1995

Effects of Parameters of Spectrally Remote Frequencies on Binaural Processing

Anthony N. Grange
Loyola University Chicago

LOYOLA UNIVERSITY OF CHICAGO

EFFECTS OF PARAMETERS OF SPECTRALLY REMOTE FREQUENCIES ON BINAURAL PROCESSING

A DISSERTATION SUBMITTED TO THE FACULTY

OF THE GRADUATE SCHOOL IN CANDIDACY

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

BY

ANTHONY N. GRANGE

CHICAGO, ILLINOIS

MAY, 1995

ACKNOWLEDGEMENTS

I would like to thank those who served on my committee for their advice, patience, and understanding during the dissertation process. The environment in which this work was done, including the accessibility and willingness on their part to discuss any matters at all points along the way, was invaluable.

I would also like to thank Shauneen, my first (and current) wife, whose impending wrath provided additional motivation at every step.

Thanks to my parents, who stressed the importance of an education, and provided the means for obtaining one.

Special thanks to Annie, who always kisses me when I come home.

TABLE OF CONTENTS

TABLE OF CONTENTS (cont.)

LIST OF FIGURES

Figure Page

ABSTRACT

Three experiments and a computer simulation of human subjects' behavior were conducted to gain insight into how processing of interaural time (or phase) differences at low frequencies is affected by the interaural parameters of other spectral regions.

The task of the first two experiments was to detect an interaural phase shift of a 20-Hz wide band of noise centered on 500 Hz. In the first experiment, the target band was surrounded by diotic noise of varying bandwidth. The noise outside of the diotic band was interaurally uncorrelated. Twenty statistically independent samples of each condition, each with an identical target interaural phase shift, were stored in a computer, and detection performance was measured for each sample. Percent correct ranged from chance to over 90% for all subjects in all conditions. The data were accounted for by a comparison of the cross-correlation coefficient of a filter pair centered on the target with the cross-correlation of surrounding filters.

A computer simulation of an experiment that measured IPD detection thresholds using an identical stimulus (Grange, 1991) was run, to see if behavior could be modelled assuming a fixed binaural processing bandwidth. The simulation data did not fit the behavioral data well, and it was argued that a fixed processing band was inadequate to describe binaural processing; however, there were several potential cues that were not incorporated into the simulation.

In the second experiment, interaural parameters of off-target frequencies varied over

time. Detection performance improved when the interaural phase of the remote frequencies initially differed from that of the target band, suggesting that listeners can differentially weight spectral regions on the basis of their interaural parameters. A general model was suggested that includes spectral weighting prior to combination of interaural information across frequencies.

The final experiment was a replication of a tone-in-noise detection experiment (Sondhi and Guttman, 1966) using a more rigorous psychophysical procedure. The interaural configuration of the noise was similar to that used in the previous two experiments, and the discussion includes an interpretation of the data in terms of the previously mentioned model.

CHAPTER 1

GENERAL INTRODUCTION

In an everyday listening environment, the acoustic signal that the auditory system receives can be quite complex. There may be multiple sound sources, and the complex waveforms produced by each may overlap both spectrally and temporally. The signals from each sound source are combined so that what arrives at the two ears are not separate sets of components representing different sources, but a summation of all acoustic energy across sources. One of the primary tasks of the auditory system is sound source determination, which involves grouping together the various frequency components that arise from a particular sound source, segregating them from all other sources (Yost, 1991a). One way that this may be accomplished is on the basis of common interaural values that components arising from a sound source will have. The purpose of this dissertation is to investigate a listener's ability to process interaural time differences of a low-frequency target, the influence of the spectral parameters of frequencies surrounding the target on this process, and the role of interaural information in sound source determination.

Several steps are required for sound source determination to proceed. The complex waveform must first be separated into its individual frequency components. This is achieved by the bandpass characteristics of elements in the peripheral auditory system (see Moore, 1988, for a review). Each component must then be analyzed in terms of various parameters (e.g., interaural time, amplitude, and relatively slow temporal fluctuations in amplitude). Individual components arising from the same sound source must then be integrated, presumably on the basis of commonalities among these parameters. Attributes of sound that provide cues for sound source determination can be categorized into two broad classes - monaural and binaural. Monaural cues are those which can result in grouping of components based on information at one ear; for example, if the spectral components comprising one speaker's voice had a pattern of relatively slow amplitude fluctuations that differed from the pattern of another (simultaneously present) voice, segregation of the two sets of components could proceed on that difference (Yost and Sheft, 1989; Bregman, 1990).

Binaural cues are those that arise from a comparison of the waveforms at the two ears. Two cues of this nature are particularly important for localization of a sound source in the horizontal plane - interaural differences of level (IDL's) and interaural differences of time (IDT's). IDL's exist because an object in the path of acoustical energy will reflect some of that energy; therefore, the level of the sound will be attenuated behind the object. The degree to which sound is reflected depends on the relationship between the size of the object and the wavelength of the sound - a short wavelength will be reflected more than a long wavelength. An object the size of a human head will not appreciably attenuate frequencies below about 2000 Hz, leaving IDT's as the dominant cue at low frequencies.

Interaural differences of time (IDT's) occur when the distance from the sound source to one ear is longer than the distance to the other ear, as occurs for sounds off from the midsagittal plane (see Yost and Hafter, 1987, for a review). This introduces a latency between the left-ear and right-ear neural response to a particular instance of the stimulus waveform, since the waveform reaches one ear before the other. The interaural delay in firing times can be used as a cue to sound source location for low frequency stimuli. At higher frequencies, however, the period of the waveform decreases to the point where the maximum ITD that could possibly occur (determined by the speed of sound and the distance between the ears) is greater than the period. At these frequencies, any temporal relationship between spikes arriving from the two ears is ambiguous, since there are multiple spatial locations from which a sound resulting in that relationship could have originated. If, for example, the period of the waveform were equal to the IDT resulting from a source to one side of the listener, the binaural auditory system would receive spikes from each ear simultaneously, a situation that would also arise if the sound source were directly in front of the listener. Therefore, at higher frequencies, the information arising from IDT's is ambiguous, and provides no information about the location of sound sources. While the Duplex Theory is correct for pure tone stimuli, such sounds are rarely encountered in a real listening environment. It has also been found that IDT's in the *envelopes* of high frequency complex stimuli can also serve as cues to lateralization (McFadden and Pasanen, 1976). While localization and sound source determination are not synonymous, localization may be used as a cue for source determination, though spatial separation does not guarantee that separate auditory images will be formed (Hartmann, 1988).

Models of Lateralization

Most of the data arising from studies of the processing interaural differences of time have been collected using sounds presented over headphones, since this allows independent manipulation of interaural time and level. The term "lateralization" is used when referring to stimuli presented over headphones to differentiate the percept from that of "localization." Sounds presented over headphones result in images within the listeners head (intracranial images), while sounds presented over loudspeakers (or naturally occurring sounds) are perceived as being external.

In general, models of lateralization propose that the left-ear and right-ear waveforms are passed through banks of bandpass filters, a process that occurs in the auditory periphery. The outputs of frequency matched pairs of filters are then cross-correlated (see Colburn & Durlach, 1978), which involves multiplication of the left and right filter outputs and summation of the cross-products. A physiologically plausible mechanism of cross-correlation has been proposed by Jeffress (1948), in which a set of coincidence detectors receive the left and right outputs from each set of frequency matched peripheral filters. A different delay is added to one of the channels for each coincidence detector, so that the most active coincidence detector will be the one in which the delay introduced internally offsets the interaural delay in the stimulus.

A cross-correlation function (cross-correlation values for a range of interaural delays) is computed for each frequency pair by delaying filter output relative to the other, then multiplying and summing the products, for a range of delays. The cross-correlation value (for the discrete case) at frequency f and interaural delay τ is given by

$$R(f,\tau) = \Sigma\ [l(f,i) * r(f,i-\tau)], \text{ for } i=1,\mathrm{T} \qquad (1)$$

where T is the number of points in each waveform (sampling rate * duration). Equation

1 gives the average cross-correlation over the entire duration of the waveform. Binaural models generally make use of *running* cross-correlation, in which a value is computed for each moment t, with the cross-products summed over a limited temporal window, Ω, so that past auditory events do not continue to affect the predicted position of the intracranial image. Running cross-correlation values are given by

$$R(f,t,\tau) = \Sigma\ [l(f,t-i) * r(f,(t-i)-\tau) * e^{-i/\Omega}], \text{ for } i=0,t. \qquad (2)$$

The values of the running cross-correlation can be plotted in (f,τ) space for a given moment in time. For a sinusoidal stimulus, the cross-correlation function is cosinusoidal, and lateralization is predicted by the position of the peak of the cross-correlation function for the filter pair centered on the stimulus frequency.

Various transformations of the running cross-correlation values have been proposed to account for several characteristics of sound localization. For example, a central weighting function that places greater weight on those cross-correlation peaks nearer interaural delays of 0 μs has been postulated (Jeffress, 1972; Stern and Colburn, 1978) to account for better localization acuity for sound sources directly in front of the listener (see Blauert, 1983). In models based on coincidence detection networks (Colburn, 1977), a greater concentration of neural units near interaural delays of 0 μs serves an identical purpose (Stern & Colburn, 1978; Stern et al., 1988). This weighting function also deemphasizes peaks more distant from the $\tau = 0$ μs axis, so that cross-correlation functions produced by sinusoidal inputs do not predict multiple lateral positions of the tone, each located along the τ axis at $(1/\text{freq}) + \tau$. A spectral weighting function that emphasizes information from frequencies near 600 Hz has also been proposed

(Shackleton et al., 1992; Stern et al., 1988) to account for the dominance of that spectral region in determining laterality of intracranial images (Bilsen & Raatgever, 1973).

There are slight differences in the ways that the particular models compute lateralization indices, depending on the way interaural information is combined across channels. Shackleton et al. (1992), propose a simple "across-frequency integration", in which the weighted cross-correlation functions are combined into a summary cross-correlogram. The position of the largest peak value in the summary function is used to predict the position of the intracranial image. Shackleton et al. also note that the *average* of all cross-correlation peaks, weighted by their height, could also be used as a predictor. The weighted-image model of Stern et al. (1988) weights images by two aspects of the trajectory of peaks across frequency in the (f/τ) plane; straight trajectories (produced by stimuli in which all components have a common interaural time) and trajectories which are nearer the $\tau = 0$ μs axis are weighted more heavily. As Shackleton et al. point out, the straightness of a trajectory will directly affect the height of the peak in a summary cross-correlogram, so that qualitatively, the predictions from each are similar.

One inadequacy of current lateralization models is that they are generally designed to predict the location of a single intracranial image. Under certain circumstances, however, a listener may report hearing more than one intracranial image. For example, multiple images can arise when a pure tone is presented 180° out-of-phase. Subjects forced to indicate a single intracranial location for such stimuli choose locations to the far right and far left with equal frequency (Yost, 1981), as if they were selecting between two images. Sayers (1964) found similar results, except that subjects occasionally indicated that the image was located in the center of their heads, but that it was very diffuse. The cross-correlation function for this stimulus shows two peaks of equal height, at equal distances from the midline. Either of the perceptions mentioned above could be predicted by either the presence of two peaks in the summary cross-correlogram, or by the average position of all peaks present. If, on the other hand, multiple images arise as a result of conflicting interaural values in different spectral locations, a model that averaged interaural information across frequency to make a prediction about the location of a single image would, of course, fail. The stimuli used in the present series of experiments usually produces multiple auditory images (a noise image and a pitch-like image) in different intracranial locations, although there are few systematic investigations into the relative position of the images.

Since most models of lateralization use cross-correlation values in (f/τ) space, most analyses in this dissertation will rely on cross-correlation functions as an analytic tool. Further, since the stimuli used in the current set of experiments only contain interaural time (or phase) differences, mechanisms proposed for prediction of image position due to interaural level differences will not be addressed.

In order for ITD's to be used as a basis for grouping, the auditory system must be able to independently assess the interaural parameters at various spectral locations. Consider a situation in which a 3-tone complex containing frequencies of 200, 350, and 500 Hz is presented with an interaural delay to the left channel, and a second 3-tone complex containing frequencies of 270, 420, and 570 Hz is presented with an interaural delay to the right channel. It is important to keep in mind that all 6 components are present at both ears. If the stimulus is to be separated into two auditory groups on the basis of interaural time, then it is important that the interaural delay of the 350 Hz component be assessed independently of the 270 and 420 Hz components, which contain a different interaural delay. As a precursor to an investigation of the role of ITD's in sound source determination, it is then important to have an understanding of the ability of listeners to process these interaural differences in the presence of other, potentially disruptive, components.

A good deal of research has been aimed at determining a listener's ability to process ITD's at low frequencies when there are other frequencies present (e.g., Dye, 1990; Stellmack & Dye, 1993; Trahiotis & Bernstein, 1990). The general research strategy has been to find the minimum ITD necessary for a listener to discriminate the target stimulus from an identical stimulus presented diotically (no interaural differences). Various other components are added, and target interaural thresholds are measured again and interpreted in terms of interference on ITD detection caused by the non-target frequencies. Although the results of these studies are not interpreted directly in terms of sound source determination, they indicate how the resolution of the binaural auditory system at one frequency is affected by energy at other frequencies. In these experiments, the task is to discriminate two temporally distinct stimuli on the basis of interaural time differences. It must be kept in mind, however, that these experiments do *not* directly assess the extent to which interaural parameters are being processed *independently* of those at other frequencies. More specifically, the results of such experiments could be explained by the lateralization of an intracranial image associated with the target component(s), or by the intracranial position of an image resulting from *averaging* the interaural parameters of both the target and the distractors. In terms of the percepts arising from such stimuli, subjects often report basing their judgements on the lateralization of a single intracranial image, even though the stimulus components may have differing interaural parameters (Dye, 1990; Stellmack and Dye, 1993; Woods & Colburn, 1992)

If the spectra of two or more sound sources overlap, then the ability to group the different sets of components on the basis of interaural differences necessarily depends on how well the binaural system can independently process these differences across frequencies. Measures of cross-spectral interference in the processing of interaural parameters are important in that they indicate a limit on the analytic ability of the auditory system. Ideally, the system would be able to filter the acoustic signal into very narrow bands, and extract the information concerning various cues independently. This, of course, is not the case. Processing of any stimulus parameter value (interaural time, level, amplitude modulation, etc.) can be affected by the presence of off-target energy, or particular aspects of the off-target component. For example, threshold levels for detecting a tone based on interaural cues may be affected not only by the *presence* of another component, but by the interaural parameters of the interfering component (Bernstein, 1991).

The present dissertation is a combination of three behavioral experiments and a computer simulation of behavioral data. The intent is to gain insight into the mechanisms of binaural processing at low frequencies, under conditions in which there are more than one set of interaural phase differences in the spectrum of the stimuli. The first experiment measured a listener's ability to detect a narrowband interaural phase shift in different reproducible noise samples. The only difference between the different noise samples was the starting phases of the components (all interaural parameters were identical). Analysis of the samples in terms of the cross-correlation of the outputs of a bank of bandpass filters as a function of frequency and interaural time should reveal relationships that correlate with the subjects' responses; i.e., if the samples that led to good performance differ in some systematic way from the samples that led to poor performance, then that difference is likely to be an important aspect of binaural processing.

The computer simulation of data from an earlier experiment (Grange, 1991) is an attempt to account for subjects' performance using a simple mechanism based on the cross-correlation of a number of peripheral filter outputs. The decision statistic used is similar to the summary cross-correlograms used to predict lateral position of intracranial images (Shackleton et al., 1992). The simulation gave an estimate of the noise bandwidth that is effective in serving as a background against which dichotic pitches are heard.

The second experiment addresses the idea that the way in which components are grouped prior to the introduction of a target interaural phase shift can affect subjects' performance. It is proposed that detection of the signal (a narrowband interaural phase shift) proceeds on the basis of the changes in the summary cross-correlation function that the signal causes. This experiment explores the possibility that the summary cross-correlation function is more influenced by those channels that initially contain common interaural values. In all conditions, there are extraneous interaural changes, temporally coincident with the target (in both signal and non-signal intervals), that ideally should be ignored; it is proposed that this should be achieved more readily if the spectral regions containing the extraneous changes are initially given low weights.

The final experiment makes use of an MLD paradigm to investigate binaural processing. Prior studies using similar stimuli have suggested that a model based on processing of the outputs of a single pair of peripheral filters is insufficient to explain subjects' data (Sondhi and Guttman, 1966); put another way, it has been suggested that the binaural critical band is considerably wider than the monaural critical band. However, Kohlrausch (1988) points out that these conclusions can be traced to (1) the assumption of rectangular filters in calculations of filter width, and (2) the failure to emphasize the difference in the auditory cues available in monaural and binaural tasks. He concludes that a single peripheral filter pair can indeed account for much of the data. Even so, an explanation based on a single bandwidth does not account for perceptual aspects of the task. Discussion of the results from this experiment will be based on the assumption that multiple spectral channels are being processed and compared, and a general (non-quantitative) model will be suggested, along with future experiments to further test these ideas.

CHAPTER 2

EXPERIMENT I

Introduction

Dichotic pitch is a name given to a category of phenomena in which the perception of pitch is generated purely as a result of the central comparison of the signals at the two ears; that is, neither monaural signal alone provides a basis for the perception of the pitch. If diotic noise is interaurally altered over a narrow band of frequencies, (e.g., an interaural phase sweep through 360°, or a constant interaural phase/time shift), then a pitch is generally perceived at the center frequency of the narrow band (see Raatgever and Bilsen, 1986).

The existence of dichotic pitches provides evidence that the auditory system can perceptually segregate different regions of the spectrum based entirely on interaural phase/time cues. It is generally thought that processing of these interaural cues proceeds on the basis of a cross-correlation network in which frequency-matched left and right peripheral filters are cross-correlated, creating activity within f/τ (frequency/interaural delay) space. The existence of dichotic pitches depends on the comparison (and separation) of the neural representation of the cross-correlation function of the filter pair centered on the segregated spectral region with the cross-correlation function of other spectral regions.

In an attempt to discern the spectral range over which such comparisons are made, Grange (1991) measured thresholds for detection of an interaural phase shift of a 20-Hz portion of lowpass (cutoff = 1500 Hz, target CF = 500 Hz) noise. A wider band surrounding the target band was presented diotically, and the interaural phase of each component outside of the diotic band was selected randomly (producing interaurally uncorrelated noise). Diotic bandwidths of 900, 700, 500, 300, 200, 150, and 125 Hz were tested. Conditions were also run in which components outside of the diotic band were removed.

Examples of the stimulus interaural configuration and the data from that experiment are shown in Figure 1. Thresholds for detection of the target interaural phase shift began to increase when the diotic bandwidth decreased from 500 Hz to 300 Hz, if interaural phases outside of this band were randomized. Thresholds remained relatively constant if components were simply removed.

Grange (1992) suggested that detection of the target interaural phase shift proceeded on the basis of the decorrelation that the target caused between the left and right peripheral filters centered on the target frequency (and not a shift in the peak of the cross-correlation function), compared with the filter channels immediately surrounding the target. This conclusion was based on the observation that the effect of even a large interaural phase shift of a 20-Hz band is a decrease in the height of the cross-correlation peak, and not a shift along the interaural time axis (since most of the noise passing through the peripheral filter is still diotic).

Figure 2 shows cross-correlation values for a sample of each of these conditions as a function of filter CF at an interaural delay of 0 μs, for diotic bandwidths of 700, 500,

Figure 1

The upper portion of the figure shows examples of the noise interaural configuration of a signal and non-signal interval. The lower portion shows thresholds for detecting the target interaural phase shift as a function of the width of the diotic band surrounding the target.

Figure 1

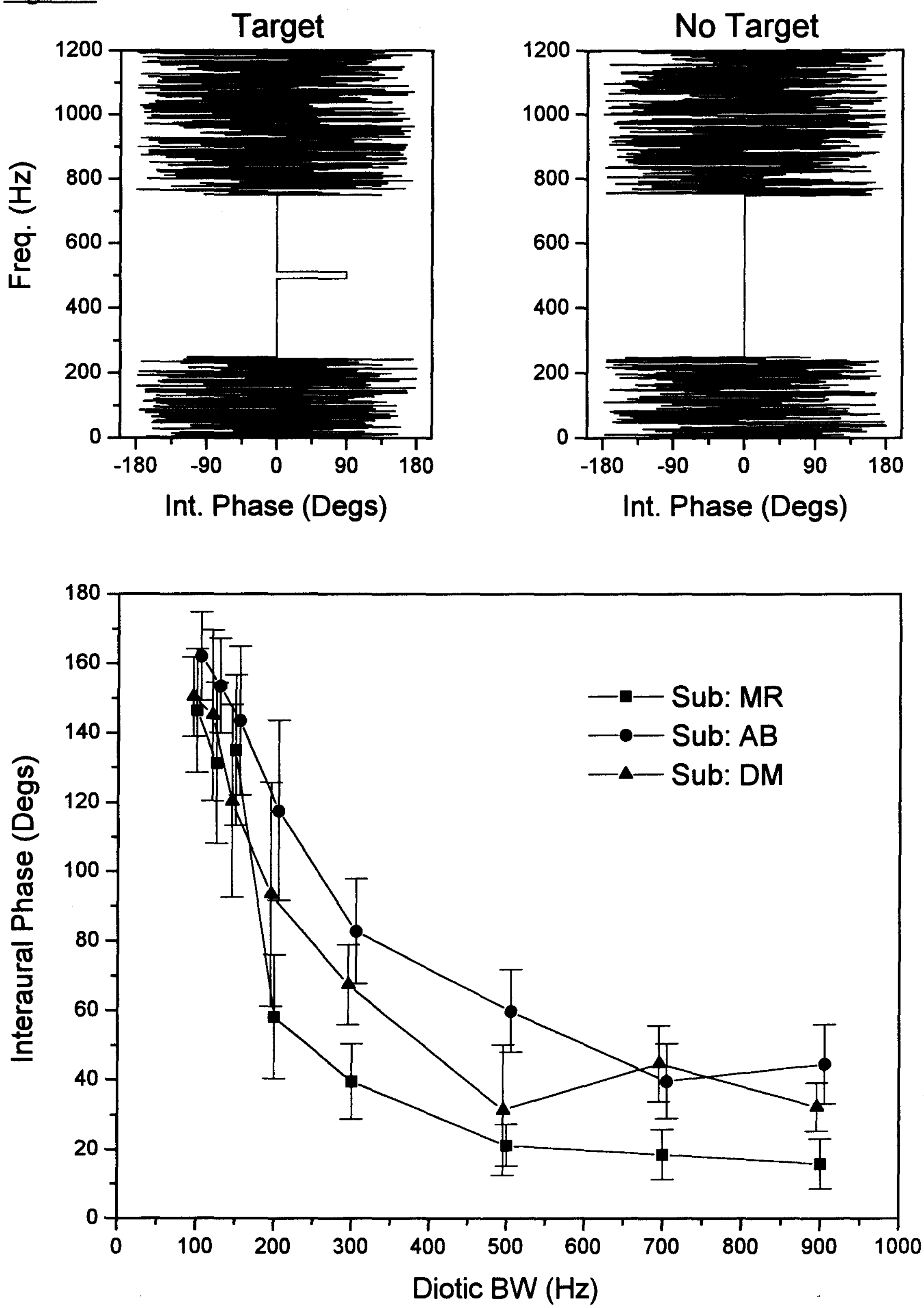
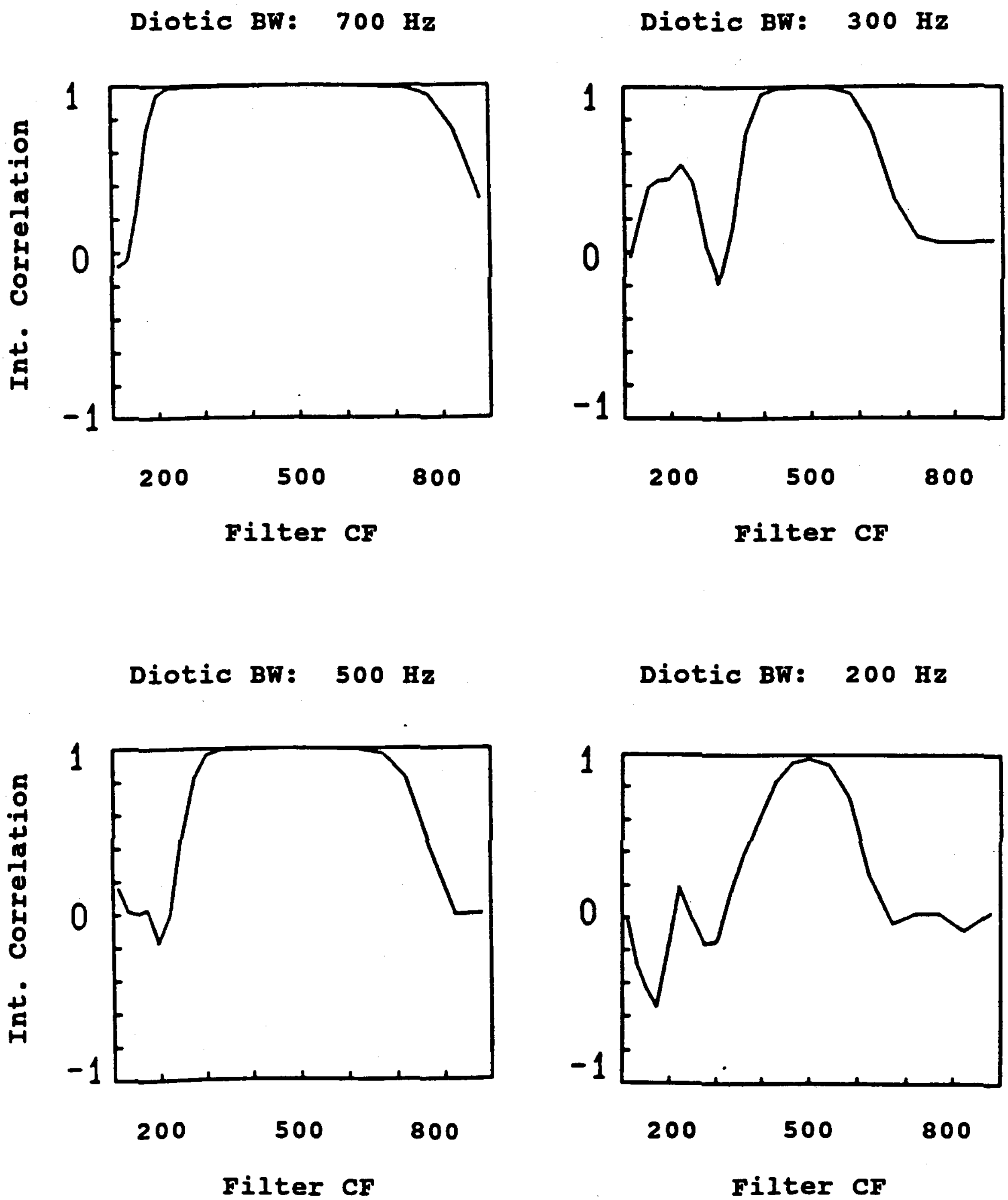

Target
No Target
Freq. (Hz)
1200
1000
800
600
400
200
0
-180
-90
0
90
180
Int. Phase (Degs)
Interaural Phase (Degs)
180
160
140
120
100
80
60
40
20
0
Sub: MR
Sub: AB
Sub: DM
0
100
200
300
400
500
600
700
800
900
1000
Diotic BW (Hz)

Figure 2

Cross-correlation coefficients (the value of the cross-correlation function at $\tau = 0$ μs) of frequency-matched left and right bandpass filters are plotted as a function of the CF of the filter pair. Each panel represents a different diotic bandwidth of the stimulus (interaural phases outside of the diotic band were randomized).

Figure 2


Diotic BW: 700 Hz
Diotic BW: 300 Hz
Diotic BW: 500 Hz
Diotic BW: 200 Hz
Int. Correlation
1
0
-1
200
500
800
Filter CF

300, and 200 Hz. Filter channels within about 50 Hz of the target begin to be decorrelated as a result of the uncorrelated noise as the diotic bandwidth decreases from 500 to 300 Hz (at which point thresholds begin to increase). Presumably, as thesurrounding "reference" filter correlations decrease, a greater target decorrelation is necessary for detection, analogous to what was demonstrated by Gabriel and Colburn (1981) for broadband decorrelation.

The outputs of bandpass filters with diotic noise as input signals are perfectly correlated; this cross-correlation value does not change over time. However, if interaural differences of phase are introduced, then (1) the cross-correlation value will decrease, and (2) the cross-correlation value will vary over time. The variation in cross-correlation values results from differences in the time domain output of the filters, which can be approximated by sinusoids with a frequency corresponding to the CF of the filter, with random fluctuations in instantaneous phase and amplitude (Jeffress, 1972). For diotic noise, the instantaneous phase and level of each filter output will be identical, and therefore perfectly correlated. If the phases of any noise components are changed in one signal relative to the other, however, the variations in instantaneous phase and level at each filter output will not be identical, resulting in slowly changing *interaural* phase and level (and therefore correlation). If detection of a narrowband interaural phase shift is based on a comparison of the cross-correlation of the filter pair centered on the target with surrounding channels, then detectability should vary over time (as the cross-correlation value does). Put another way, particular samples of the stimulus may lead to better detection of the target, depending on the cross-correlation of that sample.

Grange (1991) used an adaptive procedure to estimate thresholds, in which component starting phases were selected at random from trial to trial. Several conditions were run in a 2-AFC psychophysical procedure, with a constant interaural phase shift based on threshold estimates obtained from an adaptive procedure, in order to confirm that phase shift led to approximately 70% correct. Subjects reported that on some trials, the pitch was quite strong, while on others, they did not perceive a pitch at all (the target was always present in one of the two intervals). The present experiment is an investigation of these sample-specific effects.

Methods

The initial step in this experiment was to find thresholds for detecting an interaural phase shift of a 20-Hz section of equal amplitude noise low-pass filtered at 1200 Hz. The target band was centered on 500 Hz, and was surrounded symmetrically by a diotic spectral band, which was either 100, 125, 150, or 900 Hz wide. Outside of this diotic band, the interaural phase for each frequency component was selected randomly from a rectangular distribution. Thresholds were obtained using an adaptive procedure with a two-down, one-up rule estimating the target phase shift leading to 70.7% correct (Levitt, 1971). Data were collected in blocks of 50 trials. A 2-AFC task was used in which each interval consisted of 250 ms of forward fringe and 250 ms of observation segment. Each interval was gated on and off using a 10-ms $\cos^2$ ramp. The forward fringe consisted of noise with the same interaural parameters as a non-signal interval (diotic within a band around 500 Hz, random interaural phases outside of this band).

After each subject's threshold was found for all 4 diotic bandwidths, 20 different

250-ms samples of each condition were generated for each subject, with a target interaural phase shift at the nominal threshold value. Twenty different non-signal samples for each condition were generated as well. To produce the 20 different samples, a 16384 point inverse FFT routine was used to generate 1 s of the stimulus for the left channel. Random starting phases were chosen from a rectangular distribution, and the power spectrum was flat. The appropriate phase shifts were made for the right channel stimulus, and another inverse FFT was done. Twenty starting points were chosen at random, and the 250-ms samples were stored in the computer.

Data were collected in blocks of 60 trials. On a given trial, the subject was instructed to indicate which of two intervals contained the narrowband interaural phase shift. In each 500-ms interval, the first 250 ms served as a forward fringe. The same non-signal sample (selected arbitrarily) served as the forward fringe for all trials in a given condition. In each block of 60 trials, each of the 20 samples for a particular condition were presented 3 times, in random order. Each signal-present sample was always paired with the same non-signal sample (e.g., when the signal sample labelled "3" was presented, the non-signal interval would be the non-signal sample labelled "3"). Subjects were allowed to practice with feedback before each block, but during the actual test block, the feedback was turned off. Thirty-two blocks of each condition were run, so that P(C) scores for each sample were based on 96 presentations.

It was expected that the average P(C) across all 20 samples of a condition would be approximately 70%. In cases where a subject's average P(C) for a condition deviated greatly from this, a different set of samples at an adjusted target interaural phase shift was generated. When possible, the same set of samples was used for more than one subject, so that comparisons could be made between listeners. That is, if subjects nominal thresholds found with the tracking procedure were similar, they would receive a common set of samples.

Stimuli were generated on a Masscomp minicomputer interfaced with a 16-bit d/a converter set to a rate of 16384 points per second. After being attenuated (Leader LAT-46), signals were sent to Crown amplifiers which drove the earphones (Sony MDR-V6). Subjects, three Loyola University students who were paid an hourly wage, performed the task in 1.5-hour sessions, usually listening to 400-600 trials a day.

Results

Figures 3 - 6 show the percent correct (PC) for detecting the presence of the target interaural phase shift for each of the 20 different samples for each diotic bandwidth, for each subject. When the same set of stimulus samples for a condition was used for more than one subject, data for those subjects are plotted on the same graph. For all subjects, in all conditions, performance ranged from approximately 90% correct to near chance (50% correct) even though the interaural configuration was identical for all samples within a condition. In conditions where two or more subjects heard the same set of samples, there was a strong correspondence between subjects on each particular sample. Figures 7 - 8 show an average P(C) of two subjects from Figs. 2.3 for the sample set in each condition on which multiple subjects were tested (analyses of the stimuli will proceed in these cases)[1].

Figures 3 - 6

P(C) is rank-ordered from high-performance to low-performance for each of the 20 samples used in each condition. When two or more subjects used the same set of samples, their data are plotted on the same graph. The specific target interaural phase shift for the subject(s) is shown at the top of each graph. Diotic bands of 900, 150, 125, and 100 Hz are shown, respectively.

Figure 3

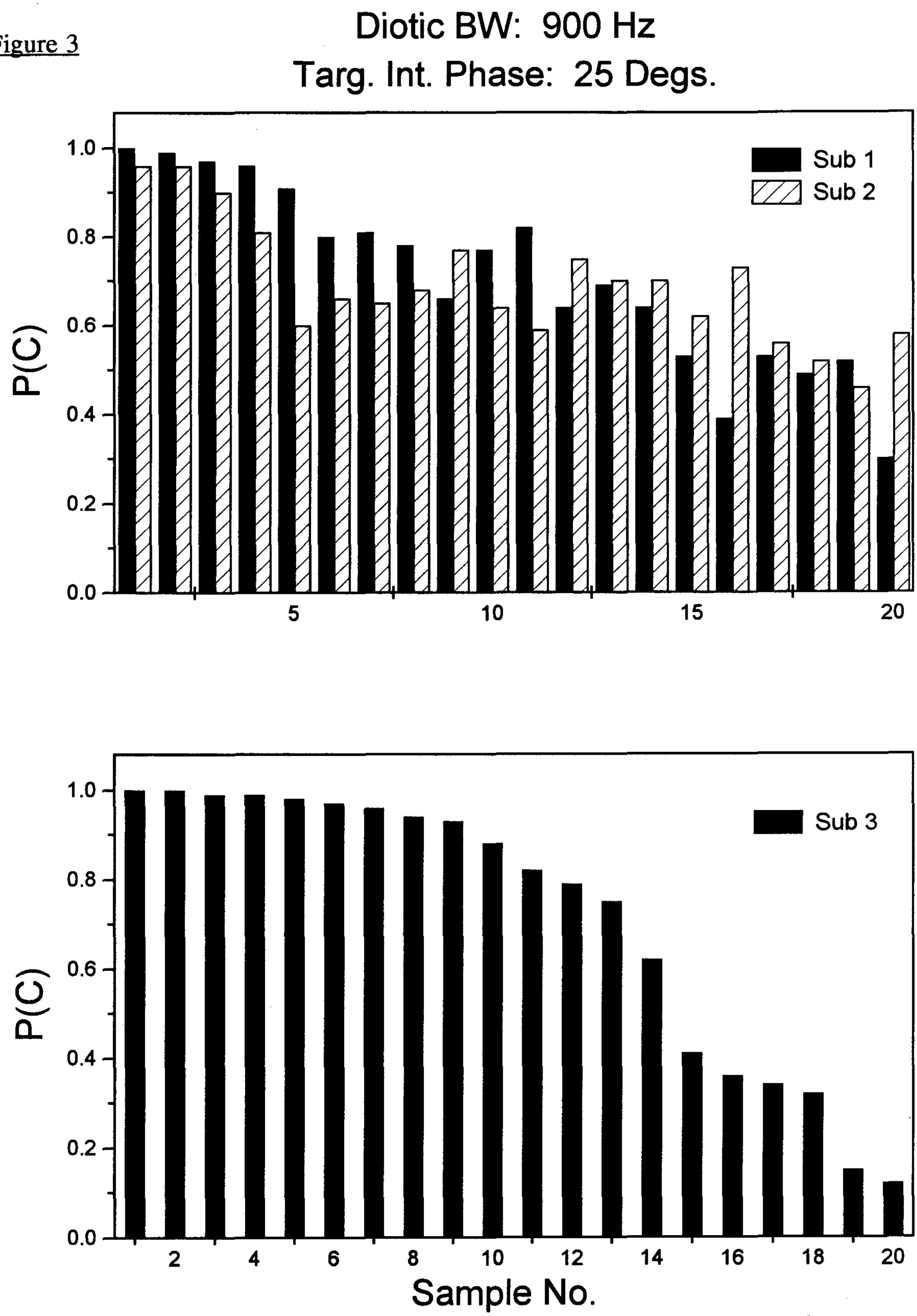

Diotic BW: 900 Hz
Targ. Int. Phase: 25 Degs.
Sub 1
Sub 2
Sub 3
P(C)
1.0
0.8
0.6
0.4
0.2
0.0
5
10
15
20
2
4
6
8
12
14
16
18
Sample No.

Figure 4

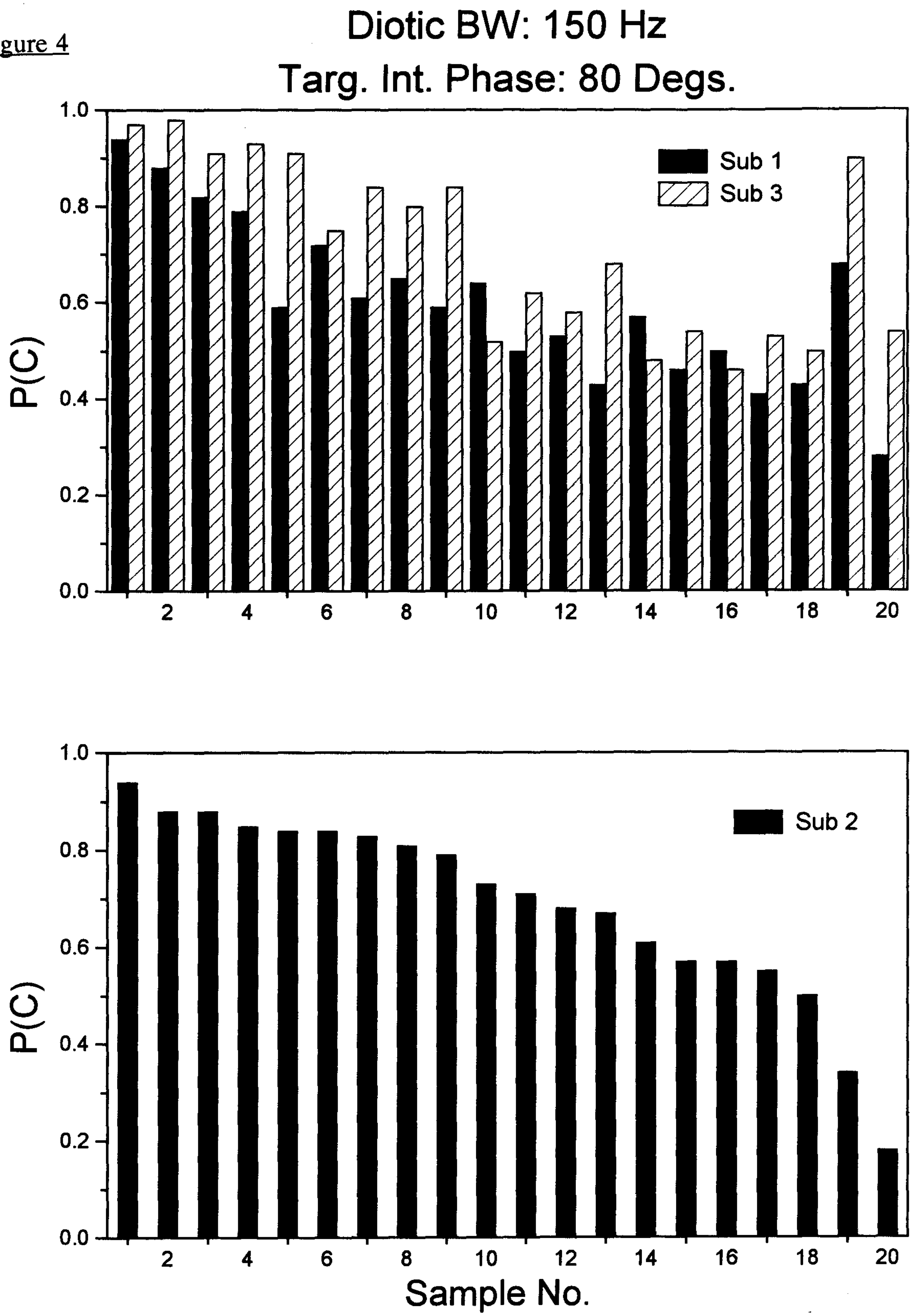

Diotic BW: 150 Hz
Targ. Int. Phase: 80 Degs.
Sub 1
Sub 3
Sub 2
P(C)
Sample No.
1.0
0.8
0.6
0.4
0.2
0.0
2
4
6
8
10
12
14
16
18
20

Figure 5

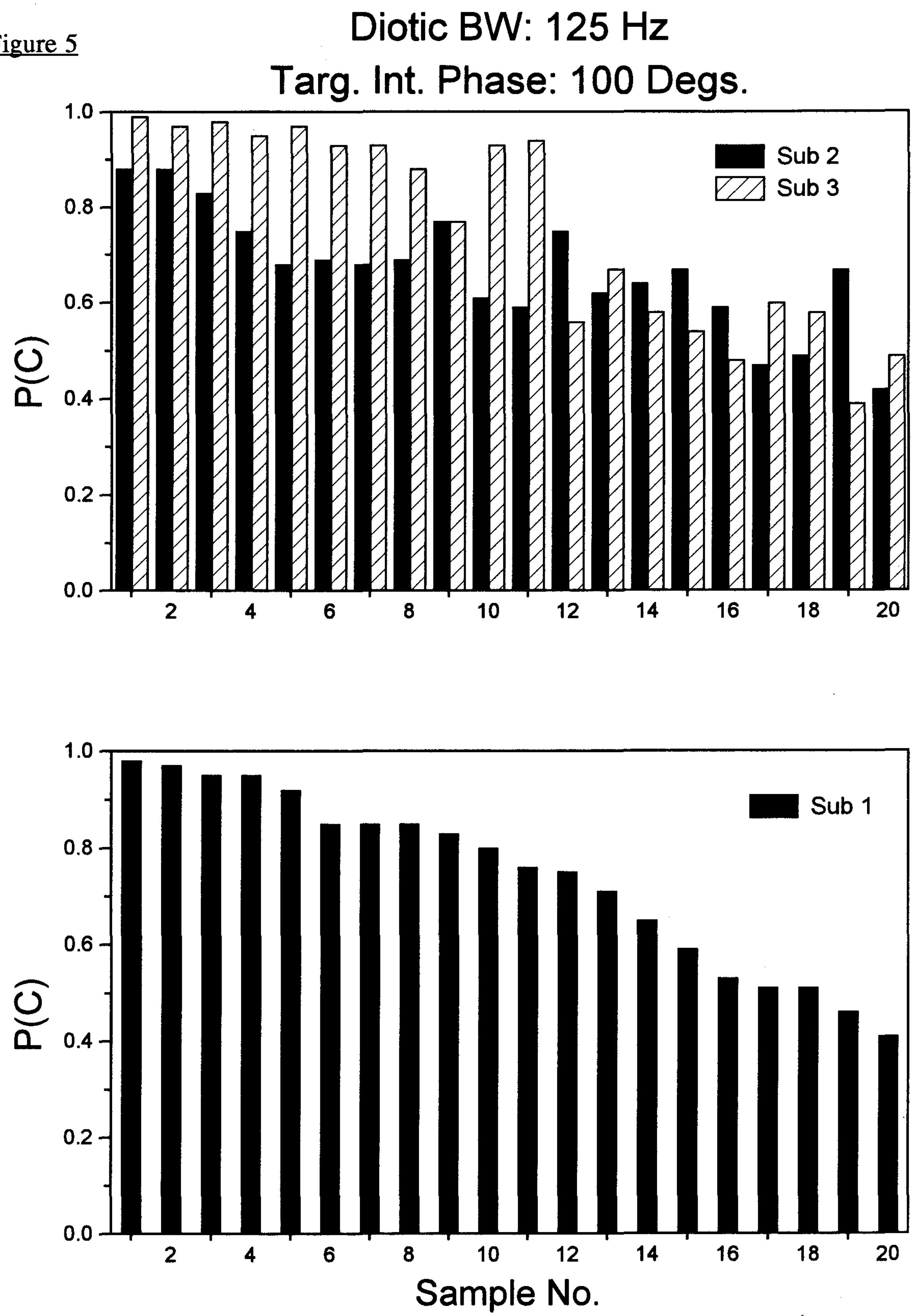
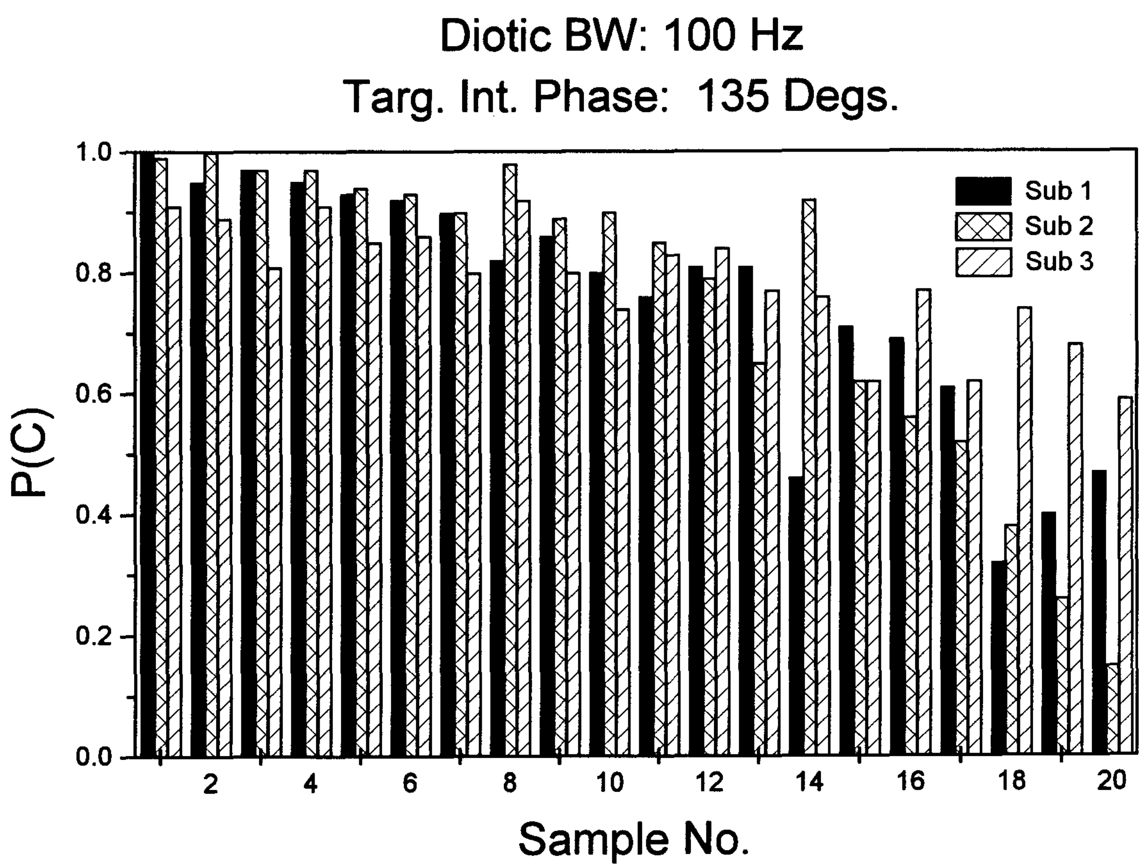

Diotic BW: 125 Hz
Targ. Int. Phase: 100 Degs.
1.0
0.8
0.6
0.4
0.2
0.0
P(C)
Sub 2
Sub 3
2
4
6
8
10
12
14
16
18
20
Sub 1
Sample No.

Figure 6


Diotic BW: 100 Hz
Targ. Int. Phase: 135 Degs.
P(C)
1.0
0.8
0.6
0.4
0.2
0.0
Sub 1
Sub 2
Sub 3
2
4
6
8
10
12
14
16
18
20
Sample No.

Figures 7 - 8

Same format as Figs. 3 - 6, but data for the sample sets that at least 2 subjects received are averaged across those two subjects and plotted. Sample 19 in the 150 Hz. condition was categorized incorrectly and is not included in further analyses.

Figure 7

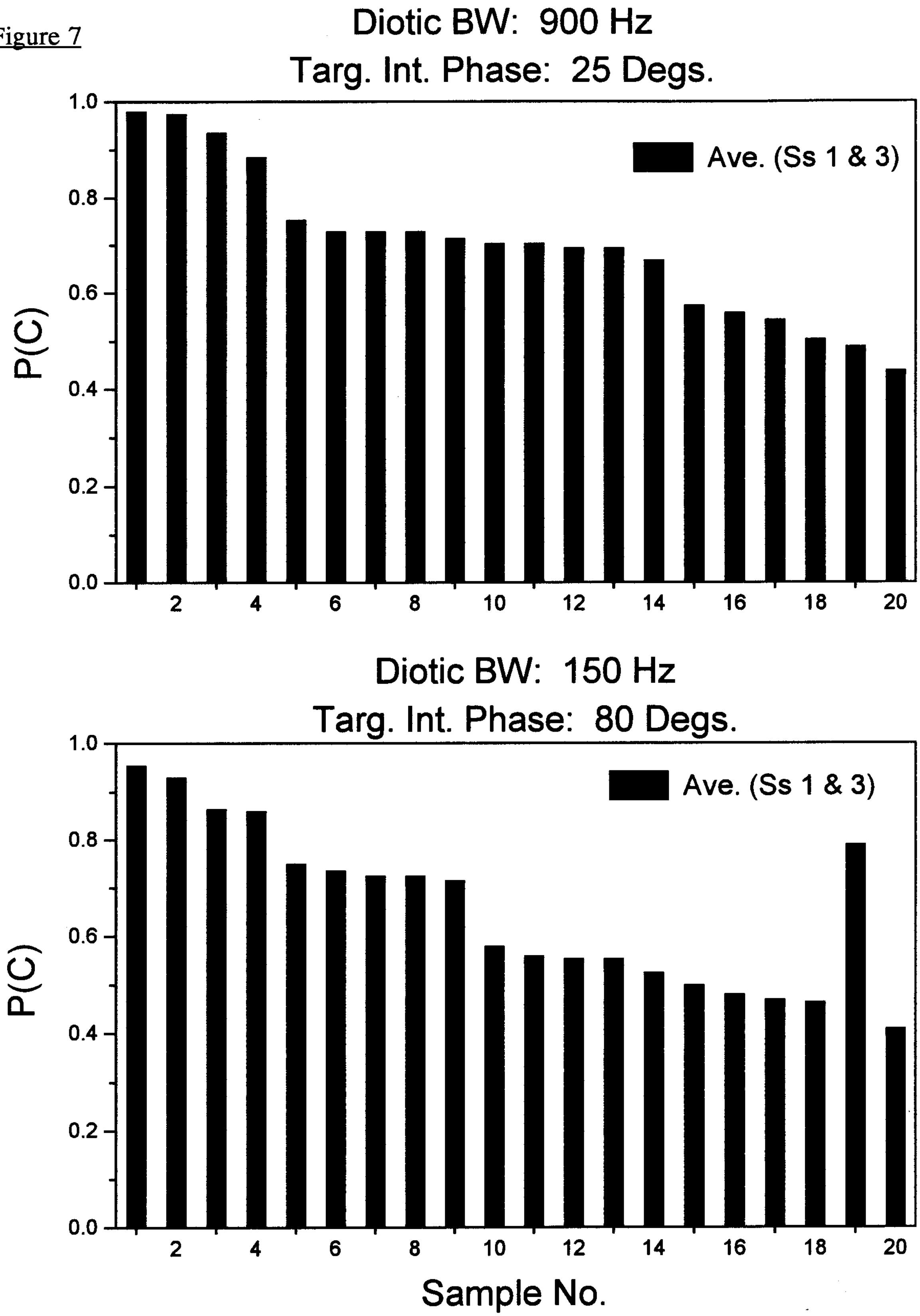
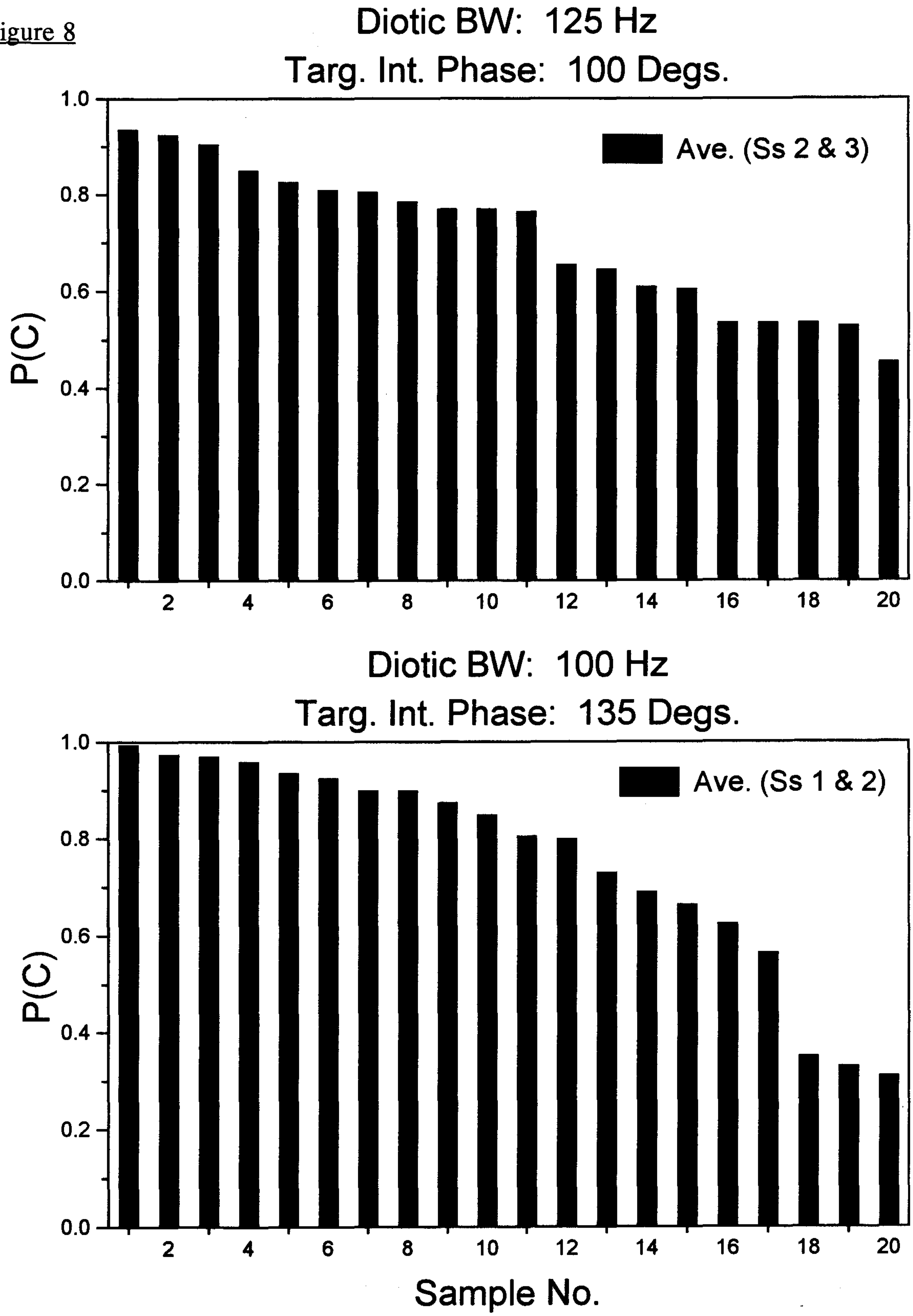

Diotic BW: 900 Hz
Targ. Int. Phase: 25 Degs.
Ave. (Ss 1 & 3)
P(C)
0.0
0.2
0.4
0.6
0.8
1.0
2
4
6
8
10
12
14
16
18
20
Diotic BW: 150 Hz
Targ. Int. Phase: 80 Degs.
Ave. (Ss 1 & 3)
P(C)
Sample No.

Figure 8



Diotic BW: 125 Hz
Targ. Int. Phase: 100 Degs.
Ave. (Ss 2 & 3)
P(C)
1.0
0.8
0.6
0.4
0.2
0.0
2
4
6
8
10
12
14
16
18
20
Diotic BW: 100 Hz
Targ. Int. Phase: 135 Degs.
Ave. (Ss 1 & 2)
P(C)
Sample No.

Discussion

If performance does proceed on the basis of comparing cross-correlation coefficientsof filters centered on the target frequency with the coefficients of filters surrounding it, then it would be predicted that the particular samples that led to good performance were relatively more decorrelated at the target than those leading to poor performance.

For each condition, the three samples that led to the best performance and three samples that produced chance performance were selected for further analysis. The analyses were done on samples selected from the set that at least two listeners were tested on, in each condition. The left and right signals from each sample were passed through a bank of band-pass filters[2] (simulating the auditory periphery), and the outputs of frequency-matched filters were cross-correlated. Figures 9 - 16 shows cross-correlation coefficients as a function of filter CF for the four different conditions. The left panels for each condition are cross-correlation spectra for the "high performance" samples, and the right panels represent "poor performance" samples. It can be seen that the left panels show a relative "dip" in the cross-correlation values at the target frequency, which is absent in the right panels. Figure 17 shows cross-correlation spectra for samples of non-target intervals. In these graphs the functions have a single peak which is maximal in the center of the diotic band (similar to "poor performance" signal samples).

---

[1] Sample no. 19 the 150-Hz condition was rank-ordered incorrectly and is not included in subsequent analyses.

[2] Filter bands used to simulate the auditory periphery throughout this dissertation are gamma-tone filters, based on the model of Patterson and Nimmi-Smith (1987).

Figures 9 - 16

Interaural cross-correlation coefficients are shown for the output of frequency-matched left and right bandpass filters, as a function of the CF of the filter pair. The three panels in each figure are generated from stimuli containing the target interaural phase shift. Figures 9, 11, 13, and 15 represent stimuli from the 900, 150, 125, and 100 Hz conditions, respectively, that led to high performance ($P(C) > 90\%$), while figures 10, 12, 14, and 16 represent stimuli that led to poor performance ($P(C) < 60\%$). The plot for sample 150_080_19 is not include because it was categorized incorrectly as a poor performance sample.

Figure 9

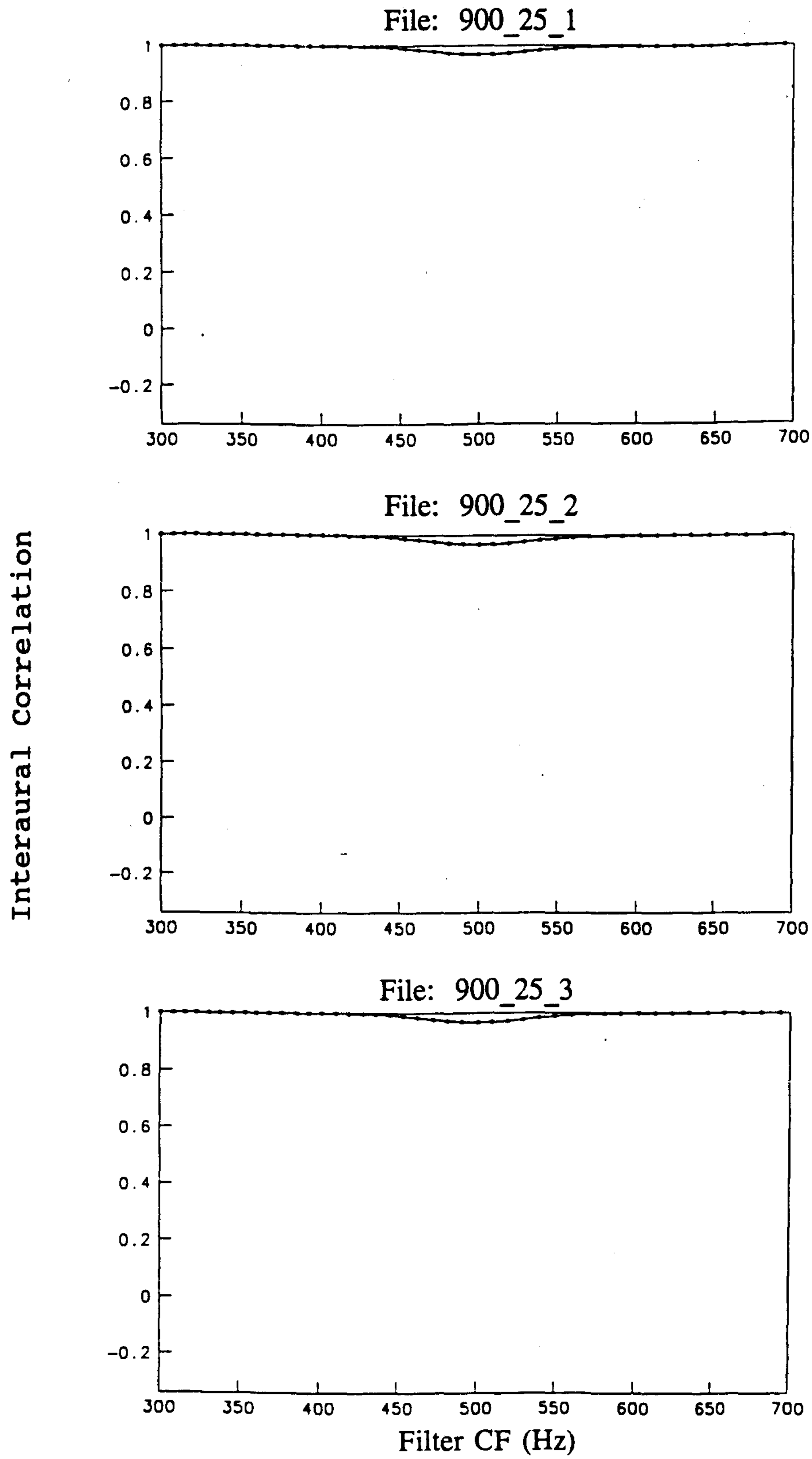

File: 900_25_1
File: 900_25_2
File: 900_25_3
Interaural Correlation
1
0.8
0.6
0.4
0.2
0
-0.2
300
350
400
450
500
550
600
650
700
Filter CF (Hz)

Figure 10

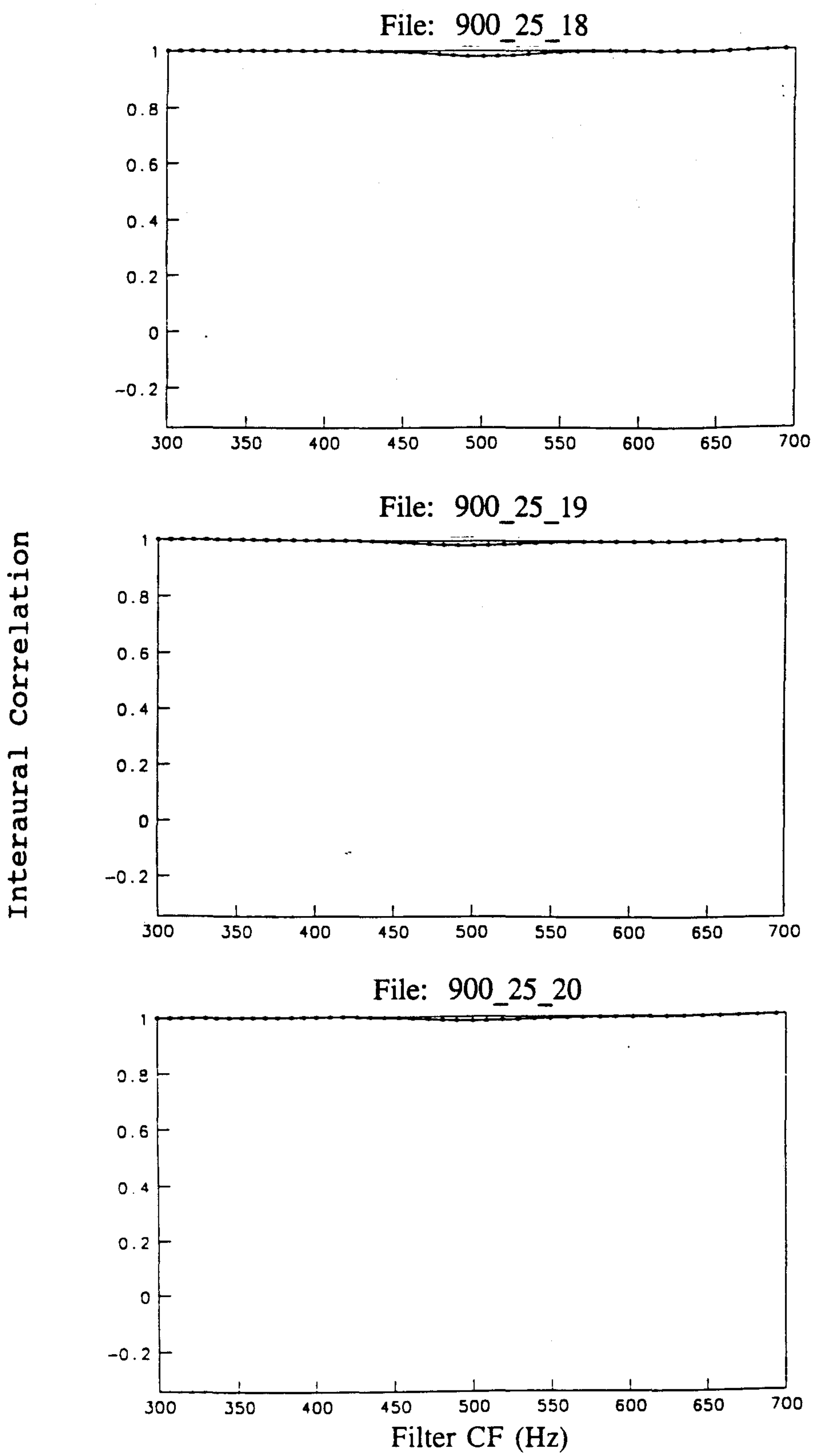

File: 900_25_18
1
0.8
0.6
0.4
0.2
0
-0.2
300
350
400
450
500
550
600
650
700
File: 900_25_19
1
0.8
0.6
0.4
0.2
0
-0.2
300
350
400
450
500
550
600
650
700
File: 900_25_20
1
0.8
0.6
0.4
0.2
0
-0.2
300
350
400
450
500
550
600
650
700
Interaural Correlation
Filter CF (Hz)

Figure 11

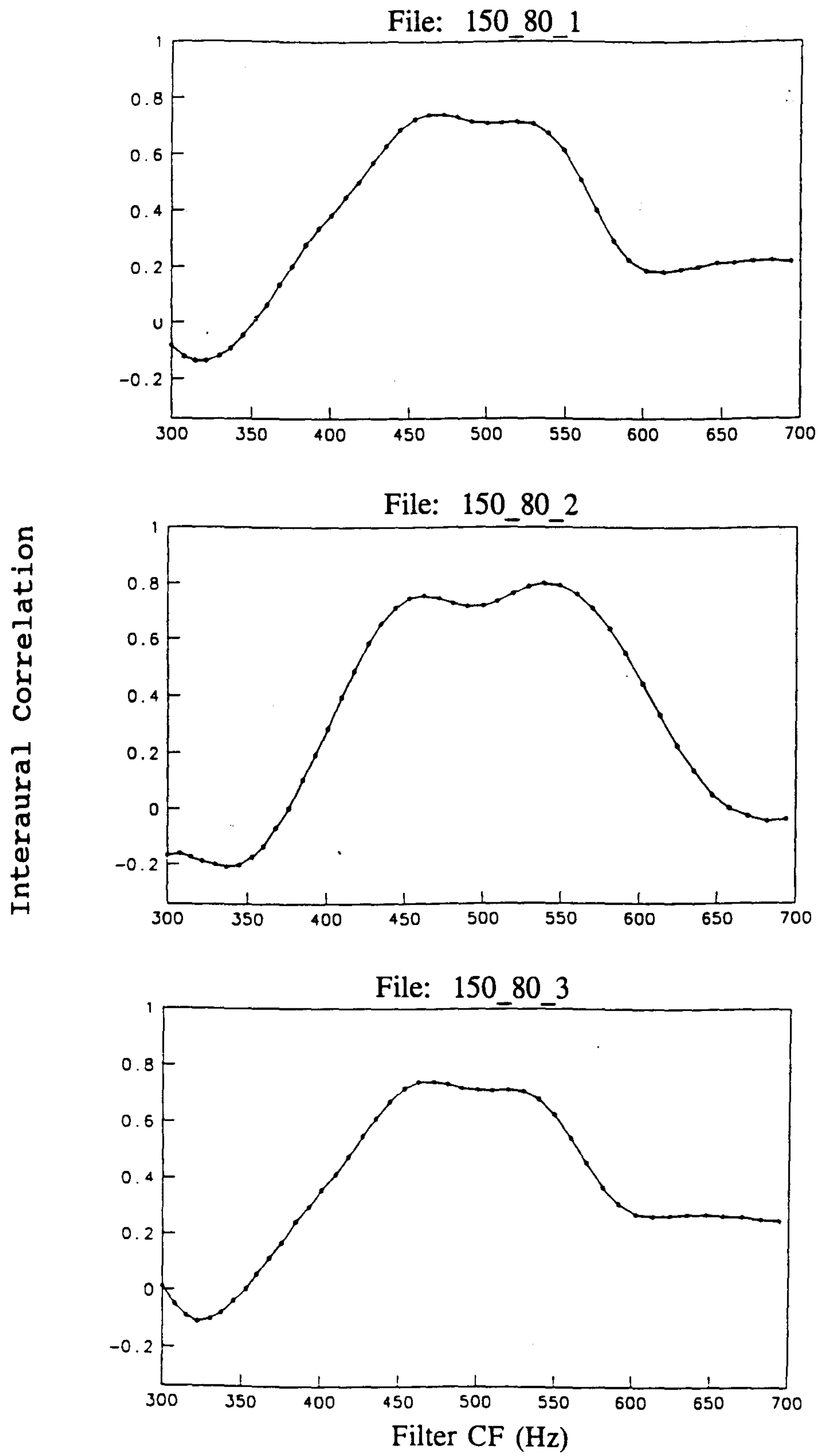

File: 150_80_1
1
0.8
0.6
0.4
0.2
0
-0.2
300
350
400
450
500
550
600
650
700
File: 150_80_2
1
0.8
0.6
0.4
0.2
0
-0.2
300
350
400
450
500
550
600
650
700
File: 150_80_3
1
0.8
0.6
0.4
0.2
0
-0.2
300
350
400
450
500
550
600
650
700
Interaural Correlation
Filter CF (Hz)

Figure 12

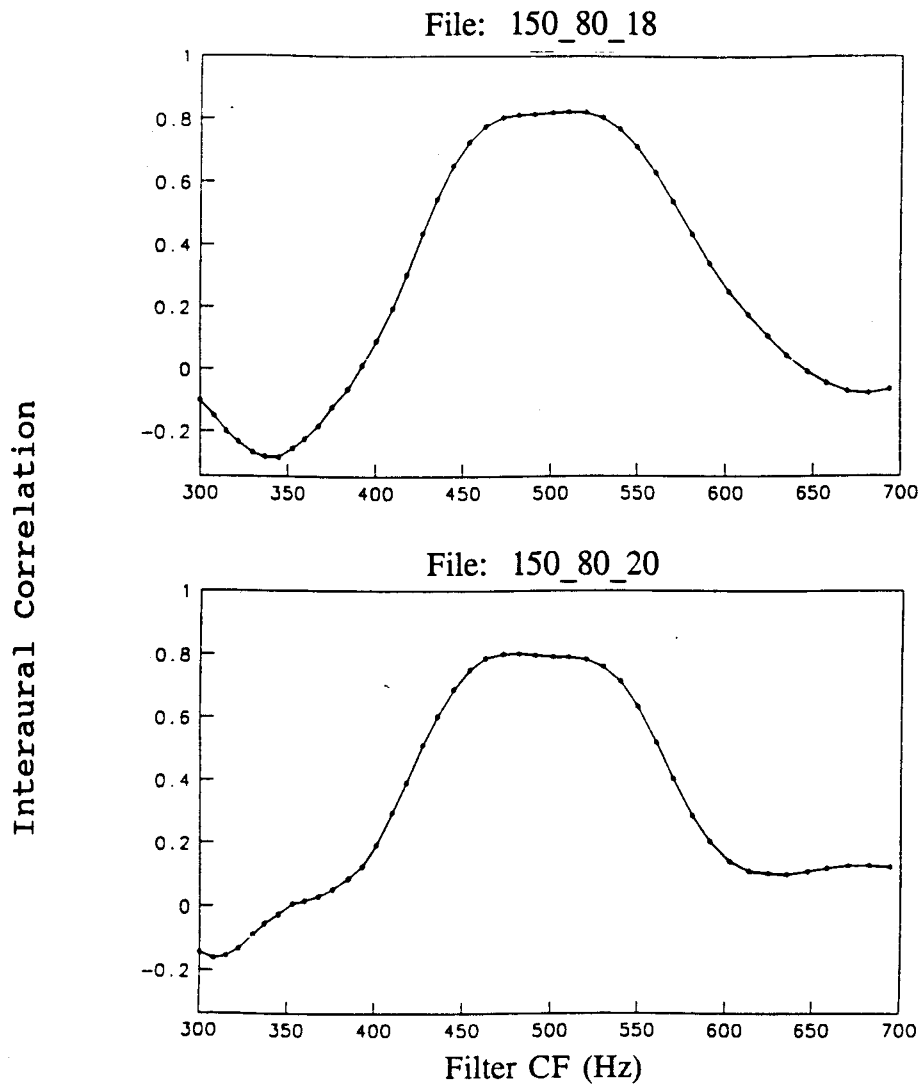

File: 150_80_18
1
0.8
0.6
0.4
0.2
0
-0.2
300
350
400
450
500
550
600
650
700
File: 150_80_20
1
0.8
0.6
0.4
0.2
0
-0.2
300
350
400
450
500
550
600
650
700
Interaural Correlation
Filter CF (Hz)

Figure 13

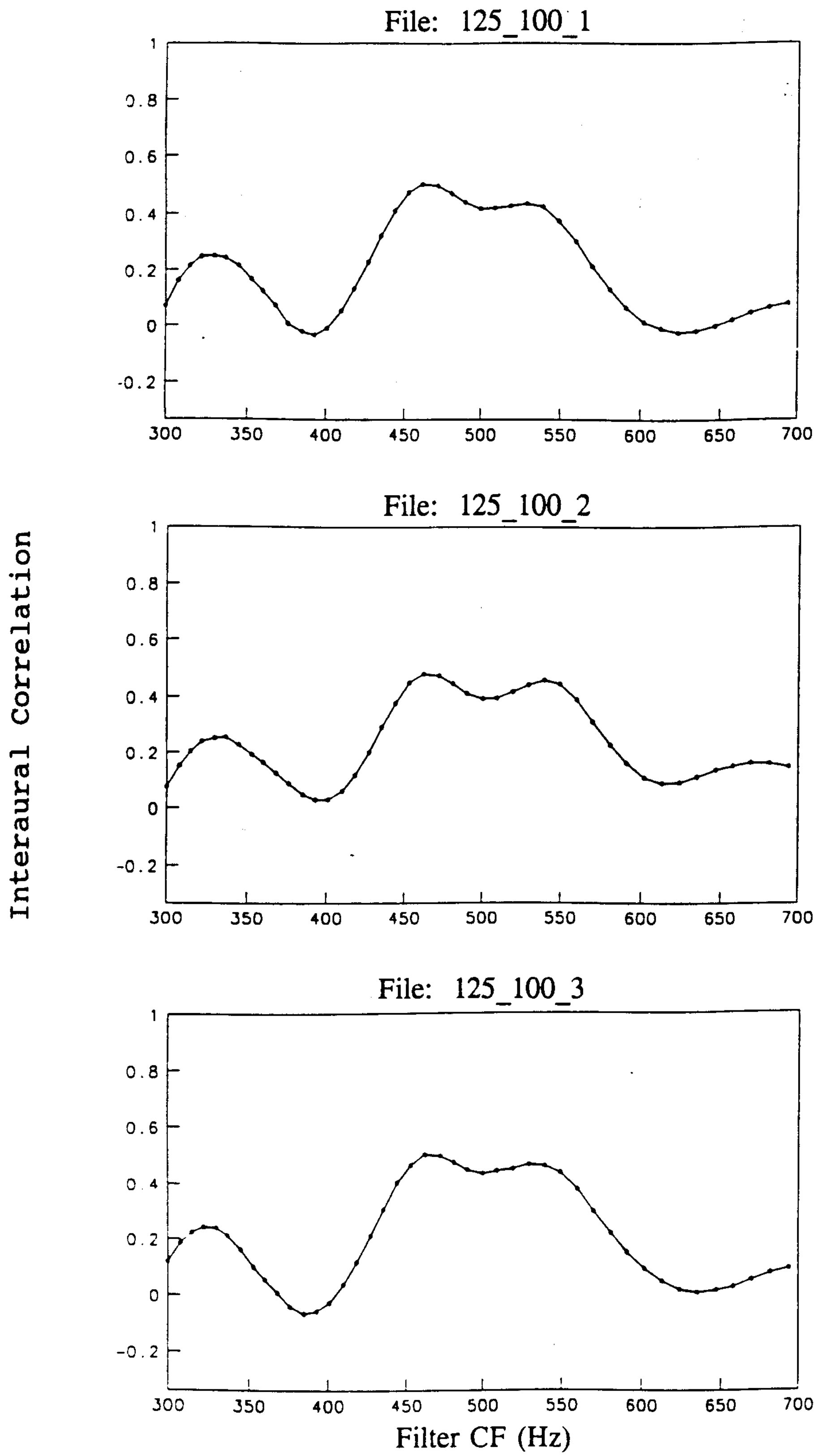

File: 125_100_1
File: 125_100_2
File: 125_100_3
Interaural Correlation
Filter CF (Hz)
1
0.8
0.6
0.4
0.2
0
-0.2
300
350
400
450
500
550
600
650
700

Figure 14

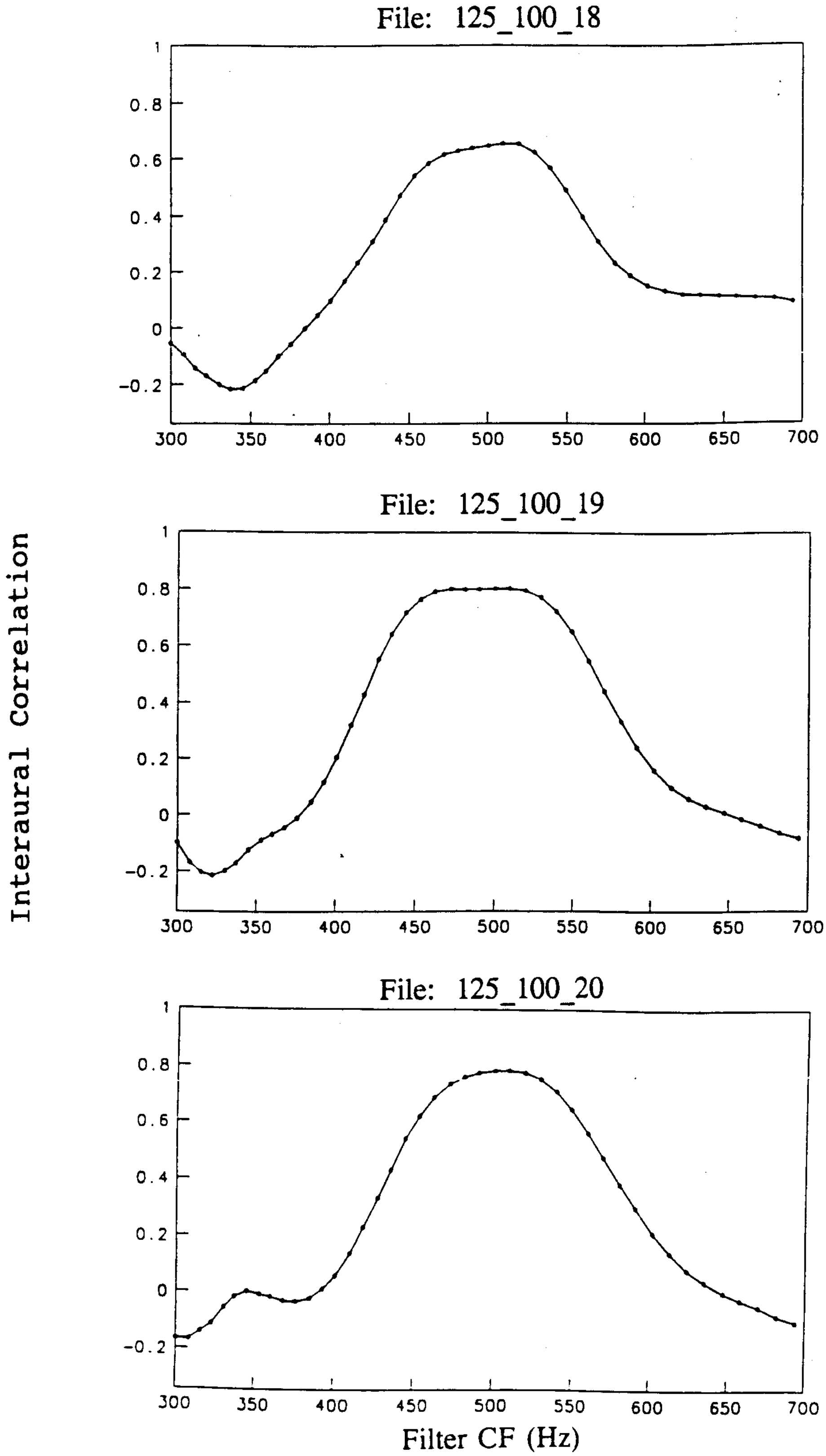

File: 125_100_18
File: 125_100_19
File: 125_100_20
Interaural Correlation
Filter CF (Hz)
1
0.8
0.6
0.4
0.2
0
-0.2
300
350
400
450
500
550
600
650
700

Figure 15

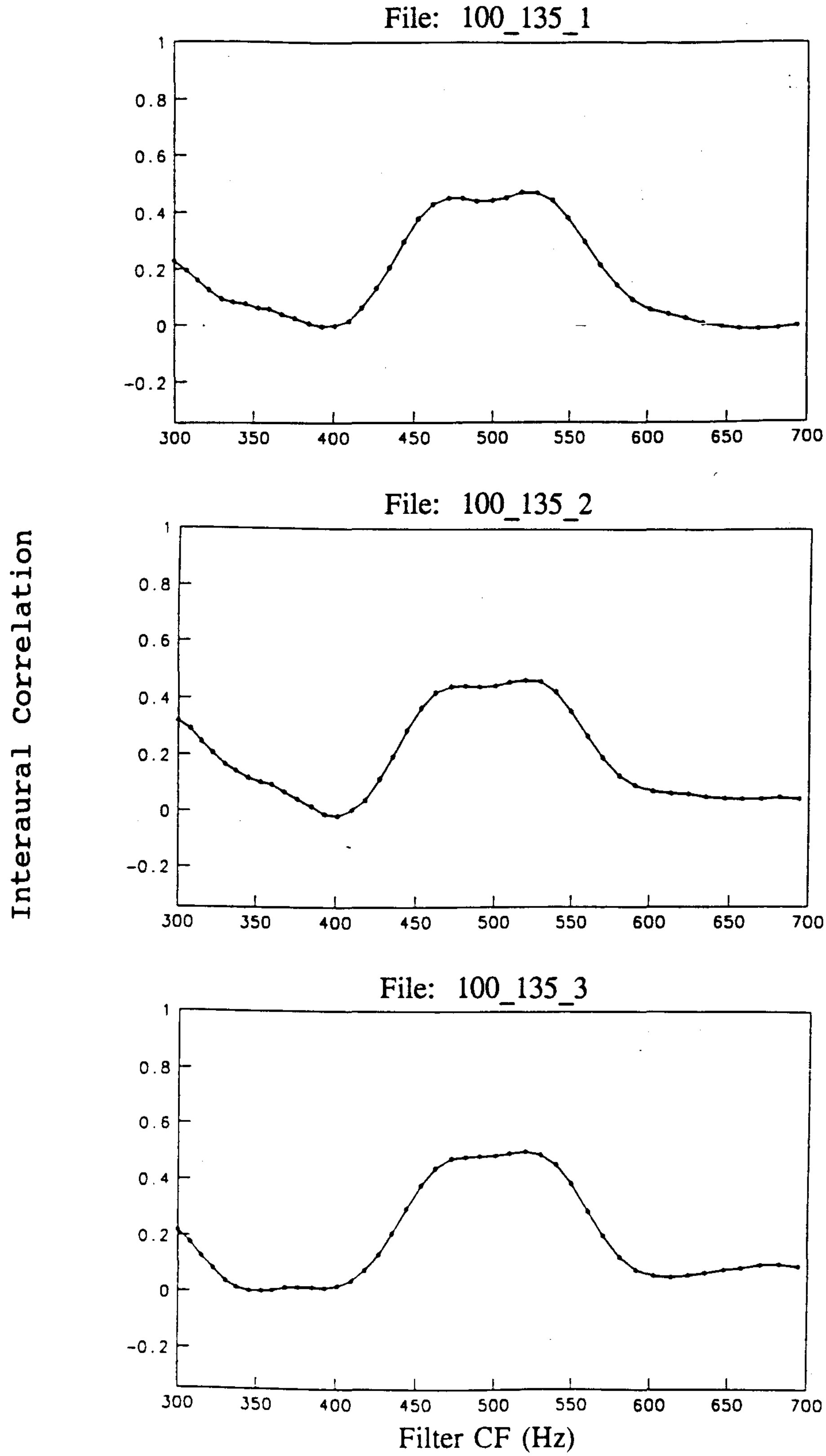

File: 100_135_1
1
0.8
0.6
0.4
0.2
0
-0.2
300
350
400
450
500
550
600
650
700
File: 100_135_2
1
0.8
0.6
0.4
0.2
0
-0.2
300
350
400
450
500
550
600
650
700
File: 100_135_3
1
0.8
0.6
0.4
0.2
0
-0.2
300
350
400
450
500
550
600
650
700
Interaural Correlation
Filter CF (Hz)

Figure 16

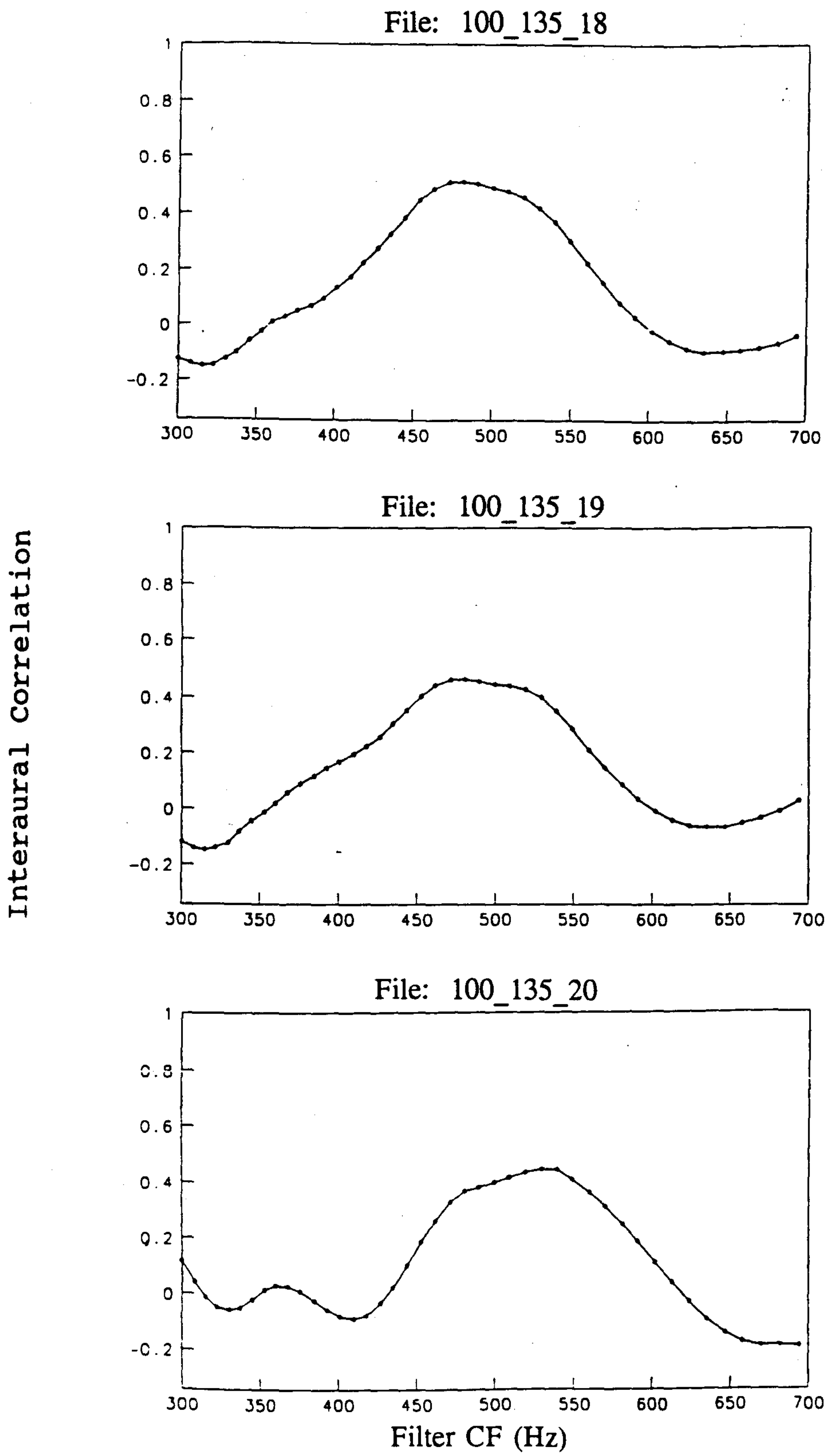

File: 100_135_18
File: 100_135_19
File: 100_135_20
Interaural Correlation
1
0.8
0.6
0.4
0.2
0
-0.2
300
350
400
450
500
550
600
650
700
Filter CF (Hz)

Figure 17

Same format as Figs 9 - 16, except that the stimuli used for generation did not contain the target interaural phase shift (non-signal intervals).

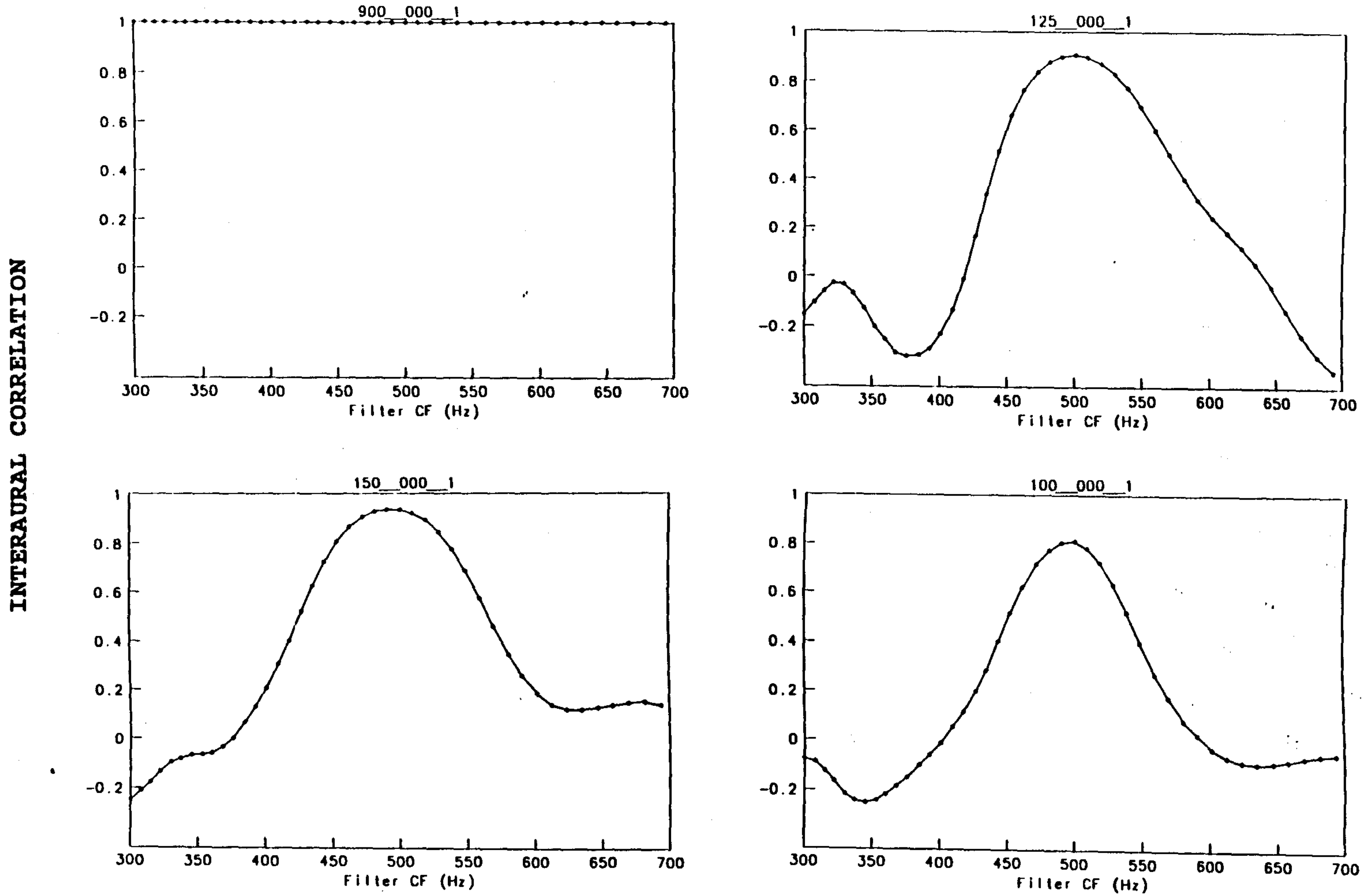

INTERAURAL CORRELATION
900__000__1
125__000__1
150__000__1
100__000__1
1
0.8
0.6
0.4
0.2
0
-0.2
300
350
400
450
500
550
600
650
700
Filter CF (Hz)

Figure 17

In this sense, detection of the target interaural phase shift can be thought of as a sort of interaural correlation profile analysis. There is a confound present, however, in that even though the cross-correlation value of the filter pair centered on the target appears to lower relative to the values of the surrounding filters, it is also lower than the cross-correlation value of the *same* filter pair during a non-target interval. While it is argued on the basis of the perceptions associated with target and non-target intervals thatdetection must be based on cross-spectral cues, there is a possible single-filter explanation. As suggested above, and as shown in Fig. 2, there is a narrow range of diotic bandwidths between 500 and 300 Hz over which thresholds begin to change, but interaural correlation at the target frequency appears to remain unaffected by the outer uncorrelated bands. Of course, the specific filter shape used in the generation of such figures will affect the estimated bandwidth over which cross-spectral comparisons are made. Further experiments will include the use of spectrally discrete bands of noise, which can be manipulated independently (in terms of cross-correlation).

CHAPTER 3

COMPUTER SIMULATION

Introduction

Grange (1991) measured thresholds for detection of a 20-Hz wide interaural phase shift within a wideband noise. The noise was diotic for some bandwidth (900, 700, 500, 300, 200, 150, or 100 Hz) around the target band, and the interaural phase of each component outside of this band was randomized. Data for three subjects in this experiment were shown in Figure 1. As the diotic bandwidth was narrowed, a greater target interaural phase shift was required for detection. It was suggested that detection of the target might be accomplished by a comparison of cross-correlation values between filters centered on the target and filters surrounding the target; i.e., a recognition of the relative minimum seen at the CF of the target in Figs 9 - 16. Presented here are results of a computer simulation of the data which was run with the goal of assessing whether the trends in the behavioral data can be reproduced assuming that cross-correlation values are compared over a constant, fixed spectral region (regardless of the stimulus configuration). Previously, it has been suggested that a "binaural critical band" exists that is somewhat wider than the monaural critical band with the same center frequency (Sondhi and Guttman, 1966). This is also based on the finding that, in bandnarrowing experiments using an MLD paradigm, thresholds begin to decrease at a wider noise bandwidth when binaural information is available (the N_oS_π conditions; Bourbon and

Jeffress, 1965).

Several simplifying assumptions are made for the simulation. First, since the noise surrounding the target is diotic, having cross-correlation peaks at 0 μs, only correlation values at midline will be used in the computations (in essence, using the Pearson product-moment correlation instead of the entire cross-correlation *function*). Second, since the presence of the target causes a decorrelation in a narrow region of the spectrum, it is assumed that an average of cross-correlation coefficients across the spectrum will reflect that decorrelation.

Methods

Diotic bandwidths of 300, 200, 150, and 100 Hz were tested. A tracking procedure (two down, one up rule) was used in conjunction with a 2-AFC task, identical to the procedure in the behavioral experiment. On each experimental trial, two sets of noise were generated (each set consisting of the left and right channels of a single interval) via an 4096-point inverse FFT program on a Masscomp minicomputer. For each set, random starting phases were selected for the left channel. Within the diotic band, starting phases for the right channel were identical, and starting phases outside of this band were selected randomly. Of course, one set contained the target interaural phase shift (different starting phases were selected for signal and non-signal intervals).

In order to simulate internal noise within the auditory system, independent noise of an identical bandwidth was added to each channel, in both the signal and non-signal intervals. Conditions were tested at noise levels of 0, -6, and -12 dB (re: stimulus noise level). The left and right noise signals for each interval were then passed through a bank of gamma-tone filters, simulating the characteristics of the auditory periphery. The output of each pair of frequency matched filters (left and right) were then cross-correlated, and the average cross-correlation across the spectrum was computed. The spectral shaping of the weighting function was accomplished by multiplying the coefficient of each filter pair according to a weighting function. The cross-spectral weighting function was Gaussian in shape, with a mean at 500 Hz (the target frequency). The width was varied with standard deviations of 1, 50, and 200 Hz (in linear frequency). The spectral weighting function implemented here should not be confused with the spectral weighting function suggested by Raatgever and Bilsen (1971), which places more weight on the region around 600 Hz, and is based on lateralization judgements of phase-inverted noise. The resulting average from the two observation intervals was then compared, with the interval producing the lower of the two averages selected as the signal interval. A step-by-step description of the simulation is shown in Figure 18.

Threshold estimates were obtained from a block of 50 trials, in which an even number of reversal points were averaged, excluding the first two reversals. If fewer than 10 reversals occurred, then the entire block was discarded. Final thresholds for each condition were based on at least 4 blocks.

Results

Figure 19 shows simulation interaural phase shift thresholds as a function of the diotic bandwidth of the stimulus. Each panel represents a different "internal" noise level, with the parameter in each panel representing a different standard deviation of the

Figure 18 A step-by-step description of the computer simulation is shown (see text).

Figure 18

1. Generate left and right waveforms for signal and non-signal intervals (Interval 1 will contain signal):

Interval 1	Interval 2
$\mathbf{L_{(t)}}$ & $\mathbf{R_{(t)}}$	$\mathbf{L_{(t)}}$ & $\mathbf{R_{(t)}}$

2. Add independent noise to both channels in both intervals:

$$\mathbf{L_{(t)} = L_{(t)} + N1_{(t)}} \qquad \mathbf{L_{(t)} = L_{(t)} + N2_{(t)}}$$
$$\mathbf{R_{(t)} = R_{(t)} + N3_{(t)}} \qquad \mathbf{R_{(t)} = L_{(t)} + N4_{(t)}}$$

3. Filter each waveform through a bank of bandpass filters:

$$\mathbf{L_{(cf,\, t)}} \qquad \mathbf{L_{(cf,\, t)}}$$
$$\mathbf{R_{(cf,\, t)}} \qquad \mathbf{R_{(cf,\, t)}}$$

4. Cross-correlate frequency matched pairs of filters for each interval:

$$\mathbf{CC\mathit{r}_{(cf)} = \mathit{r}[L_{(cf,\, t)} \;\&\; R_{(cf,\, t)}]} \qquad \mathbf{CC\mathit{r}_{(cf)} = \mathit{r}[L_{(cf,\, t)} \;\&\; R_{(cf,\, t)}]}$$

5. Multiply by spectral weighting function:

$$\mathbf{Weighted\ CC\mathit{r}_{(cf)} = CC\mathit{r}_{(cf)} * WF_{(cf)}} \qquad \mathbf{Weighted\ CC\mathit{r}_{(cf)} = CC\mathit{r}_{(cf)} * WF_{(cf)}},$$

where $WF_{(cf)}$ is Gaussian with a mean of 500 Hz and a standard deviation of 1, 50, or 200 Hz.

6. Average weighted cross-correlation values across frequency:

$$\mathbf{AVE_{CC} = \Sigma\ CC\mathit{r}_{(cf)}\ /\ \#\ of\ filters} \qquad \mathbf{AVE_{CC} = \Sigma\ CC\mathit{r}_{(cf)}\ /\ \#\ of\ filters}$$

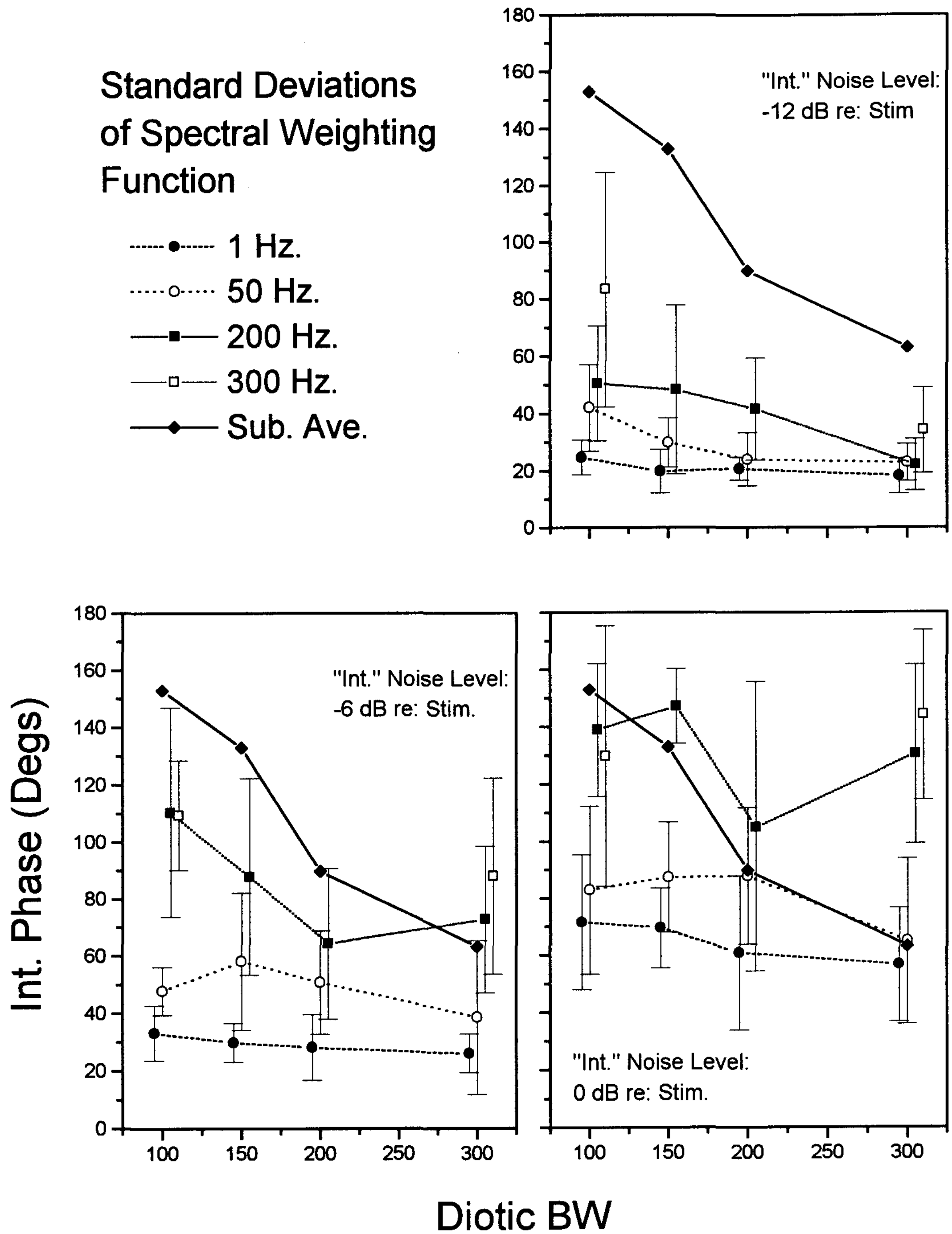
7. If Interval 1 (AVE_{CC}) > Interval 2 (AVE_{CC}) then
   RESPOND "TARGET PRESENT INTERVAL 2"

   If Interval 1 (AVE_{CC}) < Interval 2 (AVE_{CC}) then
   RESPOND "TARGET PRESENT INTERVAL 1".

Figure 19

Threshold interaural phase shifts for the computer simulation are plotted as a function of the diotic bandwidth of the stimulus. The parameter is the standard deviation of the Gaussian spectral weighting function, and each panel represents a different level of added independent noise. The diamond symbols in each graph represents the behavioral data from Grange (1991).

Figure 19



Standard Deviations of Spectral Weighting Function
1 Hz.
50 Hz.
200 Hz.
300 Hz.
Sub. Ave.
"Int." Noise Level: -12 dB re: Stim
"Int." Noise Level: -6 dB re: Stim.
"Int." Noise Level: 0 dB re: Stim.
Int. Phase (Degs)
Diotic BW
0
20
40
60
80
100
120
140
160
180
100
150
200
250
300

spectral weighting function. The human subjects' data for the corresponding conditions are represented by the solid line in each panel (the 500, 700, and 900 Hz conditions are omitted in the graphs).

In the conditions where the "internal" noise was -12 dB re: stim. level (upper righthand panel), the simulation outperformed the subjects in all conditions, regardless of the width of the spectral weighting function. As expected, thresholds were not affected by diotic bandwidth for a weighting function SD of 1 Hz. Thresholds increase slightly for wider SD's as a function of diotic bandwidth, as the spectrally remote uncorrelated bands are given more weight, but the increase is not as steep as the subjects' data.

At "internal" noise levels of -6 and 0 dB, thresholds overall begin to increase. For spectral weighting function SD's of 200 Hz, the simulation thresholds increase as a function of the stimulus diotic bandwidth more rapidly, approximating the shape of the behavioral data. Unexpectedly, at both of these internal noise levels, there is an *increase* in thresholds as the stimulus diotic bandwidth increases from 200 to 300 Hz; however, variability in the threshold estimates are quite large.

Discussion

If there were no internal noise at all, it would be predicted that there would be an interaction between spectral weighting bandwidth and stimulus diotic bandwidth; that is, if the spectral weighting function were narrow, the interaural correlation of the noise in the spectrally remote regions would have no effect on thresholds. However, as the weighting function becomes wider, the outer uncorrelated noise will serve to introduce more variability into the distributions of average cross-correlation values. Therefore, thresholds should increase as an inverse function of diotic bandwidth, with the rate of increase becoming greater with wider weighting functions. This pattern is generally seen in the simulation data at an internal noise level of -12 dB.

The effect of adding internal noise should be to cause an overall increase in thresholds, since the distributions of average cross-correlation values are more variable. This can also be seen across the three different panels of the graph. However, the interaction between internal noise level and stimulus condition is less intuitive. First, even though the internal noise is the same overall bandwidth as the stimulus, outer spectral regions of the stimulus are already uncorrelated, so at these frequencies, adding the noise essentially has no effect. Therefore, adding internal noise simply decorrelates the diotic band. As the level of the added noise increases, the *difference* between correlation values of the diotic region and the uncorrelated regions of the stimulus decreases. For the 50 and 200 Hz wide spectral weighting functions, it is this difference that caused the increase in thresholds as the diotic bandwidth of the stimulus decreased. When the difference between the diotic (inner) and outer spectral bands is reduced, the effect of diotic bandwidth should also be reduced. Therefore, as level of the internal noise increases, thresholds as a function of stimulus diotic bandwidth should flatten out. It does appear that the error bars for a particular spectral weighting function overlap a great deal, due to the increase in variability caused by greater internal noise levels and wider spectral weighting functions, making conclusions difficult at best.

It appears that, given the assumptions made about the basis of detection in the behavioral experiment, the data are not well modelled by a system with a fixed bandwidth over which interaural information (in the form of cross-correlation functions) is combined. Several factors must be considered, however. In the behavioral experiment, each interval was preceded by a temporal fringe having an interaural configuration identical to a non-signal interval, so that the *onset* of the target may have provided an additional cue for detection that was not accounted for in the simulation.

Further, it was assumed that information about the presence of the target was represented as a decrease in the *average* cross-correlation across frequency, which served as the decision statistic for choosing between the two intervals. In the behavioral experiment, even though a two-interval task was used, a dichotic pitch can be heard (and detection can occur) with only one interval. Clearly, the perception of dichotic pitch is based on segregation of the target channel and surrounding channels, and a simulation based on averaging may not capture the true basis of detection.

Finally, the simulation assumed that behavior was based only on cross-correlation values at midline (interaural delay = 0 μs). This assumption was based on the fact that for diotic noise, the peaks of cross-correlation functions will be located at midline. If behavior is influenced by activity elsewhere within (f/τ) space, it would not be represented in the simulation. This is certainly possible, even probable, since the perceived location of the pitches is off of the midline.

Still, an alternative to a fixed bandwidth over which binaural information is combined can be suggested. Common interaural values are thought to be a possible cue to grouping of spectral components (Bregman, 1990). By this argument, the components within the diotic band would be "grouped" together on the basis of the position/height of the cross-correlation peaks of the peripheral filters centered on those frequencies. Detection would take place on the recognition that the cross-correlation function of a narrow spectral region differs from the (common) functions of the surrounding regions. This would suggest that a description of the pattern of cross-correlation functions across the spectrum would need to be established prior to detection of a deviation (caused by the presence of the target) from this pattern. Detection of the interaural manipulation could be modelled as *statistical outlier detection.* In this procedure, an average cross-correlation value (at $\tau = 0$ μs, since that is location of the peaks along the τ axis) might be computed across a range of frequencies, along with a measure of variability of those cross-correlation values. A particular cross-correlation value would be determined an outlier (and produce a dichotic pitch) if its probability of occurrence was below some criterion, given the variability across the spectrum. One simple formulation of this concept is shown in equation (3):

$$\text{If } r_j < r_{ave} - (X * SD_r), \text{ then } r_j \text{ is an outlier,} \quad (3)$$

where r_j is the cross-correlation value of the j^{th} filter pair, SD_r is the standard deviation of the cross-correlation values across N filters, given by

$$SD_r = [\Sigma(r_i - r_{ave}) / N]^{.5}, \text{ for } i = 1, N, \quad (4)$$

and X is the number of standard deviations defined as the outlier criterion.

Two factors could decrease the probability of a given cross-correlation value being determined an outlier. One is that variability among values across the range of frequencies being averaged increases, so that SD_r increases. This is consistent with the notion of a constant spectral weighting function, as assumed in the simulation. The other is a decrease in N, as would occur if the spectral region over which the averaging took place was narrowed. This is consistent with the idea that the activity in the interaural configuration of the stimulus (the range of consistent cross-correlation coefficients) determines the spectral width of the noise background against which dichotic pitches are heard.

Unfortunately, both of these predict that as the width of the diotic band decreases, thresholds for detecting the target interaural phase shift should increase. Given the variability in the simulation data, it does not seem warranted to reject a model based on a fixed bandwidth. Instead, the next experiment investigates the possibility that the binaural system does group spectral regions on the basis of interaural parameters prior to detection of the target.

CHAPTER 4

EXPERIMENT II

Introduction

Numerous studies have found that the presence of frequency components spectrally distant from a target (distractors) can interfere with the processing of the interaural parameters of the target (e.g., Dye, 1990; Stellmack and Dye, 1993; McFadden and Pasanen, 1976). One explanation for such cross-spectral interference is that the interaural information about the separate components is integrated across different peripheral channels, so that the interaural value available to the decision making stage of processing is some weighted combination of the individual values.

If this was true, it might be predicted that conditions leading to *perceptual* segregation of the target and distractor components would reduce the amount of interference provided by the distractors. There is evidence that this is true in some conditions; however, in other cases perceptual segregation seems to have no effect on interaural processing. Buell and Hafter (1988) found that when a harmonic relationship existed between a target tone and a distractor tone (leading to perceptual fusion; Bregman, 1990), thresholds for detecting an ITD at the target were elevated compared to when the tones were inharmonic. Studies using onset asynchronies between a distractor tone and a target tone, however, have found no such difference in thresholds, even though subjects do report "hearing out" the target when the asynchrony is present

(Stellmack and Dye, 1993; Woods and Colburn, 1992).

A model tested by Woods and Colburn (1992) suggested that separate components are grouped together by assigning weights to the various components, and extracting overall values (including spatial position, based on interaural values) about the group by combining individual information according to the weights assigned. Using tonal targets and distractors, they elicited perceptual segregation using an onset asynchrony between target and distractor, and assumed that the binaural system assigned weights of 1.0 to the target and 0.0 to the distractor. In these conditions, it was predicted that performance should be equivalent to conditions in which the target was presented alone. In conditions where all components were gated on simultaneously (and perceptually fused), they predicted that weights would be equal across the spectrum, and therefore that interference should occur.

In the present experiment, a noise stimulus will be used to investigate the effect of *changes* in the interaural parameters of off-target regions, with particular interest in the original interaural position of those regions. It has been observed previously that a diotic temporal fringe preceding an observation interval potentially containing a narrowband interaural phase shift can lead to better detection performance; that is, the temporal fringe increases the likelihood that a listener will hear a dichotic pitch. This has been attributed (in an MLD paradigm) to the added presence of an onset dichotic cue, while with no fringe, only ongoing dichotic cues are present (Gilkey, Simpson, and Weisenberger, 1990; Yost, 1988).

In this experiment, both forward and backward fringes will temporally surround an observation interval in which a target narrowband interaural phase shift may occur. At the onset of the observation interval, an additional, extraneous interaural phase shift will occur in outer spectral regions, regardless of whether the target was presented. A listener would ideally "ignore" the phase shifts at the spectral regions remote from the target.

The main variable of interest in the experiment is the initial interaural position of outer spectral regions, which will contain extraneous interaural changes that coincide temporally with the onset of the target. The target will once again be an interaural phase shift of a narrow band, surrounded by diotic components of a given bandwidth. If the binaural auditory system can weight different spectral regions according to their interaural parameters, and if the entire spectrum is diotic in the forward fringe, all components should be weighted equally. If the outer spectral bands have an interaural position other than midline, however, they should be given less weight (assuming that the diotic region, which contains the target, is given maximum weight).

In general, models of lateralization suggest that the perceived position of a stimulus depends on the position of peaks in the (f/τ) plane. These can be accurately represented by a summary cross-correlation function, which is essentially generated by collapsing functions across the entire spectrum (Shackleton et al., 1992; Stern et al., 1988). For completely diotic noise, this summary function will have a large peak at midline. Introduction of a narrowband interaural phase shift will cause a decrease in the height of the summary peak, and presumably this is the onset cue for detection of that phase shift.

If, however, extraneous phase shifts at other frequencies also occur, they will also serve to decrease the height of the peak, if they were included in the summary function (that is, if they were weighted heavily). In this case, a greater decorrelation caused by the *target* phase shift should be needed for detection. On the other hand, if the non-target frequencies are given low weight, and therefore do not contribute greatly to the summary cross-correlation function, then phase shifts at these frequencies should not affect target thresholds. The current experiment compares conditions in which non-target spectral regions are either diotic or interaurally phase shifted 90° during the temporal fringe, to examine whether the binaural system can differentially weight regions on the basis of interaural parameters.

It is important to recognize that previous experiments that induced perceptual segregation of a target and interfering components used *monaural* cues to achieve this. An implicit assumption is that the spectral weights generated based on monaural cues can be applied to any binaural mechanism. In the present experiment, the basis of a spectral weighting function lies in the binaural processor itself.

Methods

Thresholds for detecting an interaural phase shift of a 20-Hz wide section of noise low-pass filtered at 1 kHz were measured. The 20-Hz wide target band was centered on 500 Hz. A wider section of noise symmetrically surrounding the target band was presented diotically for the entire duration of the stimulus; the width of this inner band was either 750, 500, or 250 Hz. The interaural phases of the frequencies above and below the diotic inner band varied during the course of each presentation.

The stimulus was made up of three temporal segments - a forward fringe, an observation segment (during which the target phase shift occurred on signal intervals), and a backward fringe. In one set of conditions, each component in the outer bands was diotic during both the forward and backward temporal fringes, and phase shifted 90° (either to the left or right) in the observation segment. In the other set of conditions, the 90° phase shift (left or right) of the outer bands occurred in the temporal fringes, and the outer bands were diotic in the observation segment. Figure 20 shows the two conditions when the outer phase shift is to the left. All four conditions were tested separately. The duration of the observation segment also varied. In one set of conditions, the observation segment was 250 ms, with the temporal fringes being 225 ms each. In another set, the observation segment was 100 ms, and the fringes were 300 ms, so that total interval duration was always 700 ms.

A 2-AFC adaptive procedure with a two-down, one-up rule was used to estimate the interaural phase shift leading to 70.7% correct (Levitt, 1971). In the signal interval, the target band was interaurally phase shifted during the observation segment, while in the non-signal interval, the target band remained diotic for the entire stimulus duration. Thresholds for a block of 50 trials were based on a minimum of 8 reversals (with the first two being ignored), and final thresholds for a condition were based on at least 6 blocks (300 trials).

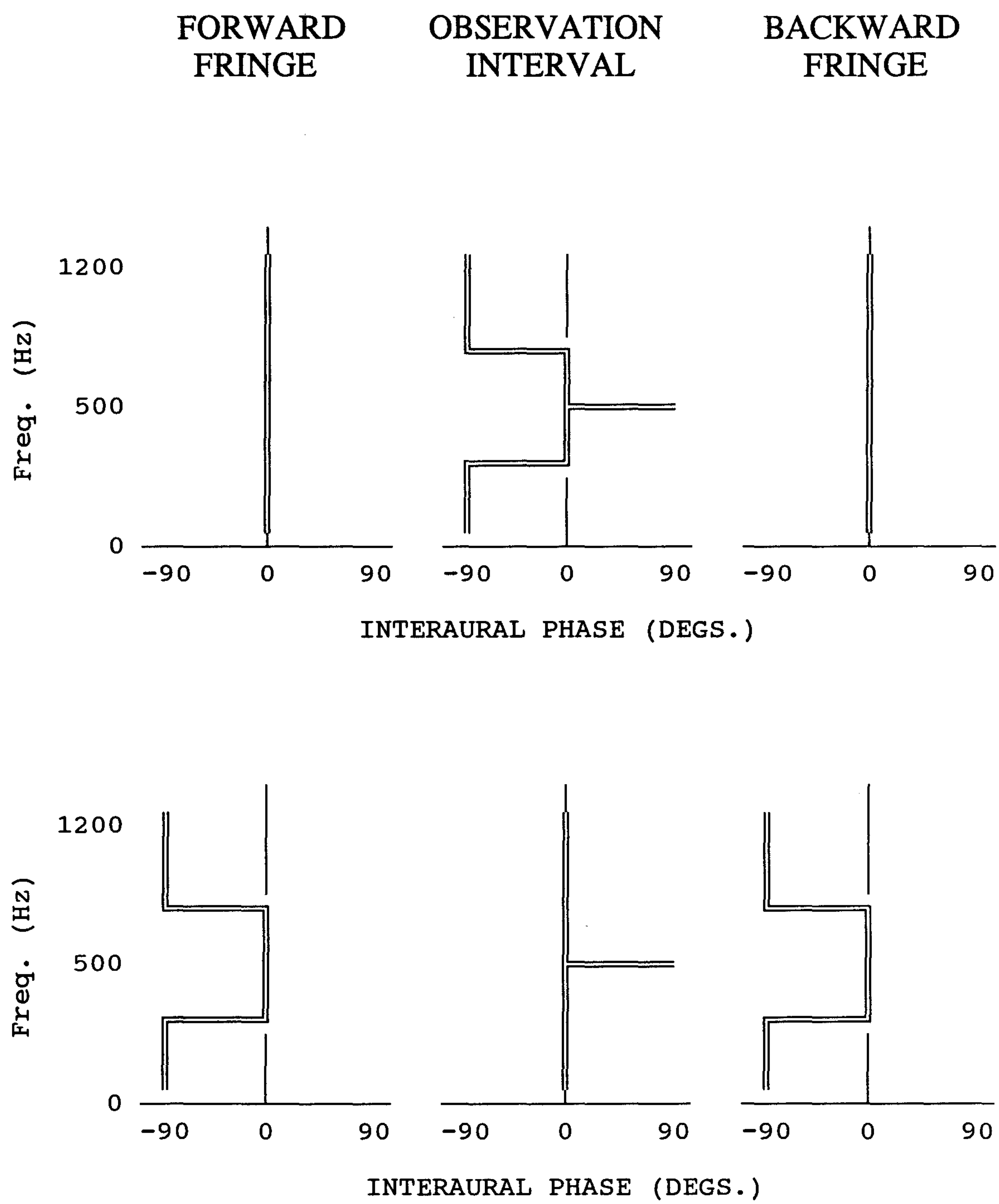
Stimulus generation was achieved via an inverse FFT routine on a Masscomp minicomputer, with the time-domain waveform played out through a 16-bit d/a converter at a rate of 4096 points per second. The noise had a flat power spectrum, and was low-

Figure 20

Examples of two of the stimulus interaural configurations are shown. In each panel, interaural configurations are shown for each of the three temporal segments of an interval (forward fringe, observation interval, and backward fringe). The top panel represents conditions in which the spectrally flanking bands were interaurally phase shifted in the observation interval, and diotic in the temporal fringes. The bottom panel shows the opposite configuration with the flanking bands phase shifted in the fringes.

Conditions were run with the extraneous phase shifts to the right as well as to the left, and the target phase shift was always to the right.

Figure 20


FORWARD FRINGE
OBSERVATION INTERVAL
BACKWARD FRINGE
1200
500
0
Freq. (Hz)
-90
0
90
INTERAURAL PHASE (DEGS.)
1200
500
0
Freq. (Hz)
-90
0
90
INTERAURAL PHASE (DEGS.)

pass filtered at 1 kHz (Rockland Series 2000). A one-second sample of noise with random starting phases for each component was generated for the left signal. Another 1-s sample (with the appropriate phase shifts) was generated thatserved as the right signal for the fringes or the non-signal observation segment (depending on condition). The right waveform for a signal interval was generated anew each trial, using the same starting phases as a non-signal right waveform, except for the 20-Hz target band. On each trial, a starting point in the appropriate set of waveforms for each segment was chosen randomly. The three temporal segments were joined without gating, but onset and offset of the entire stimulus were shaped with 10-ms $\cos^2$ ramps. Stimuli were attenuated (Leader LAT-45) and sent to Crown amplifiers, which drove the headphones (Sony MDR V6).

Subjects were three Loyola University students with no known hearing deficiencies. Data were collected in 1.5-hour sessions, with approximately 300 trials being presented each session.

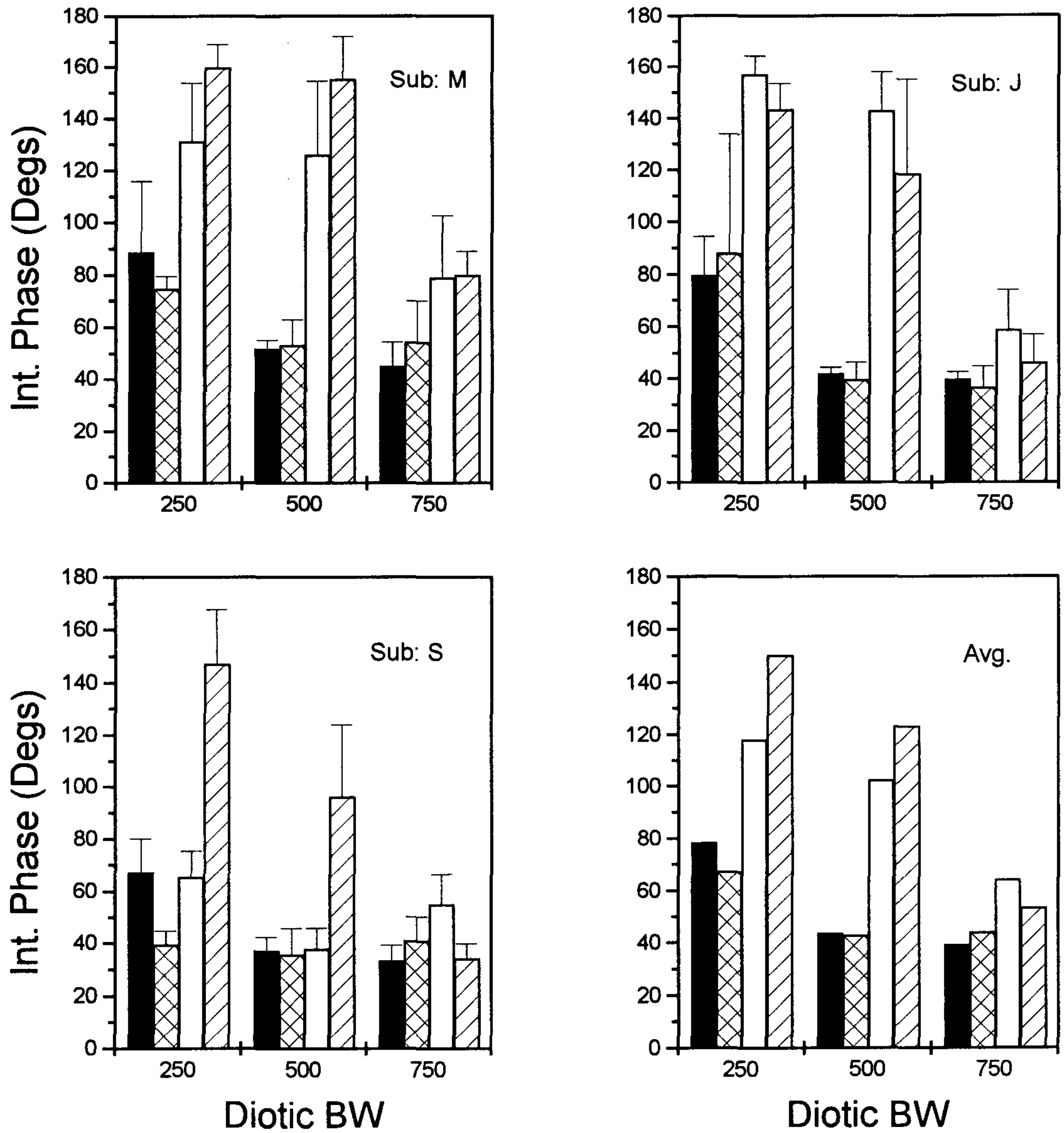
Results

Figures 21 and 22 show the data for three subjects along with the average results. Figure 21 represents conditions in which the observation interval was 100 ms, and Figure 23 shows the data for the 250-ms observation intervals. Each panel represents a different subject, and thresholds for detection of the target interaural phase shift are plotted as a function of the bandwidth of the diotic region surrounding the target. The four bars at each diotic bandwidth represent the different temporal location/direction combination for the spectrally flanking phase shifted bands.

Figure 21

Threshold interaural phase shifts are plotted for each diotic bandwidth. Each panel represents a different subject (the lower right panel represents the average of all three subjects). At each diotic bandwidth, the four bars represent the different direction/temporal position combinations of the extraneous interaural phase shifts. These data were obtained with fringe durations of 300 ms, and an observation interval duration of 100 ms.

Figure 21



Duration of temporal segments (Ffringe, Obs, Bfringe):

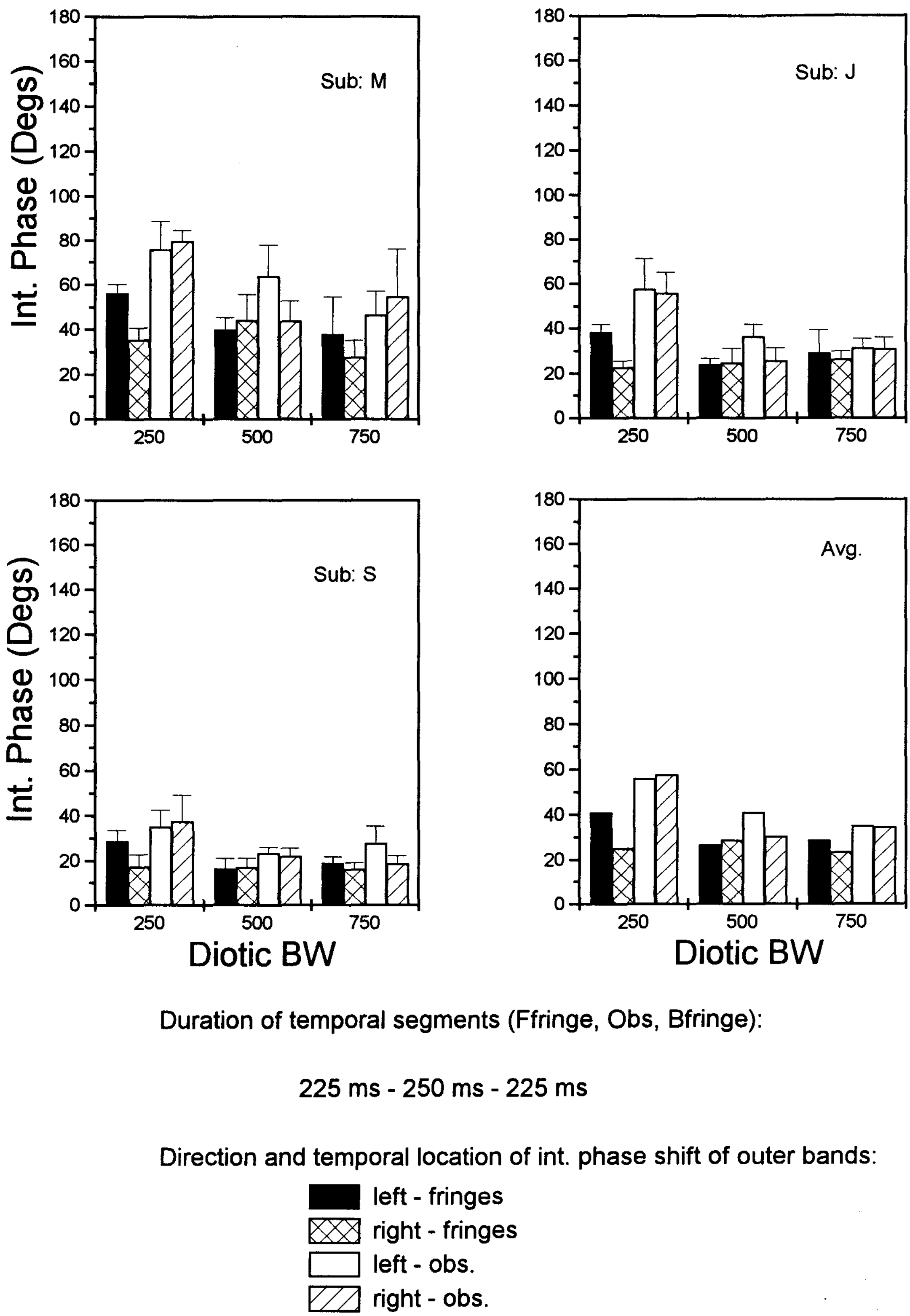
300 ms - 100 ms - 300 ms

Direction and temporal location of int. phase shift of outer bands:

Figure 22

Same format as Figure 21, except that the fringe durations were 225 ms, and the observation interval duration was 250 ms.

Figure 22


Sub: M
Sub: J
Sub: S
Avg.
Int. Phase (Degs)
180
160
140
120
100
80
60
40
20
0
250
500
750
Diotic BW
Duration of temporal segments (Ffringe, Obs, Bfringe):
225 ms - 250 ms - 225 ms
Direction and temporal location of int. phase shift of outer bands:
left - fringes
right - fringes
left - obs.
right - obs.

For observation interval durations of 100 ms, thresholds are generally higher when the extraneous interaural phase shifts are located in the observation interval, as opposed to the temporal fringes. This suggests that, during the fringes, if the entirespectrum is diotic, all components are grouped together, and therefore the outer phase shifts cannot be ignored at the onset of the observation interval. This effect is reduced by several other factors, however. Increasing the bandwidth of the diotic noise surrounding the target reduces the difference between thresholds in the two sets of conditions, as does increasing the duration of the observation interval. There does not appear to be a strong effect of direction of the extraneous phase shifts, except for Subject S, for whom thresholds only increased when the outer phase shift was in the same direction as the target. It is also important to note that, for the 100-ms observation interval duration, thresholds for conditions when the outer phase shifts are in the temporal fringes increase at the narrowest diotic bandwidth (the importance of this will be explained in the discussion section).

Discussion

While a quantitative model is not advanced here, some general predictions can be made if certain assumptions are made. First, it is assumed that the binaural auditory system has access to the pattern of activity in a cross-correlation network (the f/τ plane), in which the outputs of frequency-matched filters from each periphery are cross-correlated. It is also assumed that the activity in this network is represented as an average cross-correlation function, collapsed across frequency. At the onset of the forward temporal fringe, frequency bands are assumed to be weighted according to the height of the corresponding cross-correlation function at a particular interaural delay, then combined into a summary cross-correlation, with the contribution of each individual cross-correlation function dependent on the weight assigned to it. More specifically, it is assumed that prior to averaging across frequency (formation of a summary cross-correlogram), the height of the cross-correlation function at each frequency *at a particular* τ is assessed, and greater weight is assigned to frequencies with higher values at that τ. Since both the narrow target band and the surrounding frequencies are diotic during the forward fringe, it is assumed that the cross-correlation values at $\tau = 0$ μs are evaluated, and therefore, frequency bands with cross-correlation peaks off of midline are weighted less heavily.

Detection of the target narrowband interaural phase shift presumably proceeds on the basis of the changes in the summary cross-correlation function that the target onset causes. With these assumptions in mind, each independent variable can be discussed separately.

Effects of Fringe Interaural Configuration

Since the target band and the surrounding frequencies were always diotic during the forward fringe, that spectral region will always be weighted heavily, and therefore influence the characteristics of the summary cross-correlation. At the onset of the observation interval, the target interaural phase shift will cause a change in those characteristics (e.g., a decrease in the height of the peak). This change is modelled as the detection cue. In conditions where the flanking spectral bands are interaurally phase shifted, they are not weighted heavily, and the interaural phase shift that occurred in these bands (simultaneously with the target) should not have affected the summary cross-correlation. Therefore, because they are deemphasized when weights are originally assigned, thresholds should not be affected by these extraneous changes.

In conditions where the flanking spectral bands are diotic during the temporal fringes, a maximum weight would be assigned to these outer bands as well. At the onset of the observation interval, the interaural phase shift in these bands would decrease the height of the summary cross-correlation, necessitating a larger target interaural phase shift for detection.

It is possible that this pattern of results could be accounted for simply by the interaural configuration of the noise in the observation interval alone. That is, if the observation interval was presented in isolation (no fringes), the two conditions would be (1) detection of the target against a background of completely diotic noise (corresponding to conditions when the fringes contained the spectrally flanking phase shifts) and (2) a diotic band of either 250, 500 or 750 Hz surrounding the target, with a 90° phase shift outside of this band. It might be expected that thresholds between the two sets of conditions would differ more as the diotic bandwidth became narrower, based on the results of Grange (1991); this is indeed what occurs. However, in the present experiment, if only the noise interaural configuration in the observation interval affects performance, then thresholds in conditions where the noise is completely diotic (labelled left-fr. and right-fr. in Fig. 4.2) during the observation interval should be identical across diotic bandwidths. Thresholds are elevated, however, at the 250-Hz diotic bandwidth, suggesting that the interaural configuration in the temporal fringe is in fact affecting performance.

Effects of Diotic Bandwidth

If the transition between maximum weighting of diotic spectral regions and minimum weighting of non-diotic regions were abrupt (that is, the weights went from 1.0 to 0.0 as a step function), then thresholds might be expected to improve as the bandwidth of the diotic region surrounding the target decreases, since the relative decorrelation caused by the target would increase. However, if the transition from weights of 1.0 within the diotic band to 0.0 outside of this band were gradual, then one would expect that even in conditions in which the outer spectral bands are interaurally phase shifted during the forward fringe, that frequencies at the transition between the diotic and phase shifted regions would be given a non-zero weight and included in the summary cross-correlation. By increasing the width of the diotic band, the number of frequency bands included in the summary cross-correlation is increased, but the added bands are diotic. As the diotic bandwidth increases, though the influence of frequencies at the transition areas remains, the *relative* influence of those regions on the summary representation is decreased (a greater percentage of the frequencies included in the summary remain diotic throughout the interval). Therefore, as the width of the diotic band increases, thresholds would be expected to decrease.

Effects of Observation Interval Duration

Differences in performance due to the temporal location of the spectrally flanking interaural phase shifts decrease as the observation interval increases in duration from 100 to 250 ms. This would not be predicted if the only cues available were the dynamic changes in the summary cross-correlation caused by the target phase shift. To account for these differences, it is useful to distinguish between *onset* binaural cues and *ongoing* binaural cues. Yost (1985) has shown that at longer signal durations, the MLD for simultaneously presented signal and masker is equivalent to that obtained when the masker is continuous. Though onset and ongoing cues may provide qualitatively different types of information, given enough time, the information from ongoing cues can lead to performance comparable to that attained with onset cues. In the present experiment, it is suggested that, when changes occur in the interaural configuration of the stimulus, the auditory system attempts to assign a new set of weights, a process assumed to be *non*instantaneous. The longer observation interval may provide adequate time for a new set of weights to be generated, excluding the interaurally phase shifted outer bands, and for ongoing binaural cues to be utilized in detecting the target interaural phase shift.

One interesting aspect of this idea is that, according to this model, onset cues would not be expected to contain information about the spectral location of the target, while ongoing cues would. It was suggested that the onset of the dichotic target in this experiment was detected by a change in the peak of the summary cross-correlation function; this summary function is by definition collapsed across frequency. If, for example, the forward fringe was entirely diotic (all frequencies weighted maximally), then the onset of a narrowband phase shift *anywhere* in the spectrum should have the same effect on the summary function. Therefore, uncertainty of spectral location should not affect thresholds. For detection based solely on ongoing cues, however (if no fringe is present), then the auditory system must identify a particular spectral location that differs from all others, since the summary cross-correlation function would not change greatly. This possibility (frequency specific vs. frequency non-specific cues) will be explored in future experiments.

An alternative account for the data based on the number of sources present across intervals can also be considered. In conditions in which the extraneous phase shifts are in the temporal fringes, the noise can be thought of as arising from two sound sources, one at midline, and one elsewhere (for this discussion, all components with a 90° phase shift are treated as arising from a common source). During the observation interval, the interaural configuration will be consistent with two sources (when the target is present) or one (when the entire noise is diotic). Therefore, from a "source-counting" viewpoint, the task can be thought of as a same-different task, with the number of sources either going from two to one, or remaining at two in both the fringe and the observation interval. When the extraneous phase shifts are in the observation interval, however, there is one "source" present in the fringe, and either two sources (when the target is not present) or three (when the target is present) in the observation interval. In these conditions, there is always a *change* in the number of sources, necessitating counting of sources, and not simply monitoring of number of sources across temporal segments.

Unfortunately, it is difficult to assess this hypothesis on the basis of the present data. It could be argued that even though the number of sources remains constant when the extraneous phase shifts are in the fringe and the target is present, the sources in each temporal segment are not the same; i.e., inner band + outer bands vs. entire noise +

target band. A further problem in addressing a "source-counting" hypothesis is that the manipulation of the target interaural phase shift affects the salience of one of the "sources". A more appropriate manipulation would be the number of sources present. If, for example, only one of the outer spectral regions (either above or below the diotic band) contained the extraneous phase shift during the fringe, and the other region contained it during the observation segment, then a "source counting" approach should produce good performance (since there would be two sources during the fringe, and either two or three during the observation segment). On the other hand, if detection is based on changes in a summary cross-correlation function, then performance should be relatively poor, since there would always be an interaural phase shift of initially diotic spectral regions during the observation segment. These conditions will be run in future experiments.

CHAPTER 5

EXPERIMENT III

Introduction

The final experiment measures thresholds for detection of a tonal target against a background of noise of varying interaural configurations. It is well known that if a tonal target is presented interaurally out-of-phase with a background noise, detection performance improves relative to conditions in which both share a common interaural phase. The simplest example of this is the difference in detection thresholds between the N_oS_o and N_oS_π conditions, where the subscripts denote the interaural phase of the noise (N) and the signal (S). The difference in thresholds is known as the masking-level-difference (MLD), since the addition of binaural cues provides a release from masking of the signal.

Since binaural mechanisms are responsible for both the MLD and dichotic pitch detection, it is interesting to consider how the ideas put forth to describe the data from the earlier experiments can also describe MLD data, in conditions in which the masking noise has frequency dependent interaural parameters (as did the noise in the previous experiments presented here).

The existence of the MLD is well explained by several models (e.g., the vector model, Jeffress, 1972; the E-C model, Durlach, 1972). In general, the models hold that in the N_oS_o condition, detection is based purely on monaural mechanisms, and can be modelled as energy-increment detection. In the N_oS_π condition, addition of the signal leads to decorrelation between the left and right signals, which provides an additional cue for detection. The vector model views this in terms of the outputs of left and right bandpass filters, which can be approximated by sinusoids of the frequency that the filters are centered on. For the noise alone, the vectors of the left and right filter outputs are identical. If a signal is added out of phase, left and right vectors will be altered in different ways such that the vector length and phase angle will be different in each channel. The absolute difference of the vector phases represents the interaural phase difference between the two signals; it is this difference that provides the additional cue for detection. In addition, the disparate length of the resultant vectors (representing the magnitude of each signal) represents an interaural level difference, which likewise may serve as a detection cue (Hafter and Carrier, 1969).

The E-C model suggests that the binaural auditory system equalizes the left and right signal by adjusting the time and level of the waveforms, and then subtracting one signal from the other. For a diotic signal and masker, the subtraction is nearly perfect (random "noise" is introduced to the timing and level of each signal so that the model does not overpredict human performance). For a dichotic signal, however, subtraction of the two waveforms always leaves remaining energy, due to the fact that the signals cannot be perfectly equalized. The remaining energy (in the binaural processor) is the source of the improvement in detection. Colburn and Durlach (1978) have argued that these (and other models of binaural hearing) are instances of a general model based on cross-correlation activity between left and right banks of bandpass filters. It is important to recognize that either model (and a general model based on cross-correlation) can also predict the existence of dichotic pitches, since the narrowband interaural manipulation will result in non-identical vectors (for filters passing the interaurally altered band), or alternatively, imperfect cancellation (beyond that caused by time and amplitude jitter).

Sondhi and Guttman (1966) measured MLD's using a noise with an interaural configuration similar in many respects to the noise used in the previous experiments presented here. Specifically, noise within some band surrounding the signal was diotic, and noise outside of this band was 180° interaurally out-of-phase. This noise configuration is denoted as $N_{\pi o \pi}$. A configuration of $N_{o \pi o}$ was also used, and detection thresholds were measured for S_o and S_π signals, for various widths of the "inner" band. Of course, as the inner band increased in width from 0 to wideband in the $N_{o \pi o}S_o$ condition, the task essentially changed from N_oS_o to $N_\pi S_o$, and thresholds decreased as a function of the inner bandwidth. When the signal was presented interaurally out-of-phase, that is, for $N_{o \pi o}S_\pi$, thresholds increased as the inner band became wider. The compelling aspect of the data was the asymmetry of the two curves; as the inner bandwidth in the $N_{o \pi o}S_\pi$ condition increased from 0 Hz, thresholds increased sharply, approaching the $N_\pi S_\pi$ value quickly. However, in the $N_{o \pi o}S_o$ condition, as the inner bandwidth increased from 0 Hz, thresholds decreased relatively gradually, reaching the $N_\pi S_o$ level at a wider inner bandwidth.

Sondhi and Guttman argued that this asymmetry showed that MLD's could not simply be based on the outputs of a single pair of filters. Their conclusions, however, were based on predictions made using the EC model (Durlach, 1972) assuming rectangular filters. The problem can be stated as "how can an increase in inner bandwidth affect thresholds for an S_π signal but not for an S_0 signal, if the filter width is identical"? To explain the apparent paradox, two aspects of the task must be considered. First is the binaural advantage provided by the two signals, as the width of the inner band varies. At the widest inner bandwidth (for $N_{\pi 0 \pi}$ noise), where the entire noise is diotic, S_π detection proceeds on the basis of the decorrelation that the signal causes, whereas S_0 detection can only proceed on the basis of a level increase (monaural cues). As the width of the inner band is narrowed, however, the interaural correlation of the noise begins to decrease, due to the presence of dichotic noise in the tails of the filters. There is, theoretically, a binaural cue available for both signals in this situation; the *increase* in interaural correlation caused by the S_0 signal, or the *decrease* caused by the S_π signal. For relatively high reference correlations, the effect of an S_0 signal is minimal, and thresholds do not improve (until the inner band is extremely narrow). The change in interaural correlation caused by the S_π signal remains relatively high, however, simply because the difference between signal and noise correlation is greater.

The second aspect of the task is the changing reference correlation across inner bandwidths. In the N_0S_π condition, thresholds are lowest, due to the fact that the noise correlation is 1.0, and the signal provides the most decorrelation that a tonal signal can; that is, the binaural cues are largest. A narrow band of N_π noise centered on the signal will tend to decorrelate the left and right signals a great deal, since the (approximately triangular) filters are influenced predominately by that spectral region. Note that if the filters were rectangular, placing N_π noise anywhere in the filter band would have the same effect. Since the reference correlation has now decreased from 1.0, the additional decorrelation caused by the signal is less effective in these conditions, based on the finding that Δ correlation thresholds increase as the reference correlation decreases from 1.0 (Pollack and Trittipoe, 1959). In the N_0S_0 condition, however, the addition of the signal does not change the interaural correlation, and therefore the task is based on monaural mechanisms alone. If a narrow band of N_π noise centered on the target is added, it will, of course, decrease the reference correlation (from 1.0), but the binaural cue arising from the addition of the signal is an *increase* in interaural correlation. The effect of adding the S_0 signal on interaural correlation will increase gradually as the reference correlation decreases, and therefore thresholds decrease gradually. Presumably, the dependence of Δ correlation on reference correlation is related to the variability of interaural correlation, which increases inversely with reference correlation (Gabriel and Colburn, 1981).

Given the similarities between the noise interaural configurations of the first two experiments and that of Sondhi and Guttman (1966), and since both types of tasks assess binaural processing as a function of the interaural parameters of off-target frequencies, it is reasonable to expect a common process underlying both. The present experiment uses the MLD paradigm with a noise background similar to that of Sondhi and Guttman (1966), and is intended, to some extent, as a replication of that study. The MLD's found by Sondhi and Guttman were substantially lower than generally reported (7 dB vs. 15-18 dB), perhaps due to the psychophysical methods used (method of adjustment). For example, if the psychometric functions for N_0S_π and N_0S_0 had differing slopes, and the method used tended to estimate threshold toward one asymptote of the function, then MLD's would be different than those obtained using a method basing threshold estimates on the center of the psychometric function. Considering the differences between their MLD's and those reported elsewhere (Green and Yost, 1975), it is possible that the asymmetry seen in the MLD's as a function of inner bandwidth between $N_{o\pi o}S_o$ and $N_{o\pi o}S_\pi$ conditions is also due to the specific methodology used. For the present experiment, a more rigorous psychometric procedure will be used.

An additional difference between these conditions and those of Sondhi and Guttman is that the interaural phases for the inner and outer bands of noise will be -90° and 90° (as opposed to 0° and 180°), so that differences in the location of the cross-correlation peaks along the interaural time axis (re: $\tau = 0$ μs) between the inner and outer bands will be minimized. A reasonable approach to the task would be to place greater weight on components having cross-correlation peaks in a location that would be changed most by the addition of the signal (that is, if the signal were S_{-90}, initially weight the N_{90} components heavily, so that the summary cross-correlogram will be most changed at the signal onset). An assessment of whether the auditory system can accomplish this would be confounded by an *a priori* greater weighting at more central τ's, if the noise components were either 0° or 180° interaurally out of phase. Therefore, the noise at any particular frequency will be either -90° or 90° out of phase. A more in-depth analysis of this approach will be undertaken in the discussion section.

<u>Methods</u>

Thresholds for detection of a 500-Hz tone presented against low-pass filtered equal-amplitude noise (cutoff = 1200 Hz) were measured. The noise was partitioned spectrally into three segments, distinguished by the interaural phase of the components in each: an inner band, centered on 500 Hz, in which components were interaurally phase shifted by -90° (90° left-leading); two outer bands, adjacent to the inner band, in which components were interaurally phase-shifted by +90° (90° right-leading). The width of the inner band 700, 500, 300, 200, or 100 Hz. At each width of the inner band, thresholds for detecting the target tone were found with the tone at each of four interaural phases (-90°, 0°, 90°, and 180°). The interaural configuration of the masking noise is shown at the top of the data figures (Figures 23 and 24).

A two-interval, forced choice procedure was used. Each interval was comprised of a 500-ms burst of noise, and in the signal interval, the tone was added during the last 250 ms. Therefore, the first 250 ms served as a forward fringe. These durations were selected to ensure that both onset and ongoing dichotic cues were available (see Yost, 1985, for discussion of duration effects). Both the noise and the tone were gated using 10-ms $\cos^2$ onset/offset ramps. Trials were run in blocks of 50, using an adaptive procedure (two-down, one-up rule) tracking on level of the tonal target. The initial step size was 4 dB, and was reduced to 2 dB after 2 reversals. Thresholds for each block were based on at least 8 reversals, with the first two reversals being discarded. Final thresholds for each condition were based on at least 6 blocks (300 trials total).

Stimuli were generated on a Masscomp minicomputer interfaced with a 16-bit d/a converter. The noise was generated using a 4096 point inverse FFT. Starting phases for the left noise components were selected randomly from a rectangular distribution, and starting phases for the right noise components were either advanced or delayed by 90° relative to the left channel, depending on the frequency of the component. The power spectrum of the noise was flat. Signals were output at a rate of 4096 points per second, low-pass filtered at 1200 Hz (Rockland Series 2000), attenuated (Leader LAT-45), amplified (Crown), and presented to the listeners over earphones (Sony MDR-V6).

Another set of conditions was run using identical methods except that the center frequencies of the target tone and the inner noise band were 250 Hz, and that the inner band was presented at widths of 400, 300, 200, 100, and 50 Hz.

Subjects were Loyola University students (two male, one female) in their early twenties. None had any known hearing deficiencies. Data were collected in 1.5-hour sessions, with an average of 300-500 trials being presented in a session.

Results

Figures 23 and 24 show the data from 3 subjects and the average across subjects for the CF's of 500 and 250 Hz, respectively. The top section of each figure represents the interaural configuration and phase of the signal, and each panel in the lower section shows each subjects' thresholds for detection of the target as a function of the interaural phase of the signal. Thresholds are plotted as dB re: $N_{90\backslash-90\backslash90}S_{-90}$ for the widest inner bandwidth, which was assumed to be equivalent to $N_{-90}S_{-90}$, wherethresholds are expected to be greatest. The parameter in each panel is the inner bandwidth.

At the widest inner bandwidths, a large MLD is seen. Thresholds are highest for S_{-90} signals, lowest for S_{90} signals, and slightly higher for the intermediate phases of S_0 and S_{180} (in terms of vector addition of peripheral filter outputs, these two intermediate

Figure 23

In the top portion, the interaural configuration of the masking noise ($N_{90/-90/90}$) is shown. In the bottom portion, thresholds for detection of a 500 Hz tone are plotted as a function of the interaural phase of the target. Thresholds are in dB re: S_{-90} when the -90° portion of the noise is at its widest. The parameter is width of the -90° portion of the noise. Each panel represents a different subject, and the lower right panel is the average of all three.

Figure 23

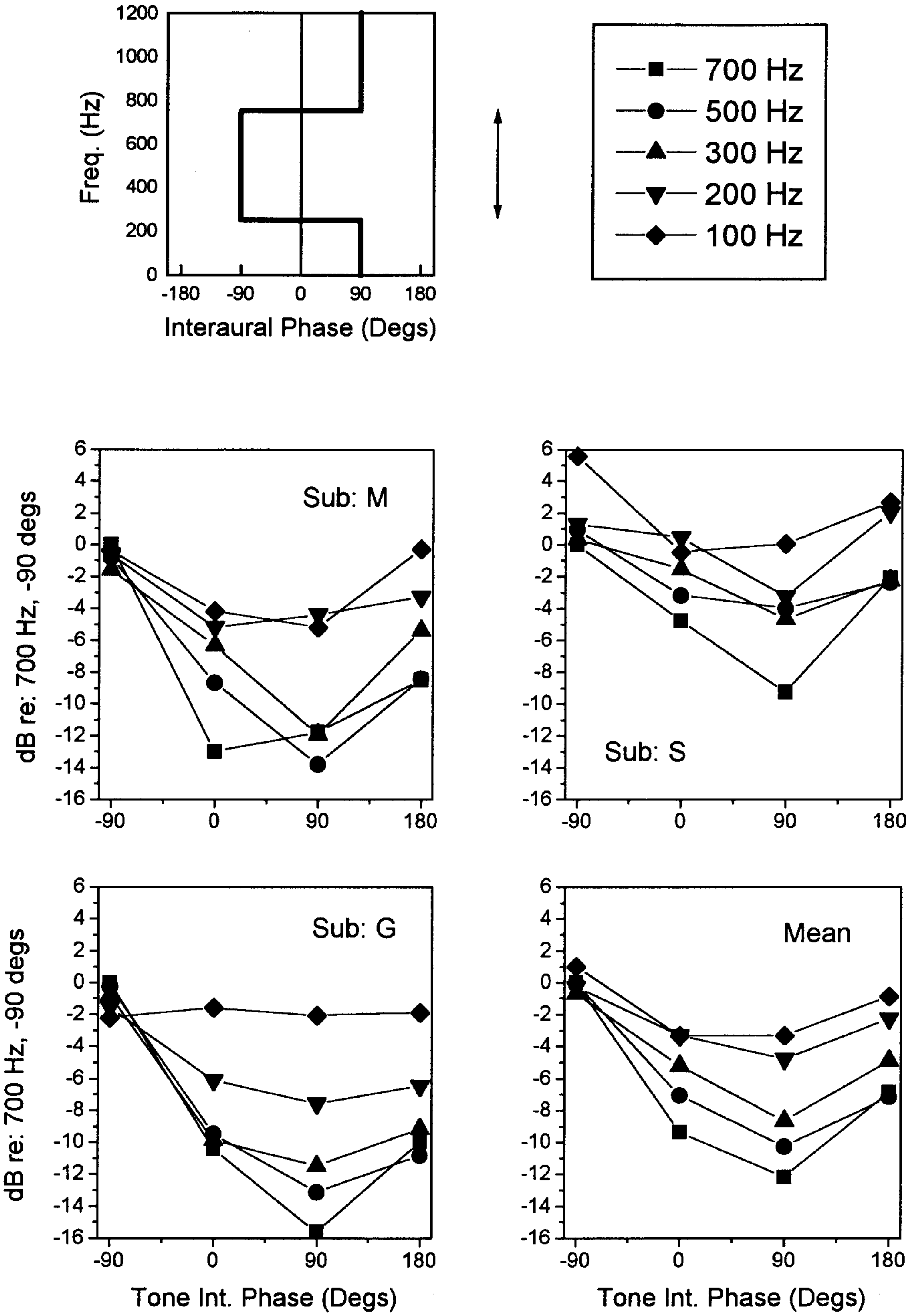
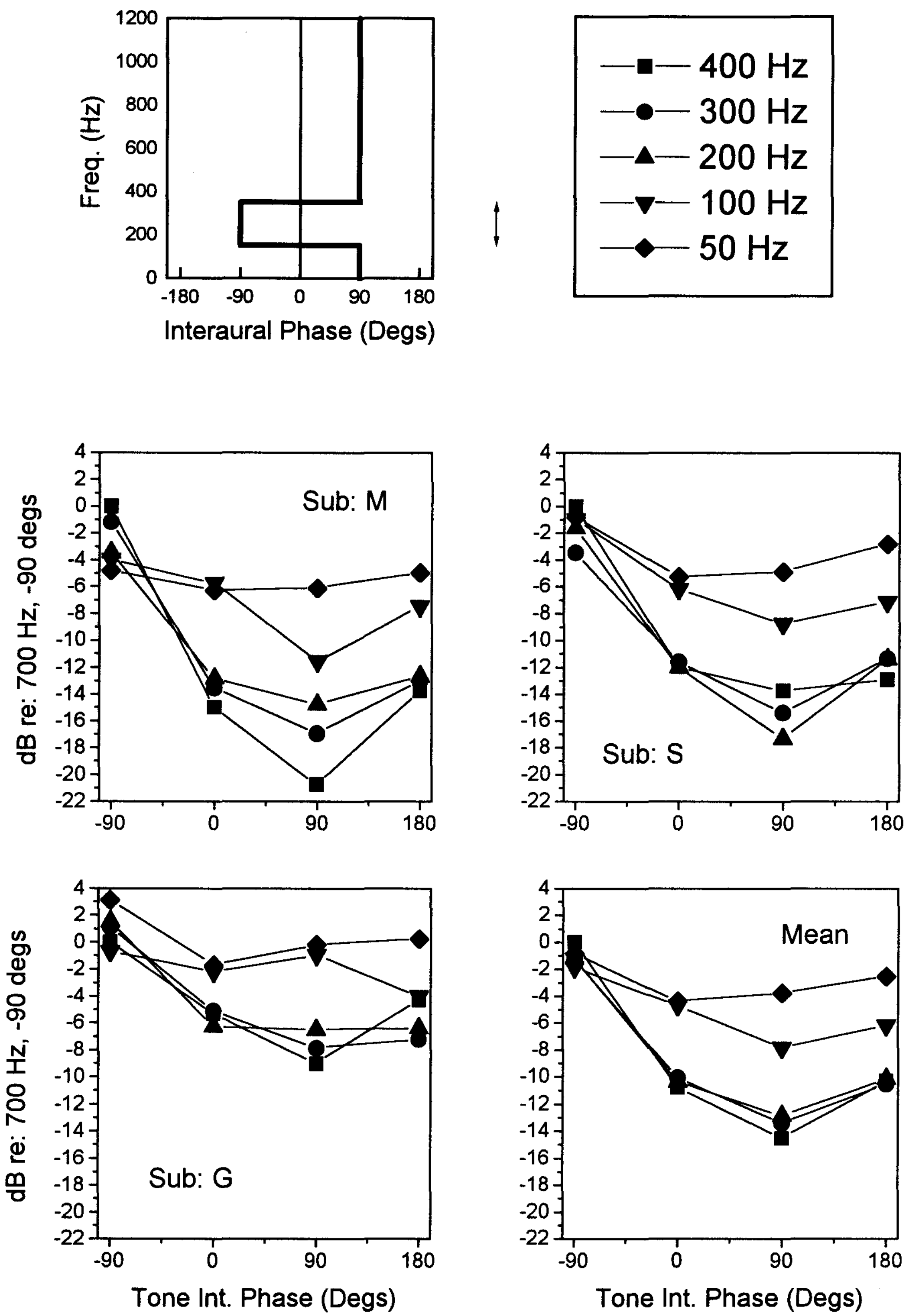

1200
1000
800
600
400
200
0
Freq. (Hz)
-180
-90
0
90
180
Interaural Phase (Degs)
700 Hz
500 Hz
300 Hz
200 Hz
100 Hz
Sub: M
Sub: S
Sub: G
Mean
dB re: 700 Hz, -90 degs
6
4
2
0
-2
-4
-6
-8
-10
-12
-14
-16
-90
0
90
180
Tone Int. Phase (Degs)

Figure 24

Same as Figure 23, except that the signal frequency and the CF of the -90° portion of the masking noise are both 250 Hz.

Figure 24


1200
1000
800
600
400
200
0
Freq. (Hz)
-180
-90
0
90
180
Interaural Phase (Degs)
400 Hz
300 Hz
200 Hz
100 Hz
50 Hz
Sub: M
Sub: S
Sub: G
Mean
dB re: 700 Hz, -90 degs
Tone Int. Phase (Degs)

conditions have equivalent effects for the conditions in which the noise passing through the filters is entirely N_{-90} or N_{90}; this is not true for conditions in which noise passing through the filters contains both interaural phases. However, since thresholds for these two conditions are approximately equal in all conditions, these differences will not be addressed). For both CF's, the MLD's are approximately 12 to 13 dB, which is consistent with previous estimates (Durlach and Colburn, 1978).

As the width of the inner band decreases, the MLD also decreases. However, since the MLD only reflects changes in the difference between $N_{90\backslash-90\backslash90}S_{-90}$ and $N_{90\backslash-90\backslash90}S_{90}$, and not changes in the absolute threshold of either of these conditions, it is helpful to consider each condition separately (since thresholds for intermediate signal positions of S_0 and S_{180} as a rule fall between thresholds for these two conditions, they will not be emphasized here). More specifically, the entire decrease in the MLD is due to an increase in thresholds when the signal is S_{90}. In other words, for S_{-90} signals, decreasing the width of the inner band had no effect on detection thresholds. Of course, if it were further narrowed beyond the limit in these conditions (to 0 Hz), thresholds would eventually drop 12-13 dB, since this is essentially the same condition as when the inner band was at its widest, with an S_{90} signal. This pattern is identical to that found by Sondhi and Guttman (1966).

Discussion

These data can be accounted for by the activity within a single auditory filter pair centered on the target (as described above, and in Kohlrausch, 1984). While this provides a straightforward explanation for this set of data, an explanation similar to that put forth to account for the data from Experiment II can be considered. The main point of such an argument is that a cross-spectral comparison must be made for perceptual segregation to occur. When the noise alone is presented to a listener, the amount of energy passing through each peripheral filter is equal (disregarding instantaneous differences). Therefore, there is no basis for segregation of the spectral region destined to contain the signal tone from the rest of the noise spectrum, and the perception is simply that of noise (although the interaural phase spectrum provides a basis for segregation of the inner and outer bands, and faint edge pitches have been reported at 180° phase transitions in noise; Hartmann, 1988). When a tonal target is added, however, the perception is that of a tone against a background of noise, which inherently involves a comparison (and segregation) of activity across different spectral regions. Once again, there is a question of over what spectral range such comparisons take place, and whether that range is determined by the configuration of the stimulus or by a fixed aspect of the auditory system (a binaural "filter").

Assuming that the listener assigns weights to spectral regions on the basis of interaural parameters (which determine the position of the cross-correlation peaks) during the forward fringe preceding the onset of the target, there are two interaural phase values at which maximum weight might be placed; at N_{-90} (the noise surrounding the target) or N_{90} (the outer bands of noise). Ideally, the maximum weights would be placed in such a manner that the addition of the signal (after the temporal fringe) would cause the greatest change in the summary cross-correlation function. In the case of $N_{90/-90/90}S_{90}$, of course, maximal weight should be given to those components with an interaural phase of -90° (for the noise used by Sondhi and Guttman, maximal weight should be given to components with an interaural phase of 0°). If maximum weights were placed on frequency regions having this interaural value during the fringe, then it would be expected that thresholds would increase gradually, as the width of the effective background was decreased (as suggested in the second experiment). This pattern is seen in the data, where thresholds increase from -13 dB (re: $N_{-90}S_{-90}$ condition) to 0 dB as the inner band is decreased in width from 700 to 100 Hz, for a CF of 500 Hz. Of course, if the weights were placed maximally on those frequency regions having a phase shift of -90°, and the signal was also -90°, then no binaural advantage is gained, and thresholds would be expected to remain relatively constant.

For S_{-90} signals, assuming the optimal binaural approach is to weight the spectral regions most different from the signal (in interaural phase) maximally, so that the addition of the signal causes the greatest relative change in the left and right signals, then it might be suggested that the greatest weight be placed on the spectral regions having an interaural phase of 90° (the outer bands). However, if during the fringe, the outer regions are heavily weighted (having an interaural phase of 90°), and the inner region is minimally weighted (having an interaural phase of -90°), then the onset of the signal will have a relatively small effect, since the spectral region it occupies has been "ignored". Alternatively, if the inner band was weighted heavily, detection would be based primarily on monaural cues. In either case, the bandwidth of the inner band should have little effect on detection thresholds. This is also reflected in the data, where $N_{90/-90/90}S_{-90}$ thresholds do not improve as the inner band becomes narrow.

This interpretation is also consistent with the data of Bernstein (1991), who measured MLD's for an 800 Hz tone against a background of continuous noise, in the presence of an interferer tone of 400 Hz, that was either I_o or I_{pi} and was either gated on and off with the target or was continuous. When the interferer was gated on and off with the target, more interference (higher thresholds for detecting the target) was caused by the I_{pi} interferer than by the diotic interferer. This would be expected, since the onset of a diotic interfering tone would not affect interaural correlation of the masking noise, while the dichotic interferer would. Further, if the dichotic interferer were presented continuously, less interference should be seen, since that spectral region will be assigned less weight. This is apparent in Bernstein's (1991) finding of approximately a 5-dB difference in thresholds between the two conditions.

CHAPTER 6

GENERAL DISCUSSION

The general findings of this dissertation can be summarized as follows:

Exp I: The detectability of an interaural phase shift of a narrow section of noise varies over time, due to the temporal variability in the cross-correlation function of the left and right peripheral filters centered on the phase shifted region, relative to the frequency regions immediately surrounding it. Therefore, the detectability of an interaural phase shift within any particular sample of noise will depend on the cross-correlation functions produced by that noise sample.

Computer Sim. I: A simulation using integration of cross-correlation values across a fixed frequency range did not account for the behavioral data of Grange (1992), where thresholds for detecting an interaural phase shift of a narrow band increased as uncorrelated noise was brought closer and closer to the target.

Exp II: The bandwidth of noise that determines thresholds for detecting the onset of an interaural phase shift within a narrow region of the noise is determined by the prior interaural configuration of the noise, not by a fixed "binaural critical band".

Exp III: A model based on these findings can also qualitatively account for data from MLD experiments initially assumed to suggest a wider binaural critical band, using cross-spectral comparisons, as opposed to a single filter explanation, which does not acknowledge the perceptual aspects of the task.

The data from Experiment I showed that performance of subjects required to detect an interaural phase shift within a 20 Hz band of noise range from chance to above 90% correct. The noise was diotic within a spectral region surrounding the target band, and uncorrelated outside of the diotic region. The width of the diotic region had no effect on the variability of performance across different stimulus samples. A visual inspection of cross-correlation coefficients at $\tau = 0\ \mu s$ suggested that performance was based on a decorrelation of the peripheral filters centered on the target frequency relative to the cross-correlation values of filters immediately surrounding the target.

The data from Experiment II and the computer simulation of data from Grange (1992) suggests a weighting mechanism that places greater emphasis on frequencies with common interaural parameters. Thresholds in Experiment II increased when extraneous interaural phase shifts occurred at frequencies initially identical in interaural phase to the target frequency. Thresholds were less affected when the frequencies containing the extraneous phase shifts initially differed in interaural phase from those components within the target band. This is consistent with the idea that less weight was being placed upon those frequency regions during the forward fringe. The results of the computer simulation are not inconsistent with this conclusion. A fixed-width binaural weighting function (binaural critical band) did not fit the behavioral data of Grange (1992) well. This might be a consequence of the fact that only a Gaussian weighting function was used in the simulation; however, it could also be the case that binaural processing is based on a bandwidth determined by the interaural characteristics of the noise.

Hall, Tyler, and Fernandes (1983) made a similar argument to account for differences found in binaural detection tasks using a band-narrowing vs. a notched-noise masker. Computation of filter widths based on thresholds from each condition gave disparate estimates of processing bandwidth, the band-narrowing conditions leading to much wider estimates (consistent with Bourbon and Jeffress, 1965). Hall et al. suggested that the number of peripheral auditory filters processed by the binaural auditory system is determined by the amount of energy present in each filter, and that as the width of the noise increases (in the band-narrowing conditions), the relative number of filters with large interaural differences (caused by the presence of the masker) being processed is reduced.

The general model put forth in the discussion of Experiment II can be summarized as follows: (1) frequency regions are weighted by the height of the peaks in the (f/τ) plane at a particular interaural delay. (2) Frequency regions are combined according to the weights assigned in this process to produce a summary representation of the (f/τ) plane. (3) After this combination, an interaural alteration will cause a change in this summary representation if the frequency region containing the interaural change was weighted heavily initially. Detectability of interaural changes, therefore, will depend on initial weighting of spectral regions. (4) Weights of spectral regions are updated over time, so that changes in the interaural parameters of off target frequencies can be "ignored" (weighted less heavily), with increasing duration of target presentation.

This model can account for binaural changes caused by the addition of an out-of-phase tonal signal as well. Experiment III was in part a replication of Sondhi and Guttman (1966), who measured thresholds for detecting a tone in a background noise that was diotic within some inner band, and 180° out-of-phase outside of this band. In the present experiment, only inner bands wider than estimates of peripheral bands were used, since the emphasis was on cross-spectral interactions. Further, the inner band and outer bands were -90° and 90°, respectively, to minimize the perceptual "diffuseness" of the outer band, often reported for N_π noise. Thresholds for detection of an S_{-90} signal did not change as a function of inner bandwidth, since binaural cues were not being used. For an S_{90} signal, however, the binaural advantage was lost as the inner bandwidth decreased from 700 Hz to 100 Hz, with a large increase in thresholds seen between 300 and 100 Hz (for a signal and inner band CF of 500 Hz). This is consistent with the estimates of Grange (1992), in which thresholds for detecting an interaural phase shift began increasing as uncorrelated noise began affecting filters within about 50 - 100 Hz of the target.

This model, and particularly the data from Exp. II, suggest that binaural cues can serve as the basis for auditory "grouping". It appears that the auditory system segregated (assigned weights to) spectral regions on the basis of interaural parameters of different frequencies. Having done this, interaural changes at spectral regions given low weight interfere less with detecting changes at a heavily weighted spectral region. Evidence from other studies that perceptual segregation can lead to a release from cross-spectral interference is sparse, however. Woods and Colburn (1992) and Stellmack and Dye (1993) found that when a target tone was perceptually segregated due to an onset asynchrony between it and distractor tones, there was essentially no release from the interference caused by the distractors (ITD thresholds did not decrease with perceptual segregation). Yost (1991a) has suggested that auditory grouping mechanisms are essentially monaural, so that the combination of individual frequency components into an object may occur prior to and independently of localization of that object. One obvious difference between Exp. II and the studies of Woods and Colburn (1992) and Stellmack and Dye (1993) is the spectral density of the stimuli. The auditory system may interpret 2 or 3 component complexes as belonging to the same auditory object, since real sound sources rarely consist of a single frequency. Therefore, though an onset asynchrony may cause a tone to "pop out," the separate mechanism determining lateralization may still be influenced by all of the components present.

REFERENCES

Bernstein, L. R. (1991). Spectral interference in a binaural detection task. J. Acoust. Soc. Am., 89, 1306-1313.

Bernstein, L. R., and Trahiotis, C. (1993). Spectral interference in a binaural detection task: Effects of masker bandwidth and temporal fringe. J. Acoust. Soc. Am., 94, 735-742.

Bilsen, F. A., and Raatgever, J. (1973). Spectral dominance in binaural lateralization. Acustica 28, 131-132.

Blauert, J. (1983). Spatial Hearing. Cambridge, Mass: MIT Press.

Bourbon, W. T., and Jeffress, L. A. (1965). Effect of bandwidth of masking noise on detection of homophasic and antiphasic tonal signals. J. Acoust. Soc. Am., 37, 1180.

Bregman, A. S. (1990). Auditory Scene Analysis: The Perceptual Organization of Sound. Cambridge, Mass: MIT Press.

Colburn, H. S. (1977). Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise. J. Acoust. Soc. Am., 61 (2), 525-533.

Colburn, H. S., and Durlach, N. I. (1978). Models of Binaural Interaction. In E. C. Carterette and M. P. Friedman (Eds.), Handbook of Perception IV: Hearing. New York: Academic.

Cramer, E. M., and Huggins, W. H. (1958). Creation of pitch through binaural interaction. J. Acoust. Soc. Am., 30, 412-417.

Durlach, N. I. (1972). Binaural Signal Detection: Equalization and Cancellation Theory. In J. V. Tobias (Ed.), Foundations of Modern Auditory Theory: Vol. II. New York: Academic.

Durlach, N. I., and Colburn, H. S. (1978). Binaural Phenomena. In E. C. Carterette and M. P. Friedman (Eds.), Handbook of Perception IV: Hearing. New York: Academic.

Dye, R. H. (1990). The combination of interaural information across frequencies: Lateralization on the basis of interaural delay. J. Acoust. Soc. Am., 88, 2159-2170.

Gabriel, K. J., and Colburn, H. S. (1981). Interaural correlation discrimination: I. Bandwidth and level dependence. J. Acoust. Soc. Am., 69, 1394-1401.

Gilkey, R. H., B. D. Simpson, and J. M. Weisenberger (1990). Masker fringe and binaural detection. J. Acoust. Soc. Am., 88(3), 1323-1332.

Grange, A. N. (1992). The effects of the interaural parameters of the background noise on dichotic pitch detection. Unpublished master's thesis, Loyola University of Chicago, Chicago, Il.

Green, D. M., and Yost, W. A. (1975). Binaural Analysis. In W. D. Keidel and W. D. Neff (Eds.), Handbook of Sensory Physiology: Hearing. Berlin: Springer-Verlag Berlin.

Hafter, E. R., and Carrier, S. C. (1970). Masking-level Differences Obtained with a Pulsed Tonal Masker. J. Acoust. Soc. Am., 47 (2), 1041-1047.

Hartmann, W. M. (1988). Pitch Perception and the Segregation and Integration of Auditory Entities. In G. M. Edelman, W. E. Gall, and W. M. Cowan (Eds.), Auditory Function: Neurobiological Bases of Hearing. New York: Wiley & Sons.

Jeffress, L. A. (1972). Binaural Signal Detection: Vector Theory. In J. V. Tobias (Ed.), Foundations of Modern Auditory Theory: Vol. II. New York: Academic.

Jeffress, L. A. (1948). A place theory of sound localization. J. Comp. Physiol. Psychol. 41, 35-39.

Kohlrausch, A. (1988). Auditory filter shape derived from binaural masking experiments. J. Acoust. Soc. Am., 84, 573-583.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. J. Acoust. Soc. Am., 49, 467-477.

McFadden, D. and E. G. Pasanen (1976). Lateralization at high frequencies based on interaural time differences. J. Acoust. Soc. Am., 59, 634-639.

Moore, B. C. J. (1988). Dynamic Aspects of Masking. In G. M. Edelman, W. E. Gall, and W. M. Cowan (Eds.), Auditory Function: Neurobiological Bases of Hearing. New York: Wiley & Sons.

Patterson, R. D., and Nimmi-Smith, I. (1987). An efficient auditory filterbank based on the gammatone function. Paper presented at a meeting of the IOC Speech Group on auditory modelling at RSRE.

Raatgever, J. and R. A. Bilsen (1986). A central spectrum theory of binaural processing. Evidence from dichotic pitch. J. Acoust. Soc. Am., 80(2), 429-441.

Sayers, B. (1964). Acoustic-image lateralization judgements with binaural tones. J. Acoust. Soc. Am., 36, 923.

Shackleton, T. M., R. Meddis and M. J. Hewitt (1992). Across frequency integration in a model of lateralization. J. Acoust. Soc. Am., 91, 2276-2279.

Sondhi, M. M., and N. Guttman (1966). Width of the spectrum effective in the binaural release of masking. J. Acoust. Soc. Am., 40(3), 600-606.

Stellmack, M. A. and R. H. Dye (1993). The combination of interaural information across frequencies: The effects of number and spacing of components, onset asynchrony, and harmonicity. J. Acoust. Soc. Am., 93, 2933-2947.

Stern, R. M., and H. S. Colburn (1978). Theory of binaural interaction based on auditory-nerve data. J. Acoust. Soc. Am., 64(1), 127-140.

Stern, R. M., A. S. Zeiberg, and C. Trahiotis (1988). Lateralization of complex binaural stimuli: A weighted-image model. J. Acoust. Soc. Am., 84, 156-165.

Trahiotis, C. and Bernstein, L. R. (1990). Detectability of interaural delays over select spectral regions: Effects of flanking noise. J. Acoust. Soc. Am., 87, 810-813.

Woods, W. S. and H. S. Colburn (1992). Test of a model of auditory object formation using intensity and interaural time difference discrimination. J. Acoust. Soc. Am., 91, 2894-2902.

Yost, W. A. (1991a). Auditory image perception and analysis: The basis for hearing. Hear. Res., 56, 8-18.

Yost, W. A. (1991b). Thresholds for segregating a narrow-band from a broadband noise based on interaural phase and level differences. J. Acoust. Soc. Am., 89 838-844.

Yost, W. A. (1985). Prior stimulation and the masking-level difference. J. Acoust. Soc. Am., 78, 901-907.

Yost, W. A. (1981). Lateral position of sinusoids presented with interaural intensive and temporal differences. J. Acoust. Soc. Am., 70(2), 397-409.

Yost, W. A. and E. R. Hafter (1987). Lateralization. In W. A. Yost & G. Gourevitch (Eds.), Directional Hearing. New York: Springer-Verlag.

Yost, W. A., and S. Sheft (1989). Across-critical-band processing of amplitude-modulated tones. J. Acoust. Soc. Am., 85, 848-857.

VITA

Anthony N. Grange received a Bachelor of Science degree in psychology from Idaho State University in Dec., 1989, and a Master's degree in Psychology from Loyola University of Chicago in May, 1992. He held a graduate assistantship in the Psychology department at Loyola from August, 1990 through May, 1993. He was also a recipient of the Augmentation Award for Science and Engineering Research Training (Air Force Office of Scientific Research) between September, 1993 and August, 1994.

Anthony has been a member of the Acoustical Society of America since 1990. Presentations to the society include "Detection of interaural decorrelation in multitone complexes" (Stellmack, Grange, and Aguilar, 1990; JASA 88, S96-S97), "The effects of randomizing interaural phases of components spectrally distant from the target on detection of a dichotic pitch" (Grange and Yost, 1992; JASA 91, 2414), "Effects of different samples of reproducible noise on detection of a 20-Hz wide interaural phase shift centered on 500 Hz" (Grange and Yost, 1993; JASA 93, 2350), and "Effects of interaural configuration of wideband forward and backward fringes on detection of an interaural phase shift of a narrow portion of wideband noise" (Grange and Dye, 1994; JASA 95, 2916-2917).

DISSERTATION APPROVAL SHEET

The dissertation submitted by Anthony N. Grange has been read and approved by the following committee:

Raymond H. Dye, Ph.D., Director
Associate Professor, Psychology
Loyola University Chicago

William A. Yost, Ph.D.
Director and Professor of Hearing Sciences
Loyola University Chicago

Patricia Tenpenny, Ph.D.
Assistant Professor, Psychology
Loyola University Chicago

Stanley E. Sheft, Ph.D.
Research Assistant Professor
Parmly Hearing Institute
Loyola University Chicago

The final copies have been examined by the director of the dissertation and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the dissertation is now given final approval by the committee with reference to content and form.

The dissertation is, therefore, accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

March 3, 1995 — Date

Raymond H. Dye, Jr. — Director's Signature