness level was measured. Pitch shifts with loudness at levels below 120 dB were much less extreme, generaly below 10%. Two observers showed no pitch shifts at any loudness levels. This work gave credibility to the large pitch shifts reported by Stevens (1935), but also showed the extremes of subject variability that must be recognized. Single subject research on this topic is clearly not acceptable.

In 1943, Thurlow simulated a loudness change by adding a tone of the same loudness and frequency to the opposite ear. The object was to determine the frequency change necessary in the monaural tone to match the pitch of the binaural tone. At both 200 Hz and 400 Hz the binaural stimulus was perceived to be lower in pitch. The 200 Hz stimulus showed a much larger drop in pitch. In a second experiment tones of different frequencies were added to the opposite ear. Pitch changes in the target tone due to the contralateral tone were in the directions of Stevens Rule for all contralateral frequencies. Target and contralateral tones ranged from 120 Hz to 4600 Hz. The largest pitch shifts were found when the contralateral tone was close in frequency to the target tone.

In one of Thurlow's (1943) control experiments subjects were asked to simulate yawning to see if bilateral muscle contractions would effect the perceived pitch. There were pitch shifts for some subjects, but the directions of the shifts were inconsistent. A similar finding was reported in a letter by Corey (1950) indicating a consistent rise in pitch for a 400 Hz pure tone when subjects clenched their jaw tightly. The degree of pitch shift depended upon the amount of force exerted.

Morgan, Garner, & Galombos (1951) used frequencies from 125 Hz to 8000 Hz and found pitch shifts matching Stevens Rule. The pitch shifts were small relative to Stevens (1935) and Snow (1956) but individual results were not shown. Inter-subject variability was large so they chose to show only medians and inter-quartile ranges for their data.

A2. Modern pure tone research

Small & Campbell (1961) were unable to match Stevens Rule for a 100 Hz pure tone. They found that the pitch rose as intensity increased from 30 dB to 60 dB SL.

Cohen (1961) examined the pitch-intensity problem with respect to the pitch matching abilities of the subjects for pure tones at identical intensities. His results cast doubt on any previous research showing large changes in pitch with intensity (i.e. Stevens, 1935; and Snow, 1936). Pitch matching errors were nearly as large as the measured pitch shifts with intensity. Correcting for those errors, however, resulted in shifts following the general directions of Stevens Rule, but the size of those shifts rarely exceeded 2% of the starting frequency. Cohen concluded that subject variability should be watched more closely. He suggested that there is a more complex relationship between intensity and pitch than revealed by Stevens Rule.

Cohen lamented the problem of large inter-and intra-subject variability but failed to point out that in some cases changes in the ability of a subject to match pitches must parallel the size of the pitch shifts due to intensity changes. It might be the case that the actual pitch shifts are difficult to observe simply because both the variability and the shifts have the same underlying mechanism.

In a study by Terhardt (1974) pitch shifts following Stevens Rule were found. Terhardt wished simply to present one more piece of evidence to clarify the status of the pitch-intensity effect. He concluded that pitch shifts with intensity are real and not just artifacts as stated by Cohen (1961).

The most comprehensive study of the effect of intensity on the pitch of pure tones was the work of Verschuure & van Meeteren (1975). Seven loudness levels of the comparison (variable frequency) tone were completely crossed with the same seven levels of the test tone, all at frequencies from 500 Hz to 8000 Hz. Unfortunately only 2 subjects were used. Their results agreed, in general with Stevens (1935), but their conclusions specified qualifications to supplement Stevens Rule. 1) Averaged data followed Stevens Rule, but inter-subject variations make many observed pitch shifts statistically insignificant. The standard deviation should be used with caution. 2) Individual subjects' data deviated from Stevens Rule primarily by being nonmontonic.

A3. Early complex tone research - pre-1960

Fletcher (1934) conducted two experiments to examine the effect of intensity on stimuli with multiple frequency components. In his first experiment he showed that two 100 dB loudness level tones, matching standard 40 dB loudness tones of 200 Hz and 400 Hz, sounded successively were perceived to be an octave apart. When sounded simulataneously, however, they were judged quite discordant. His second experiment showed that the change induced by intensity in the pitch of a 200 Hz sine tone was 5 times greater than the induced change in a complex tone that had a perceived pitch of 200 Hz, but no spectral energy at that frequency. When these two stimuli were combined they did not sound discordant, but the periodicity pitch of 200 Hz was strengthened. At the high loudness level the combined signal shifted downward in pitch only slightly more than the original complex signal.

This might be interpretted as showing that a complex tone is much less susceptible to pitch changes with intensity. The higher frequency components, by Stevens Rule, shift less than, or in the opposite direction from, the low frequency components. The average shift of the entire complex tone is less than the shift of any one of the lower frequency components taken alone as a pure tone. A less passive interpretation might be that components of a complex tone interact in some way to resist pitch shifts.

Lewis & Cowan (1935) utilized the logic of Fletcher's (1934) experiment and asked trained musicians to play musical intervals on their violins and cellos at pianissimo and fortissimo levels. The two tones of each interval were played successively and the frequencies of the resulting tones were measured. No changes in the frequencies of the tones played were found. They concluded that complex tones were not subject to changes in pitch as intensity changed. This agreed with the experiment by Fletcher (1934) using a two tone complex.

In one of Thurlow's (1943) control experiments an auxillary tones was added to the monaural pure tone stimulus. If the extra tone was at low intensity no pitch shift was found in the original stimulus. If the extra tone was at the same level as the original stimulus a pitch shift was found. The pitch shifts followed Stevens Rule for all frequencies of the auxillary tone, whether above or below the original tone in frequency. The pitch shifts observed were smaller than the shifts seen when the auxillary tone was added to the opposite ear (see description of Thurlow's pure tone research in section II.A1. above).

A4. Modern complex tone research

Small & Campbell (1961) looked at pitch shifts for filtered pulse trains. All stimuli produced pitches of approximately 100 Hz, but 2 of the 3 different pulse trains had very little energy at that frequency. They were unable to observe monotonic pitch shifts for any of the pulse trains as stimulus intensity was increased.

In 1975, Terhardt published the results of research using complex tones having a fundamental frequency of 200 Hz and all harmonics up to 8000 Hz. By filtering out portions of the harmonics he was able to show that the pitch shift of the complex tone, relative to a pure tone at 200 Hz, depended upon the pitch shifts specified by Stevens Rule for the remaining harmonics. The pitch of a complex tone having upper harmonics of 200 Hz from 2800 Hz to 8000 Hz shifts up as intensity increased.

B. Summary of previous research

The effect of intensity on pitch has been noted and studied for more than 150 years. Recent studies (within the last 50 years) have shown the following general rules to apply:

1) The pitches of low frequency pure tones (less than 1000 Hz) go down with increasing intensity, (Fletcher, 1934; Stevens, 1935; Snow, 1936; Thurlow, 1943; Morgan,etal.,1951; Terhardt, 1974; Verschuure & van Meeteran, 1975).

2) The pitches of high frequency pure tones (greater than 2000 Hz) go up with increasing intensity, (Stevens, 1935; Morgan et al., 1951; Terhardt, 1974; Verschuure & van Meeteran, 1975).

3) The pitches of middle frequency pure tones (800 Hz to 2000 Hz) change very little with changes in intensity, (Fletcher, 1934; Stevens, 1935; Thurlow, 1943; Morgan et al., 1951; Cohen, 1961; Terhardt, 1974; Verschuure & van Meeteren, 1975).

4) Intersubject variability is extremely large and makes averaged pitch shifts appear statisticaly insignificant, (Cohen, 1961; Verschuure & van Meeteren, 1975).

5) Complex tones show very small changes in pitch with intensity (Fletcher, 1934; Lewis & Cowan, 1936; Small & Campbell, 1961; Terhardt, 1975), but the direction of observed shift is a function of the frequencies of the individual harmonics Terhardt, 1975).

C. Variability of Pitch-Intensity Effects Over Time.

Three of the reports described in the previous sections reported on changes in pitch-intensity effects when subjects were tested more than one time.

Thurlow (1943) specifically tested for changes in pitch-intensity effects over time. In 4 sessions, each separated by at least one day, subjects repeated measurements of pitch matching behavior. Percent pitch change was plotted for each of 5 subjects in each of the 4 sessions. The largest reported change from one session to the next was for subject C. At 200 Hz, subject C showed a pitch shift with intensity of -6% during session 1 and -3% during session 2. Changes of 2 percentage points were not uncommon for the 5 subjects at this frequency. At 400 Hz the changes from session to session were not nearly as large, but the amount of pitch shift with intensity was smaller to begin with.

Morgan, Garner, & Galombos (1951) did not study time effects specifically, but reported that two sets of data from one subject (an author) taken 3 years apart looked so different the two data sets were treated as if from two separate subjects.

Cohen (1961) included a test-retest condition in his study. Subjects participated in identical experiments 2 days apart. When combined with three other conditions in a 4-way analysis of variance test-retest was the only condition not producing significant results. In fact, all interactions involving the test-retest condition were also non-significant. The conclusion is that no changes in pitch-intensity effects were found to have occurred over the time period of this test.

D. Models and explanations

Zurmuhl (1930) explained the decrease in pitch for low frequencies at high intensity using Helmholtz' (1863) resonance model for the basilar membrane. Strongly vibrating units would be tenser and therefore resonate at a higher frequency. A high intensity low frequency tone would therefore match a lower than normal element of the basiler membrane. The perceived pitch would thus drop to the characteristic pitch of the lower unit. This explanation fails to deal with the effects at high frequencies, however.

Stevens (1935) attributed the pitch shift to changes in "the point of maximal stimulation on the basilar membrane" away from the region of greatest sensativity when the intensity increases. He did not, however, indicate why it would shift. Stevens & Davis (1938) attributed the shift in the maximal stimulation point to an asymmetric spread in the mechanical disturbance of the basilar membrane as its motion becomes nonlinear. They point out, however, that this only happens at relatively high intensity levels. At moderate or low intensities the point of maximum response along the basilar membrane is clearly distinguishable. For greater intensities, the surrounding portions of the basilar membrane may continue to increase their response after the initial maximum reaches saturation. At this point the "overall resonant properties" of the ear would determine the direction of the skew in excitation growth.

Thurlow (1943) concluded that the binaural pitch shift he observed was evidence against "mediation of pitch by the position of maximal stimulation" at the peripheral level. Based on this same spread of excitation phenomenon and Thurlow's results Wever (1949) concluded that the pitch-intensity effect is really only an illusion.

Bekesy (1960) briefly examined this problem attempting to determine whether a mechanical change in the response of the basilar membrane occurred at high intensities. Using a 200 Hz pure tone as the stimulus, the pattern of response of the basilar membrane did, indeed, shift toward the apex at high intensities. He attributed this to an increase in the stiffness of the membrane. The shift occurred only at extremely high intensities, at or near the level of the threshold of feeling. He concluded that the pitch-intensity effect at low frequencies must be a largely neural phenomenon. This mechanical shift might, however, explain the extreme pitch shifts found by Snow (1936) at 120 dB loudness levels.

Terhardt (1974) cites Zwicker & Feldtkeller's (1967) work on masked thresholds of pure tones masked by narrow-band maskers as evidence that the principal excitation produced by pure tones shifts toward higher frequencies as intensity increases. This effect only works at high frequencies. At low frequencies, the partial masking effect due to the decreasing sensitivity of the auditory system (equivalent to a low-pass noise) decreases as the intensity of the pure tone grows and so the pitch decreases, (see the review of Egan & Mayer, 1950, below).

Evans (1977) showed that there is a shift in the critical frequency of single neural fibers in the cochlea as intensity level increases. Fibers with a critical frequency below 1000 Hz show an upward shift as the intensity of the stimulus increases. The shift in critical frequency for fibers with a critical frequency above 1000 Hz is downwards. At low frequencies, for example, this means that a higher frequency is required to maximally drive a given fiber. If this single fiber cues the sensation of a given pitch, then the pitch sensation of the particular input frequency is now mediated by a fiber that would otherwise generate a pitch sensation for a lower frequency. The perceived pitch of the input tone has been effectively shifted downward.

Analogously, Rhode (1971) reported that the maximum mechanical response of a particular point on the basilar membrane was produced by lower frequencies as intensity increased. He obtained his data in the 6000 Hz to 9000 Hz frequency region at 70, 80, and 90 dB and so, by the same analysis as for the Evans (1977) results above, Stevens Rule is matched.

III. Chronology of Threshold Microstructure Investigations

A. Introduction.

The aim of this project is to evaluate models of pitch extraction. Our chosen method depends upon a relationship between threshold microstructure and pitch shifts with intensity.

Verschuure & van Meeteren (1975), reviewed in the previous section, showed that Steven's Rule is correct in general, but violated in detail. With increasing intensity, low frequency tones become lower in pitch and high frequency tones become higher in pitch, but individual subjects show non-monotonic structure in their individual response patterns of pitch change with intensity. We take the point of view that these non-monotonic details are associated with fine structure in the sensitivity patterns along the basilar membranes of individual listeners. We believe that we can learn about these sensitivity patterns by measuring the detailed audiograms of listeners. This section reviews the previous research on threshold microstructure.

B. Threshold Microstructure Existence

The earliest article describing threshold microstructure was by Elliott (1958). This work showed the fine structure in the audiograms of two subjects, measured at 10 Hz intervals from 400 Hz to 3000 Hz. They showed changes in threshold of nearly 10 dB within 30 Hz. Other listeners, data not shown, also showed the same degree of structure in their audiograms. The peaks and valleys of threshold were, however, located at different frequencies for different listeners. In fact, the opposite ears of individual subjects showed microstructure located at different frequencies from the structures of the first ear tested in the same subject. Elliott concluded that these audiogram "ripples" were a universal phenomenon.

In research attempting to relate diplacusis to threshold microstructure, van den Brink (1971) also reported observing audiogram fine structure. It is not clear from this paper at what intervals the threshold measurements were taken, but the plots of threshold microstructure show structures approximately the same size as reported by Elliott.

Thomas (1975) reported finding threshold microstructure with valley to valley widths corresponding to critical bandwidth. Measurements were made from 200 Hz to 5000 Hz. The reported intensity difference from a peak to an adjacent valley was typically 12 dB. Thomas also reported that the perceived pitch of near threshold sine tones jumped quantally as the frequency moved from one valley region to the next.

Kemp (1979b) described an investigation of threshold microstructure as part of the body of evidence supporting his theory of evoked cochlear mechanical response described below. He reported threshold microstructure matching the results of both Elliott and Thomas. Using a method of adjustment procedure he reported that frequency intervals of 0.5% and intensity intervals of 5 dB were adequate for resolving the microstructure in the audiogram. Most of his microstructure study was carried out by measuring perceived loudness. The peaks in the loudness curves corresponded exactly with the frequencies having the lowest thresholds. The mean separation between loudness maxima for four subjects showed roughly equal separation intervals (in Hz) for frequencies below 1 KHz. Above 2 KHz the mean separation between peaks (in Hz) increased geometrically with frequency.

Kemp was able to shift the loudness maxima by 5% by changing the ambient displacement of the ear drum. By plugging the ear canal and applying pneumatic pressure the inward or outward displacement of the ear drum could be changed. The shifts in the loudness maxima lasted as long as the pressure was applied.

Wilson (1980) was able to modify the microstructure of the audiogram by changing the middle ear pressure. Pressure changes were "induced hydrostatically by body tilt". Microstructure was gradually reduced and finally disappeared (the ratio of peak to valley intensities went to 1.0) with little change in mean threshold as the body was tilted from upright to horizontal. As the body was tilted even further the threshold fine structure reappeared, but with maxima and minima interchanged. When the subject was fully inverted the mean threshold increased by about 20 dB, but the locations of the structures were unchanged. Wilson only reported data from a narrow frequency range including only one or two threshold ripples, so it is not clear exactly how widespread this effect is.

Van den Brink (1980) also reported microstructure changes with body position. While not showing the threshold curves directly, he showed the difference in the thresholds from the two conditions. He found the changes to be small, but almost identical in the two ears. Diplacusis was found to change in the same way.

Long (1980) made a study of masking effects and threshold microstructure. She reported that masked threshold ripples disappeared as a simultaneous masker was increased to 40-50 dB. Above this level the audiogram remained flat. Long also used non-simultaneous maskers. She found that while the audiogram microstructure disappeared when the masker level reached 30-40 dB, it gradually reappeared, with maxima and minima interchanged, as the masker level was increased still further.

Cohen (1981; preliminary results reported in Cohen & Schubert, 1979) also reported detection threshold microstructure. The ratio between the maxima and minima were in accordance with the findings of all earlier investigators (10-15 dB within 30 Hz).

The frequency ranges investigated were narrow, only 100 to 200 Hz, but were still wide enough to judge the width of the ripples. Again, the reported data matched earlier results.

C. Threshold Microstructure Stability

Most of the investigators reporting threshold microstructure retested their subjects to measure the stability of the audiogram ripples observed initially. The general result was that the microstructure of the audiogram is remarkably stable. Small day to day variations in the locations of maxima and minima were observed, but no large restructuring of the audiogram was reported.

Elliott (1958) reported the audiogram structures were stable over 2 months. Thomas (1975) found that threshold maxima had a standard deviation of less than 10 Hz for stimulus frequencies in the range 200 Hz to 5000 Hz. He did not indicate the time interval between measurements. Van den Brink (1971) reported that repeated measurements of threshold microstructure did "not reproduce as nicely as the pitch matching in the diplacusis measurements", but there appeared to be a "satisfactory conformity" between measurements.

Kemp (1979b) measured the locations of loudness function maxima daily for 3 weeks. The mean standard deviations for maxima locations over the 3 weeks were 0.5% (relative to the maxima frequency) for maxima below 1 KHz and 0.25% for maxima above 2 KHz. Subjects monitored intermittently over 4 years showed no changes in excess of 0.5%.

Cohen (1981), in a two week delayed retest of selected threshold points, found relatively small changes in threshold microstructure. The locations of peaks and valleys remained constant, but individual threshold values changed. Some by as much as 7 dB. In a two-month retest for one subject, however, relatively large changes in microstructure location were observed. Cohen only measured thresholds at 4 different frequencies so it is difficult to determine the real change in the structure.

IV. Stimulus Encoding, Excitation Patterns, and Pitch Extraction

The form of the information supplied to the pitch extraction mechanism is crucial to the manner in which it will perform its task. In section I we stated that we assume that the central neural activity pattern, from which the pitch is extracted, is a pattern in space. In this section we describe the processes that generate that central spatial activity pattern. Knowledge of these processes is relevant as part of the chain of logic allowing us to link the characteristics of the peripheral auditory system (threshold microstructure) to predictions of central pitch extraction behavior as measured with a pitch matching task, (pitch shifts with intensity).

We define the neural activity pattern as the pattern of activity of a collection of neurons. Each individual neuron or small group of neurons responds to a different region of the frequency spectrum. The function of the pitch extraction mechanism is to observe the output of this neural spectrum analyzer and generate a sensation of pitch. The first part of this section describes how that spectrum analysis is carried out. The last part of this section describes models of the pitch extraction process.

Acoustic stimuli to be processed by the auditory system are patterns of air pressure variation occurring in time. For the audible range of acoustic stimuli those air pressure fluctuations occur at frequencies anywhere from 20 to 20,000 Hz. For a sine tone lasting 1 second, however, 20 to 20,000 separate events are not perceived, but a single pitch that remains roughly constant for the entire second.

It is the function of the auditory periphery to encode those pressure variations into the spatial neural activity pattern processed by the pitch extraction mechanism.

Mathematically an acoustic signal may be represented in either of two ways, in the "time-domain" or in the "frequency-domain". The representations carry the same information, but each reveals different aspects of the same signal. A question central to theories of pitch perception is whether the transformation of the acoustic signal from the time-domain representation to the frequency-domain representation takes place before or during the pitch extraction process. While the manipulation of the information in either domain is mathematically equivalent, the physical realization of those manipulation processes may be vastly different. The behavioral and perceptual effects to be observed at the level of the pitch extraction depend upon those physical processes.

Helmholtz' book <u>On the Sensations of Tone</u> (1863) marks the beginning of the modern era of auditory research. According to Holmholtz' theory the stretched elements of the basilar membrane vibrate in sympathy with the frequency components of the incoming sounds. In this way the frequency components are coded at specific places; each connected to different nerves. Helmholtz, however, did not originate the idea that the acoustic wave was coded as a spatial activity pattern. Wever (1949) described the work of DuVerney in 1683. DuVerney, a French physician, had a resonance theory of the basilar membrane very similar to Helmholtz' theory. It was Helmholtz' book though, serving as a timely and eloquent integration of previous research, that won him most of the credit for originating the resonance-type of place theory.

The frequency selectivity of the stretched elements of the basilar membrane was extremely fine according to Helmholtz. Critics pointed out that this leads to those elements "ringing" for an unacceptably long time after the tone itself has stopped. Helmholtz was forced to conclude that these stretched elements must have considerable damping in order to eliminate the persistence. This, however, greatly reduced the frequency selectivity of the elements and led to a spread in the number of elements resonating to a single input frequency. Thus, the excitation pattern was born.

Helmholtz could not reconnect this excitation pattern with Muller's doctrine of specific nerve energies that he had extended to mean that specific nerves or groups of nerves corresponded to each different perceivable pitch (Nordmark, 1970). It remained until 1900 for Gray's principle of maximum stimulation to offer a resolution to this problem (Wever, 1949). According to Gray, the sensation of pitch is derived from the maximally stimulated segment of the basilar membrane. Impulses from the surrounding nerves are suppressed. The maximum referred to is only a relative, or local, maximum. Multiple maxima are allowed explaining the ability to hear more than one pitch in two simultaneously played tones. The limits of pitch discrimination are determined by the ability to discriminate between two close maxima in the excitation pattern. This early model of pitch extraction is referred to as the Peak Detection model for the remainder of this paper.

Von Bekesy (1960), in a compendium of his earlier work, showed that the Helmholtz resonance model of basilar membrane motion is a

specific case of a more general model of the mechanical action of the basilar membrane. Using direct observation of the basilar membrane von Bekesy was able to find more accurate parameters for this general model. He found that the motion of the basilar membrane is more closely represented by what appeared to be a traveling wave.
Von Bekesy observed that the point of greatest displacement of the basilar membrane varied with frequency. The point of maximum displacement is close to the apex of the cochlea for low frequencies and close to the oval window for high frequencies.

Von Bekesy's travelling wave theory easily replaced Helmholtz resonance theory of basilar membrane motion. The amount of excitation in any particular hair cell along the basilar membrane varies with the degree of displacement of its location due to the traveling wave.

Von Bekesy's travelling wave theory had problems with frequency resolution, much the same as Helmholtz' resonance theory. Galambos & Davis (1943) found that individual nerve fibers responded to much more specific frequency regions than would be predicted by the broad displacement maxima observed by von Bekesy. Zwislocki (1948, cited in Nordmark, 1970) reported that separate maxima do not appear in the basilar membrane displacement from two tones unless those tones are at least an octave apart. Because of these discrepancies between the single fiber studies and displacement calculations based on von Bekesy's work, models of the mechanical motion of the basilar membrne continue to be diveloped (c.f. Allen, 1980, a & b). The result of this effort has been to greatly increase the predicted frequency selectivity of basilar membrane displacement maxima.

Another approach to the problem of frequency selectivity was to investigate neural coding of aspects of basilar membrane motion other than simple maximum displacement. Von Bekesy' himself was aware of the frequency selectivity problem and suggested that the neural mechanisms might code the derivative of basilar membrane displacement instead of the displacement itself (von Bekesy, 1960). Von Bekesy invested a great deal of effort into the problem of lateral inhibition and selectivity (Bekesy, 1967). He considered the increased selectivity to be an entirely neural process.

Huggins & Licklider (1951) investigated a variety of neural coding schemes that might sharpen the neural fiber frequency selectivity while maintaining von Bekesy's basilar membrane displacement calculations. They calculated maximum basilar membrane displacements for two simultaneously sounding tones-200 Hz and 1600 Hz with the higher tone 30 dB less intense than the lower tone. The calculated maximum displacement from the 1600 Hz tone was completely masked by the motion of the membrane due to the 200 Hz tone. Two pitches, however, are easily heard in this stimulus in behavioral tests. Huggins & Licklider found that the negative of the second derivative of the basilar membrane displacement yields two separate maxima at approximately the same locations as displacement maxima from the two tones sounded separately. They concluded that neural coding of the second derivative of basilar membrane displacement is a mechanism that can account for some of the discrepancy between single fiber results and von Bekesy's displacement model.

Nordmark (1970) disagreed with Huggins & Licklider. He pointed out that the interaction effects of the two tones caused those second derivative maxima to be shifted enough to result in perceivable pitch shifts. He claimed that those pitch shifts had not been seen in behavioral data. These pitch shifts, however, have been shown to exist. Thurlow (1943) was the earliest. More recently Terhardt & Fastl (1971) and Terhardt (1973) have reported pitch shifts in the simultaneously sounding tones. It is not clear, though, that these reported pitch shifts are due to neural coding of the second derivative of basilar membrane motion.

Green (1976) reviewed the studies showing that the direction of shear between the cilia of the hair cells and the tectorial membrane forms a pattern that changes with the frequency of the input tone. This direction is mainly across the width of the basilar membrane on the portion of the basilar membrane between the stapes and the location of maximum displacement. It is mainly along the length of the membrane from the point of maximum displacement to the apex. Since hair cells are sensative to shear in specific directions this shear pattern may affect the neural excitation pattern.

Efforts such as those by Huggins & Licklider (1951) have been abandoned. More recent observations of basilar membrane displacement using the Mossbauer technique (Johnstone & Boyle, 1967) and the capacitive probe technique (Wilson & Johnstone, 1975) have shown that the displacement maxima are much more sharply defined than originally thought.

The structure of the excitation pattern may be approached from an alternate direction. By observing the responses of individual auditory nerve fibers to input signals at various frequencies, amplitudes, and phases a description of an excitation pattern to any arbitrary input may be constructed as a collection of individual neuron responses.

For describing the activity of individual auditory nerve fibers, the work of Galombos & Davis (1943), Kiang and collaborators at MIT and Rose and his collaborators at Wisconsin are the most popular references. Galombos & Davis were first to show that individual auditory nerve fibers respond best to a narrow range of frequencies. The one frequency to which the fiber responds at the maximum rate is called the characteristic frequency or best frequency. A fiber's best frequency may be obtained in two ways. One is to outline a response area for a single fiber by measuring a fiber's response to a constant intensity signal presented at different frequencies. The other way is to find the response threshold of the fiber for a variety of frequencies. The frequency that has the lowest threshold for the fiber is the characteristic frequency.

The excitation pattern to be approximated is defined along what we have called the tonotopic axis. This tonotopic axis is a device to describe an ordering of single fibers or small groups of fibers, each having different best frequencies. The unit along the tonotopic axis might be a discrete quantity,simply assigning an ordinal value to the fibers as arranged by best frequency. We expect, however, that the distribution of fibers with different best frequencies is not uniform across the range of audible frequencies. Zwislocki (1965) showed that neurons are spaced approximately uniformly with respect to the log of the fibers' best frequencies. Von Bekesy (1960) showed the place of maximum displacement of the basilar membrane is roughly a uniform distribution along the length of the basilar membrane with the log of the frequency of the signal. Therefore, we have chosen the unit along the tonotopic axis to be "z", equal to the log of those best frequencies. Z is a continuous variable corresponding to the center of a useable range of fibers with similar best frequencies. This collection of fibers with similar best frequencies make up a minimal patch along the basilar membrane. The widths of these minimal patches are assumed to be equal widths in log(best frequency). We further suppose these widths to correspond, on the average, to the widths of neural excitation patterns from pure tone signals presented at threshold intensity levels.

The amount of excitation at any given coordinate is a double average. It is an average over the firing rates of all of the fibers within the minimal patch along the basilar membrane represented by z and it is an average over time of the driven firing rates of those fibers. The driven firing rate is the firing rate of a fiber above its spontaneous firing rate.

To complete the excitation pattern approximation a variety of parameters must be specified. The response of a fiber to input at its best frequency must be specified for various intensity levels. The response of a fiber to input not at its best frequency, at various frequency intervals and for various intensity levels, must be specified. The spread of mechanical displacement is already included in this approximation because the 'off-frequency' responses of fibers are measured with an intact basilar membrane. The fact that the fiber is in a section of the basilar membrane not vibrating at maximum displacement is an inseparable component of its off-frequency response characteristics.

One model of pitch extraction has already been discribed. The Peak Detection model of Gray specifies that the pitch of a tone is determined by the z coordinate along the tonotopic axis with the highest activity level. One problem with this theory stems from the limited response range of auditory nerve fibers to intensity. The firing rate of single auditory nerve fibers saturates when the input signal is greater than 60 dB above the threshold intensity level (Green, 1976). A high intensity signal causes a wide segment of the basilar membrane to vibrate at large amplitude. Many fibers are stimulated at a level more than 60 dB above threshold. The excitation pattern then has many frequency locations responding maximally. This creates a plateau rather than a peak in the excitation pattern. If the peak detection model is the mechanism of the pitch extraction process then we expect frequency selectivity in pitch perception to decrease with increasing intensity for high intensity signals. Two high intensity tones that are close in frequency stimulate large overlapping regions of the excitation pattern to saturation level. Pitch discrimination ability under the peak detection model is expected to decrease. In fact, pitch discrimination improves somewhat as intensity increases (Wier, Jesteadt, & Green, 1977). This argues against a simple peak detection model of pitch extraction.

An alternative model is the Centroid model. In this model the pitch extraction process computes the centroid of the excitation pattern. The centroid is computed by taking the sum of all the z coordinates represented by fibers stimulated above threshold; each weighted by its excitation level. This sum is divided by the sum of the weights. The perceived pitch corresponds to the centroid of the excitation pattern. There are problems with this model, however. As the intensity of a signal increases more fibers will fire. According to this model we expect the increase in the number of contributing fibers to increase the specificity of the pitch. This results in improved pitch discrimination at high intensities. Based upon the single unit response curves of Rose, et al, we expect the number of fibers responding to increase dramatically over this intensity range. Wier, et al, (1977) find only a very small increase in pitch discrimination for two sine tones when the signal intensity is increased from 40 dB to 80 dB.

Another model for the pitch extraction process involves the point of maximum slope in the excitation pattern. The extraction mechanism would locate the z coordinate along the tonotopic axis at which the amount of activity is most different from its neighbors. This search would be restricted to one end of the excitation pattern. Rose, et al (1971) showed response areas of single neurons that indicated that excitation patterns expand along the tonotopic in only one direction as the signal increases in intensity. The end of the excitation pattern to be evaluated for maximum slope would have to be the stationary end of the pattern. The pitch extraction mechanism causes a sensation of pitch corresponding to the z coordinate along the tonotopic axis having this largest slope. This model does not suffer from the problems of the first two models. The spread of the excitation pattern does not affect this mechanism. Looking for a single point in this fashion, however, leads to very volatile predicted pitch sensations. Due to internal noise this single point varies considerably.

Whitfield (1978) proposed that the process of assigning a location along the tonotopic axis to an excitation pattern use the simplest mechanism possible. His model finds the boundary point between a region of the excitation pattern where all cells are below threshold and a region where all cells are above threshold. These boundary points are the points along the tonotopic axis where the excitation pattern begins and ends. The pitch extraction mechanism uses the locations of these end points, in some unspecified way, to determine the perceived pitch of the signal. This model suffers from the same problem as the maximum slope model. Internal noise will lead to large errors in determining the end points. This will lead to unstable pitch sensations. Whitfield claimed that this problem of noise is eliminated by an internal squaring process that accentuates the excitation pattern boundaries.

There are other models of pitch extraction. The models listed above do not exhaust the possibilities. Tests of the Peak Detection, the Centroid, and the Whitfield models are described below. As a set they provide a reasonable starting place to evaluate the logic of the experimental paradigm.

V. Data Collection

A. Introduction

Our experimental paradigm supposes that there is a relationship between microstructure in the sensitivity function for hearing and the non-monotonic deviations from Stevens Rule found in measurements of pitch change with intensity. We believe that through this relationship we can learn about the mechanisms of pitch extraction.

The first part of our experimental paradigm is the collection of measurements relating to the sensitivity function of hearing, and the collection of measurements of pitch shifts with changes in intensity. These data are used in the last part of the paradigm to evaluate models of pitch extraction via computer studies. This section describes the data collection. Section VI below describes the computer modelling studies.

B. Data Collection Protocol

Information about the microstructure in a subject's sensitivity function was obtained by making detailed measurements of the subject's threshold of hearing. We assumed that the threshold microstructure would be different for the two ears of a single subject (van den Brink, 1971) so we only took measurements from one ear. For frequencies at which a low threshold was measured we assumed that the corresponding point along the tonotopic axis was very sensitive. Points along the tonotopic axis that were relatively insensitive were revealed by high thresholds at the corresponding frequency.

We wanted to measure pitch shifts with intensity (P-I) in an intensity regime where sensitivity microstructure is known to affect perception. Kemp (1979) and others have shown that the microstructure of the threshold is highly correlated with measures of perceived loudness for stimuli near threshold. This correlation decreased as the stimuli of the loudness experiment were increased in intensity. When the stimuli reach approximately 50 dB above threshold the correlation was no longer significant. Based on this result we chose to collect the P-I data using stimuli less than 50 dB above threshold.

The next priority was to elicit pitch shifts that were large enough to measure reliably. Ideally, based upon Kemp's results both tones should be as close to threshold as possible to maximize the effects of sensitivity microstructure. If the two tones are too close in intensity, though, the pitch matching errors will be larger than the P-I effect (Verschuure & van Meeteren, 1975). We chose to use a 30 dB intensity difference. This provided a large intensity difference and allowed both tones to be less than 50 dB above threshold. Specifically, intensities of 40 dB SPL and 10 dB SPL were used. 10 dB SPL was chosen as close to threshold, but not so quiet that reliable pitch matching cannot be done. By chance, the 40 dB level was the intensity Fletcher (1934) proposed as a standard for pitch matching experiments.

Studies described above (Chapter II, section III.C. Threshold Microstructure Stability) showed that the microstructure of the auditory threshold changes over time. For this reason we measured the threshold microstructure both before and after the pitch-intensity measurements were made. We expected any changes in the thresold microstructure to be reflected in the variations in the pitch-intensity measurements. Measuring thresholds before and after the pitch matching task allowed us to monitor any changes in microstructure that may have occurred during the course of the experiment.

Detailed threshold measurements were made in the one octave range from 1200 Hz to 2400 Hz. Changes in pitch with intensity were made at 24 equally logarithmically spaced frequencies from 1392 Hz to 2101 Hz. These sampled frequencies do not span the entire range from 1200 Hz to 2400 Hz because we expected that the pitch shifts might be affected by sensitivity microstructure not exactly at the same location along to tonotopic axis.

These frequency ranges were divided into smaller ranges for the testing procedure. The threshold measurements were done in 3 overlapping ranges: 1200 Hz to 1600 Hz, 1496 Hz to 1978 Hz, and 1800 Hz to 2400 Hz. The pitch shift measurements were taken in 2 contiguous ranges of 12 frequencies each: 1392 Hz to 1696 Hz and 1726 Hz to 2101 Hz. These frequency ranges are illustrated by Figure 11. These divisions of the frequency regions served two purposes. The first was to enable the threshold measurements to be taken relatively close in time to the pitch shift measurements. The second was to relieve subject boredom.

The following experimental protocol was followed: (See Figure 12).

a. Measure threshold from 1200 Hz to 1600 Hz.

b. Measure threshold from 1496 Hz to 1978 Hz.

c. Measure pitch shifts with intensity for the 12 low frequencies; the set of 12 is measured 5 times.

d. Re-measure threshold from 1200 Hz to 1600 Hz.

e. Re-measure threshold from 1496 Hz to 1978 Hz.

f. Measure threshold from 1800 Hz to 2400 Hz.

g. Measure pitch shiftss with intensity for the 12 high frequencies; the set of 12 is measured 5 times.

h. Re-measure threshold from 1800 Hz to 2400 Hz.

This protocol took 5 to 6 hours to complete. Each threshold measurement took 15 minutes. Each measurement of pitch shifts for the 12 frequencies of a set took about 20 minutes. Each of these 15 to 20 minute measurement sessions was followed by at least 5 minutes of rest. The entire protocol was always run as a complete "marathon".

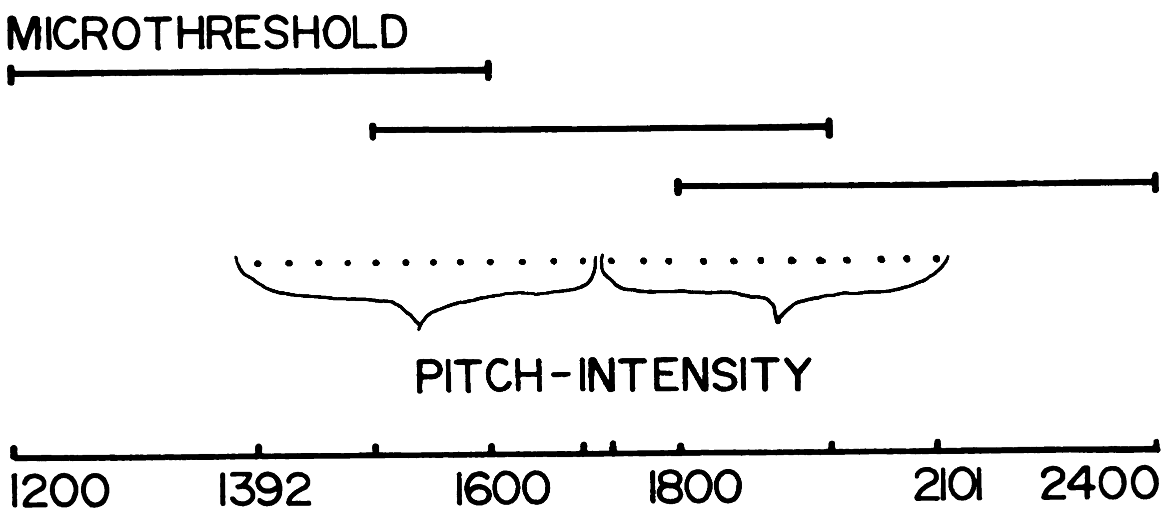

Figure 11. Frequency ranges

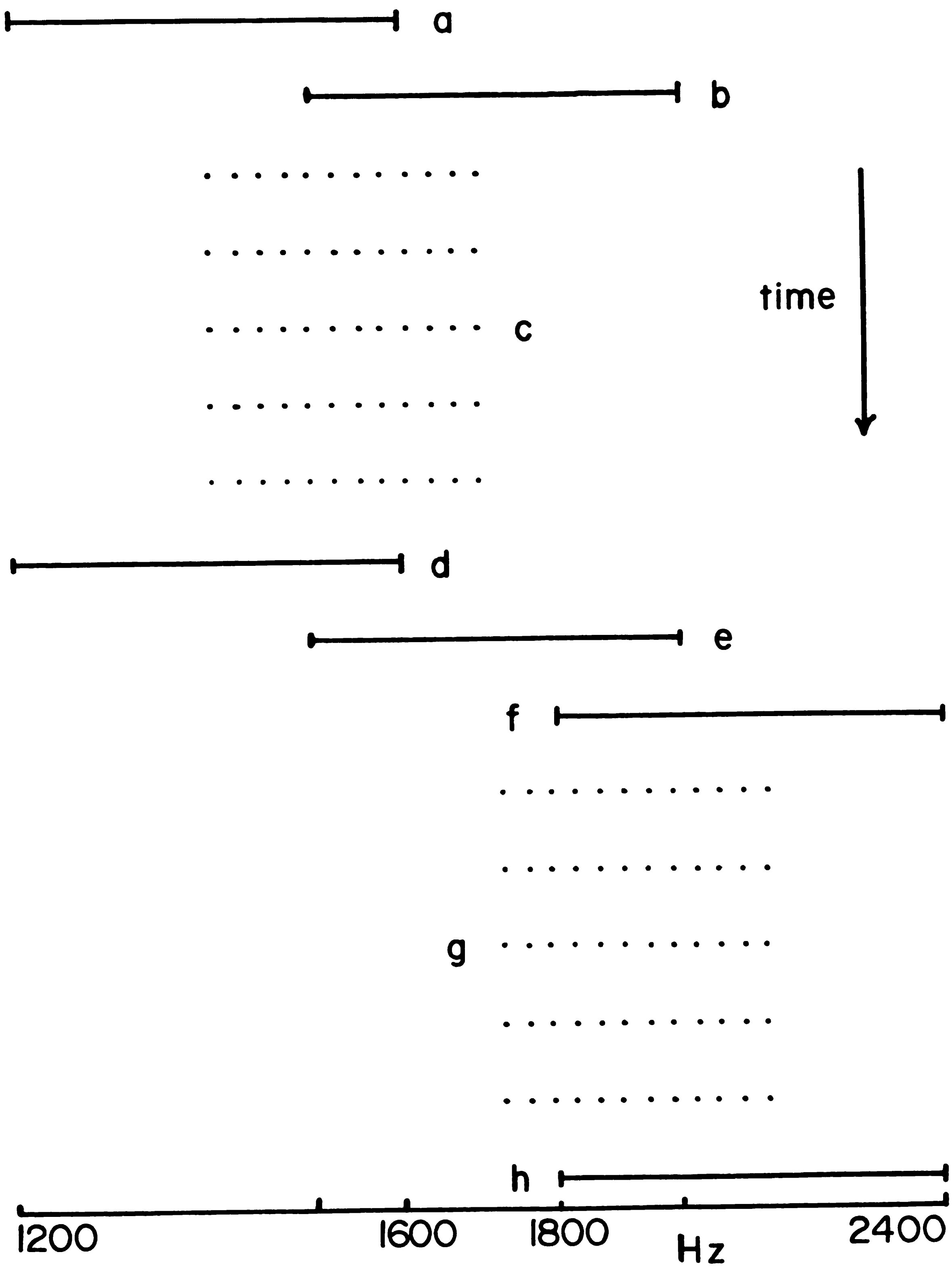

Figure 12. Data collection protocol

C. Apparatus and Subjects

All tones were presented monaurally. The subjects were seated in a sound-proof booth and listened through TDH-49 headphones with MX41 cushions. The signal source was a Wavetek model VCG 116 voltage controlled function generator. The frequency and amplitude of all signals were controlled by a microcomputer. The microcomputer controlled the sequencing and timing of all signals in the experiment and also collected the responses.

Only sine wave signals were used. The tones were high-pass filtered (6 dB/octave) to flatten the response characteristics of the TDH headphones as measured with a flat plate coupler. The tones were turned on and off with a sine-squared edge lasting 6 milliseconds.

Three subjects were used, J, M, and W. M was the author. All 3 had experience in psychoacoutic listening tasks.

D. Threshold Microstructure-Procedure.

Threshold of hearing curves for pure tone stimuli were measured for a single ear of each subject. These detailed measurements were taken using a computer controlled von Bekesy Audiometer tracking procedure (von Bekesy, 1960).

The signal used was a pulsing sine tone. The tone pulsed with a period of 500 ms; 250 ms on, 250 ms off. This pulsing was to eliminate adaptation effects. The period of the pulsation was chosen to be long enough to exceed temporal integration time for this frequency range (200 to 250 ms; Plomp & Bouman, 1959). The intensity of the tone changed in 1 dB steps every 500 ms. These intensity changes were coincident with the gating of the pulses.

The response required of the listener was very simple. The subject pressed a button whenever he heard the tone and released the button when he did not. While the subject heard the tone and was pressing the button the intensity of the tone decreased. When the intensity of the tone got below threshold the subject no longer heard it and released the button. While the button was released the intensity of the tone increased until, when finally above threshold intensity again, the subject heard the tone and pressed the button. This caused the intensity to begin to decrease again. By this process the intensity of the tone oscillated back and forth across the actual threshold intensity for the subject.

The frequency of the tone in the standard procedure is changed continuously. In our computer-controlled version, the frequency was changed every 20 ms. Each time the frequency was changed it was increased by 0.00063% (1.91% per minute). The frequency steps were logarithmically equal. The size of the steps was chosen so that 15 minutes was required to sweep the tone across the entire frequency range being tested; an increase in frequency of 33% for all 3 ranges. This rate of change in frequency was slow enough that the pitch of the tone seemed constant.

By changing the frequency of the tone the intensity threshold crossed by the repeating oscillations of the stimulus intensity was at a different frequency for each oscillation. By changing the frequency slowly the frequencies for which threshold was estimated were very closely spaced.

The level of the tone during an entire session was plotted on a strip chart recorder. A portion of a typical response curve is shown in Figure 13. This curve oscillates across the intensity threshold for the subject. At the peaks in the oscillation the subject began to detect the tone and pressed the button. At the valleys in the oscillation the subject no longer heard the tone and released the button. The exact threshold was never observed directly. The threshold at any given frequency was approximated by taking the arithmetic mean of the intensities at the peak and valley that straddled that point. About 60 threshold values were estimated from each response curve. The frequencies of these thresholds were equally logarithmically spaced. The interval between each measurement point was about 0.45%. This is equivalent to a frequency interval of about 5.45 Hz around 1200 Hz and a frequency interval of about 10.9 Hz around 2400 Hz.

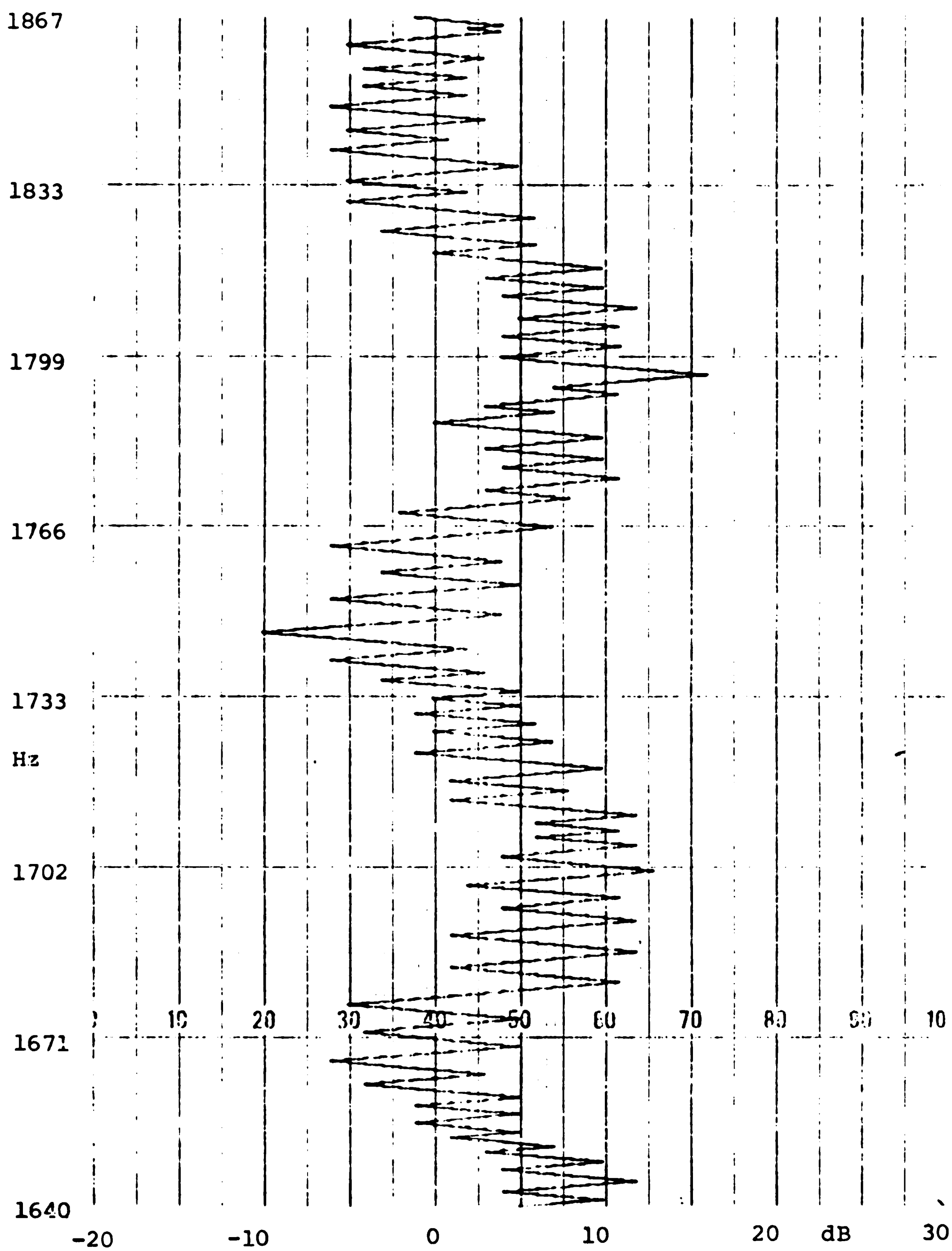

Figure 13. Example output from von Bekesy audiometer

E. Threshold Microstructure - Results

The data from the Von Bekesy Audiometer runs were coded as described above. Only the runs following the pitch-intensity trials were used (steps d, e, and h from the protocol). The overlap at the ends of the frequency regions were eliminated by using the data from the middle frequency range only. This was done after verifying that any structure occurring in each frequency range was present in both of the data sets covering that overlap region. A uniform constant was added to all measured thresholds in the low or high frequency range data when necessary to eliminate discontinuities in the overall threshold curves. This was justified because our ultimate objective was to correlate the structure of these curves with the pitch-intensity measurements. The absolute levels were relatively unimportant.

We estimated the sensitivity along the tonotopic axis as exactly the negative of the threshold along a log-frequency axis. These negative threshold curves for all 3 subjects are plotted in Figure 14. Data for subjects J and M fit along the same zero line because of the large difference between overall levels of sensitivity.

Subject J is relatively insensitive compared to the other two subjects. Subject J shows less sensitivity in the range of approximately 1500 Hz to 1900 Hz. J shows a large peak at approximately 1400 Hz. There are also well defined peaks at about 1600 Hz, 1800 Hz, 1970 Hz, 2100 Hz, and 2180 Hz. These peaks are 7.5 to 15 dB higher than surrounding levels. There is a full 20 dB difference between the most sensitive point (about 1450 Hz) to the

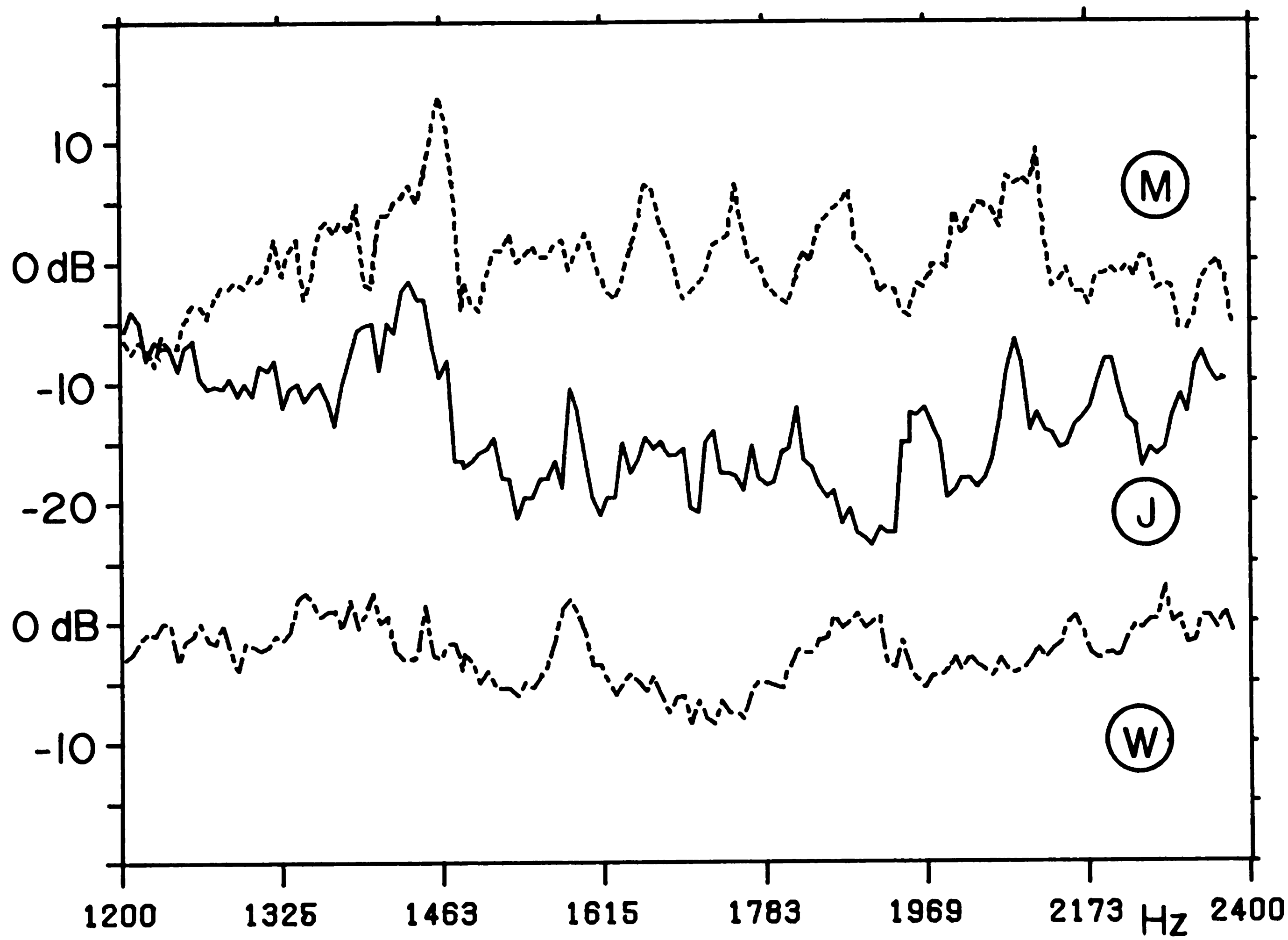

Figure 14. Sensitivity curves

least sensitive point (about 1900 Hz).

Subject M is the most sensitive subject in this frequency range. Subject M has well defined peaks at approximately 1350 Hz, 1460 Hz, 1550 Hz, 1650 Hz, 1750 Hz,1850 Hz and 2000 Hz. These peaks are 6 to 17 dB higher than surrounding levels. There is a 23 dB difference between the most sensitive point (about 1400 Hz) and the least sensitive point (about 1240 Hz).

Subject W has a relatively flat sensitivity curve compared to subjects J and M. There are only two well defined peaks, at 1600 Hz and 1900 Hz. These peaks are only about 10 dB higher than surrounding levels. Subject W tended to have smaller oscillations in intensity in the raw data plots. This may have ben the part of the reason for the relatively smooth sensitivity curve.

There is a no systematic relationship between data for the different subjects. None was expected. Each subject has unique sensitivity microstructure.

Subject W executed the marathon protocol 3 times. These 3 times were separated by 2 month intervals. Figure 15 shows these data. The top curve is the sensitivity curve from the first threshold measurement in each frequency range. The next curve (identical to the curve for subject W in Figure 14.) is from the second threshold measurement in each frequency range taken during the first marathon. The third curve is from an execution of the marathon two months later. The bottom curve is from the third execution of the marathon, 4 months after the first. The bottom 3 curves are all from second measurements in each frequency range. The similarities here are remarkable. The first two curves, both taken within 1 to 2 hours, are no less different than the first curve and the fourth curve, taken 4 months apart. Clearly sensitivity microstructure is very stable over long periods of time.

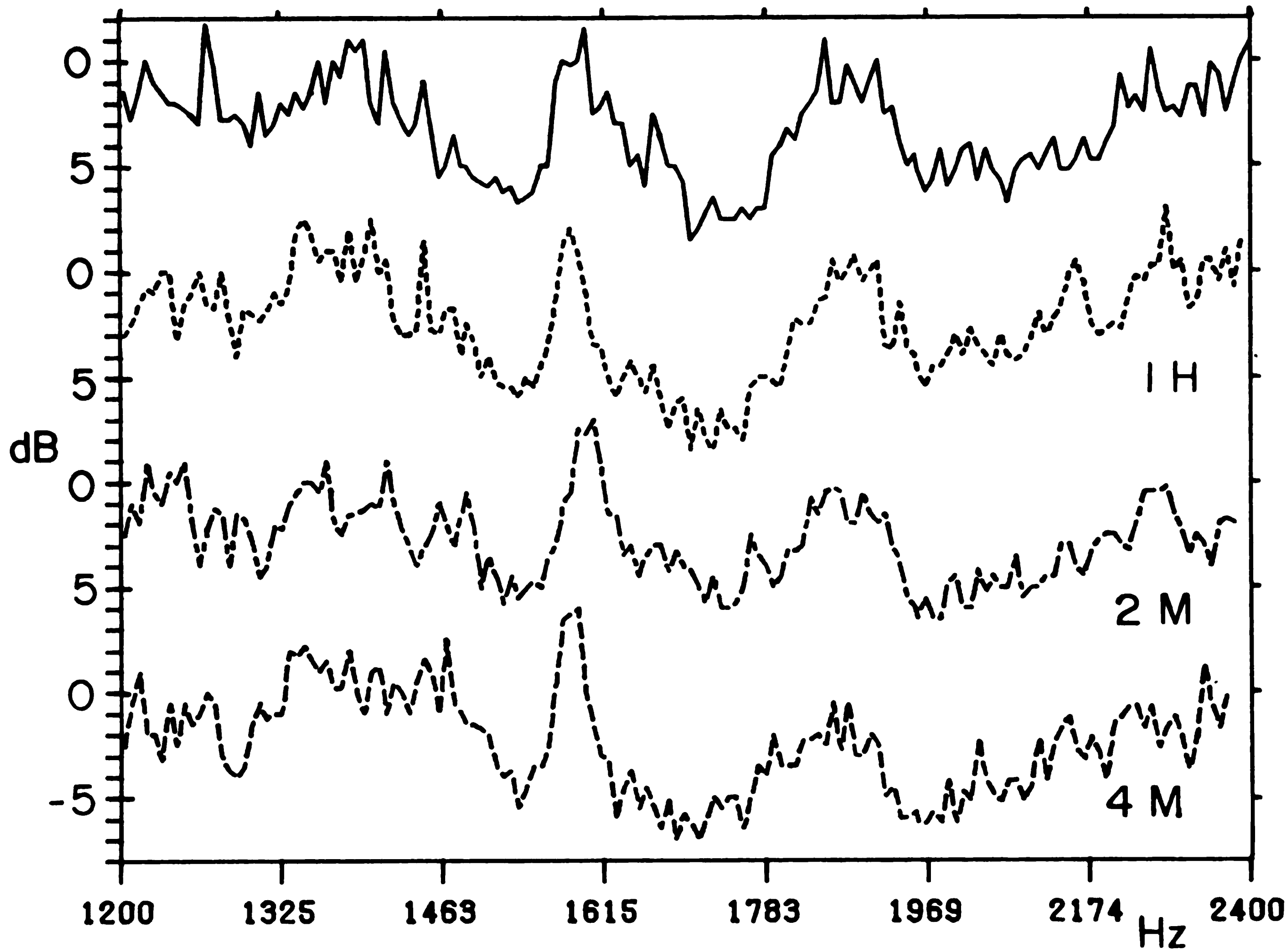

Figure 15. Sensitivity curves for subject W at delays of approximately 1 hour, 2 months, and 4 months

F. Pitch-Intensity-Procedure

Measurements of the P-I effect were made at 24 different frequencies using a pitch matching procedure. The 24 frequencies used were at equal logarithmic steps from 1392 Hz to 2101 Hz.

On each trial the subject was given unlimited exposure to a 4 part stimulus presentation sequence. During the first interval the standard tone was presented at about 40 dB SPL. During the third interval the matching tone was presented at about 10 dB SPL. The second and fourth intervals were silent. Each of the four segments lasted 500 ms. The frequency of the standard tone remained constant during the entire trial. The subject adjusted the frequency of the matching tone until satisfied that the two tones had the same pitch. The segments were marked by separate lights of different colors.

Twelve different frequency standard tones were presented during each run. 5 runs were done using each set of 12 frequencies. During each run the tones were presented in a different random order. The final frequency of the matching tone minus the frequency of the standard tone was the measurement of the pitch shift.

Tones presented to subjects J and W were at slightly higher intensities. Subject J matched the pitches of tones at 45 and 15 dB SPL. Subject W matches the pitches of tones at 43 and 13 dB SPL. These higher intensities were used in order to ensure that all tones were above threshold when presented. Preliminary experiments showed that pitch matches had extremely high errors when the lower intensity tone was below threshold as measured by the Von Bekesy tracking task. The difference in intensity between the standard and matching tones was maintained at 30 dB.

G. Pitch-Intensity-Results.

Figure 16 shows the results of the pitch-intensity measurements for all 3 subjects. As for the sensitivity curves there is no systematic relationship between the data for the individual subjects. The maximum amount of change between adjuacent points along these curves is roughly comparable for all 3 subjects. This is not consistent with the sensitivity microstructure data. Subject W's sensitivity was much less variable than J and M . There are only three locations showing these large jumps, however, about 1500 Hz, 1700 Hz, and 2000 Hz.

The pitch-intensity data from the three marathons subject W executed is shown in Figure 17. Like the sensitivity microstructure, these data are remarkably similar. Only two features of the curves seemed to change during the 4 months. The P-I curve at 2 months did not have the prominent peak at about 1700 Hz found in the other two curves. The curve of P-I effects at 4 months shows a single peak between 1750 Hz and 2000 Hz whereas the two earlier data sets show what might be called a jagged plateau across that range.

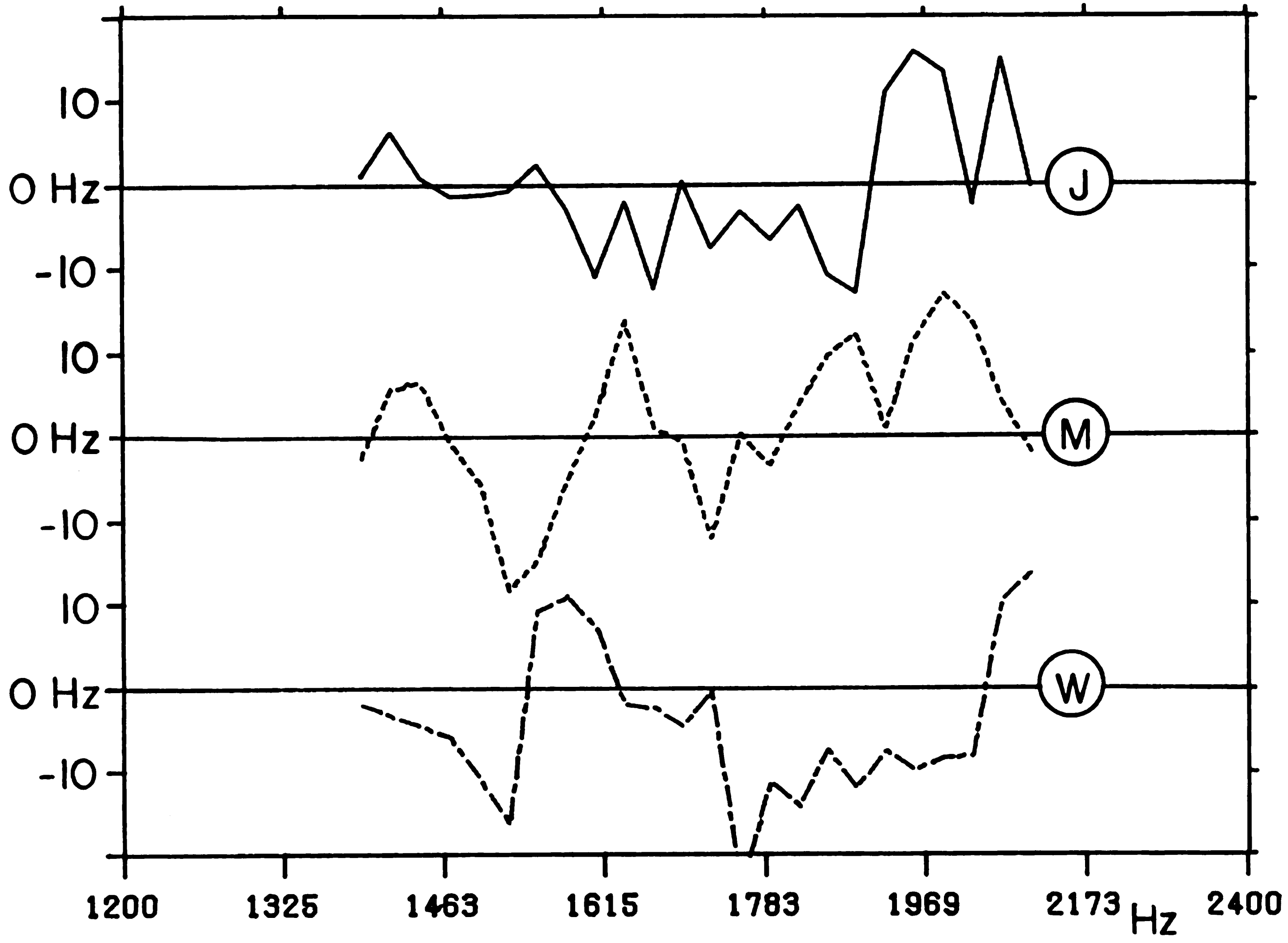

Figure 16. Pitch-Intensity data

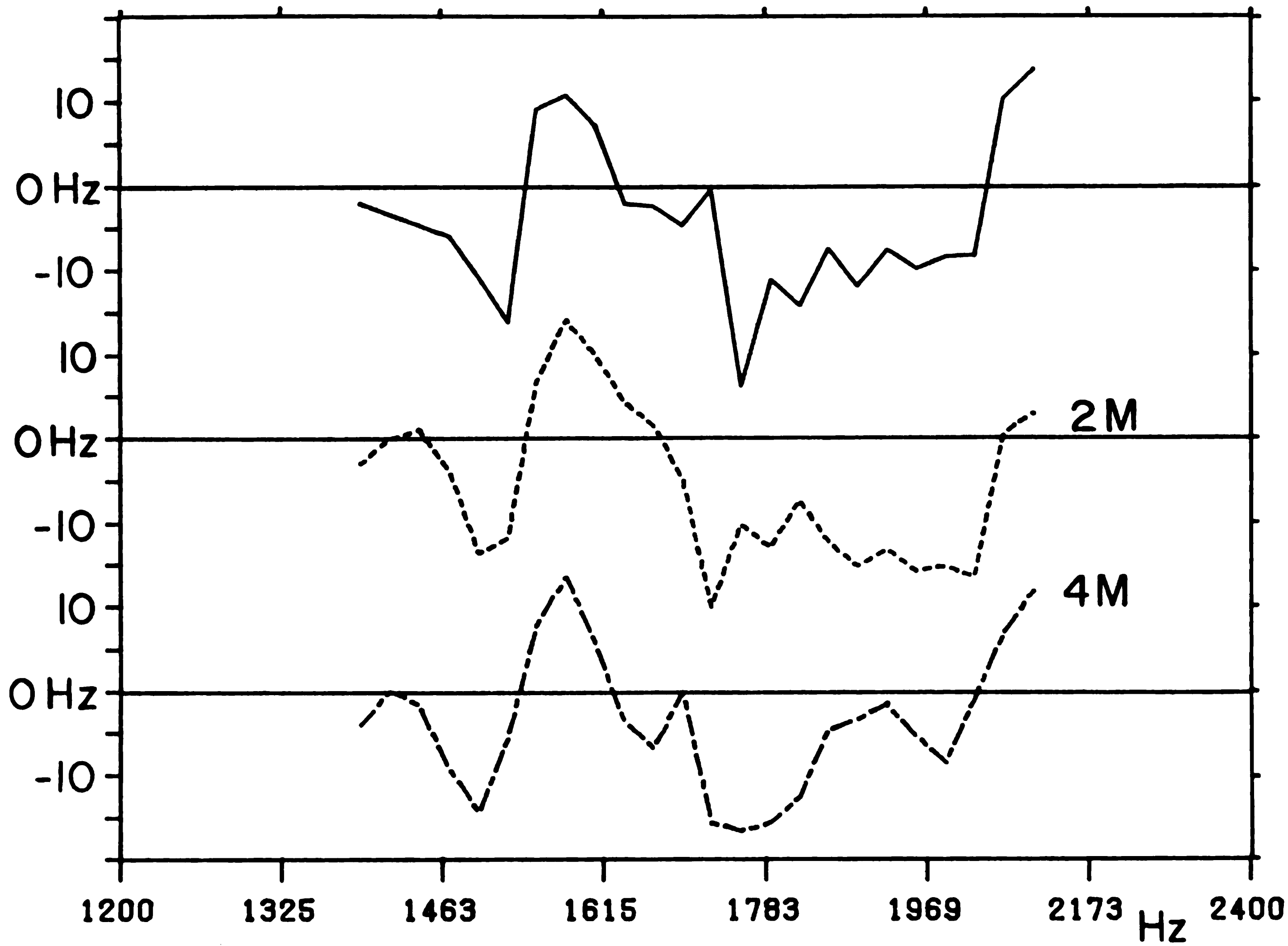

Figure 17. Pitch-Intensity data for subject W at delays of 2 months and 4 months

H. Data Collection-Discussion

Figures 18,19, and 20 show the data from the two procedures for each subject. The bar at the top left in each figure shows the estimated error for the sensitivity function. This was calculated by sampling the size of the intensity oscillations. The length of the bar is two standard deviations. The error bars on the P-I curve in each figure (the bottom function) represent two standard deviations also.

A number of points may be made in comparing the two measures of auditory function. First, the size of the errors relative to the structure in the curves is consistent. The large changes in sensitivity and the changes in amount of pitch shift with intensity within narrow frequency ranges are larger than the average error in either case. This shows that the procedures used are reasonable. The structure we expected to find in these curves is real and not just the result of subject variability. The similarity in the size of the structure relative to the size of the error also means that these measurement procedures are equivalent in accuracy.

These similarities in error relative to observed structure also mean that the magnitude of the structures we found are roughly equivalent. Our notion that sensitivity microstructure is related to P-I structure is reasonable in that the magnitude of the sensitivity microstructure matches the magnitude of the pitch shifts.

There appears to be a correlation between the actual level of sensitivity and the P-I error. In many cases where the amount of error in the P-I function changes dramatically from one point to the

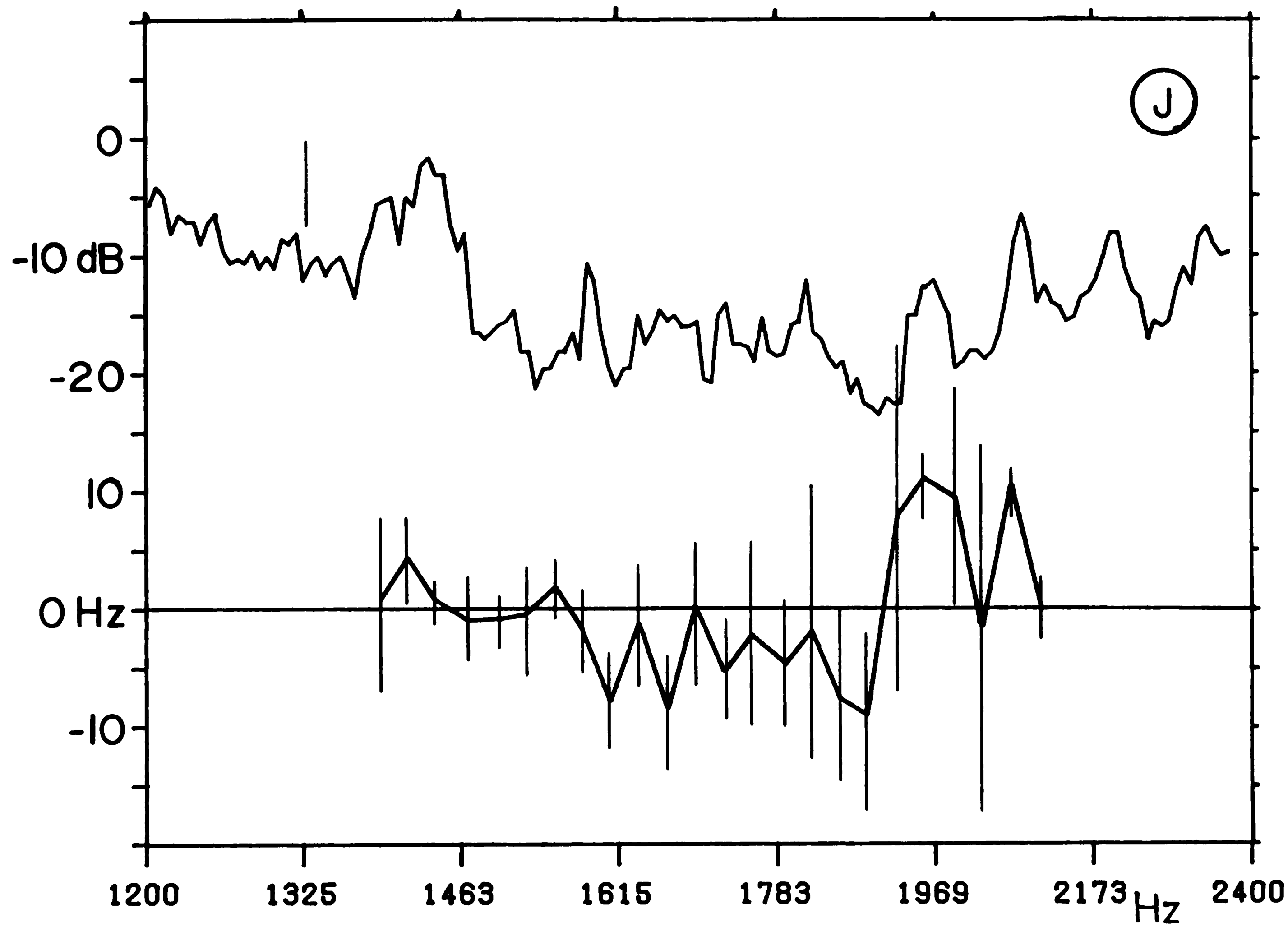

Figure 18. Data for subject J

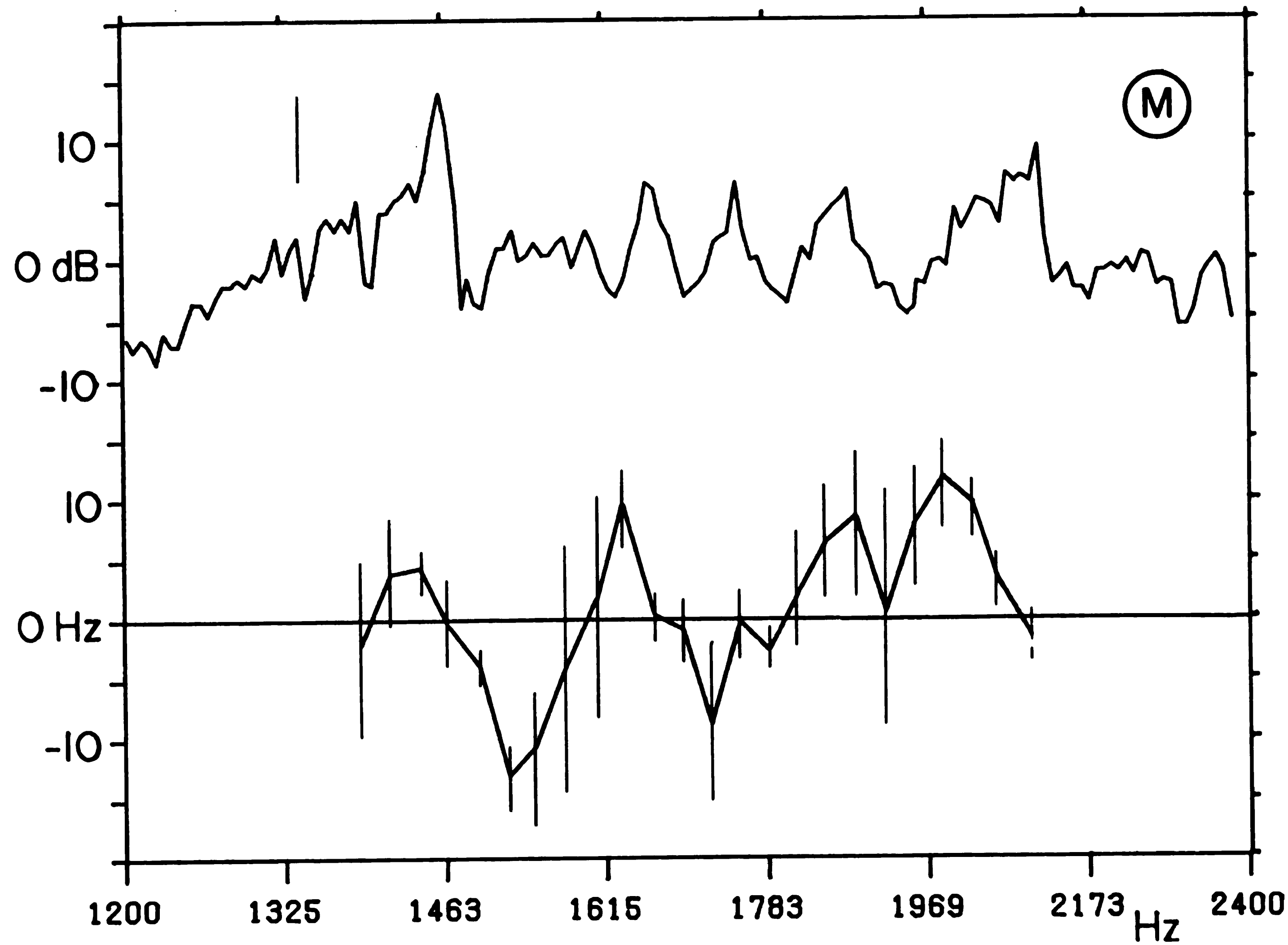

Figure 19. data for subject M

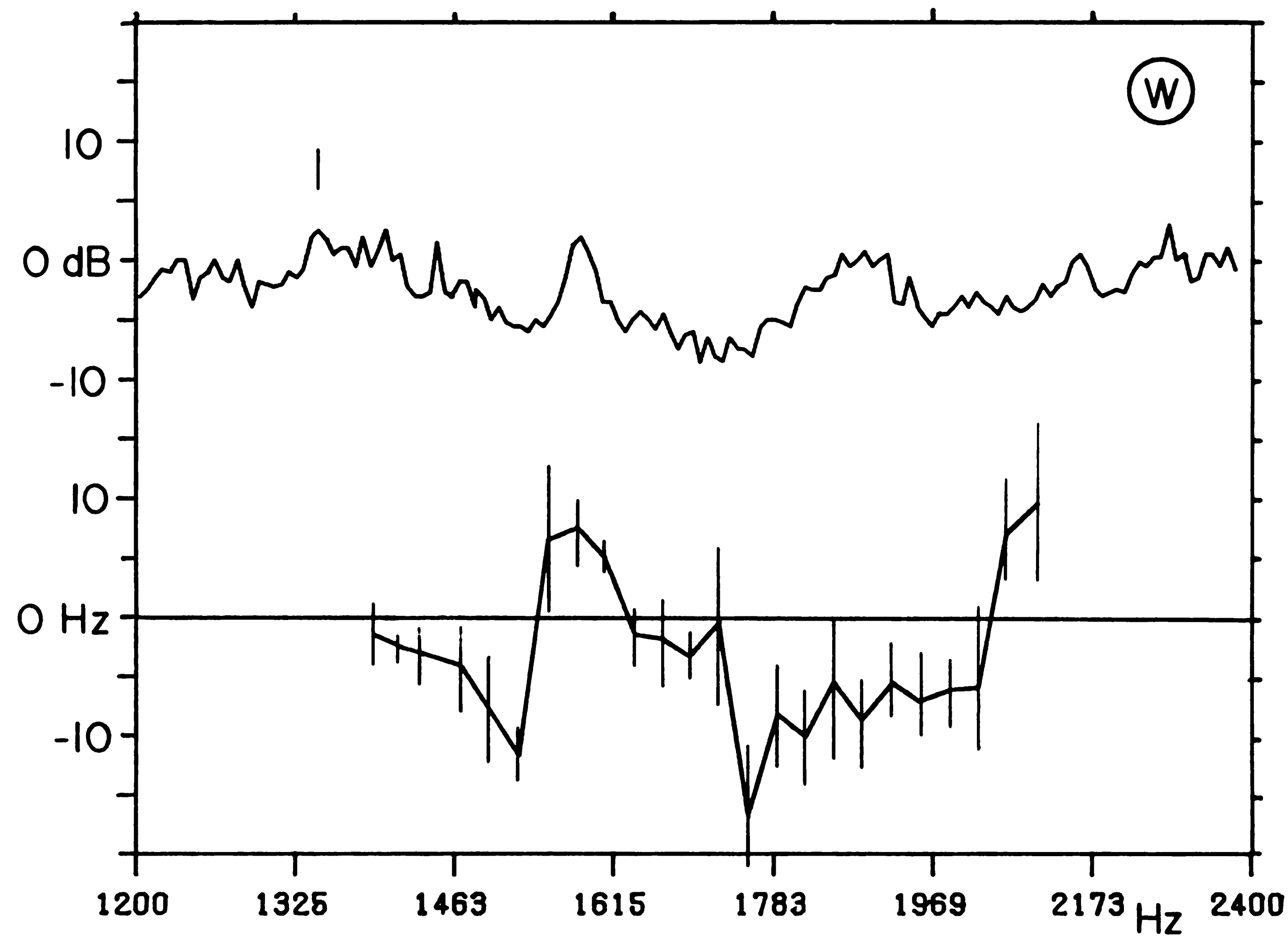

Figure 20. Data for subject W

next these changes are exactly coincident with a large change in sensitivity. An especially good example of this may be seen in the data for subject J at the 6 highest frequency P-I points. Because of his low sensitivity the low intensity tone may have been below the threshold resulting in extremely large errors for some of these points. The two points in that last 6 that have small errors, however, are at frequencies exactly corresponding to peaks in the sensitivity curve.

There is an explanation of the relationship between sensitivity micro-structure and P-I effects that we consider trivial. If the effect of the peaks in the sensitivity structure is simply to boost the effective intensity of the low intensity tones while having little influence on the effective intensity of the high intensity tones, then the structure in the P-I curve is due simply to changes in the relative intensity difference between the two tones. At sensitivity peaks the effective intensity difference might only be 20 dB instead of 30 dB. This would cause a change in the amount of pitch shift, resulting in structure in the P-I function matching the structure in the sensitivity function. Based on this model of the relationship and Stevens Rule for high frequency tones (high intensity increases pitch) the following predictions about the shape of the P-I curve can be made:

1. The average pitch shift will be positive.

2. At sensitivity peaks there should be a minimum in the P-I curve. This is due to a decrease in the effective intensity difference. The minimum in P-I should be zero or still positive.

3. Sensitivity minima will correspond to P-I peaks. The peaks should be positive.

The actual data reveals two aspects inconsistent with this explanation of P-I structure: 1) the average pitch shift is approximately zero and 2) the peaks in the P-I curve are more often associated with sensitivity maxima than minima. If one argues that this frequency range falls in the no-shift regime of Stevens Rule then the peaks of the sensitivity curve should be associated with zeroes of the P-I curve. This is still not the case.

VI. Modelling

A. Introduction

The purpose of this investigation was to evaluate models describing the pitch extraction process of the auditory system. The approach used in this evaluation was to predict the behavior of the individual subjects on the P-I task under each of the different models. We expected one model to make obviously superior predictions, predictions that closely matched the actual measured responces of the subjects.

The evaluation procedure was the same for each model. We assumed that the perceived pitch of a tone, P, is some function, p, (the extraction process) of the excitation pattern, E.

$$P = p(E)$$

The excitation pattern is determined by the frequency of the input signal, f, and the intensity of the input signal, I. The pitch of the 40 dB tone then is:

$$P_{40} = p(f_{40}, 40)$$

The pitch of the 10 dB tone is:

$$P_{10} = p(f_{10}, 10) \quad .$$

When the perceived pitches of the two tones are equal:

$$P_{10} = P_{40} \ .$$

This means that for equal pitches

$$p(f_{40}, 40) = p(f_{10}, 10).$$

the frequency of the 40 dB tone, f_{40}, was fixed and known. The only unknown quantity was f_{10}. Each model was used to calculate P_{40}. Then values of f_{10} were tried in the calculation of P_{10} using the model being evaluated, until P_{10} equalled P_{40}. The predicted pitch shift was

$$f_{10} - f_{40}$$

which matched the formula for the measured pitch shifts from the data collection. The models were used to predict the pitch shifts for all 24 frequencies from the P-I procedure. The similarity between the predicted and the measured pitch shifts was the gauge of the success of the particular model.

This evaluation procedure did not depend upon the actual values assigned to P. With this method we continued in the realm of classical psychophysics. It was not necessary to determine what the actual perception was. It was only necessary to find stimulus conditions (mathematically defined) that generated an equivalent sensation along the same dimension (P).

B. Computed Excitation Pattern

Colburn (1973) modelled the excitation, driven firing rate, at each point along the tonotopic axis, E(z), by the following formula:

$$E(z) = \frac{A(f)H(z-\log f)}{\{K+A^2(f)H^2(z-\log f)\}^{1/2}}$$

A(f) is the amplitude of the signal at frequency f. H is a filter function describing the decrease in the excitation level at location z as a function of the distance along the tonotopic axis from z to the point where the best frequency of the fiber matches the input frequency (log f). The maximum of this function is at the point where z=logf. This function is not necessarily symmetric ie.

$$H=2^{-m(z-\log f)}$$

$$H(z-\log f) \neq H(\log f-z)$$

K is a parameter making the function non-linear corresponding to saturation at high intensities.

Our model differs from Colburn by our addition of the sensitivity function of the tonotopic axis to the equation.

$$E(z) = \frac{A(f)H(z-\log f)S(z)}{\{K+A^2(f)H^2(z-\log f)S\ (z)\}^{1/2}}$$

For this model, in the limit of infinite input amplitude, A(f), all points along the tonotopic axis will have an equal output level of 1. We called this the constant saturation model.

An alternative model for excitation that we also tried was

$$E(z) = S(z)\ \frac{A(f)H(z-\log f)}{\{K+A^2(f)H^2(z-\log f)\}^{1/2}}$$

In this model the dynamic range of the excitation at each point is identical, but multiplied by the sensitivity of point. This means that for two points along the tonotopic axis having the same input, both will be equally saturated but the more sensitive point will still have a higher drive firing rate. This was called the variable saturation model.

A final addition to the model was that the amplitude of the signal reaching point z, (A(f)H(z-log f)), had to be greater than the threshold at that point. Otherwise E(z) was set to 0.0.

C. Pitch Extraction Models

1. Peak Detection: The most obvious way to implement is a simple search for the peak of the excitation pattern E. This was not done. Instead a function that predicted the location of the peak was used. The prediction was based on the first and second derivatives of the sensitivity curve at the point z=log f and the filter slope parameters of H. The basic idea was to use an analytic formula to calculate the value of z that resulted in the derivative of E being zero.

2. Centroid: In this model the centroid of E was calculated by the following formula:

$$P = \frac{\sum_{i=1}^{N} Z_i \, E\,(Z_i)}{\sum_{i=1}^{N} E\,(Z_i)}$$

This model weights each point along the tonotopic axis by the level of excitation at that point. Using Colburn's model of the excitation pattern, this model would be reducible to analytic form. The filter slope parameters would determine the shift of the centroid away from the peak. With sensitivity included in the computation of the excitation an analytic form is no longer possible.

3. Whitfield: In this model (Whitfield, 1978) a search is made for the z coordinates closest to the center of the filter function, H, at which the excitation pattern changes from 0.0 to a value greater than 0. The mean of these two points is the computed value of P.

4. Whitfield Centroid: Whitfield's notion of a simple decision

process based on a point on the excitation pattern being above or below threshold led to an alternative centroid model. In this modification all points along the tonotopic axis that are above threshold are given equal wieght.

$$P = \frac{\sum_{i=1}^{N} (Z_i \mid E(Z_i) > 0)}{\sum_{i=1}^{N} (1 \mid E(Z_i) > 0)}$$

This gives more weight to points of low excitation than the centroid model. The FORTRAN implementation of all models is shown in Appendix B.

D. Parametric Variations

These models and the excitation calculation have many parameters. The filter function has two parameters, one for each condition $z = \log f$ and $z \log f$. The excitation has the parameter for saturation, K. It is reasonable to introduce a threshold higher than the measured threshold. This corresponds to the fact that at the level of absolute threshold a tone may be reliable detected, but not produce a sensation of pitch (Plomp & Bowman, 1959). It has been shown that the best frequency of single fibers changes with increasing intensity (Rose, etal, 1971). Based on this, the location of the maximum of the filter function H may be considered to be a function of intensity as well as frequency.

Clearly there is a large parameter space within which to test these models of pitch extraction. Not all models were tested in all variations. The models and their evaluation was an evolving process. Some of the variations outlined above were not included in the testing procedure until late in the process. In those cases models previously tested (e.g. the Peak model) were not retested with the new parameter variations.

E. Computer Modeling - Best Results

Data from only one subject, W, is shown in this section. The results for subject W are typical of the 3 subjects. Showing the results for only 1 subject facilitiates comparison between the models.

1. <u>Peak Detection</u>: This model was tried using both a symmetric and an asymmetric filter with the variable saturation excitation pattern. A relatively high Pearson product-moment correlation r=.5 between the predicted pitch shifts and the actual pitch shifts was found when the center of the filter was offset to 1 semitone higher frequencies. An asymmetric filter was used. The filter parameters used resulted in the excitation decreasing from the center toward lower frequencies at 180 dB/octave and 240 dB/octave toward higher frequencies. This very narrow filter caused the predicted pitch shifts of the appropriate magnitude. The correlation was slightly lower (r=.46), however, because the points that were incorrectly predicted before were now even more incorrect. The results for subject W are shown in Figures 21 and 22 as examples.

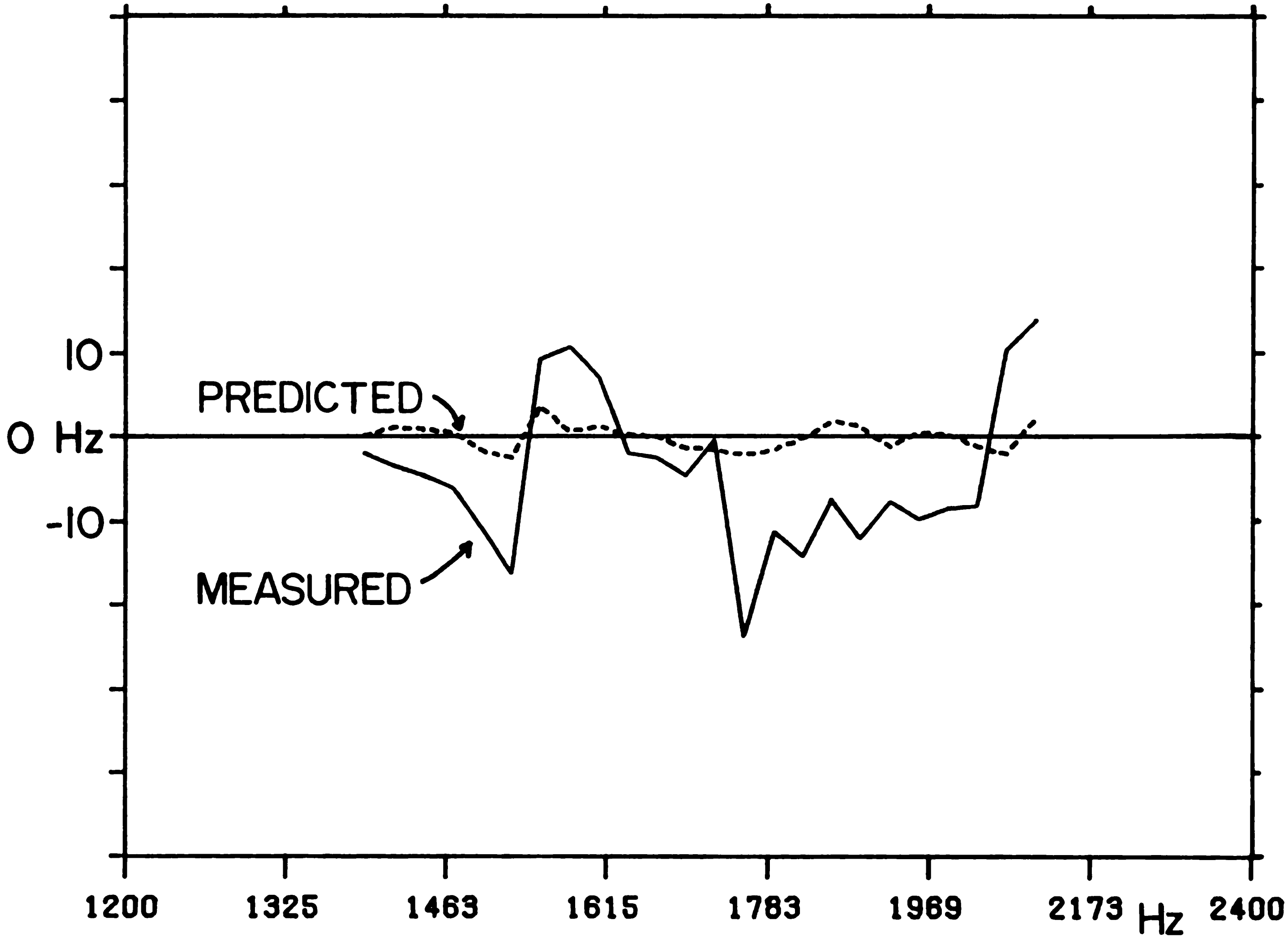

Figure 21. Peak model results. Filter parameters equal 60 and 80

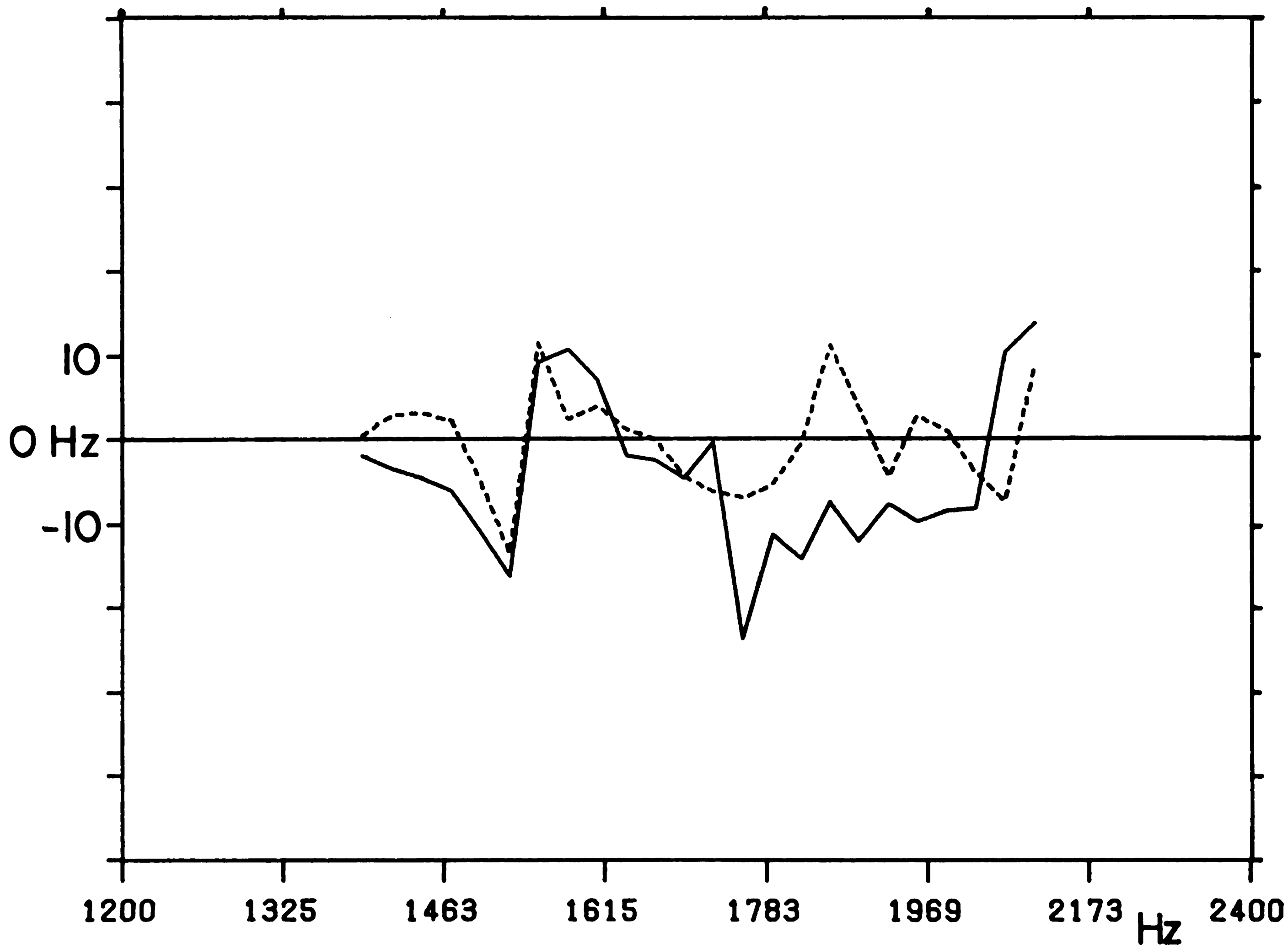

Figure 22. Peak model results. Filter parameters equal 40 and 40

2. Centroid: This model was tried using the most parametric variations of all the models. This model did not produce pitch shift structure of the proper magnitude under any conditions. Ths predicted pitch shifts calculated were always within a very small range This range could be shifted up (decrease the lower slope parameter) or down (decrease the upper slope parameter), but the range of predictions remained the same. A typical set of predictions is shown in Figure 23. The filter parameters were 120 dB/octave for lower frequencies and 180 dB/octave for higher frequencies. The variable saturation model was used. The resulting correlation was r=.27, but clearly these predictions are not very good.

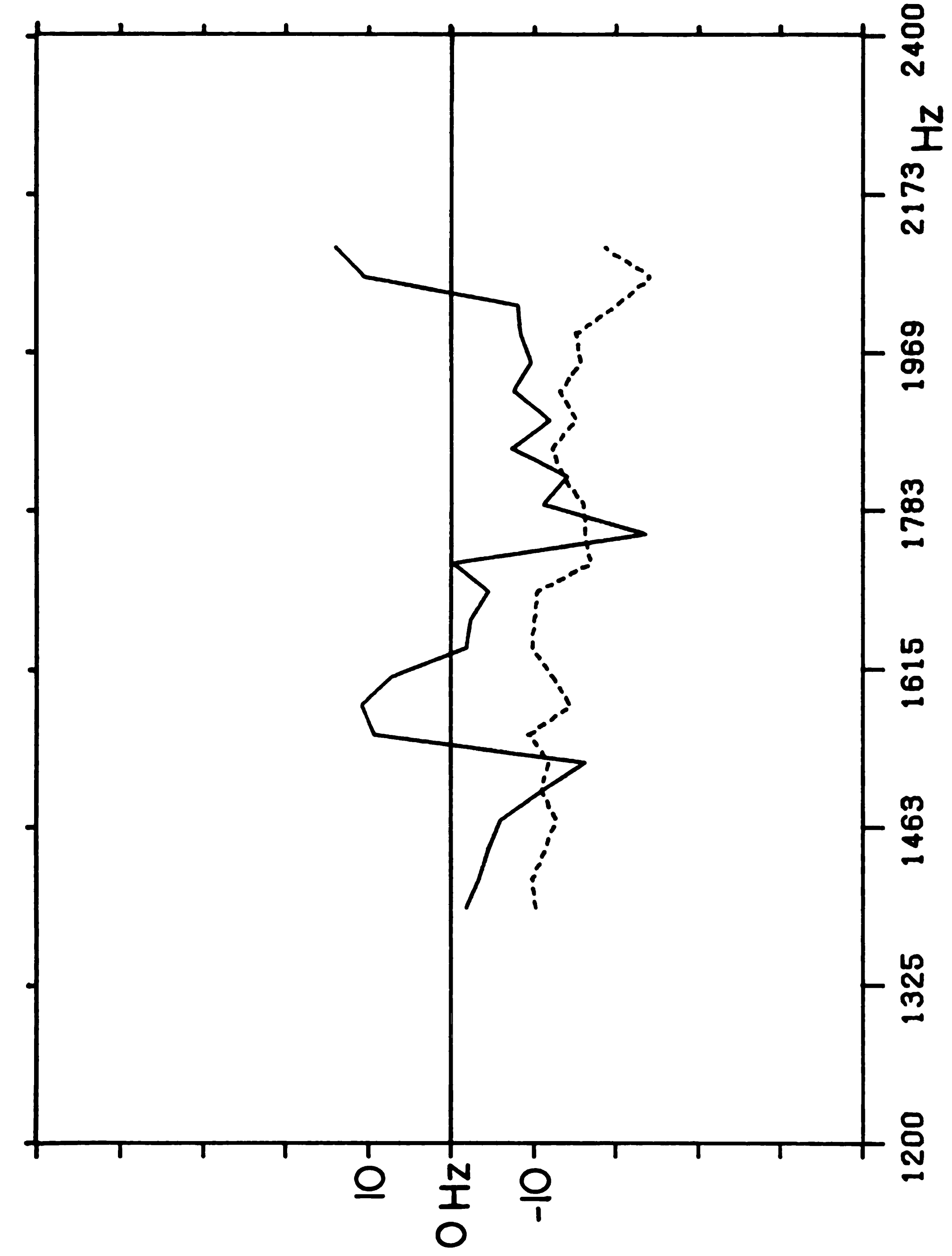

Figure 23. Centroid model results

3. Whitfield: The Whitfield model was not very successful. Using the constant saturation excitation pattern with no offset and both symmetric and assymmetric filter slopes typical predictions only had a correlation of r=.03. The magnitude of the predicted shifts was smaller than required as well. A typical set of predicted pitch shifts is shown in Figure 24.

An alternative to finding the two edges of the excitation pattern closest to the center of the filter (H) was to choose the edges furthest from the center. This alternative was tried when it was observed that because of the structure of the sensitivity curve the excitation pattern, even for a sine tone, was not always continuous along the tonotopic axis. This modificaton increased the size of the predicted shifts and improved the correlation of the predicted to measured shifts to as high as r=.32. This best fit was found with asymmetric filter slopes (240 dB/octave to lower frequencies and 300 dB/octave to higher frequencies). A separate threshold for pitch was also used that was higher than threshold for detection. A typical result for this model is shown in Figure 25. Clearly this model does not work well either.

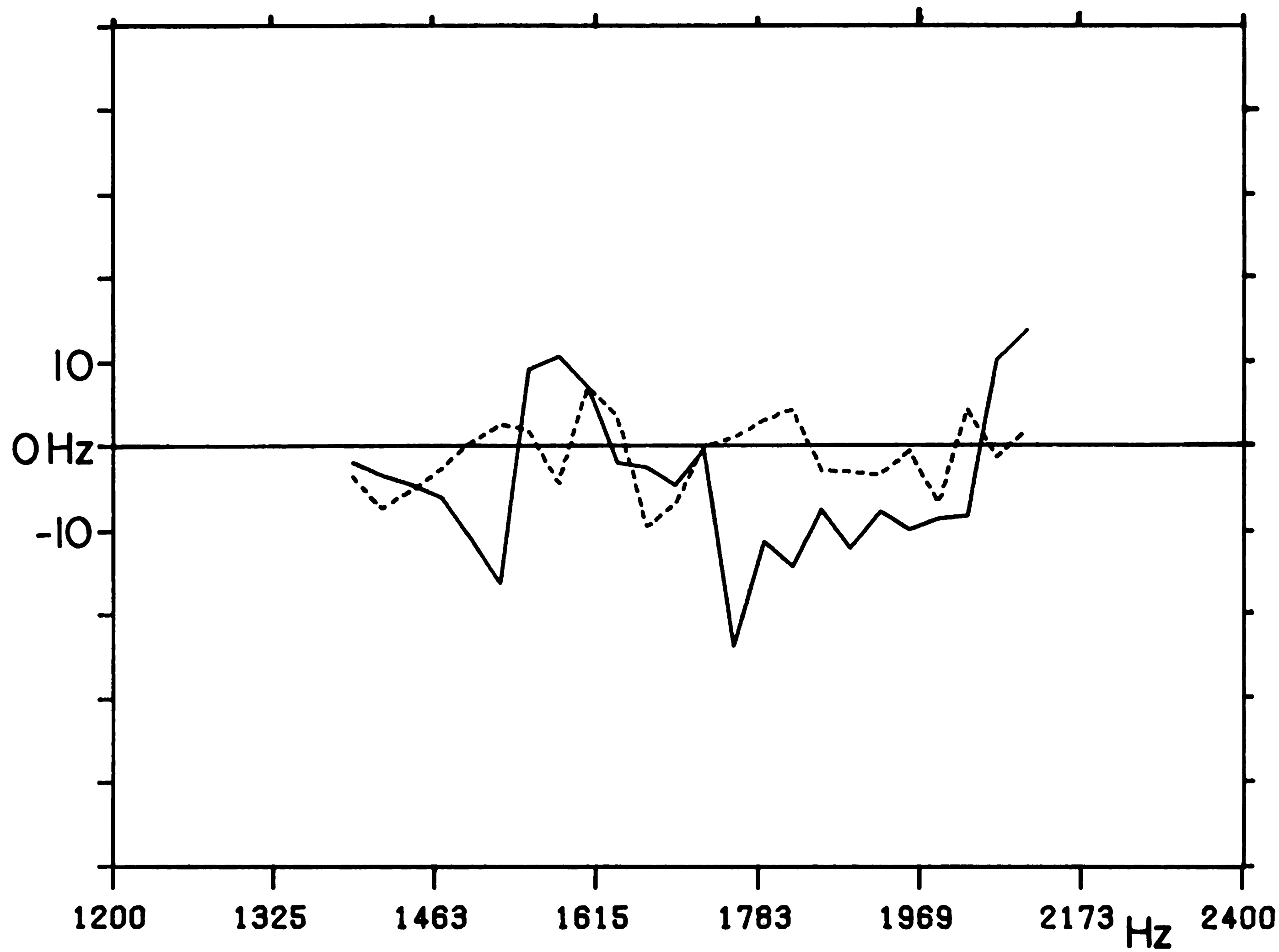

Figure 24. Whitfield model results (a)

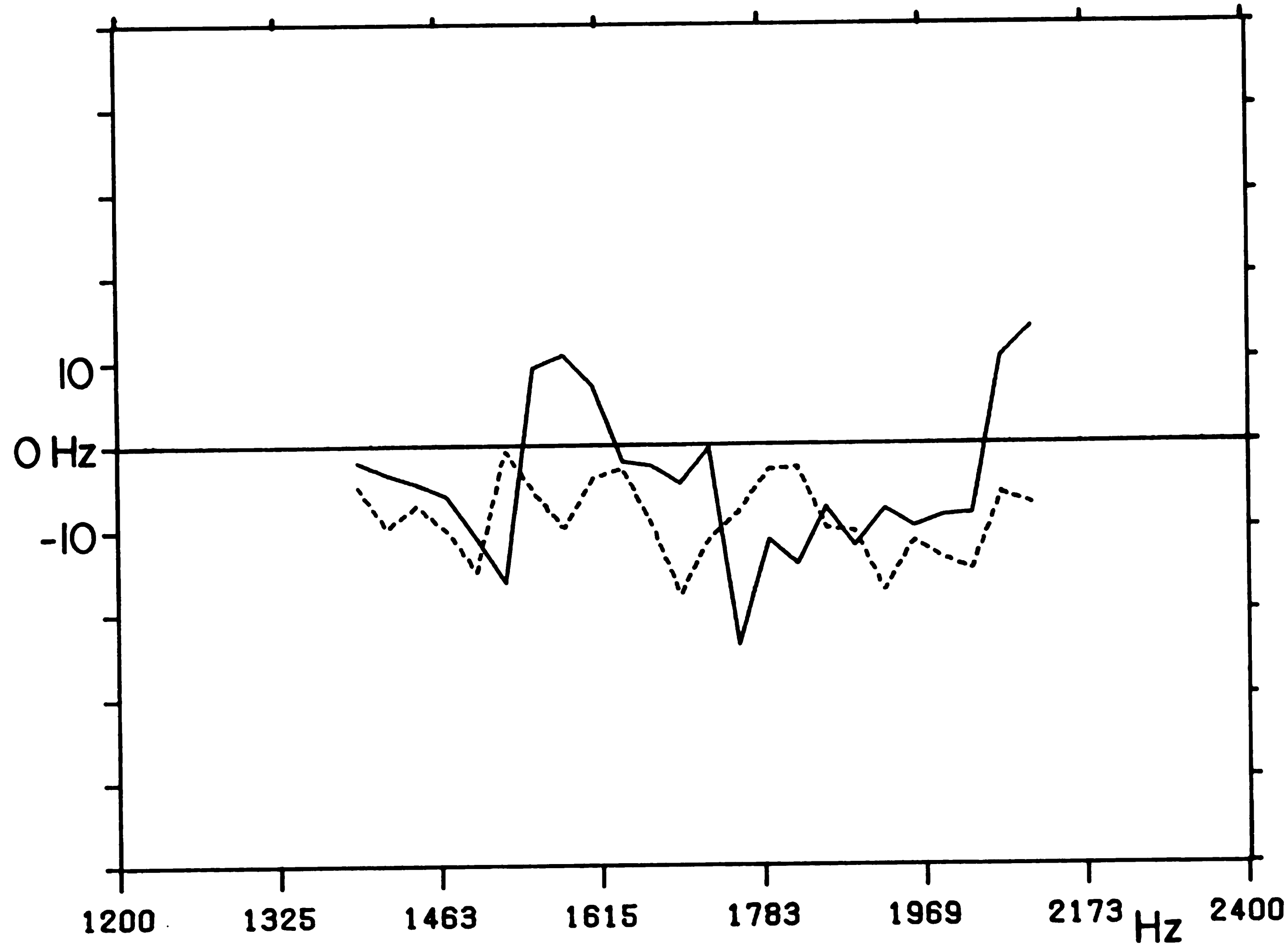

Figure 25. Whitfield model results (b)

4. <u>Whitfield Centroid</u>: This model predicted pitch shifts of the proper magnitude. Using the variable saturation model, a separate pitch threshold, and an offset of the center of the filter of 1 semitone, a correlation of r=.45 was found (Figure 26). This result was found using symmetric filter slopes of 120 dB/octave. Still this model fails. It predicts a peak in the pitch curve that does not exist.

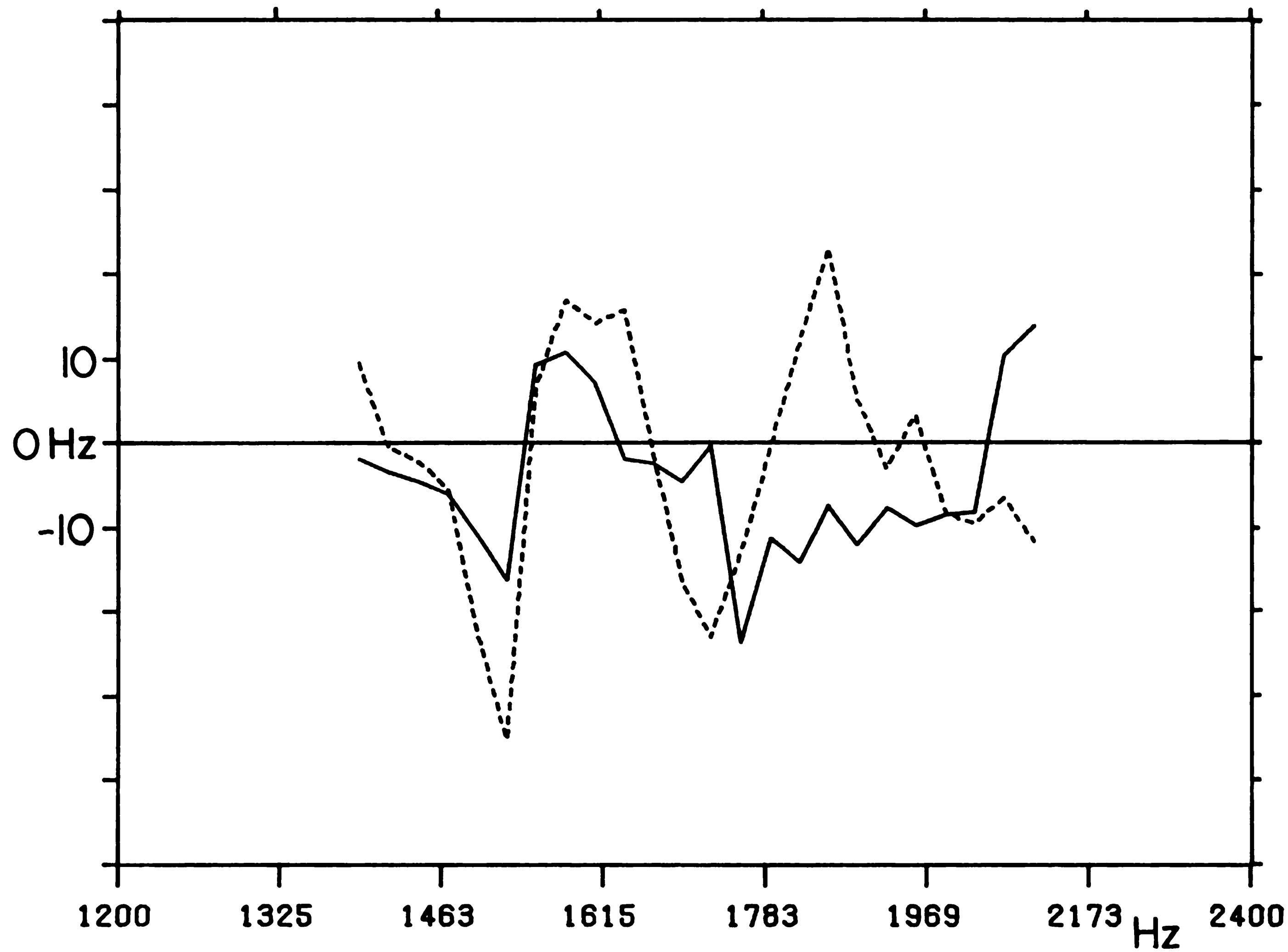

Figure 26. Whitfield centroid model results

F. Effects of Parametric Variations

The filter parameters had different effects depending on the model being tested. In the Peak Detection model the filter parameters determined the size of the predicted shift. Extremely sharp filter slopes resulted in very small predicted shifts in perceived pitch. In the Centroid model the filter slopes served to bias the predicted shifts in one direction or another. If the lower filter slope was less steep than the upper slope then the predected shifts were biased toward more negative pitch shifts. This biasing was observed, but to a lesser extent for both of the Whitfield models. This biasing was smaller because both the Whitfield models are less sensitive to changes in the structure of the excitation pattern.

Introducing an offset in the filter center had varying effects depending upon the model also. For the Peak Detection model this parameter shifted the entire prediction curve to the left (positive offsets) or to the right (negative offsets). For the other models the effect was more complicated. Rose, et al, (1971) showed that the shift in best frequency of a single fiber varied with intensity. For this reason independent offsets were tried for the calculation of for the two intensities were tried for the calculation of the patterns. In the frequency range of this experiment the shift in best frequency was always to higher frequencies. The offsets tried were all either identical for the calculation of the two tones or the pattern for the 40 dB tone was offset more (to higher frequencies) than the pattern for the 10 dB tone. When both tones were offset identically the curve of predicted pitch shifts simply moved to the left relative to the measured shifts. When the louder tone excitation pattern was offset more than the quiet tone excitation pattern the curve of predicted pitch shifts moved upward as well.

The saturation parameter k was found to have very little effect on the model predictions when maintained at a realistic value. This parameter was adjusted to force the excitation pattern to begin to show saturation at specific intensities. A reasonable value for this intensity is 40 to 50 dB above threshold (Green,1976). This meant that this parameter had very little effect on even the peaks of the excitation patterns for the 10 and 40 dB tones used in this test.

The higher threshold for pitch had unpredictable effects. The structure of the excitation pattern vaired from tone to tone. For this reason the effects of a higher threshold varied depending upon the specific features of the excitation pattern. The Centroid model was affected less that the Whitfield model due to the low weights given to the z coordinates at the edges of the pattern. The Whitfield model was affected the most because the predicted pitch of the tone depends directly upon the edges of the excitation pattern.

VII. Discussion

This project was an attempt to learn about the central processes involved in the extraction of pitch from a central excitation pattern. The approach used depended upon assumptions about the changes in the central excitation pattern caused by the intensity of the stimulus and the listener's sensitivity to different frequencies. The pitch of a sine tone is known to change with the intensity of the signal. Verschuure & van Meeteren (1975) showed that the changes in pitch are not strictly monotonic as claimed in Stevens Rule (Stevens, 1935). We take the point of view that the deviations from Stevens Rule are due to the variations in listener sensitivity. Alterations in the central excitation pattern caused by the intensity changes and the sensitivity function of the individual cause the pitch extraction process to generate different pitch sensations even though the frequency of the stimulus remains constant.

Our method of investigation of the central pitch extraction processes was as follows. Information about a listener's sensitivity function was used in modelling the shape of the central excitation pattern in response to sine tones. Each of the pitch extraction models to be tested was used to predict pitch shifts with intensity based on the model excitation pattern. The generated predictions were compared to measurements of shifts in pitch with intensity changes from the same listener.

Information about the listener's sensitivity function was taken from detailed measurements of the threshold of hearing. These threshold measurements were shown to have structure that was stable over time. The structure was unique to each individual. Kemp (1979b) showed that threshold structure was highly correlated with perceived loudness for low intensity tones. We assumed that the threshold of hearing was a measure of sensitivity. Relatively high measured threshold at a given frequency was indicative of low sensitivity. Low thresholds were indicative of high sensitivity at the particular frequency.

The model for the excitation pattern was derrived from a model by Colburn (1973). The Colburn model for excitation at a single location along the tonotopic axis was modified to include the sensitivity of the listener. Through this model the excitation level at any point in the excitation pattern was calculated.

Four models of the pitch extraction process were tested. The Peak Detection model generated predicted shifts in the location of the peak of the excitation pattern. These shifts in the location of the peak were translated directly into shifts in perceived pitch. The Centroid model generated estimates of perceived pitch from the excitation pattern by finding the centroid of the pattern. All tonotopic coordinates with excitation levels above threshold were weighted by that level. The sum of those weighted coordinate values was divided by the sum of the weights to yield the centroid. The Whitfield model estimated perceived pitch by finding the mean of the two boundaries of the excitation pattern. A boundary was defined as the tonotopic coordinate where a region of below threshold excitation levels along the tonotopic axis ended and a region of above threshold levels began. The extent of the excitation pattern along the axis was the distance between those two boundaries. The Whitfield Centroid model estimated perceived pitch by finding the mean of all tonotopic coordinates where the excitation level was above threshold. All points were given equal weight.

These models were evaluated by comparing the predicted shifts in pitch with intensity to real data. The evaluation was done by visual inspection of plots of both predicted and measured pitch shifts and by Pearson product-moment correlation scores. None of the models was very good. The two best models were the Peak Detection model and the Whitfield Centroid model. The two worst mdels were the Centroid model and the Whitfield model. All models failed by having low correlations between the predicted and measured pitch shifts.

Both the Peak Detection and the Whitfield Centroid models failed in the same way. As shown in the figures (Figure 22 and 26) both models predicted a peak in the pitch shift function at about 1900 Hz. The sensitivity curve for subject W (Figure 20) showed a peak at this frequency also. There was no peak in the measured pitch shifts, however. This was the typical failing of all of the models; predicting peaks in the pitch shift curve where none was measured.

Given the failure of all the models, what should be tried next? The first thing to note is that the two successful models provide conflicting cues to the next step. The Peak Detection model gave no weight to the tails of the excitation pattern. The Whitfield Centroid model gave equal weight to the center and the tails of the excitation pattern. These models worked equally well, failing in the same ways. A next step might be to give more weight to the tails of the excitation pattern than the center. That is essentially the Whitfield model (no weight to the center) and it failed quite completely. An alternative next step then might be to evaluate a model that only uses one tail of the excitation pattern (e.g. the maximum slope model or a Whitfield type model that only utilizes the lower boundary to determine perceived pitch). Clearly other models of the pitch extraction process need to be evaluated.

The problems with these models may, however, lie not with the extraction stage, but with the computation of the excitation pattern. Many alternatives are available at this stage. Some type of contrast enhancement effect might be appropriate at the level of the excitation pattern. The structure of the sensitivity curve might be magnified by that process. This would result in more predictions of peaks in the pitch shift curve where none existed in the data, however. It is clear that contrast enhancement is a characteristic of hearing (von Bekesy, 1967), though. Perhaps with the proper model of the extraction process contrast enhancement can be incorporated into this scheme.

The data of Kemp (1979b) suggests another change in the calculation of the excitation pattern. Kemp observed that the structure of the threshold curve was reflected in perceived loudness at low intensity levels. As the intensity increased the correlation decreased and disappeared at about 40 to 50 dB above threshold. This suggests that the sensitivity should affect the calculated excitation level less at high intensities. Perhaps contrast enhancement is also a function of intensity. Frequency discrimination improves somewhat at higher intensities (Wier, et al, 1977). Decreased enhancement at low intensities may help eliminate the problem of predicting inappropriate structure in the pitch shift curve.

Both of the more successful models worked the best when the center of the filter function was offset upward by one semitone. It is not clear whether this reflects a real property of the pitch extraction process or is required to make up for inadequacies in the models tested. If it is a real property of the system more attention needs to be paid to the shifts in best frequency with intensity shown by Rose, et al (1967). This might be another level dependent property of the excitation pattern calculation.

The notion that sensitivity microstructure exists at all has been questioned. If there is no structure in sensitivity then the logic of the entire paradigm must be doubted. In 1948, Gold (cited by Wit & Ritsma, 1980) proposed that under certain circumstances the auditory system emits acoustic energy. In a series of papers by Kemp (1978;1979a,1979b,; Kemp & Chum, 1980) the idea of emissions from an active process along the basilar membrane has been further developed. Kemp's basic idea is that an acoustic signal moving into the inner ear is partially reflected by rapid impedance changes along the basilar membrane. The reflected signal is transmitted back through the middle ear and into the ear canal. These acoustic emissions are strong enough to be detected. Sometimes the emissions are loud enough to be noticed by people near by (Zurek, 1980).

Because of another impedance gradient at the oval window a stimulus of the proper frequency will create a standing wave resonance in the inner ear. Kemp's theory states that minima in threshold microstructure occur at these resonance frequencies. Resonances of this type occur for frequencies having an integer number of periods corresponding to the time delay required for this echo. The threshold of hearing is lower, not because of increased sensitivity, but because the ear actively increases the intensity of the input sound by resonating, (Kemp, 1979a). The separation of those resonances, around 1000 Hz, is roughly 80 to 100 Hz. This matches well with the microstructure reported by ourselves and most other investigators. Idiosyncrasies in the intervals between threshold microstructure maxima and minima are attributed to idiosyncrasies in the mechanical properties of the basilar membrane.

Wilson (1980) agreed with this acoustic re-emission theory. He explained changes in the threshold microstructure with body position as changes in the impedance characteristics of the middle ear. Van den Brink (1980) also attributed changes in diplacusis and threshold microstructure to changes in impedance properties as body position changed. He observed that both the middle ear and the inner ear impedances change with body position.

This theory is by no means complete. Wit & Ritsma (1980) point out that current models of cochlear mechanical function predict echo delay times much shorter than observed. This theory is currently very much studied and numerous investigators are working on its exact mechanism (e.g. Rutten, 1980; Kemp & Churn, 1980; and others).

If the Kemp echo theory is correct, then the observed structure in the pitch shift curves is not due to sensitivity structure. There is no sensitivity structure. The observed pitch shift structure is due to changes in the relative loudness of the two tones as described in section V above.

This alternate explanation may be tested by the following experiment: Measure pitch shifts with percieved loudness changes. Equalize all the high intensity tones for loudness, and find the equiloudness curve for the low intensity level. Using the same experimental procedure have the listeners match the pitches of two tones having a uniform loudness difference. According to Kemp, all structure is now to be eliminated. The pitch shifts should match Stevens Rule or be zero. If the sensitivity model is correct then there is still structure in the excitation pattern that modifies the perceived pitch. This results in structure in the pitch shift curve.

With this experiment we can check the assumptions of our paradigm and test some of Kemp's ideas at the same time.

APPENDIX A.

APPENDIX A.

Main program

```
PROGRAM PITCH      74/175   OPT=1                              FTN 4.8+538

      PROGRAM PITCH(INPUT,OUTPUT)
      REAL FTONE(50,2),ZTONE(50,2),AMP(2)
      REAL E(200),XPT(50)
      REAL XPOINT(200),YPOINT(200,10)
      REAL PRED(50)
      DIMENSION THR(200),VT(200),FS(200),XM(200) ,P(50),PFREQ(50)
      DIMENSION RAW(3,100),PFREQ(3,100),SFREQ(3),EFREQ(3),ADJ(3)
      INTEGER TITLE(3,100),NNTHR(3),NTHR(3),PRIME
      REAL POINTS(50,10),RANGE(4)
      REAL FQ(200),SENS(200)
      INTEGER ITITLE(144),ICHAR(10),IMAG4(5151)
      ALOG2(X) = ALOG(X)/ALOG(2.0)
      ALNT = ALOG(2.0)
      START = ALOG(1200.0)/ALNT
      READ 901,(ITITLE(I),I=1,144)
901   FORMAT(72A1)
C*********  READ DATA AND SET UP ARRAYS        ***********
1     READ 100,NSET,PRIME
      IF(EOF(5LINPUT).NE.0) STOP
100   FORMAT(2I5)
      DO 5 II=1,NSET
      READ 101,(TITLE(II,I),I=1,80)
101   FORMAT(80A1)
      READ 102, SFREQ(II),EFREQ(II),NNTHR(II),ADJ(II)
102   FORMAT(2F8.1,I3,2F8.1)
      XS=ALOG(SFREQ(II))/ALNT   $     XE=ALOG(EFREQ(II))/ALNT
      ANTHR=NNTHR(II)   $    DX=(XE-XS)/(ANTHR-1.0)
      N=NNTHR(II)
      READ 103, (RAW(II,I),I=1,N)
103   FORMAT(12F6.2)
      DO 4 J=1,N
      RAW(II,J)=RAW(II,J)+ADJ(II)
      X=XS+DX*(J-1)
4     PFREQ(II,J)=X
5     CONTINUE
      NTH=,
      DO 8 II=1,NSET
```

```
      NTHR(II)=0
      N=NNTHR(II)
      DO 7 I=1,N
      IF(RFREQ(II,I).LT.RFREQ(PRIME,1)) GO TO 6
      M=NNTHR(PRIME)
      IF(RFREQ(II,I).GT.RFREQ(PRIME,M)) GO TO 6
      IF(II.NE.PRIME) GO TO 7
    6 NTH=NTH+1     $    NTHR(II)=NTHR(II)+1
      XM(NTH)=RFREQ(II,I)
      FQ(NTH) = 2.0**XM(NTH)
      THR(NTH)=RAW(II,I)
      SENS(NTH) = -THR(NTH)
      FS(NTH)=10.0**(-RAW(II,I)/20.0)
      YPOINT(NTH,1) = FS(NTH)
    7 CONTINUE
    8 CONTINUE
      J=NTH-1    $    N=0
      RSUM=0.0
      DO 9 I=1,J,2
      RSUM=RSUM+XM(I+1)-XM(I)
    9 N=N+1
      RSUM=RSUM/N
      DO 10 I=1,J
   10 WT(I)=(XM(I+1)-XM(I))/RSUM
      WT(NTH) = WT(NTH-1)
      NCNT = 0
      DO 50  I=1,NSET
      SFREQ(I) =  2.0**XM(NCNT+1)
      NCNT = NCNT + NTHR(I)
   50 EFFEQ(I) = 2.0**XM(NCNT)
      READ    104, NSHFT ,AL1,AL2
  104 FORMAT(I3,3X,2F5.1)
      PRINT 401
      DO 20  I=1,NSET
      ATHR=NTHR(I)
      PRINT 16    $    PRINT 106, (TITLE(I,J),J=1,80)
  106 FORMAT(# #,80A1)
```

```
PROGRAM PITCH       74/175  OPT=1                              FTN 4.8+538

   20    PRINT 15, SFREQ(I),EFREQ(I),ATHR
         PRINT 16
         DO 30 I=1,NTH
         AI = I
   30    PRINT 15, AI,FQ(I),YM(I),THR(I),FS(I),WT(I)
  105    FORMAT(F7.1,2X,F5.1)
   15    FORMAT(1H ,10F12.5)
   16    FORMAT(1H0)
         PRINT 16
         ASHFT=NSHFT $ ANTH=NTH $ PRINT 15,ANTH,ASHFT $ PRINT 16
         READ   105, (PFREQ(I),P(I),I=1,NSHFT)
         PS=0.
         DO 11 I=1,NSHFT
         PRINT 15,PFREQ(I),P(I)
         PS=PS+P(I)**2
   11    CONTINUE
         A1=10.**(AL1/20.) $ A2=10.**(AL2/20.)
C******  DATA NOW READY    ***********
         ICHAR(1) = ≠A≠
         RANGE(1) = 0.0
         RANGE(2) = 1.0
         RANGE(3) = 0.0
         RANGE(4) = 0.0
         DO 198 I=1,NTH
         XPOINT(I) = (ALOG(FS(I)/1200.0))/ALNT
  198    CONTINUE
         CALL USPLT
        1(XPOINT,FS,200,NTH,1,1,ITITLE,RANGE,ICHAR,1,IM/G4,IER)
         DO 199 ITR=1,NSHFT
         FTONE(ITR,1) = PFREQ(ITR)
         ZTONE(ITR,1) = (ALOG(PFREQ(ITR)))/ALNT
         XPT(ITR) = ZTONE(ITR,1) - START
         FTONE(ITR,2) = PFREQ(ITR) + P(ITR)
         ZTONE(ITR,2) = (ALOG(FTONE(ITR,2)))/ALNT
  199    CONTINUE
         AMP(1) = A1
         AMP(2) = A2
```

```
107   FORMAT(2F5.2)
108   FORMAT(3I5)
      READ 108,IT1,IT2,IT3
      READ 107,TSIZE
      READ 107, AK
      READ 107, CUT
      READ 108,ML1,ML2,ML3
      READ 108,MU1,MU2,MU3
      READ 108, ICF1S,ICF1E,ICF1I
      READ 108,ICF2S,ICF2E,ICF2I
      READ 107,CFSIZE
      DO 400 IC1=ICF1S,ICF1E,ICF1I
      CF1 = +IC1/CFSIZE
      ICF2E = IC1
      DO 400 IC2=ICF2S,ICF2E,ICF2I
      CF2 = +IC2/CFSIZE
      DO 400 IT=IT1,IT2,IT3
      T = IT
      T = T/TSIZE
      DO 400 IML=ML1,ML2,ML3
      AML1 = IML
      AML2 = AML1
      DO 400 IMU=MU1,MU2,MU3
      AMU1 = IMU
      AMU2 = AMU1
      ERS = 0.0
      SGNERR = 0.0
      PRINT 401
      DO 200 ITR = 1,NSHFT
201   TR = ITR
      F = FTONE(ITR,1)
      Y = ZTONE(ITR,1)
      EC = 0.0014
      Y = Y + CF1
      DO 205 J = 1,NTH
      X = XM(J)
      D = X - Y
```

```
            IF(D.GT.0.0) GO TO 202
            H = 2.0**(D*AML1)
            GO TO 203
 202        H = 2.0**(-D*AMU1)
 203        CONTINUE
            B = AMP(1)*FS(J)*H
            BB = AMP(1)*H
            EX = 0.0
            IF(B.GT.CUT) EX=(B/(SQRT(BB**2.0+AK)))*WT(J)
            YPOINT(J,1) = EX
 205        E(J) = EX
             CALL USANMX(E,NTH,1,ELMN,ELMX)
 C          CALL PITCH EXTRACTION SUBROUTINE
            Y = ZTONE(ITR,2)
            DY = 0.02
            DYSV = DY
            DO 250 K=1,50
            DO 210 J=1,NTH
            X = XM(J)
            D = X-Y
            IF(D.GT.0.0) GO TO 208
            H = 2.0**(D*AML2)
            GO TO 209
 208        H = 2.0**(-D*AMU2)
 209        CONTINUE
            B = AMP(2)*FS(J)*H
            BB = AMP(2)*H
            EX = 0.0
            IF(B.GT.CUT) EX=(B/(SQRT(BB**2.0+AK)))*WT(J)
 210        E(J) = EX
 C          NOW FIND CENTER OF SECOND TONE
 C          CALL PITCH EXTRACTION SUBROUTINE
 211        ER = ABS(ZCENT2-ZCENT1)
            IF(ER.LT.EC) GO TO 251
            DY = ABS(DYSV)
            IF(ZCENT2.GT.ZCENT1) DY = -DY
            IF((DY*DYSV).LT.0.0) DY = DY/2.0
```

```
      IF(ABS(DY).LT.0.00015) GO TO 251
      DYSV = DY
      Y = Y + DY
  250 CONTINUE
  251 REPS = K
      F2 = 2.0**(Y-CF2)
      DF = F2 - F
      PRED(ITR) = DF
      ED = DF - P(ITR)
      ERS = ERS + ED**2.0
      IF(SIGN(1.0,P(ITR)).NE.SIGN(1.0,DF)) SGNERR = SGNERR + 1.0
      POINTS(ITR,1) = P(ITR)
      POINTS(ITR,2) = DF
      IF((IT1.NE.IT2).OR.(ML1.NE.ML2).OR.(MU1.NE.MU2)) GO TO 207
      DO 301 IJK=1,NTH
      YPOINT(IJK,2) = E(IJK)
  301 CONTINUE
      RANGE(1) = 0.0
      RANGE(2) = 1.0
      RANGE(3) = ELMN
      RANGE(4) = ELMX
      ICHAR(1) = 1H1
      ICHAR(2) = 1H2
      ICHAR(3) = 1HL
      ICHAR(4) = 1HO
      CALL USPLT
     1(XPOINT,YPOINT,200,NTH,4,1,ITITLE,RANGE,ICHAR,1,IMAG4,IER)
  207 PRINT 15,F,ZTONE(ITP,1),ZTONE(ITP,1)+CF1,ZCENT1,FLOAT(N1)
      PRINT 15,F2,Y-CF2,Y,ZCENT2,FLOAT(N2),REPS,DF,P(ITR),ED,TR
  200 CONTINUE
      PRINT 15
      PRINT 16
      PRINT 15,AL1,AL2,AK
      PRINT 15,AML1,AMU1,CUT
      PRINT 15,AML2,AMU2,T,CF1,CF2
      PRINT 16
      CORREL = CORR(PRED,F,NSHFT)
```

```
PROGRAM PITCH        74/175  OPT=1                                    FTN 4.8+538

            CALL NORM(PRED,P,NSHFT,POINTS(1,1),POINTS(1,2),PRFM,PFM)
            PRINT 15,FRS,PS,SGNERR,CORREL,PFM,PRFM
            ICHAR(1) = ≠A≠
            ICHAR(2) = ≠P≠
            RANGE(3) = -50.0
            RANGE(4) = 50.0
            CALL USPLT
           1(XPT,POINTS,50,NSHFT,2,1,ITITLE,RANGE,ICHAR,1,IMAG4,IER)
      300   CONTINUE
            PRINT 16
            PRINT 16
            PRINT 16
            PRINT 15,AL1,AL2,AK
            PRINT 15,AML1,AMU1,CUT
            PRINT 15,AML2,AMU2,T,CF1,CF2
            PRINT 16
            PRINT 15,FRS,PS,SGNERR,CORREL,PFM,PRFM
      400   CONTINUE
            GO TO 1
      401   FORMAT(≠1WHITFIELD4 MODEL≠)
            END
```

FUNCTION CORR   74/175   OPT=1   FTN 4.8+538

```
      FUNCTION CORR(A,B,N)
      REAL A(50),B(50),SUMA,SSUMA,SUMB,SSUMB,SUMAB
      REAL TEMP1,TEMP2,TEMP3
      INTEGER N,I
      SUMA = 0.0
      SSUMA = 0.0
      SUMB = 0.0
      SSUMB = 0.0
      SUMAB = 0.0
      DO 100 I=1,N
      SUMA = SUMA + A(I)
      SSUMA = SSUMA + (A(I)*A(I))
      SUMB = SUMB + B(I)
      SSUMB = SSUMB + (B(I)*B(I))
  100 SUMAB = SUMAB + (A(I)*B(I))
      TEMP1 = (N*SUMAB) - (SUMA*SUMB)
      TEMP2 = (N*SSUMA) - (SUMA*SUMA)
      TEMP3 = (N*SSUMB) - (SUMB*SUMB)
       CORR = TEMP1/SQRT(TEMP2*TEMP3)
      RETURN
      END
```

SUBROUTINE NORM   74/175 OPT=1   FTN 4.8+538

```
      SUBROUTINE NORM(A,B,N,C,D,AFM,BFM)
      REAL A(50),B(50),C(50),D(50)
      SUMA = 0.0
      SUMB = 0.0
      DO 100 I=1,N
      SUMA = SUMA + A(I)
      SUMB = SUMB + B(I)
  100 CONTINUE
      AFM = SUMA/N
      BFM = SUMB/N
      DO 200 I=1,N
      C(I) = A(I)-AFM
      D(I) = B(I)-BFM
  200 CONTINUE
      RETURN
      END
```

```
FUNCTION BARK        74/175   OPT=1                                    FTN 4.8+508

      FUNCTION BARK(F)
C     JASA 68.1523...F IS IN HZ
      A=0.00076*F $ B=F/7500.0 $ B=B**2
      BARK=13.*ATAN(A) + 3.5*ATAN(B)
      END
```

APPENDIX B.

APPENDIX B.

Peak program

```
PROGRAM PEAK        74/175   OPT=1                                    FTN 4.P+508

          PROGRAM PEAK(INPUT,OUTPUT)
          DIMENSION THR(200),ST(200),FS(200),XM(200) ,P(50),PFREQ(50)
          DIMENSION RAW(3,100),RFREQ(3,100),SFREQ(3),EFREQ(3),ADJ(3)
          INTEGER TITLE(3,100),NNTHR(3),NTHR(3),PRIME ,WIDTH
          REAL SMTHR(200)
          DOUBLE PRECISION   AOK,TP,TPP,GPPOG,SHFT,LN10T
          REAL INCH
          REAL POINTS(50,10),RANGE(4)
          INTEGER ITITLE(144),ICHAR(10),IMAG4(5151)
          INCH = 0.02634950814 25
          ALNT = ALOG(2.0)
C********  READ DATA AND SET UP ARRAYS       ***********
          READ 100,NSET,PRIME
    100   FORMAT(2I5)
          DO 5 II=1,NSET
          READ 101,(TITLE(II,I),I=1,80)
    101   FORMAT(80A1)
          READ 102, SFREQ(II),EFREQ(II),NNTHR(II),ADJ(II)
    102   FORMAT(2F8.1,I3,2F8.1)
          XS=ALOG(SFREQ(II))/ALNT    $    XE=ALOG(EFREQ(II))/ALNT
          ANTHR=NNTHR(II)    $    DX=(XE-XS)/(ANTHR-1.0)
          N=NNTHR(II)
          READ 103, (RAW(II,I),I=1,N)
    103   FORMAT(12F6.2)
          DO 4 J=1,N
          RAW(II,J)=RAW(II,J)+ADJ(II)
          X=XS+DX*(J-1)
    4     RFREQ(II,J)=X
    5     CONTINUE
          NTH=0
          DO 8 II=1,NSET
          NTHR(II)=0
          N=NNTHR(II)
          DO 7 I=1,N
          IF(RFREQ(II,I).LT.RFREQ(PRIME,1)) GO TO 6
          M=NNTHR(PRIME)
          IF(RFREQ(II,I).GT.RFREQ(PRIME,M)) GO TO 6
          IF(II.NE.PRIME) GO TO 7
    6     NTH=NTH+1    $    NTHR(II)=NTHR(II)+1
```

```
      XM(NTH)=RFREQ(II,I)
      THR(NTH)=RAW(II,I)
      FS(NTH)=10.0**(-RAW(II,I)/20.0)
7     CONTINUE
8     CONTINUE
      J=NTH-1    $    N=0
      RSUM=0.0
      DO 9 I=1,J,2
      RSUM=RSUM+XM(I+1)-XM(I)
9     N=N+1
      RSUM=RSUM/N
      DO 10 I=1,J
10    WT(I)=(XM(I+1)-XM(I))/RSUM
      WT(NTH) = WT(NTH-1)
      NCNT = 0
      DO 50  I=1,NSET
      SFREQ(I) =  2.0**XM(NCNT+1)
      NCNT = NCNT + NTHR(I)
50    EFREQ(I) = 2.0**XM(NCNT)
C******   MOVING AVERAGE SMOOTHING
      WIDTH = 5
      K = WIDTH/2+1
      J = NTH-WIDTH/2
      DO 500 I=K,J
      SMTHR(I) = THR(I-2) + THR(I-1) + THR(I) + THR(I+1) + THR(I+2)
500   SMTHR(I) = SMTHR(I)/WIDTH
      PRINT 401
      DO 20  I=1,NSET
      ATHR=NTHR(I)
      PRINT 15   $   PRINT 106, (TITLE(I,J),J=1,80)
106   FORMAT(/ /,80A1)
20    PRINT 15, SFREQ(I),EFREQ(I),ATHR
      PRINT 16
      DO 30 I=1,NTH
      F=2.0**XM(I)   $   Z=BARK(F)   $   AI=I
30    PRINT 15,AI,XM(I),F,Z,THR(I),SMTHR(I),FS(I),WT(I)
      READ   104, NSHFT ,AL1,AL2
104   FORMAT(I3,3X,2F5.1)
```

```
          ASHFT=NSHFT $ ANTH=NTH $ PRINT 15,ANTH,ASHFT $ PRINT 16
          READ   105, (PFREQ(I),P(I),I=1,NSHFT)
    105   FORMAT(F7.1,2X,F5.1)
     15   FORMAT(8F15.5)
     16   FORMAT(1H0)
          PRINT 16
          PS=0.
          DO 11 I=1,NSHFT
          PRINT 15,PFREQ(I),P(I)
          PS=PS+P(I)**2
     11   CONTINUE
          PRINT 16
    C******  DATA NOW READY   ***********
          READ 991,(ICHAR(I),I=1,10)
          READ 991,(ITITLE(I),I=1,144)
    991   FORMAT(72A1)
          A1=10**(AL1/20.) $ A2=10.**(AL2/20.)
          ML1 = 40  $  MU1 = 40
          AK = 100000.0
          ASTEP = 1.0
          DO 400 IDST=2,4
          DIST = ASTEP*(IDST-8)
          ICTOFF = DIST*INCH
          PRINT 401
          RANGE(1) = 0.0
          RANGE(2) = 0.0
          RANGE(3) = 0.0
          RANGE(4) = 0.0
          AMU1=MU1 $ AML1=ML1
          AMU2=AMU1 $ AML2=AML1
          PRINT 15,AL1,AL2,AK
          PRINT 15,AML1,AMU1,ICTOFF,DIST
          PRINT 15,AML2,AMU2 $ PRINT 16
          PRINT 199
    199   FORMAT(# PFREQ(I),CFREQ,P(I),PREDICTED SHIFT,ERROR#)
```

```
      ERS=0.
      SGNERR = 0.0
      II = 0  $  AOK = A1/AK  $  LN10T = ALOG(10.)/20.  $  AMM = AML1
      DO 200I=1,NSHFT
      PRINT 16
      CENTER = ALOG(PFREQ(I))/ALNT
      CENTER = CENTER + OCTOFF
      CFREQ = 2.0**CENTER
201   II = II + 1
      IF(XM(II).LT.CENTER) GO TO 201
      IF(ABS(XM(II+1)-CENTER).LT.ABS(XM(II)-CENTER))    II = II + 1
      TP = (SMTHR(II+1)-SMTHR(II-1))/(XM(II+1)-XM(II-1))
      IF(TP.GT.0.0) AMM = AMU1
      TPP = (SMTHR(II-1)-2*SMTHR(II)+SMTHR(II+1))/
     1(2.0*(((XM(II+1)-XM(II-1))/2.0)**2.0))
      GPPOG = (1-(2*AOK**2)) /(1+(2*(AOK**2))+(AOK**4))
      SHFT=(TP*LN10T)/((AMM**2)*GPPOG+(LN10T*TP)**2-LN10T*TPP)
C
      PREDF = PFREQ(I) * 2.0**SHFT
      PRESHF = PREDF-PFREQ(I)
      IF(SIGN(1.0,PRESHF).NE.SIGN(1.0,P(I)))SGNERR = SGNERR + 1.0
      ERROR = PRESHF-P(I)
      ERS = ERS + ERROR*ERROR
      PRINT 15,PFREQ(I),CFREQ,P(I),PRESHF,ERROR
      POINTS(I,1) = P(I)
      POINTS(I,2) = PRESHF
  200 CONTINUE
      PRINT 15,ERS,PS,SGNERR
      CALL USPLT(PFREQ,POINTS,50,NSHFT,2,1,ITITLE,RANGE,ICHAR,1,IMAG4
     1,IER)
400   CONTINUE
401   FORMAT(/1PEAK CALCULATION MODEL/)
      END
```

Whitfield (a) subroutine

```
SUBROUTINE ZFIT          74/175    OPT=2 TRACE   FTN [illegible]          [illegible]

             SUBROUTINE ZFIT (XM,NTH,Y,T,ZCENT)
             REAL XM(250),E(250),Y,ZCENT,A,B,C,EPS,T
             INTEGER IZL,IZU,II,NTH,NTHM1
             NTHM1 = NTH - 1
             EPS = 0.1
      1      IZL = [illegible]
             IZU = [illegible]
             DO 100 II = 2,NTHM1
             B = ABS(E(II)-T)
             IF(B.GT.EPS) GO TO 1
             A = E(II-1) - T
             C = E(II+1)-T
             IF(XM(II).GT.Y) GO TO 99
             IF((A.LT.0.0).AND.(C.GT.0.0)) IZL = II
             GO TO 100
      99     IF((A.GT.0.0).AND.(C.LT.0.0)) IZU = II
             IF(IZU.NE.0) GO TO 101
      100    CONTINUE
             ZCENT = -9999.0
             RETURN
      101    ZCENT = (XM(IZL) + XM(IZL))/2.0
             RETURN
             END
```

Whitfield (b) subroutine

```
SUBROUTINE WFIT        74/175   OPT=2 TRACE  FNDMP                 FTN 4.8+528

            SUBROUTINE WFIT (XM,E,NTH,Y,T,ZCENT)
            REAL YM(200),E(200),Y,ZCENT,A,B,C,EPS,T
            INTEGER IZL,IZU,II,NTH,NTHM1
            NTHM1 = NTH - 1
            EPS = 0.1
      1     IZL = 0
            IZU = 0
            DO 100 II = 2,NTHM1
            B = ABS(E(II)-T)
            IF(B.GT.EPS) GO TO 100
            A = E(II-1) - T
            C = E(II+1)-T
            IF(YM(II).GT.Y) GO TO 99
            IF(IZL.NE.0) GO TO 99
            IF((A.LT.0.0).AND.(C.GT.0.0)) IZL = II
            GO TO 100
     99     IF((A.GT.0.0).AND.(C.LT.0.0)) IZU = II
    100     CONTINUE
            IF((IZL.NE.0).AND.(IZU.NE.0))GO TO 101
            ZCENT = -9999.0
            RETURN
    101     ZCENT = (YM(IZL) + YM(IZU))/2.0
            RETURN
            END
```

Centroid subroutine

```
SUBROUTINE CNTRD       74/175  OPT=1 TRACE  PMDMP              FTN 4.8+528

              SUBROUTINE CNTRD (XM,E,NTH,ZCENT)
              REAL XM(200),E(200),ZCENT,SUM1,SUM2
              INTEGER NTH
        1     SUM1 = 0.0
              SUM2 = 0.0
              DO 100 II = 1,NTH
              SUM1 = SUM1 + E(II)*XM(II)
              SUM2 = SUM2 + E(II)
      100     CONTINUE
              ZCENT = 0.0
              IF(SUM2.NE.0.0) ZCENT = SUM1/SUM2
              RETURN
              END
```

Whitfield centroid subroutine

```
SUBROUTINE WHIT        74/175   OPT=1                                   FTN 4.8+538

                 SUBROUTINE WHIT (XM,E,NTH,VMN,VMX,THR,DSP,INDX,ZCENT,N)
                 REAL XM(200),E(200),ZCENT,SUM1,SUM2
                 REAL VMN,VMX,DSP(200,10),THR
                 INTEGER NTH,INDX,N
           1     SUM1 = 0.0
                 SUM2 = 0.0
                 N = 0
                 DO 100 II = 1,NTH
                 IF(E(II).GE.THR) GO TO 75
                 IF(INDX.EQ.3) DSP(II,INDX) = VMX + 10.0
                 IF(INDX.EQ.4) DSP(II,INDX) = VMN - 10.0
                 GO TO 100
          75     CONTINUE
                 IF(INDX.EQ.3) DSP(II,INDX) = VMX
                 IF(INDX.EQ.4) DSP(II,INDX) = VMN
                 N = N + 1
                 SUM1 = SUM1 + XM(II)
                  SUM2 = SUM2 + 1.0
         100     CONTINUE
                 ZCENT = 0.0
                 IF(SUM2.NE.0.0) ZCENT = SUM1/SUM2
                 RETURN
                 END
```

LIST OF REFERENCES

LIST OF REFERENCES

Allen, J.B. Cochlear modelling - 1980. Unpublished manuscript, 1980 (a).

Allen, J.B. A cochlear micromechanical model of transduction. In G. van den Brink & F.A. Bilsen (Eds.) Psychological, physiological, and behavioral studies in hearing. Delft, The Netherlands : Delft University Press, 1980 (b).

Bekesy, G. von. Experiments in hearing. E.G. Wever (Trans. and Ed.) New York : McGraw-Hill Book Co., Inc., 1960.

Bekesy, G. von. Sensory inhibition. Princeton : Princeton University Press, 1967.

Bilsen, F.A. Pitch of noise signals : evidence for a central spectrum. Journal of the Acoustical Society of America, 1977, 61, 150-161.

Brink, G. van den. Experiments on binaural diplacusis and tone perception. In R. Plomp & G. Smoorenburg (Eds.) Frequency analysis and periodicity detection. Leiden : A.W. Sijthoff, 1971.

Brink, G. van den. Cochlear mechanics as the possible cause of binaural diplacusis. In G. van den Brink & F.A. Bilsen (Eds.) Psychological, physiological, and behavioral studies in hearing. Delft, The Netherlands : Delft University Press, 1980.

Cohen, A. Further investigation of the effects of intensity upon the pitch of pure tones. Journal of the Acoustical Society of America, 1961, 33, 1363-1376.

Cohen, M.F. Temporal integration and the microstructure of the detection threshold curve. Journal of the Acoustical Society of America, 1979, 66, S8. (Abstract)

Cohen, M.F. Detection threshold microstructure and its effect on temporal integration. Private Communication, 1981.

Cohen, M.F. & Schubert, L.P. Microstructure of the detection threshold. Journal of the Acoustical Society of America, 1979, 65, S55. (Abstract)

Colburn, H.S. Theory of binaural interaction based on auditory nerve data. I. general strategy and preliminary results on interaural discrimination. Journal of the Acoustical Society of America, 1973, 54, 1458-1470.

Colburn, H.S. & Durlach, N.I. Models of binaural interaction. In E.C. Carterette & M.P. Freidman (Eds.) Handbook of perception (Vol. IV Hearing) New York : Academic Press, 1978.

Corey, V.B. Change in pitch of sustained musical note by clenching the jaws of the observer. Journal of the Acoustical Society of America, 1950, 22, 640.

Cramer, E.M. & Huggins, W.H. Creation of pitch through binaural interaction. Journal of the Acoustical Society of America, 1958, 30, 413-417.

Dallos, P. Biophysics of the cochlea. In E.C. Carterette & M.P. Friedman (Eds.) Handbook of perception (Vol, IV Hearing) New York : Academic Press, 1978.

Durlach, N.I. Note on the creation of pitch through binaural interaction. Journal of the Acoustical Society of America, 1962, 34, 1096-1099.

Durlach, N.I. Equalization and cancellation theory of binaural masking-level differences. Journal of the Acoustical Society of America, 1963, 35, 1206-1218.

Durlach, N.I. Binaural signal detection : Equalization and cancellation theory. In J.V. Tobias (Ed.) Foundations of modern auditory theory (Vol. II) New York : Academic Press, 1972.

Elliott, E. A ripple effect in the audiogram. Nature, 1958, 181, 1076.

Evans, E.F. Frequency selectivity at high signal levels of single units in cochlear nerve and nucleus. In E.F. Evans & J.P. Wilson (Eds.) Psychophysica and physiology of hearing. New York : Academic Press, 1977.

Fastl, H. Uber tonhohenempfindungen bei rauschen. Acustica, 1971, 25, 350-354.

Fletcher, H.F. & Munson, W.A. Loudness, its definition, measurement and calculation. Journal of the Acoustical Society of America, 1833,5,82-108.

Fletcher, H. Loudness, pitch, and the timbre of musical tones and their relation to the intensity, the frequency, and the overtone structure. Journal of the Acoustical Society of America, 1934, 6, 59-69.

Galambos, R. & Davis, H. The response of single auditory nerve fibers to acoustic stimulation. Journal of Neurophysiology, 1943, 6, 39-57.

Green, D.M. An introduction to hearing. Hillsdale, Hew Jersey: Lawrence Erlbaum Associates, 1976.

Guttman, N. Pitch and loudness of a binaural subjective tone. Journal of the Acoustical Society of America, 1962, 34, 1996. And Bell Telephone Laboratories Memo.

Hartmann, W.M. The effect of amplitude envelope on the pitch of sine wave tones. Journal of the Acoustical Society of America, 1978, 63, 1105-1113.

Helmholtz, H. von [On the sensation of tone] (A.J. Ellis, Ed. and trans.) New York : Dover Publications, Inc., 1954. (Originally published, 1863)

Hirsh, I.J. The influence of interaural phase on interaural summation and inhibition. Journal of the Acoustical Society of America, 1948, 20, 536-544.

Huggins, W.H. & Licklider, J.C.R. Place mechanisms of auditory frequency analysis. Journal of the Acoustical Society of America, 1951, 23, 290-299.

Jeffress, L.A. A place theory of sound localization. Journal of comparative and physiological psychology, 1948, 41, 35-39.

Johnstone, B.M. & Boyle, A.J. Basilar membrane vibration examined with the Mossbauer technique. Science, 1967, 158, 389-390.

Johnstone, J.R. Properties of ganglion cells fron the extreme basal region of guinea pig cochlea. In E.F. Evans & J.P. Wilson (Eds.) Psychophysics and physiology of hearing. New York : Academic Press, 1977.

Julesz, B. Foundations of cyclopean perception. Chicago : University of Chicago Press, 1971.

Kemp, D.T. Simulated acoustic emissions fron within the human auditory system. Journal of the Acoustical Society of America, 1978, 64, 1386-1391.

Kemp, D.T. Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea. Archives of Otorhinolaryngology, 1979, 224, 37-45. (a)

Kemp D.T. The evoked cochear mechanical response and the auditory microstructure - evidence for a new element in cochlear mechanics. In A. Hoke & E. de Boer (Eds.) Models of the auditory system and related signal processing techniques. Scandinavian audiology, 1979, Supplement 9, 35-47. (b)

Kemp, D.T. & Chum, R.A. Observations on the generator mechanisms of stimulus frequency acoustic emissions - two tone suppression. In G. van den Brink & F.A. Bilsen (Eds.) Psychological, physiological, and behavioral studies in hearing. Delft, The Netherlands : Delft University Press, 1980.

Kiang, N.Y.S. Processing of speech by the auditory nervous system. Journal of the Acoustical Society of America, 1980, 68, 830-835.

Kiang, N.Y.S., Watanabe, T., Thomas, E.C., and Clark, L.F. Discharge patterns of single fibers in the cat's auditory nerve. M.I.T. research monograph No. 35. Cambridge, MA : MIT Press, 1965.

Klein, M.A. & Hartmann, W.M. Binaural edge pitch. Journal of the Acoustical Society of America, 1981, 70, 51-61.

Lewis, D. & Cowan, M. The influence of intensity on the pitch of violin and 'cello tones. Journal of the Acoustical Society of America, 1936, 8, 20-22.

Licklider, J.C.R. The influence of interaural phase relations upon the masking of speech by white noise. Journal of the Acoustical Society of America, 1948, 20, 150-159.

Licklider, J.C.R. A duplex theory of pitch perception. Experientia, 1951, 7, 128-133.

Licklider, J.C.R. Auditory frequency analysis. In C. Cherry (Ed.) Proceedings of third symposium on information theory. New York: Academic Press, 1956.

Licklider, J.C.R. Three auditory theories. In S. Koch (Ed.) Psychology, a study of a science. New York:McGraw Hill, 1959.

Long, G.R. The microstructure of quiet and masked thresholds. Journal of the Acoustical Society of America, 1980, 68, S36.

Morgan, C.T., Garner, W.R., & Galambos, R. Pitch and intensity. Journal of the Acoustical Society of America, 1951, 23, 658-663

Nordmark, J.O. Time and frequency analysis. In J.V. Tobias (Ed.) Foundations of modern auditory theory. (Vol. 1) New York:

Academic Press, 1970.

Plomp, R. & Bouman, M.A. Relation between hearing threshold and duration for tone pulses. Journal of the Acoustical Society of America, 1959, 31, 749-758.

Raatgever, J. & Bilsen, F.A. Lateralization and dichotic pitch as a result of spectral pattern recognition. In E.F. Evans & J.P. Wilson (Eds.) Psychophysics and physiology of hearing. London: Academic Press,1977.

Rakowski, A. The magic number two: seven examples of binary apposition in pitch theory. The humanities association review, 1979, 30, 25-45.

Rhode, W.S. Observations of the vibration of the basilar membrane in squirrel monkey using the Mossbauer technique. Journal of the Acoustical Society of America, 1971, 49, 1218-1231.

Rose, J.E., Brugge, J.F., Anderson, D.J., & Hind, J.E. Phase-locked response to low frequency tones in single auditory nerve fibers of the squirrel monkey. Journal of Neurophysiology, 1967, 30, 769-793.

Rose, J.E. Hind, J.E., Anderson, D.J., & Brugge, J.F. Some effects of stimulus intensity on response of auditory nerve fibers in the squirrel monkey. Journal of neurophysiology, 1971, 34, 685-699.

Rutten, W.L.C. Latencies of stimulated acoustic emissions in normal human ears. In G. van den Brink & F.A. Bilsen (Eds.) Psychophysical, physiological, and behavioral studies in hearing. Delft, The Netherlands: Delft University Press, 1980.

Siebert, W.M. Stimulus transformations in the peripheral auditory system. In P.A. Kolers & M. Eden (Eds.) Recognizing Patterns.

Cambridge, Mass.:M.I.T. Press, 1968.

Small, A.M. & Campbell, R.A. Pitch shifts of periodic stimuli with changes in sound level. Journal of the acoustical society of America, 1961, 33, 1022-1027.

Snow, W.B. Changes of pitch with loudness at low frequencies. Journal of the acoustical society of America, 1936, 8, 14-19.

Stern, R.M. & Colburn, H.S. Theory of binaural interaction based on auditory nerve data. IV. a model for subjective lateral position. Journal of the acoustical society of America, 1978, 64, 127-140.

Stevens, S.S. The relation of pitch to intensity. Journal of the acoustical society of America, 1935, 6, 150-154.

Stevens, S.S. & Davis, H. Hearing its psychology and physiology. New York:Wiley & Sons, Inc., 1938.

Terhardt, E. Pitch shifts of harmonics, an explanation of the octave enlargement phenomenon. Proceedings of the seventh international congress on acoustics, (Vol. 3). Budapest, 1971.

Terhardt, E. Pitch of pure tones: its relation to intensity. In E. Zwicker & E. Terhardt (Eds.) Facts and models in hearing. New York:Springer-Verlag, 1974.

Terhardt, E. Influence of intensity on the pitch of complex tones. Acustica, 1975, 33, 344-348.

Terhardt, E. Pitch shift of monaural pure tones caused by contralateral sounds. Acustica, 1977, 37, 56-57.

Terhardt, E. Conceptual aspects of musical tones. The humanities association review, 1979, 30, 46-57.

Terhardt, E. & Fastl, H. Zum einfluss von stortonen und storgerauschen auf die tonhohe von sinustonen. Acustica, 1971, 25, 53-61.

Thomas, I.B. Microstructure of the puretone threshold. Journal of the acoustical society of America, 1975, 57, S26-S27. (Abstract).

Thurlow, W.R. Studies in auditory theory I. binaural interaction and the perception of pitch. Journal of experimental psychology, 1943, 32, 17-36.

Verschuure, J. & van Meeteren, A.A. The effect of intensity on pitch. Acustica, 1975, 32, 33-44.

Ward, W.D. Diplacusis and auditory theory. Journal of the acoustical society of America, 1963, 35, 1746-1747.

Ward, D.W. Musical perception. In J.V. Tobias (Ed.) Foundations of modern auditory theory. (Vol. 1.) New York:Academic Press, 1970.

Wever, E.G. Theory of hearing. New York:Dover, 1970. (Originally published, 1949)

Whitfield, I.C. The neural code. In E.C. Carterette & M.P. Friedman (Eds.) Handbook of perception (Vol. 4. Hearing). New York: Academic Press, 1978.

Wier, C.C., Jesteadt, W., & Green, D.M. Frequency discrimination as a function of frequency and sensation level. Journal of the acoustical society of America, 1977, 61, 178-184.

Wightman, F.L. The pattern-transformation model of pitch. Journal of the acoustical society of America, 1973, 54, 407-416.

Wightman, F.L., Grantham, D.W., & Fowler, C. Existence region of dichotic noise pitch. Journal of the acoustical society of America, 1977, 62, S24. (Abstract)

Wilson, J.P. Evidence for a cochlear origin for acoustic re-emissions threshold fine structure, and tonal tinnitus. Hearing research, 1980, 2, 233-252.

Wilson, J.P. & Johnstone, J.R. Basilar membrane and middle ear vibration in guinea pig, measured by capacitive probe. Journal of the acoustical society of America, 1975, 57, 705-723.

Wit, H.P. & Ritsma, R.J. On the mechanism of the evoked cochlear mechanical response. In G. van den Brink & F.A. Bilsen (Eds.) Psychophysical, physiological, and behavioral studies in hearing. Delft, The Netherlands: Delft University Press, 1980.

Zurek, P.M. Spontaneous narrowband acoustic signals emitted by human ears. Journal of the acoustical society of America, 1981, 69, 514-523.

Zwicker, E. Masking and psychological excitation as consequences of the ear's frequency analysis. In R. Plomp & G.F. Smoorenburg (Eds.) Frequency analysis and periodicity detection in hearing. Leiden:A.W.Sijthoff, 1970.

Zwicker, E. & Feldkeller, R. Das ohr als nachrichten empfanger. Stuttgart:S.Hirzel, 1967.

Zwislocki, J. Analysis of some auditory characteristics. In R.R. Luce, R.R. Bush, & E. Galanter (Eds.) Handbook of mathematical psychology, (Vol. 3). New York:Wiley, 1965.