THE EFFECT OF SPATIAL STIMULUS ATTRIBUTES ON AUDITORY TEMPORAL PROCESSING TASKS

by

Susan E. Boehnke

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia

February 12, 2003

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

0-612-79401-6

Canada

DALHOUSIE UNIVERSITY

FACULTY OF GRADUATE STUDIES

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled “The Effect of Spatial Stimulus Attributes on Auditory Temporal Processing Tasks” by Susan Boehnke in partial fulfillment for the degree of Doctor of Philosophy.

Dated: March 14,2003

External Examiner:

Research Supervisor:

Examining Committee:

[redacted]

DALHOUSIE UNIVERSITY

DATE: March 14, 2003

AUTHOR: *Susan E. Boehnke*

TITLE: *The effect of spatial stimulus attributes on auditory temporal processing tasks*

DEPARTMENT OR SCHOOL: *Psychology/Neuroscience*

DEGREE: *Ph.D.* CONVOCATION: *May* YEAR: *2003*

[redacted]

Signature of Author

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Two sounds will be given different location percepts if they are presented to different ears (EAR), if they have different interaural level differences (ILD) or different interaural time differences (ITD). If the sound sources are located in different lateral hemifields in free-field space, they will have different ITDs (coded in low-frequency channels), and frequency-dependent ILDs imposed by the head-related transfer function. The role of such spatial stimulus attributes in the segregation of spectrally identical sound sources was evaluated by exploiting the fact that temporal judgments made between sounds perceived as emanating from different sources are impaired relative to those made for single sources. Auditory temporal gap detection and sequential stream segregation paradigms were used to provide measures of this impairment.

Across 5 experiments, the ability to detect a silent interval (gap) between two noise bursts (markers) that differed only in spatial stimulus attributes was determined under binaural and monaural listening, in the free-field and over headphones. Gap detection thresholds were most elevated for markers presented to different ears, moderately and equally elevated for markers presented with different ILDs or monaurally at different levels, but not elevated if markers were presented with different ITDs. Consistent with the foregoing, free-field separation elevated thresholds for wideband and high-pass markers, for which large marker level differences were present at each ear, but not for low-pass markers which differed only in ITD. In general, gap detection performance for spatially separated markers in the free field decreased as a function of marker differences in intensive and spectral characteristics at each ear. The pattern of impairment of gap thresholds with marker spatial separation was highly correlated with the pattern of improvement in signal detection with spatial separation from a masking noise (Saberi et al, 1991; Gilkey and Good, 1995). Sequential stream segregation of repeating ABA_ sequences, where A and B were noise bursts differing only in spatial stimulus attributes, followed a similar pattern of sensitivity. Separate A and B perceptual streams were readily formed in the EAR or ILD conditions, but not in the ITD condition, for which segregation required more time to build up, and was perceptually less stable. Detection of a temporal asymmetry between A and B streams, a task reported to be sensitive to stream formation, was strongly affected by EAR of entry differences, moderately affected by ILD or monaural level differences, but not affected by ITD differences between A and B. In general, the dominant effect of spatial separation on segregation of spectrally similar sources appears to be the differential activation of the frequency channels of the two ears by each source, rather than activation of different binaural channels.

Finally, to examine temporal processing exclusively within the binaural system, a 'binaural analog' of gap detection was examined. Across a variety of interaural correlation configurations (using positively-, negatively- and uncorrelated noisebursts), thresholds for detection of a rapid temporal change, or a static change (the just noticeable difference) in interaural correlation were obtained. Detection was better for changes from positive interaural configurations than negative, mirroring the differences in the amount of binaural unmasking for the same configurations. The data provided a new measurement of the temporal integration that smears the representation of stimulus changes over time.

LIST OF ABBREVIATIONS AND SYMBOLS USED

°	Degrees; in this thesis usually refers to positions on the azimuthal plane (front of listener= 0°; back of listener=180°; lateral poles= 90° and -90°)
ρ	Rho – symbol used to denote interaural correlation in this thesis
ρ_w	Simple model of binaural temporal integration which applies the temporal integrator directly on the interaural correlation of the stimulus
$\Delta\rho$	A given change in interaural correlation, in Chapter 4 represents the JND for a change in interaural correlation
π	Pi radians; in this thesis used to denote the interaural relation of a dichotic noise. The signal at one ear is presented inverted (180° or π-radians out of phase) at the other ear
μpa	microPascals; measure of pressure used for indicating the physical intensity of a sound
AFC	Alternative forced-choice; psychophysics method in which an observer is presented with x number of alteratives and forced to choose one
A1	Primary auditory cortex, area 1
AN	Auditory nerve
ANOVA	Analysis of variance; statistics
AVCN	Anterior ventral cochlear nucleus; anatomy
BF	Best frequency; the frequency at which an auditory neuron fires maximally
BI	Binaural; presentation of a stimulus to both ears
BILD	Binaural intelligibility level difference; benefit in speech intelligibility due solely to binaural interaction processes, measured in decibels
BM	Basilar membrane; membrane in cochlea which vibrates in response to sound and provides the mechanical bases for transduction by hair cells; anatomy
BMLD	Binaural masking level difference; benefit in signal detection due solely to binaural interaction processes, measured in decibels

CF	Characteristic Frequency; the frequency for which a neuron will respond using the lowest stimulus intensity
CN	Cochlear nucleus; first acending synapse of the auditory nerve; anatomy
DCN	Dorsal Cochlear nucleus; anatomy
DNLL	Dorsal nucleus of the lateral lemniscus; anatomy
ERB	Equivalent rectangular bandwidth; measure of the bandwidth of the auditory frequency filter
F	Fisher's 'F'; statistic used to test signficance of ANOVA
FF	Free-field; sound presentation in a space with no reflecting surfaces
dB	Decibels; standardized measure of intensity
fMRI	Functional magnetic resonance imaging
Hz	Hertz; measure of the frequency of a sine wave (1 Hz= 1 cycle/second)
HRTF	Head-related transfer function; difference in spectrum of a sound source as measured at the source, and in the ear canal.
IATD	Interaural arrival time difference; difference in the onset of a sound at the two ears as a function of its location
IC	Inferior colliculus; anatomy
ICCI	Interaural correlation change interval
ILD	Interaural level difference; the difference in the intensity (or level) of a sound at the two ears the occurs due to shadowing by the head as a sound source is moved laterally
IPD	Interaural phase difference; the difference in the phase (given in proportion of a cycle, fractions of π, or degrees(max 360)) of a given frequency component at each ear that results from the IATD
ITD	Interaural time difference; global term for IATD and resultant IPD, indicated usually in microseconds
JND	Just noticeable difference; the difference limen for a sensation)
LED	Light-emitting diode
LNTB	Lateral nucleus of the trapezoid body; anatomy
LQ	'Loud-Quiet'; in this thesis this refers to a gap detection stimulus in which the leading marker is relatively louder than the trailing marker

LSO	Lateral superior olive; first brainstem site coding ILDs; anatomy
M	Monaural; presentation of the sound stimulus to a single ear
MATSS	Multi-channel Apparatus for the Temporal Sequencing of Stimuli
MGB	Medial geniculate body of the thalamus; anatomy
MNTB	Medial nucleus of the trapezoid body; anatomy
MSO	Medial superior olive; first brainstem site coding ITDs; anatomy
N	Noise
Nu	Dichotic noise that is interaurally uncorrelated
No	Diotic noise -identical in-phase noise presented to each ear
Nπ	Dichotic noise presented interaurally out of phase by π radians (180°)
NRR	Noise reduction rating
p	probability; statistics term which indicates probability difference could be due to chance
PB	Predominantly-binaural
QL	'Quiet-Loud'; in this thesis this refers to a gap detection stimulus in which the trailing marker is relatively louder than the leading marker
RL	Right-Left; in this thesis, refers to a gap detection stimulus in which the leading and trailing markers are presented from a position to the right and left of the listener's head respectively.
RR	Right-Right; both markers presented from the same position to the right
S	Signal
So	Diotic signal – identical sound presented in phase to each ear
Sπ	Dichotic signal presented 180° (π radians) out of phase to the ears
SOA	Stimulus onset asynchrony; the time between the onsets of two stimuli
SOC	Superior olivary complex; includes MSO and LSO; anatomy
SEM	Standard error of the mean; statistics
SNR	Signal to noise ratio
SPL	Sound pressure level; standardized measure for sound intensity
TB	Trapezoid body; anatomy
VCN	Ventral division of the cochlear nucleus; anatomy
VNLL	Ventral nucleus of the lateral lemniscus; anatomy

ACKNOWLEDGEMENTS

This research was generously supported by a doctoral scholarship from the Natural Science and Engineering Research Council and a fellowship from the I.W. Killam Foundation. Great thanks are due to my supervisor Dennis Phillips for his endless patience, encouragement, and grace as I carved many of my own paths, most to dead ends, but a few to new ideas. Nothing is more rewarding than when a big project ends with a wonderfully fun event, as was my thesis defense – thanks to Bruce, Dennis, Ray, Donald, Kazue and Richard for making it so. Thanks to Susan Hall for lab assistance, being a sounding board for ideas, and for being a great friend. And, to David Shore for keeping me excited about science and for saving my thesis (and my sanity) one summer night on the back porch. Thanks also to David McAlpine and especially Torsten Marquardt for teaching me the real story behind ITD coding as it unfolded in their heads. And, to Tino Trahiotis for presenting me with the puzzle that the 'percept' itself may be irrelevant to psychophysical studies of auditory perception. Also my gratitude to all the listeners, particularly SH, DA, JS, JH and BVM for the long hours. Finally, and most importantly, my deepest appreciation to all my family for continued support on all fronts – emotional, financial and intellectual.

PREFACE

GENERAL INTRODUCTION AND ORGANIZATION OF CHAPTERS

One of the main problems faced by the human auditory system is that of source segregation. In a complex environment with multiple sound sources, the auditory system is confronted by a single complex waveform (the proximal stimulus) which contains the summed signals from each source (the distal stimuli). In order to attend selectively to a given source and follow it over time (e.g. listen to one person's voice in a cocktail party), the auditory system must dissect the physical waveform arriving at the ears – the proximal stimulus - and construct perceptual representations of the multiple 'distal' sources that gave rise to it. This is referred to as the source-segregation problem, and determination of the stimulus differences that lead to perceptual segregation of different sound events is important to understanding how the auditory system solves it (Bregman, 1990). This process involves separation of sounds that are temporally overlapping (simultaneous grouping), temporally interleaved (sequential grouping), or both. In this thesis, I am primarily interested in the sequential case. In Bregman's terminology, in order to be perceptually segregated, the proximal sound components perceived as emanating from a single distal source must be grouped into a single perceptual stream. The classic paradigm to demonstrate sequential stream segregation involves temporally alternating two sounds (A and B) that differ on some dimension, typically frequency or pitch, in a sequence ABABABABA or ABA_ABA_ABA_. If the sounds are perceptually similar, they are perceived as a single rhythm – like a trill in the former sequence, or a galloping rhythm in the latter (do-rae/do-rae/do-rae vs. do-rae-do / do-rae-do). If they perceptually differ beyond some critical value, the sequence will split into two separate streams, each following their own rhythm. That is, the A and B sounds will segregate, as if they were pulses perceived as coming from different sources. Which stimulus dimensions are relevant to the formation of separate streams? The dominant stimulus dimension by which sounds sources are segregated into different auditory streams is frequency, and this is the dimension that has been most studied (Bregman, 1990).

Frequency is an obvious dimension by which sources might be segregated, because it is the dimension by which the ear initially filters the waveform. However, sources in the environment often have similar spectra, and so other dimensions could, in principle, be used for their segregation. A candidate cue to distinguish these events is the relative spatial locations of the sound sources. Surprisingly, this dimension has received relatively less study in this context (cf.Cherry, 1953; Cherry & Taylor, 1954; Roman *et al.*, 2002).

An interesting aspect of the perceptual organization of temporally interleaved sounds is that temporal judgments are impaired when they must be made between sounds which are perceived to emanate from different distal sources (Bregman, 1990, though note that 'sources' here refers to the objects creating the sound, not their respective locations which could be identical). Thus, the degradation in performance on such temporal judgments between two stimuli that vary on some dimension might be used as a marker by which to determine the relevance of that dimension in source segregation. Do spectrally similar sounds that differ only in spatial location lead to such segregation and the resultant degradation in acuity of temporal judgments? Which of the available cues to sound location are important in this regard? These are the general questions of this thesis.

The impairment of temporal judgments made between dissimilar sounds has been extensively studied in the context of auditory temporal gap detection. As a task, gap detection has not been explicitly evaluated as a probe of the dimensions used to segregate sounds. However, in form, it mimics the first transition between an A and B sound in a sequential stream segregation experiment. Furthermore, there is degradation in acuity for detecting a gap when markers are made dissimilar in frequency or are presented to different ears (Kinney, 1961; Perrott & Williams, 1971; Williams & Perrott, 1972; Collyer, 1974; Divenyi & Hirsh, 1974; Fitzgibbons *et al.*, 1974; Williams & Elfner, 1976; Divenyi & Danner, 1977; Penner, 1977; Williams *et al.*, 1978; Formby & Forrest, 1991; Formby *et al.*, 1996a; Hall *et al.*, 1996; Phillips *et al.*, 1997; Formby *et al.*, 1998a; Formby *et al.*, 1998b; Taylor *et al.*, 1999; Phillips & Hall, 2000; Grose *et al.*, 2001a; Grose *et al.*, 2001b; Lister *et al.*, 2002). These are stimulus dimensions known to result in sequential stream segregation, and the resultant loss of the ability to judge temporal relations between items in different streams (van Noorden, 1975, 1977; Bregman, 1990). Thus, gap detection may represent a simple case to examine stimulus differences leading to segregation. In gap detection, the reduction in acuity for detection of a gap with separation of gap markers in some perceptual dimension has been ascribed to activation of different 'channels' coding for that stimulus dimension. In the case of separation of gap markers in frequency or by presentation to different ears, the channels are fairly easy to conceive of – they ultimately have their genesis in the differential filtering of the markers at the auditory periphery. The task cannot be performed as a simple discontinuity detection within a given channel, as is the case with gap detection with identical markers. It has been proposed that gap detection with dissimilar markers may be based on a relative timing of the offset of activity in one channel and the onset of activity in another – a central mechanism (Phillips *et al.*, 1997; Phillips, 1999). However, gap detection thresholds are also elevated when sounds are located at different spatial locations (Phillips *et al.*, 1998). In this case, both of the markers activate both ears, and are differentiated by the way they activate each ear. A detailed study of the pattern of sensitivity of gap detection thresholds as a function of marker separation in azimuth was recently conducted (Boehnke, 1998; Boehnke & Phillips, 1999) and the results indicated that markers separated by a fixed distance within a hemifield (although discriminable on the bases of their location) did not elevate thresholds, while markers separated across the midline by that distance did. The conclusion from this study is that auditory space was coded as two hemifield-tuned spatial channels, which had skirts overlapping in the midline region. Though not tested directly, it was assumed that the tuning reflected binaural coding, and that the relative activation of the two channels might subserve sound localization, an inherently binaural process. However, the channel tuning was generated using broadband noise markers, and it is unknown if the effect of spatial separation on gap detection thresholds generalizes to markers comprised of only low or high spectral content. This is important, because spatial separation of two sounds leads to differences between those sounds at each of the ears (monaural differences), and as a function of the interaction of those two ears in the nervous system (binaural cue differences). Any of these monaural and binaural factors might contribute to the elevation in gap detection thresholds, and each has a different frequency dependence. For example, would the same pattern of gap detection thresholds with marker spatial separation be obtained if the listener only had access to one ear, or if the markers were comprised of only low, or high frequencies? Thus, in order for the two-channel model to be anything more than a description of the data obtained with wideband noise, it is necessary to determine which of the spatial stimulus characteristics that differ between the spatially separated broadband noise markers actually led to the increase in gap thresholds. That is, what are the salient stimulus differences that result from spatial separation of two sounds and lead to elevated thresholds?

The goal of the first set of experiments in this thesis was to dissect the 'channels' for auditory spatial location described by Boehnke and Phillips (1999). This was accomplished across five experiments which examined sensitivity of gap detection thresholds to spatial separation of the markers based on ear of entry, interaural time differences (ITDs), interaural level differences (ILDs), and monaural spectral-level differences that occur as a function of spatial separation over headphones and in the free-field. The second goal was to evaluate whether the pattern of sensitivity to these different spatial stimulus characteristics observed in gap detection was related to the ability with which the auditory system is able to use those cues to segregate spectrally similar sound sources. Specifically, it was hypothesized that spatial stimulus differences between markers that led to impaired gap detection would be those which would lead to the greatest segregation of sound sources in sequential stream segregation tasks. This is important because it is one of the first studies to *directly* link the emerging literature on gap detection between similar and dissimilar sounds to sequential stream segregation – a task for which a growing literature has been emerging in parallel.

The organization of the thesis is as follows. Chapter 1 will outline the filtering of a sound source at the two ears, and the neural coding of sound as a function of frequency and spatial location. The perception of one and two sounds as a function of their spatial and temporal relations will also be described. Chapter 2 will begin with a review of within- and between-channel gap detection, and continue with a description of five gap detection experiments that were performed in an attempt to deconstruct the sensitivity of gap detection to the spatial separation of wideband gap markers. Chapter 3 will begin with a review of sequential stream segregation, followed by a description of three sequential stream segregation experiments which probe the sensitivity of the stream segregation mechanism to spatial differences between noise bursts. One of the main results emerging from Chapters 2 and 3 was the relative ineffectiveness of the interaural time difference cue in source segregation, despite the fact that it is the dominant cue by which we localize sound. In an attempt to further understand the nature of temporal processing within that binaural channel, Chapter 4 reports a set of studies examining sensitivity to temporal changes in strictly the interaural characteristics of an ongoing noise (no monaural disparities). Finally, in Chapter 5, a general discussion of the relevance of these studies for processing of sound sources from different spatial locations is discussed, particularly with reference to their implications for understanding which aspects of stimulus differences arising from spatial location differences are used to solve the source segregation problem.

CHAPTER 1

INTRODUCTION TO AUDITORY PROCESSING OF SPACE

In order to understand how location differences between two sounds can be used to aid in their perceptual segregation, it is first necessary to understand the way in which the sound waveform from a source is transformed *en route* to its final perceptual elaboration. How is the waveform emanating from that object transformed from a pattern of pressure waves in the air to some neural representation that presumably underlies our perception of it? Note that the activation of the sensory epithelium (the hair cells lining the basilar membrane of the cochlea) of each ear by the proximal stimulus provides no direct representation of 'location'. Information regarding the sound objects location must be computed by the central auditory nervous system, akin to visual depth being centrally computed by binocular interactions. Ultimately, central auditory processing constructs a *representation* of a *sound object* which has attributes such as pitch, loudness, duration, and location, and as will be described, this is usually done within the constraints of the tonotopic map that emerges at transduction.

THE PHYSICAL TRANSFORMATION OF A SOUND WAVEFORM ARRIVING AT THE TWO EARS AS A FUNCTION OF SPATIAL LOCATION

Fourier's transform states that every complex sound can be characterized by the relative power of each sine wave component of which it is comprised. Since the wavelength of a given component decreases with increasing frequency, the size of an object required to diffract that component, and thus reduce its intensity in the direction of travel, decreases with frequency. That is, high frequency sine wave components are more diffracted by smaller objects than low frequency components, resulting in a greater loss of energy. In hearing with two ears, this physical reality expresses itself in a pattern of interaural intensity differences that varies across frequency as a function of the filtering that occurs by the objects encountered between the two ears – i.e. the head. This is generally referred to as the Head-Related Transfer Function (HRTF), although the transfer function is also affected by the body and particularly the shape of the pinnae (Blauert, 1997). This HRTF is measured by recording a wideband noise presented at different spatial locations through a small microphone placed in each ear canal. The recorded waveform shows an interaural intensity (or level) difference (ILD) favoring the ear nearest the sound source. Furthermore, a Fourier transformation of that waveform shows that this intensity difference generally increases with the stimulus frequency. This is visualized in Figure 1.1. The top panels show the "time waveforms" (amplitude vs time plots) of the left (blue) and right (red) ear canal recordings of a noise presented from a series of azimuthal locations. These recordings were made in the room used for many of the experiments in this thesis. The amplitude difference between the red and blue plots is the mean ILD collapsed across frequency. For the lower panels, the same recorded waverforms were subjected to Fourier analysis in order produce a plot of the sound spectrum (frequency vs. intensity, collapsed across time). Note that there is little to no ILD for frequency components below 1000 Hz, and a generally increasing ILD with increasing frequency. This function relating ILD to frequency is non-linear, however, and for a given location there are certain frequency ranges for which the head and pinnae are particularly diffractive, or are particularly good filters, as can be seen in the various peaks and notches in the spectrum across the higher frequencies.

Some of the spectral characteristics observed in the lower plots of Fig. 1.1 change with the sound source location. Of particular note is a large ILD, or 'notch' in a band of frequencies for a given sound source location. The band at which this notch occurs shifts dramatically between about 5-10 kHz as a function of the location of the sound (Middlebrooks *et al.*, 1989; Wightman & Kistler, 1989b; Wightman & Kistler, 1989a; Kistler & Wightman, 1992; Carlile & Pralong, 1994; Blauert, 1997). This notch may provide a potential monaural cue to the location of the sound (Neti *et al.*, 1992) in additional to providing a binaural cue (ILD). However, studies which artificially modify the values of these cues have demonstrated that these monaural spectral cues are of little use for localizing in azimuth, but are crucial for localizing in elevation (Wightman & Kistler, 1997a). Note that the variability in the amplitudes of low frequency components likely reflects that the room was not anechoic below 500 Hz or so, so those points reflect the summed amplitude of both the direct waveform and its echos.

A fine-grained examination of the waveforms recorded at each ear will also reveal that the sound arrives at the near ear earlier than it arrives at the far ear – that is, there is

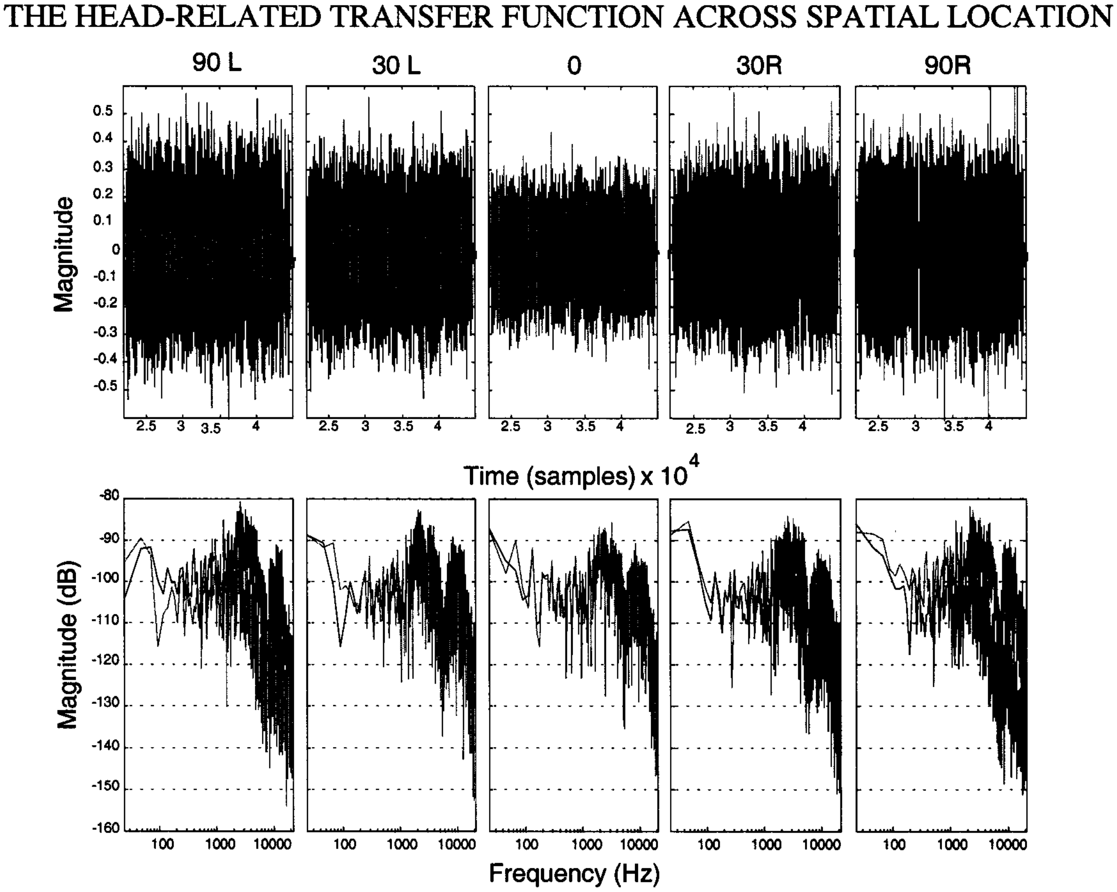


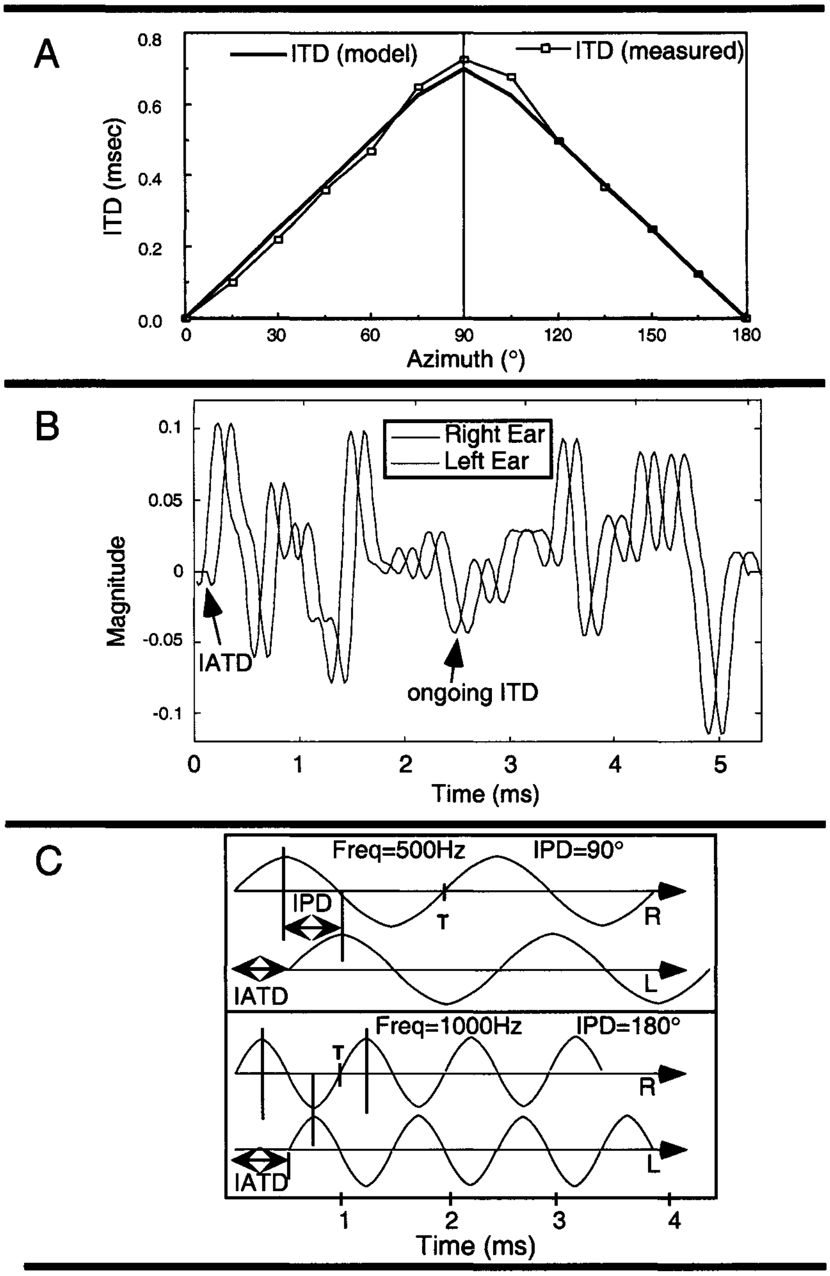
Figure 1.1

Recordings from the entrance of each ear canal were made in response to a broadband noise burst at 5 azimuthal locations. The time course (top) and spectrum (botttom) of the left (blue) and right (red) ear recordings are plotted. Note that there is a level difference between the ears (when summed across all frequencies) that favours the nearest ear (top plot), however, examination of the spectra indicates that the level difference occurs predominantly in the high frequency components. The stimulus presented (randomly generated broadband noise) is spectrally flat (would be a straight line on the spectrum plot). The pattern observed in these plots reflects that flat spectrum multiplied by the head-related transfer function.

an interaural time difference (ITD). This ITD in humans reaches a maximum at about ~600 μs for locations near the lateral poles. The ITD can be seen in the delay to the onset of the sound at the far ear (interaural arrival time difference, IATD, see Fig.1.2A), and also in a comparison of the ongoing phase of the waveforms at the two ears (Fig.1.2B). This interaural phase difference (IPD), typically expressed in cycles or degrees, is determined by the IATD and the frequency of the sound component (see Fig.1.2C). For a given IATD, a low frequency sound will have a smaller ongoing IPD than a higher frequency sound. In other words, the lower the frequency, the longer the IATD must be to generate a given IPD. While there is no ILD for the low frequency components, the IATD (and IPD) are available as a cue to sound location at all frequencies. However, the IPD is mostly useful for low frequency sound components. The ITD is zero for sounds located on the midsaggital plane (0° azimuth), increases with azimuthal location to 90° and then decreases with increases in azimuth until the source reaches the midsaggital plane at the back (See Fig. 1.2A, Woodworth, 1938; Fedderson *et al.*, 1957). In humans, this shift in ITD with source azimuth towards the lateral pole appears fairly linear and is well accounted for by a spherical model that takes into account the path length difference between the ears divided by the speed of sound (see Fig. 1.2A, adapted from Wightman & Kistler, 1993). There is evidence, however, of a frequency dependence of the ITD, in both humans (Kuhn, 1977) and cats (Roth *et al.*, 1980). For low frequencies, ITD varies inversely with frequency and this relation is not predicted by the spherical model. This means that the function of ITD across eccentricity of the sound source is multiply determined. In general, since ITD varies somewhat with both frequency and azimuth, the ITD cue in itself does not uniquely specify the sound source azimuth. ITD constancy across frequency, however, turns out to be perceptually irrelevant, as listeners judgments of spatial location are the same under natural ITD conditions (where the actual ITD varies somewhat with frequency) and conditions where the given ITD is made artifically constant for all frequencies (Kistler & Wightman, 1992).

This pattern of interaural differences as a function of spatial location is also important when considering the signal available *at each ear* in an environment with more than one source. Consider the case in which a signal (i.e. a talker) is on the left and a

THE INTERAURAL TIME (ITD) AND PHASE (IPD) DIFFERENCE



<u>Figure 1.2</u>

A. The size of the average interaural time difference across frequency as a function of azimuthal location. Measured values are similar to that obtained from a spherical model of the head (adapted from Wightman and Kistler, 1993).

B. About 5ms of a waveform presented with an interaural delay favouring the left ear. Note the onset time disparity (IATD) and the ongoing phase difference in the waveform at the two ears (IPD).

C. For a given IATD (~500μs), the ongoing IPD is twice as large for a 1000 Hz tone as it is for a 500 Hz tone, because the period of a single cycle (T(ms)) is half the duration for a higher frequency tone.

noise source is on the right. Higher frequency components of the speech will be dominant at the left ear and attenuated at the right ear, while the reverse will be true for the noise. The low frequency components for both sources will be roughly equal at the two ears. Thus, the high frequency components of the speech have an excellent signal-to-noise ratio (SNR) at the left ear, but a poor SNR at the right ear. Simply at the level of the combined waveform reaching the two ears, the benefit of spatial separation of sources in the free field is obvious. In addition to the different SNR at the two ears, each source activates different central neurons that are sensitive to interaural disparities (ILD or ITD). Thus, there are both peripheral and central differences in their representation. Although the lowest frequency sounds activate each ear at roughly the same level, they become segregated centrally via their differential activation of central neurons sensitive to ITDs.

THE NEURAL REPRESENTATION OF A STIMULUS: SEGREGATION OF SOUND FREQUENCY AT EACH COCHLEA

As described in the foregoing, the physical transformation of a sound source's spectrum as a function of spatial location both limits the information about a given source available for processing (by altering the frequency spectrum of the source due to HRTF filtering), and provides new information about the spatial location of the source, *viz.*, the temporal and spectral/intensitive disparities of the waveforms reaching the two ears. The representation of individual auditory sources in the nervous system begins with transduction at each cochlea, and then continues with further stimulus coding at subsequent stations in the ascending auditory system. The anatomical stage at which various information-processing steps occur is outlined schematically in Figure 1.3. In this Figure, monaural processing of information from the left and right ears is colored white and black respectively. Sites of binaural convergence are re-coded as light and dark grey to reflect the convergence of binaural information. In general, after the original binaural convergence in the superior olivary complex (SOC) sound sources from a given left or right hemifield are generally represented in the contralateral brain hemisphere, although there is much variability at the level of individual neurons.

SCHEMATIC OF THE ASCENDING AUDITORY SYSTEM

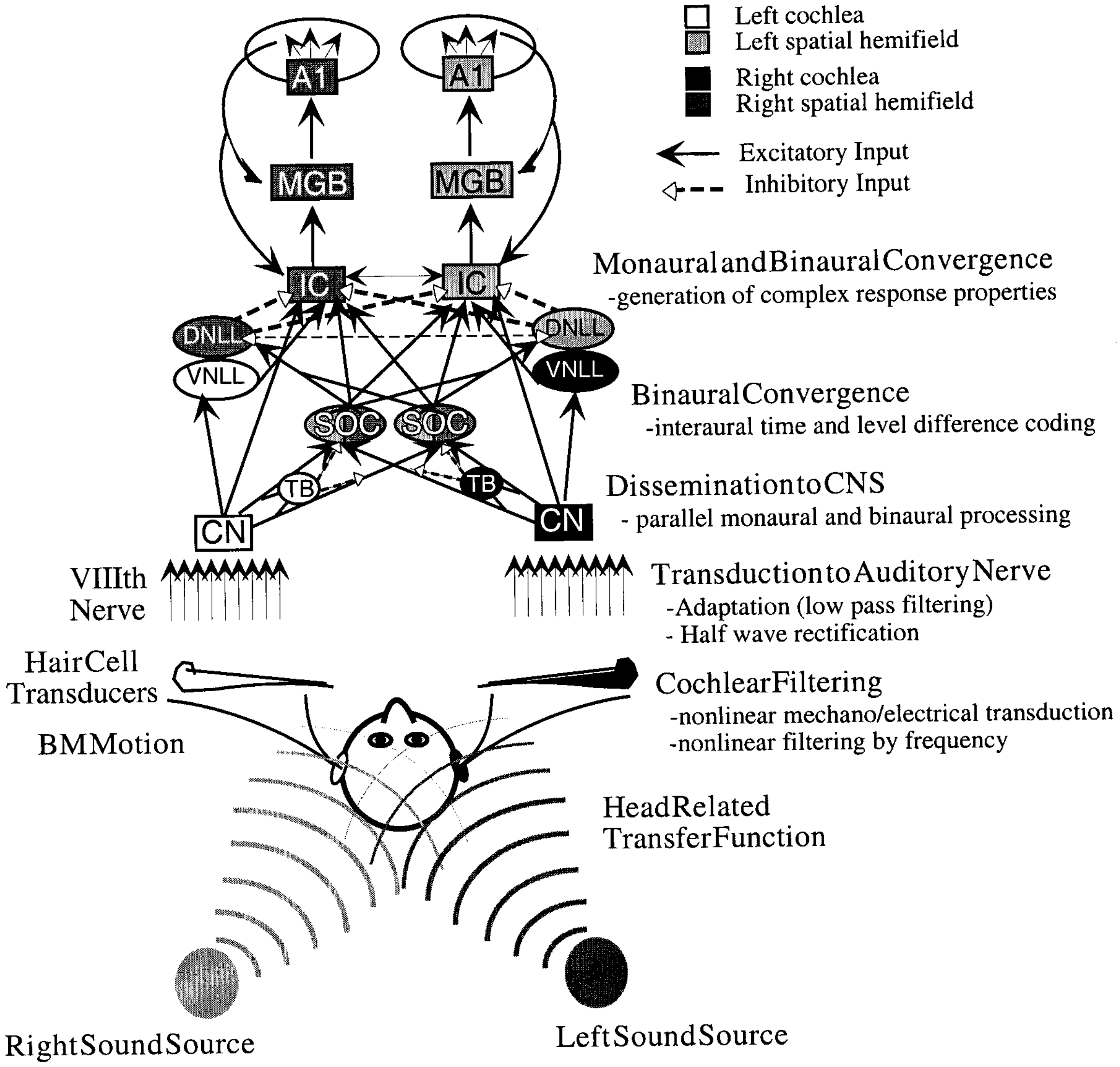


Figure 1.3

Schematic of the information processing in the ascending auditory system: source to cortex. The right and left 'ears' and the monaural nuclei representing those ears are in black and white respectively. The dark grey and light grey represent the right and left spatial hemifields respectively, the coding of which includes the consequences of binaural interaction. CN=cochlear nucleus; TB=trapezoid body; SOC=superior olivary complex; VNLL and DNLL=ventral and dorsal nuclei of the lateral lemniscus; IC=inferior colliculus; MGB=medial geniculate body of the thalamus; A1=primary auditory cortex.

Sounds are characterized by rapid temporal fluctuations. These temporal changes indicate the sound frequency, the onset and offset of sound energy, the rhythm of intensity fluctuations across the spectrum (the sound's envelope characteristics) and a host of other characteristics. Thus, it is important that such temporal information be preserved in the nervous system. The cochlea achieves this by transducing the sound waveform into a pattern of spike firing across the auditory nerve fibers. During this transduction two major transformations of the source signal occur. First, a mechanical *Fourier* transformation of the incoming signal is accomplished via the motion of the basilar membrane, upon which the hair-cell transducers lie. This essentially results in the passing of the stimulus waveform through a series of nonlinear, parallel, overlapping, and narrowband filters. Secondly, the sound intensity in each filter is subjected to a compressive non-linear process, such that the BM amplitude grows more slowly than the stimulus amplitude. The advantage of these features of cochlear processing is that by separating the wideband signal into narrow-band filters the temporal fluctuation in intensity within any given filter become slow enough to be encoded by the nervous system. Furthermore, because of the compressive non-linearity, a wide range of stimulus intensities are compressed into the smaller physiological dynamic range that can be transduced by the hair cells (Yates, 1995).

Even after cochlear filtering however, the ability of the auditory nerve to follow the temporal fine structure of a sound remains limited. The hair cell receptor potential mirrors the phase of the stimulus in its pass band, but only releases neurotransmitter into the synapse with the auditory nerve fiber during depolarization (Yates, 1995). Action potentials can thus only be generated in response to upward motions of the BM, and so the response of the auditory nerve fiber will be, at best, a half-rectified version of the receptor potential modulations. For low stimulus frequencies, the temporal variations of the waveform are slow enough to be faithfully encoded by the auditory nerve response – the response will 'phase-lock' to the stimulus (Rose *et al.*, 1967). This is illustrated in Figure 1.4, which shows the time course of the response of an auditory nerve fiber to a high (Fig. 1.4 upper) and low frequency (Fig.1.4 lower) tone. Phase locking in mammals is only significantly useful in response to stimulus frequencies below about 2000 Hz, although in some acoustically specialized animals such as owls and some bats, this limit

AUDITORY NERVE RESPONSE TO HIGH AND LOW FREQUENCY TONES

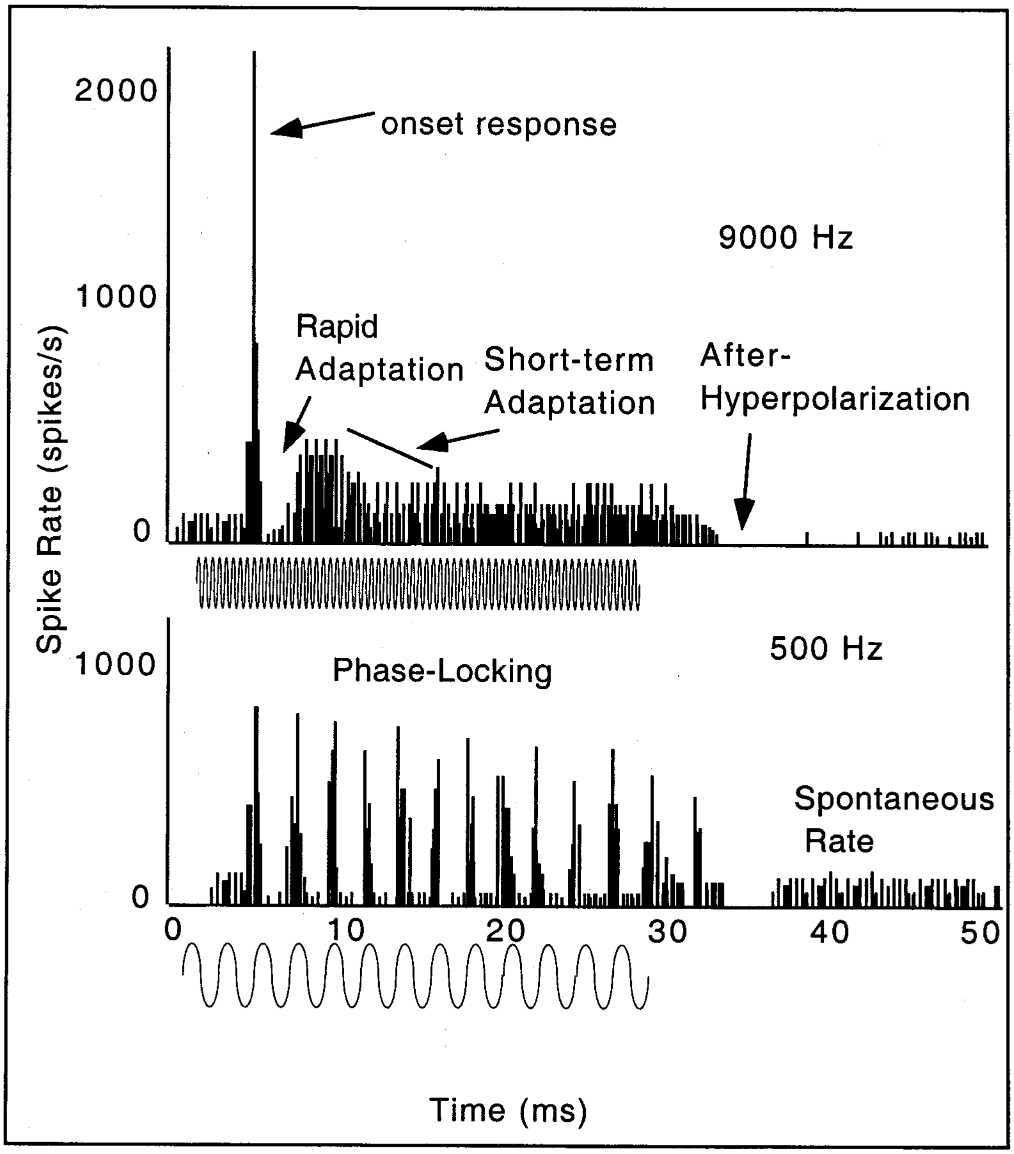


Figure 1.4
The response of the auditory nerve (high best-frequency) to a 9000 Hz tone (TOP panel) and a 500 Hz tone (BOTTOM panel). Note that the response to the high frequency tone is dominated by an onset response which resolves to a low rate, steady state response. The response to the low frequency tone is locked to the phase of the stimulus. (adapted from Ruggero, 1993)

is higher. Frequencies below this phase-locking limit may thus be represented by a temporal code. Specifically, the information encoded regarding the sound frequency is carried in the spike times, rather than in the number of spikes. The auditory nerve response to higher frequencies is relatively constant over the stimulus phase (absence of phase-locking, see Fig.1.4 upper), and thus frequency must be identified by a place-code (spatial position of the fiber in the tonotopic array). The place code for high frequencies is less ambiguous than a place code at lower frequencies. This is because the auditory filters are asymmetrical with skirts that are shallow at the low end, but very sharp at the high end. It follows that the lower the sound frequency, the larger number of auditory filters it will activate at moderate levels, and thus the less specific the 'place code'. The existence of a temporal code for low frequency sounds via phase locking provides for greater specificity.

Let us reconsider the transformation of the complex waveform comprised of those from several sound sources. Before a representation of those sources can be reconstructed in the brain, the waveform is first transformed by the HRTF, then passed through a band-pass filter bank (the cochlea), and subjected to a compressive non-linearity. This is followed by a half-wave rectification and low-pass filtering due to synaptic limitations in the generation of action potentials in the auditory nerve. At this stage, the system has deconstructed the waveform first based on the ear it arrived at, and then based on the frequency components arriving at each ear. Thus, ear and frequency can be considered the primary channels by which sound energy is sorted at the periphery, limiting the representation available for further central processing. We should of course keep in mind that there is duplication in the coding at each ear. The same sound waveform is essentially filtered in terms of frequency twice (once by each ear) but the relative intensity across the spectrum will vary as a function of the HRTF. Note that for a quiet lateralized sound, little coding of information in the high frequencies will be accomplished by the far ear.

Within a given peripheral frequency channel, the variable dimension is stimulus intensity, which fluctuates over time in most sounds. The tonotopic representation of sound in the auditory nerve array is maintained throughout the system and further stimulus features tend to be coded within the constraint of the tonotopic map. This means that additional channels for other stimulus dimensions must emerge either in some topography that is orthogonal to the tonotopic map (i.e., the dimension is represented entirely within each frequency channel) or more likely, by other, non-spatial coding mechanisms such as a temporal or population code.

The representation of a stimulus in the auditory nerve response is also influenced by adaptive processes with different time constants: a rapid adaptation (1-3ms), short term adaptation (10-40 ms), and a long term adaptation (100's of ms) (Smith, 1988). The response of an auditory nerve fiber to high and low frequency tones is illustrated in Figure 1.4 (adapted from Ruggero, 1992). The rapid adaptation results from a depleted neurotransmitter in the inner hair cell-auditory nerve synapse after the initial high frequency burst of spikes to the onset response to the sound (onset response), which leads to a brief rate notch. The spike rate then recovers and subsequently declines over 20ms to a steady-state rate due to short-term adaptation. The decrease in the size of each phase-locked response to low frequency sounds over time is also reflective of such short-term adaptation. Upon cessation of the stimulus, there is a brief period of rebound hyperpolarization during which the auditory nerve requires a certain level of stimulation to generate a new onset response. The onset response is important as it contains substantially more information than the steady state response, including coding for a greater dynamic range of intensity (Smith, 1988), and ultimately affording a greater perceptual salience (Phillips *et al.*, 2002a). The result of this pattern of adaptation is important to consider in examining the response of the system to rapid temporal changes in intensity. Such adaptive processes are likely candidates in limiting temporal acuity. In summary, the function of each auditory nerve fiber can be thought of as a monitor of the presence, amplitude and timing of energy in the fiber's pass band. As the stimulus is transformed at each station in the ascending system, the onset response comes to dominate the total neural response to the stimulus. At the level of the cortex, there is little or no ongoing response (Phillips *et al.*, 2002a). In cortex, the neural refractory period is also actively extended via a period of hyperpolarization after the onset response to the stimulus (Eggermont, 2000).

REPRESENTATION OF STIMULUS AS A FUNCTION OF SPATIAL LOCATION

As was mentioned previously, at the periphery the sound source is channeled first by ear and then by frequency. While the tonotopic organization is maintained throughout the ascending system, the representation of the two ears within each frequency channel is generally re-coded through the ascending auditory system into a binaural representation of contralateral auditory space at the level of the cortex (Irvine, 1992). That is, convergence of information from the two ears generally occurs in a fashion that preserves tonotopic organization (a 2kHz fiber from the left ear converges with a 2kHz fiber from the right ear). The anatomical pathways leading to such a contralateral representation are presented in Figure 1.3. This process begins with the dissemination of auditory nerve information through parallel information channels in the cochlear nucleus. Continued *monaural* processing occurs in the dorsal cochlear nucleus (DCN), at which there is convergence of information from different frequency channels emerging from the ipsilateral ear (Rhode & Greenberg, 1992). This may be the source of a new code for monaural cues to sound location in the sound source spectrum at each ear (Imig *et al.*, 2000). The first site of binaural interaction, and thus the first site at which interaural cues for sound location can be coded, occurs in the superior olivary complex (SOC), particularly in the Medial (MSO) and Lateral (LSO) Superior Olivary nuclei. As described in the next section, there are two ways by which a sound source differs at the two ears – an ITD for all frequency components of a sound, and an ILD for sound components above about 1.5 kHz, which generally increases with sound frequency due to the head-shadow. The coding of the interaural disparities will now be discussed, followed by a brief discussion on the coding of monaural spectral cues for sound location.

REPRESENTATION OF THE INTERAURAL DISPARITY CUES TO SOUND LOCATION

The superior olivary complex (SOC) in the brainstem is the first site of binaural interaction, receiving bilateral inputs from the cochlear nuclei. There is a general division of labor in the SOC, with the medial superior olive (MSO) receiving predominantly low frequency input, on which it operates to encode interaural phase differences (IPDs); and the lateral superior olive (LSO) receiving high frequency inputs, on which it operates to encode interaural level differences (ILDs). Recall that the ITD is a relatively fixed value for all frequencies (cf. Roth *et al.*, 1980 and pg. 4), but can be re-specified as an IPD for any given frequency. For a given ITD, the IPD (in cycles of phase, or degrees of phase) will decrease with increasing frequency (see Fig. 1.2C) because the stimulus period (ms) decreases. That is, while the location of the stimulus provides for an ITD for that stimulus that is fixed by the width of the head, this ITD results in a different IPD in each frequency channel.

The MSO receives both excitatory and inhibitory inputs that originate from spherical-bushy cells in the ventral cochlear nucleus (VCN), which in turn receive their inputs from predominantly low-CF AN fibers through an end-bulb of a Held synapse (Rhode & Smith, 1986; Irvine, 1992). This secure synapse allows for preservation of a primary-like (AN) phase-locked response to low frequencies. Low-frequency phase-locked excitatory inputs from each VCN converge on MSO cells directly, in a manner which preserves tonotopy. The preservation of the stimulus phase in the output of the VCN cells is a crucial constraint on the coding of the ongoing interaural phase *difference*. MSO neurons act as coincidence detectors of the excitatory inputs from each ear, responding maximally when the excitatory half-phases of the inputs arrive synchronously (Goldberg & Brown, 1969). Shifting the relative phase of the two inputs by delaying the sound at one ear (changing the ITD) reveals sensitivity to IPD that repeats cyclically as the delay increases. Specifically, maximum firing occurs to those IPDs for which the inputs arrive in phase, and minimum firing to IPDs in which the inputs arrive 180 degrees out of phase. The period (in ms) of the MSO output function reflects that of its phase-locked inputs – it is just slightly broader than a half-rectified sine wave of the stimulus frequency, or of the best frequency of the cell if the stimulus is noise (Yin & Chan, 1990; McAlpine *et al.*, 2001). The function tends to peak for delays in which the stimulus leads at the contralateral ear, that is, for free-field stimuli that would be located in the contralateral hemifield. Interestingly, the main peak, and subsequent peaks occurring with the next stimulus cycles (with long ITDs) tend to fall outside the physiological range of ITDs which the animal can experience (McAlpine *et al.*, 2001). This general point, that ITD functions are broader than the physiological range of the animal had been made by earlier theorists (Stillman, 1971; Phillips & Irvine, 1981; Phillips & Brugge, 1985).

The shape of this spike rate vs. ITD function, and the delay at which peak response occurs, is determined by the temporal interplay between the phase-locked excitatory and inhibitory inputs to the neuron. Before reaching the MSO, some of the contralateral VCN inputs are first transformed into a phase-locked, temporally precise glycinergic inhibition via a calyx-of-Held synapse in the medial nucleus of the trapezoid body (MNTB) (Grothe & Sanes, 1993; Grothe, 1994, 2000). Blocking these glycinergic inputs (by applying strychnine) reveals that the excitatory inputs alone tend to result in maximum coincidence detection for ITDs near 0 (Brand *et al.*, 2002). The contralateral, temporally precise, and phase-locked inhibition from the MNTB arrives at the MSO cell slightly earlier than the contralateral excitation, blocking a portion of the initial spikes that would occur due to synchronous arrival of the phase locked excitation. This has the effect of sharpening the midline slope of the function, thereby shifting the function's peak toward contralaterally leading ITDs that are outside the physiological range (Brand *et al.*, 2002). The result is that the maximum slope of the function is placed near zero ITD. A smaller proportion of the ipsilateral VCN fibers are also transformed into glycinergic inhibitory inputs via the lateral nucleus of the trapezoid body (LNTB) (Cant & Hyson, 1992), although their role is not understood. This suggests that the physical path lengths of the excitatory inputs from each ear to the MSO are functionally equal.

A necessary consequence of coincidence detection of phase-locked inputs is that the output function will broaden with decreasing frequency of the input. Given that the inhibition acts to place the slopes of these functions near zero ITD, the peak of the ITD function is dependent on the stimulus frequency. The lowest frequency neurons will have the broadest ITD functions and will peak at the longest ITDs; the highest frequency neurons have the sharpest ITD functions and will peak at relatively shorter ITDs. If ITD is transformed into cycles of the stimulus frequency (IPD), then the peaks of the function tend to occur at the same phase, about 45 degrees, regardless of the dominant input frequency of the cell (McAlpine *et al.*, 2001; Brand *et al.*, 2002). Since the peaks tend to fall outside the physiological range of ITDs, particularly at low frequencies, it is unlikely that the peak firing rate is involved in the code for sound localization (Phillips & Irvine, 1981; Phillips & Brugge, 1985; McAlpine *et al.*, 2001; Brand *et al.*, 2002). That is, the maximum firing rates that occur in response to low frequency sounds with artificially applied ITDs would typically never occur in reality – those ITDs can not be created by the conjunction of azimuth and the head size of the animal. However, the placement of the slope of the function across the midline has the effect that the maximum *change* in spike rate with changes in ITD will occur for locations near the midline. This is precisely the range for which mammals are behaviorally most sensitive (Middlebrooks & Green, 1991). This also has the consequence that each MSO is maximally activated by contralateral sound sources, suggesting that only two broadly-tuned hemifield 'channels' are generated by binaural convergence in order to process ITDs. That is, in mammals at least, there simply can be no 'map' of sound location based on peak neuronal response to specific ITDs in either MSO, only an increasing firing rate as sound location is shifted over the midline towards the more lateral positions in the contralateral sound hemifield.

A similar channel structure emerges in the coding of ILDs by neurons in the LSO (reviewed by Irvine, 1992). In contrast to the MSO, the LSO codes for ILDs congruent with sound sources located in the ipsilateral hemifield. Each cochlear nucleus sends an excitatory input to the LSO ipsilateral to the source and an inhibitory input to the LSO contralateral to the source via the MNTB. The strengths of these inputs are dependent on the intensities of the sound at the ears. For a sound source at the midline, the strength of the ipsilateral excitation and contralateral inhibition arriving at the two LSOs will be roughly equivalent and thus only partially activate the LSO cells on each side. However, for locations off the midline, the greater intensity at the ipsilateral ear *re* the contralateral ear will result in stronger net excitation in the LSO ipsilateral to the sound source. As sound source lateralization increases, output towards higher centers by the ipsilateral LSO will increase, and output to higher centers from LSO neurons contralateral to the source will decrease.

LSO cells appear to exclusively encode the ILD within a small band of frequencies centered on their best frequency. That is, the ILD is the primary determinant of that cell's azimuthal sensitivity with minimal contribution from interaural time difference and monaural spectral cues which are also available (Tollin & Yin, 2002a; Tollin & Yin, 2002b). In Tollin et al's (2002a) experiment, the azimuthal sensitivity of an LSO cell was determined using digitally-generated dichotic stimuli which simulate the HRTF for a given location, and thus provide all the spatial cues, both binaural and monaural. They then selectively eliminated one cue at a time by digitally manipulating the waveform presented. For example, they could set the ITD to zero while maintaining the ILD and spectral shape cues for a given location, or normalize the interaural level differences to zero while maintaining the ITD and monaural spectral cues. As long as the ILD within the band around the cells dominant frequency input remained valid, there was no change in azimuthal sensitivity of the LSO cell.

Notably, the LSOs project their axons predominantly to higher centers on the contralateral side of the brain. So, while sounds originating from spatial locations on the left are activating the left LSO, the sounds are activating higher centers in the right side of the brain, and vice-versa for sounds sources on the right. The fact that structures at levels rostral to the outputs of the SOC respond to spectral content in their passband coming from locations in the contralateral auditory hemifield is a basic principle of binaural processing (Phillips & Irvine, 1981; Phillips & Brugge, 1985).

While the LSO predominantly processes ILDs of high frequency sounds, a subset of neurons in the LSO's lateral limb receive low-frequency phase-locked input (Tollin & Yin, 2002a). The convergence of an ipsilateral phase-locked excitatory input, and contralateral inhibitory input results in an 'anti-coincidence' detector that is sensitive to ITDs in addition to ILDs. This is because the coincident arrival of the excitation and inhibition *prevents* the cell from responding. Assuming equal axonal path lengths from the two ears, presentation of a sound that was interaurally 180 degrees out of phase would result in maximum temporal alternation of the excitation and inhibition, allowing excitatory spikes to activate the cell in the troughs of the phase-locked inhibition. Such neurons have been described, and are referred to as 'Trough-type' neurons (Yin & Kuwada, 1983; Yin & Chan, 1990; Kuwada *et al.*, 1997; Fitzpatrick *et al.*, 2000; Fitzpatrick & Kuwada, 2001; McAlpine *et al.*, 2001). Their troughs occur at ITDs associated with the contralateral side of space while their peaks tend to fall at long ITDs associated with the ipsilateral side and well outside the physiological range. The crossed projection of the LSO to the IC and DNLL means that the peak will correspond to contralateral space at higher stations. The role of 'trough-type' neurons in binaural processing is not yet clear, but they might be involved in signaling the state of interaural correlation - particularly important in multi-source and echoic environments where the interaural correlation of distal sound sources become degraded. They may also act as a separate binaural channel to segregate uncorrelated background noise from correlated signals. This issue will receive a more thorough discussion in Chapter 4.

The contralateral projection of the LSO and ipsilateral projection of the MSO to the IC result generally in a broadly-tuned sensitivity to contralateral sound sources in neurons of each IC. The dorsal nucleus of the lateral lemniscus (DNLL) also receives most of its input from the ipsilateral MSO and contralateral LSO, in addition to a heavy input from the contralateral AVCN (low-frequency section) and the DCN (see Schwartz, 1992 for a review). The low frequency neurons found there are an exquisite reflection of the MSO inputs, showing similar IPD sensitivity (Brugge *et al.*, 1970; Boehnke & McAlpine, 2001). Subsequently, the DNLL sends inhibitory inputs predominantly to the contralateral IC, again in a fashion that maintains the tonotopic representation. This contributes to making the IC a highly convergent structure that receives excitatory binaural inputs from the SOC, inhibitory binaural inputs from the DNLL and monaural excitation and inhibition from the CN and VNLL. There is evidence that convergence of 'peak-' and 'trough-' type neurons onto IC neurons generates 'intermediate'-type cells that have ITD functions with broad, multi-lobed peaks at contralateral ITDs (McAlpine *et al.*, 1998; Shackleton *et al.*, 2000). Such convergence, however, minimally affects the placement of the maximal slope through the midline, and only serves to enhance the contralateral representation. Other convergence could generate entirely new types of binaural sensitivity, and even de novo binaural response properties unrelated to the primary response types in the SOC (see Irvine, 1992). For example, convergence might generate the 'predominantly binaural' cells (Kitzes *et al.*, 1980) which respond maximally to ILDs and ITDs associated with midline azimuths.

Projections from the IC ascend ipsilaterally through an obligatory relay at the medial geniculate body of the thalamus (MGB), which in turn projects ipsilaterally to the auditory cortex. There is a wealth of data from a variety of species on the effect of auditory cortex lesions on sound localization (Macaque: Heffner & Masterton, 1975; rat:Kelly, 1980; cat:Jenkins & Masterton, 1982; Jenkins & Merzenich, 1984; rat:Kavanagh & Kelly, 1986; Macaque: Heffner & Heffner, 1990; Macaque: Heffner, 1997). Although there are some exceptions (notably the rat), bilateral lesions of the auditory cortex result in a modestly decreased ability to distinguish left from right sources, and a complete inability to distinguish between sound locations within the left or right hemifields. Jenkins & Masterton (1982) demonstrated that unilateral lesions of the cat primary auditory cortex (A1), resulted in severe deficits in sound localization in the contralateral hemifield of space. Additional experiments indicated that azimuth was represented separately in each iso-frequency strip: ablation of a given frequency band's representation resulted in a contralateral localization deficit for stimuli only of that frequency, and ablation of all but a single band of frequency representation resulted in preserved sound localization of stimuli of the remaining frequency band (Jenkins & Merzenich, 1984). Thus, not only does the auditory cortex appear to be necessary for sound localization in azimuth, but also localization is accomplished over frequency-specific pathways in A1. This contralateral representation of auditory space in A1 was verified in primates by Heffner's studies on auditory cortex lesions in Japanese macaques. Macaques with unilateral lesions were able to localize sounds normally in the hemifield ipsilateral to the lesion, but had *no* localization acuity in the contralateral hemifield except for a region of *reduced* acuity from the midline to about 18 degrees (Heffner & Masterton, 1975; Heffner & Heffner, 1990; Heffner, 1997).

These data are congruent with those from electrophysiological studies of single units in A1, which tend to reflect the response properties generated in the brainstem. Neurons of A1 are sensitive to manipulations of ILDs (Phillips & Irvine, 1981) and ITDs (Brugge *et al.*, 1969) in dichotic stimuli. Several studies have also examined the spatial tuning of auditory cortical neurons under free-field conditions (Benson *et al.*, 1981; Middlebrooks & Pettigrew, 1981; Imig *et al.*, 1990; Rajan *et al.*, 1990; Ahissar *et al.*, 1992; Clarey *et al.*, 1995; Brugge *et al.*, 1996, 1998; Eggermont & Mossop, 1998; Middlebrooks *et al.*, 1998; Furukawa *et al.*, 2000; Recanzone *et al.*, 2000; Furukawa & Middlebrooks, 2001). In general, from a random sample of recorded units, about 50-75% are sensitive to sound location, with the remainder being insensitive or 'omnidirectional' (Middlebrooks & Pettigrew, 1981). Clarey et al. (1992) summarized the prevalence of various types of units that showed azimuth sensitivity across many studies. Of the spatially sensitive units, the largest class (50-70%) have been termed "hemifield" (Middlebrooks & Pettigrew, 1981) or "contra-field" (Rajan *et al.*, 1990) units and show maximal responses to locations throughout contralateral space with well defined borders within about 30 degrees of the mid-sagittal plane (midline). Another small (<20%) class of neurons- “ipsilateral” units (Rajan *et al.*, 1990) show the same pattern as contra units but in the hemifield ipsilateral to the sound source. Another group (~20%) are called “central” (Rajan *et al.*, 1990) or “midline-preferring”. These respond maximally to sources at or within 20 degrees of the midline, and presumably correspond to the “predominantly binaural” (PB: Kitzes *et al.*, 1980; Phillips & Irvine, 1981) neurons. Studied dichotically, PB cells respond best to binaural disparities close to zero. Finally, a very small (~2%) group of cells exist which are called “multi-peaked” (Rajan *et al.*, 1990) because they show maximum responses to more than one spatial location, although these peaks are always separated by at least 20 degrees.

“Hemifield unit” spatial selectivity for the contralateral hemifield is consistent with the contralateral deficits observed with auditory cortex lesions. Heffner (1997) observed that unilaterally lesioned macaques retained some acuity in the region from 0 to 18 degrees in the hemifield ipsilateral to the lesioned cortex. This suggests that some units in the remaining cortex may code for the near-frontal region of space, regardless of hemifield. Some could be accounted for by the small subset of “ipsi-units” found in the intact auditory cortex. The contra-hemifield units in the unlesioned cortex could be coding for regions just across the midline, since some cell have been found with borders extending beyond the midline. Perhaps this region is best explained by the representation of space near the midline afforded by the “midline-detectors” found in both cortices. This would explain why in these monkeys, this region was one of reduced acuity (firing rates of these units decrease with deviations from the midline) and was only 18 degrees wide (these units only fire in response to sources with locations within about 20-30 degrees of the midline).

The foregoing suggests that there is no systematic ‘map’ of auditory azimuthal location, such as has been observed in the midbrain (avian analogues of the IC and SC) of owls (Knudsen & Konishi, 1978) and the superior colliculus of some mammals (Middlebrooks & Knudsen, 1984). A ‘map’ in this case refers to a systematic topography across the neural tissue, in which different neurons have maximal neural firing rates in response to a specific location. The spatial sensitivity of thalamic and cortical neurons in mammals (and owls, Proctor & Konishi, 1997) is simply too broad. However, the behavioural lesion data strongly suggest that *the information that needs to be extracted* in order to localize a sound is found in the pattern of spike firing as a function of sound location in A1 neurons. Such lesion data are always ambiguous – it is not clear whether the lesioned site is the locus of the mechanism under study (in this case acuity of sound localization), or whether it is a necessary component of the pathway leading to that mechanism. Given that A1 is one of the main portals of information flow to the higher cortical areas, information regarding various features of the sound must be preserved there. If a single A1 neuron coded specifically for a single spatial location (acted like a 'feature-detector'), its ability to code other stimulus features might be limited. An alternative view is that the A1 neurons are linear integrators of information from each ear within their frequency passband (Schnupp *et al.*, 2001). Linear coding of this sort maximizes information transfer, rather than creating limited 'feature detectors', which would be important if codes for various stimulus dimensions were to be extracted from A1 by higher cortical areas. They demonstrated that spatial sensitivity of A1 neurons under binaural listening was simply the linear sum of the neuron's response to the left and right ears alone as a function of spatial location. It has been speculated that separate processing streams (ventral 'what?', and dorsal 'where') diverge out of primary auditory cortex (Rauschecker & Tian, 2000). It is generally agreed that there is a ventral stream for sound recognition, but it is not clear if the dorsal stream is for processing 'where' in space (Rauschecker & Tian, 2000) or frequency (Belin & Zatorre, 2000). Some patients with parietal lesions show impaired sound localization (Pinek & Brouchon, 1992; Griffiths *et al.*, 1996; Griffiths *et al.*, 1997). Disruption of parietal areas by transcranial magnetic stimulation shifts the perceived laterality of a dichotic noise burst (Lewald *et al.*, 2002). Also, fMRI studies have shown that sound localization or perception of sound motion engages cortical networks that include parietal areas, particularly in the right hemisphere (Griffiths *et al.*, 1998; Griffiths *et al.*, 2000; Zatorre *et al.*, 2002). In fact, behavioral performance in sound localization tasks is better predicted by responses in parietal cortex, than by those in primary auditory cortical areas (Zatorre *et al.*, 2002). With the caveat that relating behaviour to regional blood flow is problematic, the converging evidence suggests that our veridical perception of spatial location is at least somewhat dependent on secondary cortical areas, which aquire their input from primary cortical areas.

Given their spatial sensitivity, what kind of information might be extracted from output of A1 neurons in order to localize sounds? The code for sound location (wherever it is extracted) has been proposed to be based on the slopes of the spike rate vs. azimuth functions of the hemifield-tuned populations of neurons, rather than their peak firing position (e.g. Boehnke & Phillips, 1999; Delgutte *et al.*, 1999; McAlpine *et al.*, 2001). One candidate code is that sound location is based on the relative activity of the auditory areas in the left and right hemispheres. There would be equal activity between the two sides of the brain for midline located sounds and an increasing disparity in their activity rates as the sound source deviates from the midline. Such a code would be fairly insensitive to stimulus level. Alternatively, the code could be based on the absolute rate within a hemisphere. This is suggested by behavioral lesions studies in macaques (Heffner, 1997). Unilateral lesion of the auditory cortex results in preserved localization of sound sources ipsilateral to the sound source, indicating that each hemisphere may be independently capable of coding contralateral locations. It is possible that azimuthal location is indicated by the absolute discharge rate of the population of hemifield-tuned neurons (maximum activity for contralateral locations and decreasing activity toward the midline). However, such a code would be difficult to maintain across stimulus intensity, as the discharge rate of the population would be affected both by changes in stimulus level and changes in azimuthal location. Some evidence suggests that a minority of cells in each hemisphere are most sensitive to locations in the ipsilateral hemifield (Clarey *et al.*, 1992). Thus, a comparison of the relative activity of ipsi- and contra-laterally sensitive neural populations within a hemisphere might also be possible. Another possibility that may resolve the ambiguity is a comparison of the spike rate in monaural channels coding for the stimulus level with that of binaural channels coding for sound location. A problem with this solution is that very few, if any, cortical neurons are strictly monaural. All of these solutions assume that there is a general increase in cortical firing as a function of stimulus level. This is not strictly true. A subset of cortical neurons has non-monotonic spike rate vs. level functions, responding best to a given SPL (Phillips *et al.*, 1994). The existence of a code for sound level independent of generally increased activity level may make the ambiguity problem with sound source location moot. The potential codes mentioned above are essentially rate codes. A complete alternative that would render sound intensity confounds moot is to assume that the system uses a temporal code. Such codes could be based on features such as the latency to first spike and interspike intervals. Using neural network models, it has been demonstrated that the spike trains from a given auditory cortical neuron carry distinct information in its spike timing to signal individual locations *throughout* azimuth (Middlebrooks *et al.*, 1994).

While the nature of the *code* for sound location in azimuth is still under debate, in general, the spatial sensitivity observed of the strictly ascending auditory system as a whole is best characterized as a broadly-tuned contralateral sensitivity. In terms of the generation of 'channels' by which different sounds from different sources might be segregated, it would seem that separation of two sound sources into different lateral hemifields would be the most effective means of separately representing spectrally similar sources. This is because, in general each sound would be represented in neural populations in opposite hemifields. This channel 'tuning' (or spatial sensitivity of the population of neurons in each A1) is not reflective of our ability to localize sounds per se, it simply reflects how sensitivity to sound source location is represented. How the information available in that representation is used to help the organism to localize a sound source is another issue entirely.

REPRESENTATION OF MONAURAL SPECTRAL CUES TO SOUND LOCATION

The spectral shape imposed on a sound source due to head filtering results in monaural cues which can be exploited to determine the elevation of a sound and to reduce ambiguity in the cone of confusion – the region near the lateral poles where the interaural disparities are ambiguous with respect to spatial location. Analysis of the waveform available at a single ear as a function of location after filtering by the head and pinna (HRTF) reveals a number of potential cues available to the nervous system. These are evident in the waveforms recorded at each ear as a function of spatial location displayed in Figure 1.1 (see also Shaw, 1974; Shaw & Vaillancourt, 1985; Middlebrooks, 1997) . The main one is the spectral shape, in particular, a notch in the spectrum that changes in frequency (between 6-10 kHz) as a function of location. This notch, and the shift in its frequency, can be seen by following either the blue or the red waveform across spatial location. Presumably using these monaural spectral cues neural sensitivity to spatial location can be generated. In the SC of mammals the auditory space topography found in the SC can be generated based on this monaural information (Palmer & King, 1985).

While such cues are available in the ascending lemniscal system, they can only be used if the receiver has information about the spectrum of the source, i.e., if it is a familiar complex sound. Alternatively, if the sound is long enough, the receiver might be able to 'sample' the source more than once while shifting their head position in order to resolve the location (Wightman & Kistler, 1997b). Regardless, the spectrum must be compared to some template in order to localize based on monaural spectral cues. In that regard, the fact that the cue is 'monaural' does not mean that the process of localizing is any less central than localizing using the cues from binaural interaction. The neural machinery that underlies the processing of monaural spectral cues is only beginning to be unraveled. A current candidate nucleus for the *initial* processing is the dorsal cochlear nucleus (DCN). Sensitivity to spectral notches in wideband stimuli has been shown in DCN cells (Imig *et al.*, 2000; Parsons *et al.*, 2001). In cats, these cells also receive proprioceptive inputs originating from pinna movements. This suggests that pinna adjustment, presumably used to disambiguate sound source locations by changing the spectral profile, contributes to the firing pattern of DCN cells (Young *et al.*, 1995; Davis *et al.*, 1996; Kanold & Young, 2001), at least in cats. In humans, and other animals lacking mobile pinnae, any proprioceptive input is likely to arise from the neck in order to signal head movements, although this has yet to be shown. While DCN neurons are one possible site for such a spectral analysis, lesion studies would be required to show that monaural spectral cues can no longer be used after specific ablation of DCN. It is also unclear how the responses of DCN are used than used. What is known, however, is that presentation of sound stimuli moving in elevation, the perception of which is dependent on monaural cues, activates the same higher cortical 'motion' areas as motion in azimuth which relies on binaural cues (Pavani *et al.*, 2002)

In summary, there are a number of physical cues available to the system to indicate sound source location which emerge ultimately from the differential filtering of the sound waveform at each ear, and the interaural differences (ITD, ILD) that occur as a function of sound source location.

LOCALIZATION OF A SINGLE SOUND SOURCE

The preceding has outlined the physical and physiological constraints on the coding of the auditory spatial dimension. Something should be said about the acuity of sound localization behavior before we proceed. This will be limited to sound localization in azimuth, because that is the only plane examined in this thesis.

There is a general consensus that localization of low frequencies is dependent on ITDs, that localization of high frequencies is dependent on ILDs, and that neither cue is salient in the mid frequencies. Considerable psychophysical evidence (reviewed by Middlebrooks & Green, 1991) exists to support this "duplex theory" of sound localization offered first by Rayleigh (1907). The most common method for studying spatial location acuity in the free field environment is to have a listener in an anechoic room with a series of speakers located at different positions in either azimuth, elevation or both. In absolute localization paradigms listeners indicate in some way (e.g. by pointing) the absolute location in which they perceived the sound to have occurred (Stevens & Newman, 1936; Oldfield & Parker, 1984a; Makous & Middlebrooks, 1990; Middlebrooks, 1992; Wightman & Kistler, 1992; Carlile *et al.*, 1997). The mean difference between the actual location and the indicated location, or its variance is usually reported (in degrees).

Stevens & Newman (1936) first determined that azimuthal localization errors for tones were frequency dependent. Error rates averaged over azimuth were low for frequencies up to about 1 kHz and above 4 kHz, but poorer for mid-frequencies. The poor performance in the mid frequencies was taken as evidence for the duplex theory of localization, i.e. that localization is best at frequencies for which the binaural cues are most salient (ITDs for low, and ILDs for high frequencies). Other studies tested absolute sound localization in humans using broadband noise sources varying in both azimuth and elevation. Makous & Middlebrooks (1990) measured head-pointing accuracy toward the perceived source of a sound, and demonstrated that the spread of errors relative to the actual source was lowest for locations at or near the midline (SD=2°-3°) and was roughly stable at about 8° for source locations between 50° and 180° azimuth. Other studies

(Oldfield & Parker, 1984a; Wightman & Kistler, 1992) indicate a more constant magnitude of error across azimuth. Differences in results across studies probably are related to the methods of response. Collectively, the studies demonstrate generally good absolute auditory spatial acuity, with acuity near the midline probably somewhat better than more lateral positions.

There is a general consensus that interaural difference cues rather than pinna based spectral cues are used for localization in azimuth. However, for symmetrical locations in the front and back, and within the theoretical "cone of confusion" (a cone extends from each ear on the surface of which all binaural cues are the same), binaural cues are ambiguous. This has led some authors to propose that pinna-based cues at the ear ipsilateral to the source contribute to localization in azimuth by resolving front/back confusions and the positions of sound sources in the "cone of confusion" (Musicant & Butler, 1984b, 1984a, 1985). There is strong evidence for a role of spectral difference cues in localization, particularly in regions where binaural cues are ambiguous, such as elevated locations on the medial plane, locations in the "cone of confusion" and directly in front and behind the listener (Oldfield & Parker, 1986). Several studies have demonstrated that the peaks and notches in the spectrum produced by pinna filtering can signal a particular location in elevation, and that listeners actually may use the spectral components in their localization judgments. For example, accurate vertical localization only occurs for broadband stimuli, is prevented when the pinnae are occluded (Oldfield & Parker, 1984b), and is almost equally good under binaural and monaural conditions (Middlebrooks & Green, 1991).

Recently, experimenters have tried to determine the relative weights of the various cues for location as they are used by human listeners. By constructing a quantitative model that could accurately predict the responses of individual listeners based on ILD and spectrum cues, Middlebrooks (1992) demonstrated that ILDs and spectral shape cues appear to be used independently. ILDs are used for the horizontal plane, while spectrum cues were used for elevation. ILDs likely appeared as the most salient cues in this study because only high-frequency noises were used. In contrast, Wightman & Kistler (1992) used gaussian noise, which contains the full spectrum of frequencies, and demonstrated the dominance of ITDs in localization. Using a virtual environment, their stimuli were manipulated such that pinna and ILD cues signaled one direction, while ITDs signaled a different direction. Under these conditions, subjects almost always followed the ITD cue. Interestingly, when the low-frequency components were filtered out of the spectrum, the dominance of ITD disappeared and ILDs were used in localization. It was found that ILD could vary widely (up to 20 dB) with no effect on localization acuity as long as the ITD information in the low frequencies remained intact.

From the studies discussed, one can conclude that on the horizontal plane, when low frequencies are available in a signal, ITDs are used as the dominant cue for localization. In response to higher frequencies, typically ILDs are used. Finally, when localizing elevation of a source and resolving front back confusions, pinna based spectral cues are used, but are predominantly available at high frequencies, or for stimuli with high-frequency components.

PROCESSING OF TWO SPATIALLY SEPARATED SOUNDS

The results from single source localization studies have contributed greatly to our understanding of the limits of sound localization acuity and the relevant stimulus characteristics underlying it. Such experiments were conducted under idealized conditions (anechoic chambers with a single source), however, and spatial perception in humans typically occurs in environments that are more complex. More typically, sounds of interest occur amid other competing sources (i.e. other voices, music, traffic noise etc.) so we tend to localize sound absolutely for some purposes, or discrimate their locations for others.

Another common way to measure auditory spatial acuity is to present the listeners with two spatially separated sounds, and ask the listener whether the sounds came from the same or different locations. More specifically, one could ask if a second sound was to the left or right of the first, and determine the spatial separation in degrees for which the listener made a certain percentage of correct responses. Such a method determines the minimum audible angle (MAA) -the smallest angle between which two sound sources can be discriminated (Mills, 1958). Results generally indicate that MAA is smallest for stimuli directly in front of the listener and increases for peripheral locations (reviewed in Grantham, 1995). As observed in absolute localization, Mills (1958) also demonstrated that the MAA was frequency dependent. When MAAs were measured at the midline, they were low for low frequencies (1-2°), rose to a maximum around 3000 Hz (~4°), and decreased again for high frequencies (2°). The size of the MAA is particularly dependent on the position in azimuth at which it is measured. Mills (1958) noted that the MAA systematically rose with increases in azimuth. Interestingly, for high frequencies at locations near 90° azimuth, the MAA typically exceeded 40°, the largest MAA the apparatus could measure. This pattern, that spatial discrimination of two sources diminishes as sounds move to the azimuthal periphery has been replicated in several studies for both low and high-frequency tones and for wideband noise (reviewed in Grantham, 1995). Thus, our ability to *discriminate* the spatial location of two sounds, especially in the lateral hemifield, is worse than our ability to signal the location of an isolated sound.

When a signal is presented concurrently with a noise source, absolute localization accuracy for a signal decreases for signal to noise ratios (S/N) of 4 or lower (Good & Gilkey, 1996; Lorenzi *et al.*, 1999), and is dependent on the relative location of the masker. Laterally-located maskers were more detrimental than a midline-located masker to localization of a signal across a range of locations. This was attributed to a reduction in the detectability of the signal at the ear near the masker, which makes the listener functionally 'monaural' in terms of localizing the signal. That is, there is insufficient processing of the signal at the 'masked ear' to provide input into the binaural processors. This indicates that in real world listening, our localization acuity is probably worse than measured under ideal listening conditions (a single source in the free-field). Thus, to some degree, the signal must be sufficiently detectable at each of the ears in order to have its interaural differences encoded.

A selection of the several common tasks used to measure processing of two sounds is presented in Figure1.5. On the left side of the Figure is a schematic of the stimulus configurations of the different tasks, with intensity on the y-axis and time across the x-axis. In the middle panel is the question asked of the listener. The right panel describes changes in stimulus properties, including spatial changes, that alter task performance.

RELATIONS BETWEEN VARIOUS AUDITORY PSYCHOPHYSICAL TASKS

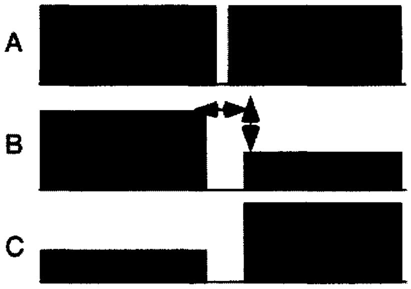
Simultaneous Masking 	Did you detect the signal in the noise?	Changing relative frequency or location of signal and noise improves performance
Forward Masking 	Did you hear the second sound?	Changing relative level, frequency or location of signal and noise improves performance
Backward Masking 	Did you hear the first sound?	
GapDetection 	Did you detect the gap in the noise?	Changing relative level (B,C), frequency or location of leading and trailing sounds worsens performance
Gap Duration Discrimination 	Which gap was larger?	Changing relative frequency or location (?) of leading and trailing sounds worsens performance
Temporal Asymmetry Detection 	Where the intervals evenly spaced? i.e. Was the rhythm even?	Changing relative frequency or location (?) of leading and trailing sounds worsens performance

Figure 1.5

Examples of the wide variety of temporal processing tasks that exist (not exhaustive), the question asked of the listener, and ways in which the task might be modified to probe 'between-channel' temporal resolution.

Spatial release from masking of a signal by a simultaneous noise

The lowest intensity, in decibels of Sound Pressure Level (dB SPL) at which a signal can be detected – signal detection threshold - increases when that signal is presented concurrently with another sound, especially if the two activate the same auditory filter. For example (see Fig. 1.5, 1st panel), if a 500 Hz signal is presented with a narrow band noise centered at 500 Hz, signal detection degrades as the bandwidth of this noise increases. At some bandwidth threshold, further increases in bandwidth do not further impair signal detection. This point of asymptote, by definition marks the width of the critical bandwidth of the filter centered on signal frequency (Scharf, 1970). Shifting the signal frequency away from the center frequency of the masker improves signal detection because the signal begins to activate channels that are not activated by the noise. Such a paradigm provides an excellent measure by which to examine stimulus dimensions which affect *concurrent* source segregation.

There is a clear improvement in signal detection, or intelligibility of a speech signal, when it is spatially separated from a simultaneous masking noise. Binaural hearing would appear to play a crucially important role in this effect. Under *headphone* listening, applying different interaural parameters to the noise and the speech signal of interest improves intelligibility (the binaural intelligibility level difference or BILD, Levitt & Rabiner, 1967; Culling & Colburn, 2000). More typically, this effect is examined for detection of a low frequency tonal signal in noise. Shifting the ITD of the signal relative to the noise, improves the signal detection threshold by up to 15 dB. Since there is no change in activity at the periphery, signal detection in this case is improved purely through the binaural interactions processing ITDs, probably at the MSO. This effect only occurs for low frequency signals. The benefit diminishes between 800 and 1500 Hz, consistent with the restriction of ITD processing to low frequency elements in the sound. Assigning different ILDs to the noise and signal also improves detection, but this release can be partially attributed to monaural factors, i.e. simply detecting the signal in the ear where the signal is louder and the noise is softer. Increases in ILD beyond 20 dB result in no further improvement in detection (Egan, 1965). The improvements with ITD are referred to as the binaural masking level difference (BMLD), which will be further discussed in the introduction to Chapter 4.

In more realistic listening conditions however, with sounds presented from speakers *in rooms or in the freefield*, the benefit of spatial separation on signal detection or intelligibility of speech occurs even during monaural listening (Saberi *et al.*, 1991; Hawley *et al.*, 1999). The largest determinant seems to be the improved signal to noise ratio (SNR) that occurs at the ear nearest the signal as the noise source is moved away from the signal (see Zurek, 1993 for a review). This improved SNR can most simply be accounted for by the fact that the noise becomes shadowed by the head and pinna and reaches the ear near the signal with less intensity, thus allowing the signal to be better encoded at that ear. The fact that the effect of spatial separation on masking release increases as the frequency of the noise and signal increase provides further evidence for a strong role of monaural SNR, resulting from the head shadow, in spatial unmasking (Gilkey & Good, 1995). It seems that this SNR effect is so overwhelming that additional contributions from binaural interaction cues (ITD and ILD) provide only a small additional benefit in the free-field listening environment (Saberi *et al.*, 1991; Zurek, 1993; Gilkey & Good, 1995). The interaural benefit may become more important, however, as the number of sources in the environment increases (Hawley *et al.*, 1999).

Masking of a signal by a temporally offset noise

An otherwise audible signal can be made inaudible when it is *preceded* or *followed* by another sound in time. This effect is referred to as 'forward' and 'backward' temporal masking respectively (see Fig. 1.5, 2nd panel). This temporal masking effect is dependent on the temporal proximity, level, and frequency of the masker relative to the signal. Specifically, signal detection improves if the masker is reduced in level, moved away from the signal in frequency or shifted away from the signal in time. Forward masking generally follows the same rules as simultaneous masking. It is most simply conceived of as a case of the masker 'persisting' longer in the neural/perceptual representation than the actual duration of physical stimulus. Plomp (1964) referred to this as the decay of auditory sensation. By varying the intensity of the second sound, and determining the shortest interval between the two sounds that could be detected, he determined the time course of this decay. Persistence makes less sense as an explanation for backward masking, as it would seem the noise can not persist backwards in time. Backward masking appears to be more cognitive in flavor or at least more plastic than forward masking, as is evidenced by the fact that backward but not forward masking is substantially reduced with practice (Moore, 1997) and is selectively impaired in some children with language disorders (Wright *et al.*, 1997). However, the two can be reconciled as being the result of a single mechanism if we imagine that the representation of the stimulus is convolved through a sliding temporal integrator, which takes a moving average of the representation of stimulus level within a given frequency channel (Penner, 1975; Plack *et al.*, 2002). This process essentially smears temporal changes in stimulus intensity by averaging the signal intensity at times before and after the signal. It can also account for the threshold for detection of a brief signal presented within a noise. Its detection is based on a short increment in intensity in the frequency channel. The briefer the signal or the gap, the smaller the increment or decrement in intensity, and thus the poorer the detectability.

An alternative to what are essentially 'persistence' theories, suggests that temporal masking and other effects might be explained by adaptation in the auditory system (Oxenham, 2001). Recall in Figure 1.4, that after the onset of a stimulus the auditory system goes through a series of adaptive processes with different time constants. Simplified, detection of a signal may be reduced in forward masking because the auditory system simply does not respond to the signal as well when it is in an adapted state. While the neural site of forward masking has been proposed to be the auditory nerve (Smith, 1977, 1979), more quantitative studies have indicated that there is less masking measured at the auditory nerve than psychophysically (Relkin & Turner, 1988). It is possible that further processes that limit neural refractory period in the auditory cortex (e.g. after-hyperpolarization, see Eggermont, 2000) might better account for the psychophysical data, although this has not been documented empirically. Both 'persistence' and 'adaptation' models lead to similar predictions and can account for much of the data obtained in studies of forward masking. In the only attempt to examine both model types with the same data, Oxenham (2001), suggested that the temporal window model might be slightly better at explaining the data, but the adaptation models did sufficiently well that adaptation could not be discounted. The most important feature required to account for the data in both models, is the inclusion of a realistic front end model of cochlear filtering which includes a basilar membrane compressive non-linearity (Oxenham, 2001; Plack *et al.*, 2002).

Spatial Separation effects on Forward Masking

Forward masking is alleviated by changing the ITD of the signal from that of the masker (Hanna *et al.*, 1982; Kollmeier & Gilkey, 1990) - essentially a BMLD effect stretched out in time. While classic forward masking paradigms have not specifically been studied in the free-field (to the author's knowledge), fairly clear predictions might be made. A brief second burst of noise will not be heard if it is preceded by a masking noise at the same location. However, if that masking noise is moved away from the signal, the signal will become more audible at the ear that is distal to the masker location, due to the head shadow. The pattern of separation should generally be predictable from the HRTF, and this idea deserves further study.

The precedence effect, or the 'Law of First Wavefront' (also called the 'Haas' effect Haas, 1951) is probably a case of temporal masking in the free-field. Precedence is the tendency to perceive only one event, when two or more transient sounds (i.e. clicks) occur in rapid sequential order, such as when the echo of a sound is present. If the reverberation time of the listening space below about 5 ms, the echo of the first sound (i.e. the second click) is not 'heard' and the two events are perceived as a single event, although the echoes do affect the timbre of the sound. The perceived location of this fused sound occurs at the location of the first sound, that is, the location information contained in the first wavefront dominates perception. Two events are distinctly heard when a gap of about eight ms is inserted (echo threshold). As the second click is spatially separated from the first click, this 'echo threshold' is reduced. Many studies have sought to explain the effect by active suppression of the spatial information of the first sound due to processes occurring in more central auditory nuclei such as the inferior colliculus (see also Litovsky & Delgutte, 2002 , for a review).

Divenyi and Blauert (1987) proposed that this effect might be related to forward masking processes, but didn't have evidence to demonstrate this. Recently, modeling studies have suggested that the second of two temporally separated transient signals (clicks) presented over headphones is degraded in the output of the cochlea, consistent with forward masking effects (Hartung & Trahiotis, 2001), or temporal integration

(Tollin & Henning, 1998, 1999). Therefore, veridical coding of the interaural parameters of the second click is degraded. These results strongly suggest that the precedence effect is largely caused by interactions between two temporally proximate transients. The effect can be predicted based on the output of the auditory nerve, without any need to postulate involvement of central mechanisms. A cochlear genesis for the effect may hold for clicks, but other processes may be involved for more complex stimuli which fuse at longer lag times. Regardless, such results also speak to the importance of understanding the representation of signals of interest in the output of the cochlea before considering higher level explanations of perceptual effects.

Detection of a temporal gap in noise

One of the features of temporal masking is that it is dependent on the temporal distance of the signal from the masker. For given masker and signal intensities, there is a minimum temporal interval between the masker and the signal required for signal detection. Alternatively, we can focus the task specifically on the minimum temporal interval between two sounds that can be detected. Gap detection, an extreme case of decrement detection, is the detection of a brief silent interval ('gap') embedded in an ongoing sound (or between two bursts of sound, see Fig.1.5, 3rd panel). In what will be referred to as a 'within-channel' case, the sounds that mark the 'gap' are identical noise bursts. Under these conditions, the minimum behaviorally detectable gap in man can be as short as a few ms for wideband markers (Plomp, 1964). If one thinks of the 'gap' as the inverse of a 'signal' presented in an ongoing noise, one might wonder if detection thresholds for a 'gap' are limited by the same mechanisms that limit signal detection. There is some evidence for this. The theoretical framework for gap detection established by Plomp (1964), and elaborated on by Penner (1977), is consistent with such a relationship to masking. The threshold for a gap was shown to increase as the level of the trailing marker was decreased relative to that of the leading marker. Below the minimum detectable gap, the persistence of the first marker (the 'decay of auditory sensation') masks the presence of the silent interval, and determines the magnitude of the onset response to the second marker. A detailed study of simultaneous masking, forward masking and gap detection using similar stimulus configurations revealed that all three tasks follow the same pattern of effect across variations in the intensity and temporal separation of the markers (Smiarowski & Carhart, 1975). This suggests that the mechanisms mediating gap detection might have some of their genesis in the mediating effects on simultaneous or temporal masking phenomena. Further support for this relationship comes from studies which show that detection of a 'gap' is degraded when it is positioned near the onset or offset of a noise (Crawley, 1999; Snell & Hu, 1999), similar to the degradation of signal detection in noise, when the signal is located near the temporal onset *or* offset of a noise masker, a phenomenon referred to as 'Overshoot' (Zwicker, 1965). In gap detection, the effect of gap placement diminishes with practice, at least in younger listeners (Snell & Hu, 1999).

There remains some debate about how exactly we 'detect a gap'. It has been suggested to be based largely on detection of short term fluctuations in stimulus level within a single-channel (Forrest & Green, 1987; Formby *et al.*, 1996a; Forrest & Formby, 1996; Florentine *et al.*, 1999) or across an array of channels (Heinz *et al.*, 1996). That is, the decrease in stimulus level relative to the marker level during the interval of silence is detected, similar to signal detection - where the increase in signal intensity relative to the noise is detected. It has also been suggested that detection might be based, at least in part, on the existence, and fidelity of the response to the onset of the second marker (Florentine *et al.*, 1999). This latter hypothesis would be predicted based on the neural coding of gap detection stimuli by single neurons in the auditory system (Zhang *et al.*, 1990; Walton *et al.*, 1997; Ison *et al.*, 2002). Particularly at higher stations, such as the cortex, the onset response to the leading marker and trailing marker are the dominant stimulus features observed in the spike trains of single units in response to a gap stimulus (Eggermont, 1995, 1999, 2000). At short gap durations, there is simply no onset to the second marker. If the onset to the second marker was of paramount importance, then we would expect that manipulations of the trailing marker onset, by making the trailing marker louder, or modifying its envelope by applying long or short rise times, would affect gap detection thresholds. In fact, increasing the trailing marker level degrades, rather than improves, gap detection (Oxenham, 2000); and applying long rise times (a feature that should diminish the onset response to the second sound) improves gap detection (Allen *et al.*, 2002). These results are more consistent with a view that gap detection is based on the effective sound level reduction associated with the gap, and that the sound level is probably best characterized as that observed in the output of a sliding temporal integrating window (Moore *et al.*, 1988; Plack & Moore, 1990).

Given this relationship of gap detection threshold to forward masking, what then are the effects of imposing differences in the spectral or spatial characteristics of the gap markers, factors known to release a signal from masking?

Frequency and Spatial Separation Effects on Gap Detection

In the temporal masking conditions, separation of the signal and masker in frequency improves detection because the masking is only occurring within a critical band. That is, the signal becomes more detectable as it becomes segregated from the masker. In the gap detection task, detection of the gap becomes *impaired* as the markers are separated in frequency (Kinney, 1961; Perrott & Williams, 1971; Williams & Perrott, 1972; Collyer, 1974; Fitzgibbons *et al.*, 1974; Divenyi & Danner, 1977; Williams *et al.*, 1978; Formby & Forrest, 1991; Formby *et al.*, 1993; Formby *et al.*, 1996a; Hall *et al.*, 1996; Phillips *et al.*, 1997; Formby *et al.*, 1998a; Formby *et al.*, 1998b; Grose *et al.*, 1999; Phillips, 1999; Taylor *et al.*, 1999; van Wieringen & Wouters, 1999; Oxenham, 2000; Phillips & Hall, 2000; Grose *et al.*, 2001a; Grose *et al.*, 2001b; Lister *et al.*, 2002; Phillips & Hall, 2002). This could be considered in two ways: as evidence for totally unrelated processes, or, that gap detection becomes impaired because it is inherently a within-channel task, and there is diminishing 'within-channel' information available upon which to perform the detection (i.e., there is reduced effective bandwidth). This latter hypothesis explains gap detection for markers which differ in frequency, but are near enough that they will still activate some common frequency channels to varying extents (Formby *et al.*, 1996a; Formby *et al.*, 1996b; Forrest & Formby, 1996; Heinz *et al.*, 1996). However, for markers with vastly disparate spectral content, or for spectrally similar gap markers presented to different ears, there are no peripheral mechanisms through which the gap can be detected as a within-channel decrease in level – the markers are effectively 'segregated' at the periphery. This segregation is probably maintained to some degree throughout the system. Phillips has argued that gap detection under conditions with markers disparate in frequency was a different task than within-channel gap detection (Phillips *et al.*, 1997; Phillips, 1999). It was proposed that such a task must be performed by a centrally located mechanism which executes a relative timing operation on the offset of activity in the channel activated by the leading marker and the onset of activity in the channel activated by the trailing marker. The observation that gap detection thresholds were equally degraded when the markers were presented to disparate frequency regions of the cochlea or different ears (Phillips *et al.*, 1997; Formby *et al.*, 1998a), or from free-field locations in opposite hemifields (Phillips *et al.*, 1998) suggested that a central timing operation was performed on the output central neural channels coding for those stimulus dimensions (referred to as perceptual channels).

The spatial case is of particular interest, and is still controversial, because free-field markers at different spatial locations activate similar peripheral channels, yet thresholds are still elevated. The tuning of the channels that were activated by gap markers at different spatial locations was determined in another experiment (Boehnke, 1998; Boehnke & Phillips, 1999) which measured gap detection thresholds as a function of spatial separation of the markers, throughout azimuth. This experiment revealed that gap thresholds rose as the gap markers were separated across the midline, but an equivalent (in degrees) marker separation within a hemifield did not elevate thresholds. It was suggested that this tuning reflected the tuning of cortical neurons, presumably binaural, to auditory spatial location. Recall from earlier in this chapter, that spatially sensitive neurons in the cortex are generally hemifield-tuned to contralateral locations. Thus, it was argued that gap detection was mediated by a relative timing operation between the output of the neural channels sensitive to spatial locations in different hemifields.

A binaural interpretation of the spatial case is controversial, because there is within-channel information available at either ear upon which the gap could be detected (Oxenham, 2000), and elevations in thresholds could be argued to be the result of level disparities in the representation of the markers in those channels at each ear: level disparities that arise as a result of the shadowing of each marker by the head. On the other hand, we know that both frequency disparities and binaural disparities (differences in ITD) result in a release of a signal from forward masking, and frequency differences between gap markers lead to an elevation in gap thresholds. Thus, insofar as there is a relationship between forward masking and gap detection (see above), segregation of the trailing marker from the leading marker by binaural disparities might lead to elevations in gap thresholds. The determination of the relative contributions of within- and between-channel processes to the effect of spatial separation of markers on gap detection will be the focus of Chapter 2.

Finally, if we extend a gap detection stimulus with leading (A) and trailing markers (B) that differ in frequency into a repeating sequence (A_B_A_B_A_B_...A_B), yet again there is a new, although possibly related perceptual task. This stimulus is similar to that used in studies of auditory sequential stream segregation (described in the preface). When A and B are similar in frequency, the sequence is perceived as coherent or integrated. That is, A and B are grouped as part of the same auditory object in time, probably because they are both activating many of the same frequency channels. However, as the frequency difference between A and B increases, the perceptual organization of the sequence goes through a process akin to fission. A and B become segregated into two separate streams. Under such conditions of streaming, there is a degradation in the ability to judge the temporal relationship between A and B. For example, listeners might have difficulty reporting whether A and B are occurring at the same time, or if they are alternating in time. This suggests a link between the processes underlying stream segregation and those underlying gap detection with dissimilar markers. In both cases, a temporal judgment made between the sounds is impaired. Little is known about the effect of spatial location differences on sequential stream segregation. However, if there is a relationship between the pattern of gap detection with dissimilar markers and stream segregation, then marker differences in spatial location that elevate gap detection thresholds should also result in stream segregation. This possibility is explored in Chapter 3. A comprehensive review of stream segregation will be reserved for that chapter.

CHAPTER 2

TEMPORAL GAP DETECTION FOR NOISE MARKERS AT DIFFERENT SPATIAL LOCATIONS

CHAPTER SUMMARY

This chapter will describe the results of five experiments examining the relative contributions of peripheral and central stimulus representations to temporal gap detection using sounds at different spatial locations as 'markers' of the gap. The impetus for conducting these experiments stemmed from a desire to further understand the nature of the sensitivity of gap detection to marker spatial separation described by Boehnke and Phillips (1999). This chapter will begin with a review of gap detection studies in which the markers are made dissimilar in a number of perceptual dimensions, including spatial location. This will be followed by a description of five experiments conducted to determine the spatial stimulus attributes that underlie this sensitivity of gap detection to marker spatial separation. The first experiment was a large parametric study in which gap-detection thresholds were obtained using leading and trailing marker locations distributed throughout 360 degrees of azimuth. This experiment was completed monaurally, that is, with one ear blocked, and the data were compared with the previously published observations in the same listeners obtained under binaural conditions (Boehnke & Phillips, 1999). This allowed comparison of thresholds obtained when one or both ears are available to perform the task for the same pairs of sound source locations. With both ears available, spatial separation across the midline resulted in the greatest degradation in performance, with minimal effect of spatial separation within a hemifield; i.e. there was a broad hemifield "tuning". When only one ear was available, there was little change in the function of gap threshold with trailing marker location for leading markers located in the hemifield of the open ear. However, for leading markers on the midline or in the blocked ear hemifield, performance *improved* under several conditions with ear blocking. This initially seemed to indicate a binaural (central) basis for the effect of spatial separation. Further examination, however, indicated that the relative level of the markers at each ear, which are location-dependent because of shadowing by the head, could account for much of the data. The binaural results are shown to parallel the results of free-field release from masking studies, in which detection is improved with separation of a signal from a noise masker (Saberi *et al.*, 1991). Moving the signal away from the masker might improve detection, not because the signal and noise activate different binaural channels, but because the S/N ratio at the ear nearest the signal is increased as a result of shadowing of the noise energy by the head.

In a second study, a subset of conditions from the first experiment was repeated under binaural conditions using noises that were either low-pass or high-pass in the frequency domain. High frequencies are more subject to the head shadow, and thus generate level differences between the markers (at either ear) as a function of spatial separation. The effect of spatial separation of gap markers was greater for high frequency noise than for low frequency noise. This implicates either interaural level differences (binaural) or level differences between the markers at either ear alone (monaural) as dominant over the binaural ITD cue in the effect of spatial separation using wideband noise markers observed in experiment 1. The results of this study also parallel those from studies examining release from masking with spatial separation using stimuli that are low or high-pass in the frequency domain.

The first two studies used a short leading marker, since the effect of spatial separation was previously found to be enhanced at shorter leading marker durations (Phillips *et al.*, 1998). Oxenham (2000) investigated the effect of spatial separation on gap thresholds by lateralizing the markers to opposite sides of the head with ILDs or ITDs. He found an effect of ILD but not ITD, and no leading marker effect for ILD. He also measured thresholds for the ILD stimulus through either headphone channel alone (i.e., louder leading marker, or louder trailing marker). These monaural conditions increased thresholds to the same degree as the ILD conditions, and did not show a leading marker duration effect. In experiment 3, the effect of leading marker duration was compared for conditions where the markers were lateralized to different sides using ILDs, presented monaurally at different levels, or presented from free-field locations in opposite hemifields. No listeners showed a leading marker duration effect for ILD or monaural level differences. Unpracticed, but not practiced, listeners showed an effect of leading marker duration with the spatially separated free-field markers. In experiment 3B, a better-controlled replication of the conditions tested by Oxenham (2000) was completed. Thresholds were obtained with a short and long leading marker duration, with markers presented dichotically over headphones to different ears, at different ILDs or ITDs, or presented monaurally with a level difference. The results confirm that separating the two sources through binaural channels coding different ITDs has no effect on gap detection acuity, consistent with the results from experiment 2 using low-frequency noise markers. Separating the sources using an interaural level difference alone (akin to that caused by the head shadow) affected gap detection acuity. The specifically binaural component of responses to the ILD stimuli could not be teased apart from the level differences available at each ear; i.e. thresholds were also elevated when a loud sound is followed by a quiet sound at a single ear.

The results of Experiment 1 did not examine the effect of leading marker duration on free-field spatial separation effects under ear-blocked conditions. There were also two possible confounds in the examination of the effect of ear-blocking in Experiment 1. The binaural conditions were performed before the ear-blocked conditions, and thus the absolute values of the thresholds could be subject to practice effects. Furthermore, ear-blocking is not completely effective for the lowest frequency components. In an attempt to resolve these issues, a fourth experiment was performed in which the free-field gap stimulus was emulated over headphones using 'virtual auditory space' methods. Gap detection thresholds were then obtained for these stimuli presented over headphones to both ears, or to the right or left headphone channel only, in a mixed order. The results for the conditions with short leading markers were congruent with the results from Experiment 1. However, the results with longer leading marker durations were not. This suggests that the effects of ear blocking on spatial effects in gap detection were, in part, dependent on the leading marker duration.

In order to delineate the factors contributing to elevated gap thresholds observed with free-field separation in the previous studies, a final study was conducted which varied all factors using the same listeners for each condition. The noise markers were wideband, high-pass or low-pass and the leading marker duration was 10, 50 or 100 ms. The leading marker was always positioned at 90° to the right, and gap thresholds were obtained for trailing markers at –90°, 0°, and 90° locations in azimuth. The results basically replicated those from previous experiments, and confirmed that the benefit observed with ear-blocking in Experiment 1 for spatially separated markers only holds for short leading markers, and probably is an expression of temporal masking phenomena.

The results of the experiments in Chapter 2 indicated that differences in purely binaural cues between gap markers had little effect on thresholds in the absence of peripheral differences between the two sounds. In fact, much of the pattern of gap thresholds obtained as a function of spatial separation between the markers could be accounted for by the marker level difference at one ear caused by the head shadow, and known effects of frequency differences between the markers resulting from the frequency-dependent shadowing by the head. In other words, within-channel temporal processes, and known effects of marker spectral differences, could account for much of the data.

INTRODUCTION TO GAP DETECTION

As described briefly in the general introduction, gap detection is one way to measure the temporal resolving properties of the auditory system. In a simple case, that of a brief silent period imposed within an ongoing piece of broadband noise, the stimulus imposes a brief decrease in activity in all frequency channels coding for that noise. Thus, it is a simple case of detecting a discontinuity in stimulus level *within any given channel.* It is typically measured using a two-alternative, forced choice task. Listeners are presented in a random order with a 'standard' stimulus that has no gap, and a 'gap' stimulus that is identical except for the presence of a gap. The listener indicates the interval of the 'gap' stimulus, and the threshold gap duration is then determined using an adaptive tracking procedure (after Levitt, 1971). Before discussing the effects on gap detection of making markers dissimilar in some dimension like spatial location, the limits on gap detection thresholds that can be unambiguously characterized as a result of 'within-channel' and 'across-channel' processes should be discussed. As a benchmark, note that the best gap detection performance (1.6-3ms: Plomp, 1964; Forrest & Green, 1987) occurs for a gap embedded in wideband markers of similar long duration.

Between-channel gap detection (Phillips *et al.*, 1997) is a paradigm based on the observation that detection of silent intervals (gaps) between sounds activating the same neural representation is up to an order of magnitude more acute than is detection of a silent period bounded by sounds activating different neural representations. If the sounds occurring immediately before and after a gap ("markers") are the same, then the minimum detectable gap can be as short as a few ms (Plomp, 1964; Fitzgibbons & Wightman, 1982; Fitzgibbons, 1983; Shailer & Moore, 1983; Florentine & Buus, 1984; Phillips *et al.*, 1997). This "within-channel" gap detection threshold is likely based on our ability to detect a decrement in level (which represents the gap at threshold) after being smeared through by a sliding temporal integrating window (Moore *et al.*, 1988). If, however, the gap's markers differ in some perceptual dimensions, then the shortest detectable gap is often lengthened to 10-50 ms.

There are a number of perceptual dimensions in audition that can be exploited in the between-channel paradigm. This poorer acuity of between-channel gap thresholds is observed whether the markers are presented to different ears, or to different regions of the basilar membrane (Phillips *et al.*, 1997; Formby *et al.*, 1998a). There is also an acuity difference if the markers are perceived to be of the same or different pitch (Chatterjee *et al.*, 1998; Oxenham, 2000), are noise markers presented from the same or different spatial locations (Phillips *et al.*, 1998; Boehnke & Phillips, 1999), are noises of different interaural level differences or are noises of different levels (loud-quiet or quiet-loud)(Oxenham, 2000).There are some cases where stimulus differences leading to a perceptual disparity between the two markers does NOT cause an elevation of gap thresholds: when the markers are distinguished by interaural time difference (ITD) (Oxenham, 2000) or perceived pitch emanating from centrally generated pitches, e.g. amplitude modulation rates or iterated rippled noises of different delays (Grose *et al.*, 2001b). These are both conditions in which activity at the periphery is basically identical at all frequency regions and at both ears. In this case, one hears a smooth perceptual shift in lateralization or pitch in the standard with no additional onset to the trailing marker as occurs in all other between-channel conditions. The perceptual shift can easily be ignored and the gap is detected by listening for the interval that has an onset to the trailing marker.

This elevation in thresholds in 'between-channel' gap detection experiments has been seen as a way to explore higher levels of auditory perceptual organization (Phillips *et al.*, 1997). However, it is important that the peripheral influences on gap detection are separated from these central influences if any conclusions can be drawn. Note that in the case where the markers are presented to different ears or two widely disparate regions of the cochlea, there are no peripheral mechanisms capable of "detecting" the gap, so detection must be mediated by a central mechanism, possibly a relative timing operation between the offset of activity in one channel and the onset of activity in another (after Phillips *et al.*, 1997). In cases where the markers are distinguished by their monaural level, peripheral mechanisms, possibly the same ones that result in temporal masking phenomena, may limit the fidelity of the gap represented at the auditory nerve and therefore the temporal fidelity of the signal available to a central (binaural) processor. The cases that fall in between these extremes are those where the markers differ in their ILD, spatial location or complex pitch. For such 'intermediate' stimuli, there is a high degree of overlap in the populations of cells responding to the two markers at early stages in the auditory system, but as these stimuli are further coded at higher levels in the system, the overlap in the cell populations responding to the two markers decreases. Most of the experiments described in this chapter concern the case of markers that differ in actual free-field spatial location - arguably the most complex intermediate case.

Following the first demonstration that markers at free-field locations in opposite hemifields elevated gap detection thresholds (Phillips *et al.*, 1998), Boehnke & Phillips (1999) performed a "between-channel gap detection" experiment with markers at a series of different free-field locations. Separation of the markers within the left or right hemifield had little effect on the size of gap thresholds, while separation across the midline resulted in elevated thresholds. Note that the changes in gap threshold were independent of absolute separation between the markers. For example, a 60° separation of markers within a hemifield did not elevate thresholds, but a 60° separation across the midline did. It was this pattern that suggested the existence of two hemifield-tuned spatial perceptual channels. This *pattern* of thresholds across marker locations, i.e., relatively undifferentiated thresholds when both markers were within a given hemifield, and a sharp increase in thresholds as the trailing marker location crossed the midline, was similar to the broad hemifield spatial tuning of central neurons (Benson *et al.*, 1981; Middlebrooks & Pettigrew, 1981; Imig *et al.*, 1990; Rajan *et al.*, 1990; Ahissar *et al.*, 1992; Clarey *et al.*, 1995; Brugge *et al.*, 1996, 1998; Eggermont & Mossop, 1998; Middlebrooks *et al.*, 1998; Furukawa *et al.*, 2000; Recanzone *et al.*, 2000; Furukawa & Middlebrooks, 2001). Since the vast majority of central auditory neurons are binaural, a binaural contribution to the genesis of the channels might be reasoned to exist by analogy (see Teller, 1984). Thus, Boehnke and Phillips (1999) assumed that the detection of the 'gap' by some decision agent is accomplished based on the output of these central (probably cortical) representations of the markers.

However, there are a number of monaural stimulus characteristics, many of which can be linked to peripheral processing limitations, which are known to elevate gap detection thresholds. As the bandwidth of the markers is decreased, gap detection thresholds become elevated, probably reflecting that: fewer channels are providing information about the presence of the level disparity, thereby making its presence more ambiguous; and decreases in bandwidth increase amplitude fluctuations in the marker, making it harder to distinguish a gap from an amplitude fluctuation (Fitzgibbons, 1983; Shailer & Moore, 1983; Eddins *et al.*, 1992). Imposing a level difference between the markers, so that the leading marker is louder (LQ) (Plomp, 1964; Oxenham, 2000) or quieter (QL) (Oxenham, 2000) than the trailing marker, impairs gap detection performance (see Fig. 1.5, 3rd panel B&C). Plomp obtained functions of performance (log Δt ms) with the level of the trailing marker (dB) and found that performance degraded linearly as a function of the level disparity. This was taken as evidence for the decay of auditory sensation – thresholds were indicative of the time for the leading marker loudness to decay to the same level as the trailing marker. Interestingly, the inverse situation – a quiet sound followed by a loud sound (QL) – appears to follow a similar pattern (Oxenham, 2000) although a systematic study of this has not been completed. In this case the trailing marker is already louder than the leading marker, so the longer gap required can not represent the time required for the quiet leading marker to decay in intensity. Thus, it may not be the decay per se that impairs gap detection, but some other function of the level disparity.

Many of these characteristics are present in a gap detection task using free-field markers at different eccentricities. Such markers will, through head and pinna shadowing, project to each ear separately, markers of different amplitude particularly across the higher frequencies. It follows from this that gap markers located in different hemifields will also have different spectral characteristics at each ear. The representation of the gap stimulus at the ear nearest the leading marker will be a broadband, relatively loud leading sound followed by a predominantly low-pass sound with relatively low intensity high frequency components (due to the head-shadowing effects). The representation at the opposite ear will be the reverse. We know independently that gap detection thresholds increase as the stimulus bandwidth is reduced (Fitzgibbons, 1983; Shailer & Moore, 1983; Eddins *et al.*, 1992), if the level of gap markers differ (Plomp, 1964; Penner, 1977; Oxenham, 2000), and if the overall stimulus level is decreased below 30 dB (Fitzgibbons, 1983; Shailer & Moore, 1983). Other known stimulus characteristics that increase gap detection thresholds include spectral separation of the markers (Kinney, 1961; Perrott & Williams, 1971; Collyer, 1974; Divenyi & Hirsh, 1974; Fitzgibbons *et al.*, 1974; Divenyi & Danner, 1977; Formby & Forrest, 1991; Formby *et al.*, 1993; Formby *et al.*, 1996a; Hall *et al.*, 1996; Phillips *et al.*, 1997; Formby *et al.*, 1998a; Formby *et al.*, 1998b; Hanekom & Shannon, 1998; Grose *et al.*, 1999; Taylor *et al.*, 1999; Phillips & Hall, 2000; Grose *et al.*, 2001a; Grose *et al.*, 2001b; Lister *et al.*, 2002). Formby (1998b) more recently demonstrated that the presence of additional frequency information in the trailing marker, but not the leading marker, increased gap thresholds relative to markers of equal frequency content. The spatial separation of markers may have thus exerted its effect on gap thresholds through these effects, rather than through spatial separation *per se* and the differences in the output of binaural interactions that arise from sounds at different locations. This view is bolstered by the observation that gap thresholds are not affected by changes in ITD (a purely binaural cue) between the markers (Oxenham, 2000).

The following five experiments are presented roughly in the order in which they were completed. Taken as a whole, the experiments probe the sensitivity of gap detection thresholds to spatial separation of gap markers for all possible spatial stimulus attributes (ear of entry, ILD, ITD, monaural spectral/level differences) over headphones and in the free-field. Table 2.1 presents a summary of spatial stimulus attributes by which gap markers can be made to differ. In experiments in which stimuli are presented over headphones, it is known that separation of gap markers by ear of entry (EAR) (Penner, 1977; Phillips *et al.*, 1997; Formby *et al.*, 1998a), ILD (Oxenham, 2000), and monaural level (Plomp, 1964; Penner, 1977; Oxenham, 2000) elevate gap detection thresholds, while separation by ITD does not (Oxenham, 2000). The failure of ITD to elevate thresholds under headphone listening could be due to the non-ecological nature of headphone listening, i.e. the fact that the spatial percepts are internalized in intracranial space, or that dichotic listening offers a unique (and non-ecological) opportunity for the listener to access information at each ear separately. It is also known that separation of wideband noise markers by free-field location elevates thresholds. This effect could be due to the ILD differences in the high frequency channels, the ITD differences in the low-frequency channels, differences in the level of the gap markers at either of the ears alone (monaural level differences). Across the five experiments in this chapter, each possibility is examined.

Table 2.1

Spatial stimulus attributes by which gap markers could be made different, and predicted results. The cues can be completely isolated when presented over headphones. In the free-field, we can manipulate which attribute is dominant by changing the marker spectral content. High-pass markers that are spatially separated will differ predominantly in interaural and monaural level, while low-pass markers will differ in ITD, but not ILD or monaural level. Across experiments 1-5, all conditions were probed.

HEADPHONE TASK	EAR	ILD	ITD	Mono LQ/QL	Diotic (within)
Gap Detection Exp. 3	Strongly Impaired	Impaired	Not impaired (Oxenham, 2000)	Impaired (Plomp, Penner, Oxenham)	Control (2-3ms)
FREEFIELD TASK	Wide Band (all) 90/-90	Hi-Pass (ILD only) 90/-90	Low-Pass (ITD only) 90/-90	Ear Block (LQ or QL) 90/-90	Same Location 90/90
Gap Detection Exp. 1,2,4,5	Impaired (Boehnke & Phillips)	Strongly Impaired (?)	Minimally impaired (?)	Impaired (?)	Control (2-3ms)

EXPERIMENT 1:

THE EFFECT OF EAR BLOCKING ON GAP DETECTION PERFORMANCE USING MARKERS AT DIFFERENT FREE-FIELD LOCATIONS

One seemingly obvious way to investigate any contribution of binaural cues in the pattern of thresholds obtained for markers with different free-field locations is to remove them by blocking one ear. In general, ear blocking of this type causes listeners to mislocalize sounds from the hemifield of the blocked ear to the hemifield of the open ear; localization of sounds within the hemifield of the open ear is relatively poor compared to binaural conditions (Oldfield & Parker, 1986; Slattery & Middlebrooks, 1994). This is probably because ear blocking does not remove interaural differences but simply creates extremely large differences in favor of the unplugged ear regardless of location of the stimulus – hence, all sounds are localized to that side (Wightman & Kistler, 1997b). Ear blocking has also been shown to cause the spatial receptive fields of cortical neurons in the cat to become even more broadly tuned across azimuth (Samson *et al.*, 1993; Samson *et al.*, 1994). If gap detection between spatial locations was based on the output of cortically-based channels, thresholds might improve with ear-blocking due the increased overlap in neural populations activated by the two markers. Blocking one ear also allows determination of the limitations on gap detection for each ear for a given pair of marker eccentricities. Furthermore, a comparison of thresholds obtained through each ear alone with those obtained with both ears should indicate whether we are limited by the 'best ear', the 'worst ear', or some combination of the two under binaural listening.

In Experiment 1, gap detection thresholds were obtained for combinations of 10 leading marker and 18 trailing marker locations distributed throughout the azimuth with one ear blocked. All listeners had previously completed these conditions under binaural listening, the data for three of whom have been reported (Boehnke & Phillips, 1999), and are used for comparison here. The data obtained under monaural conditions have been presented in preliminary form (Boehnke & Phillips, 2000). These experiments were completed with a short leading marker duration (10ms) because previous studies (Phillips *et al.*, 1997; Phillips *et al.*, 1998) had shown an optimal 'between-channel' effect with short markers.

GENERAL METHODS

The details of the stimulus creation and presentation are different for each study and will be described in detail separately for each experiment. There are, however, some apparati and procedures common to most or all of the experiments. These will be described here once for efficiency.

FREE-FIELD STIMULUS PRESENTATION

The 'free-field' room in the Hearing Research Lab at Dalhousie University Psychology Department is an Eckel sound-attenuating chamber (95-in wide, 83-in deep, and 80-in high internal measurements), whose internal walls were lined with 2-in. thick acoustic foam wedges (Auralex) (see Fig. 2.1). All experiments conducted in the free-field room, except Exp.1 and Exp. 3A (of Chapter 2), were controlled by the MATSS for Matlab presentation system (Multi-modal/Multi-channel Apparatus for the Temporal Sequencing of Stimuli for use with Matlab Software). Briefly, digital stimuli were generated and presented through Matlab software (The Mathworks) running on a Power Macintosh G3 desktop computer. Matlab is an interpretive, matrix-based programming language that is used by many auditory psychophysicists because of the ease with which digital sounds can be produced and manipulated within matrices. Signals were 'streamed' through eight ASIO (audio streaming input/output protocol, www.steinberg.org) memory buffers to an ECHO 2-input/8-output digital signal-processing card. The digital signals were converted to analog by the ECHO card and then passed through a custom-made amplification system that is comprised of 8 National Semiconductor Audio Power Amplifiers (56Watt) and 8 Dallas Semiconductor Potentiometers (65 dB range in 1 dB steps, log resistance). The amplifiers were powered by a homemade power supply. The signal was passed from the eight amplifiers to eight speakers (*Optimus* PRO x44 Diets AV 2-Way Speakers; RadioShack, 40-2080). The 10-cm speakers were 8 ohm, dual design, and 50-watt tweeter/woofer pairs encased in a metal box with a frequency response of .09 to 20kHz. The speakers were situated on telescoping microphone stands whose elevation could be varied (min=85cm; max=157cm). Listeners were seated on a chair in the middle of the array such that speakers were equidistant from the center of the

FREE-FIELD ROOM SETUP FOR STIMULUS PRESENTATION

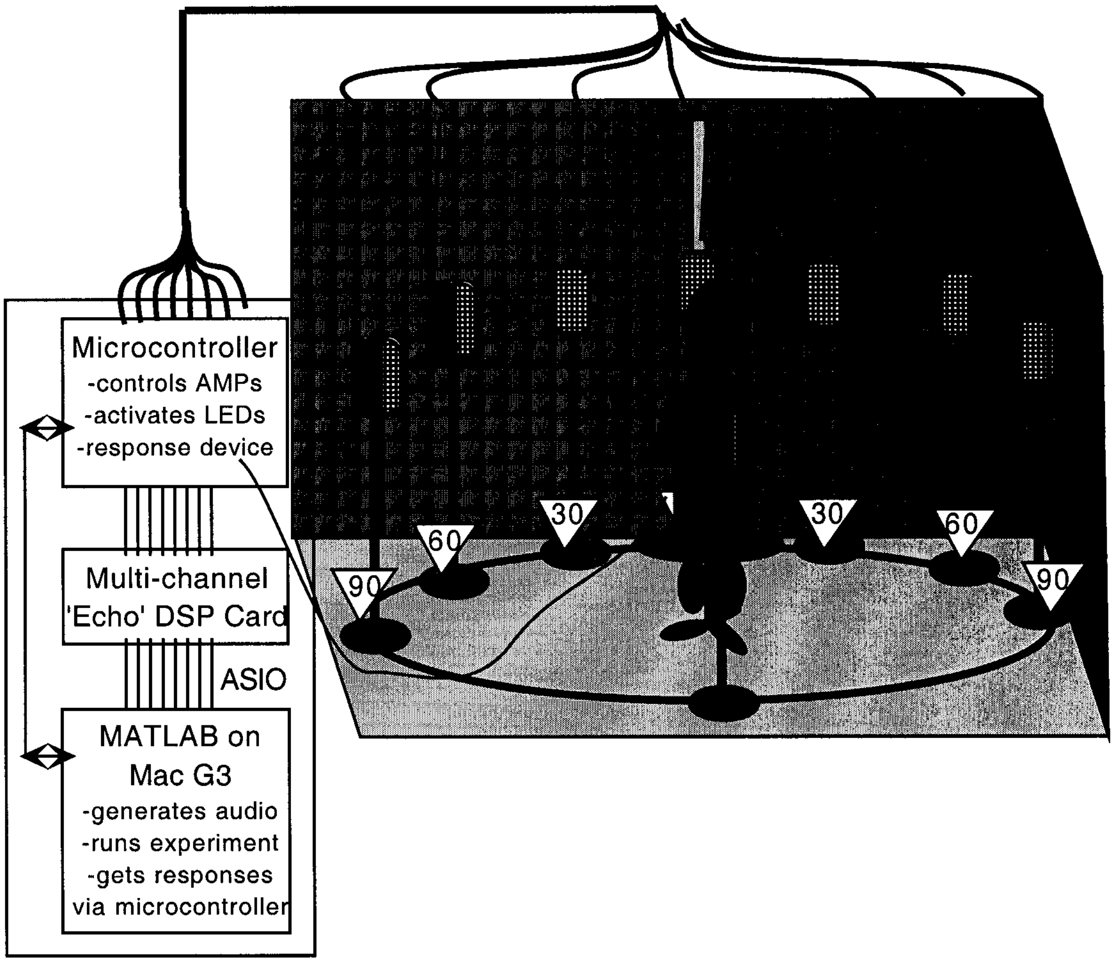


Figure 2.1
MATSS (Multi-channel/Multi-modal Apparatus for the Temporal Sequencing of Stimuli) for use with Matlab software. Schematic of the free-field room used in all experiments for which free-field stimulation was used, except Exp.'s 1 and 3a in Chapter 2. To the left is a schematic of the software and hardware components that comprise the MATSS system. The signal is generated in Matlab, sent to any combination of the 8 speakers via a multi-channel sound card and homemade, configurable amplification system. Temporally precise sequencing of sounds, LED activation, and collection of response times is co-ordinated by a PIC microcontroller.

head. Listeners rested their chin on a telescopic chin rest affixed to the chair, in order to minimize head movements. More detailed specifications are available in the Technical report entitled "*Technical specifications of MATSS for Matlab: Multi-modal/Multi-channel Apparatus for the Temporal Sequencing of Stimuli for use with Matlab Software*" available from the Hearing Research Laboratory, Dalhousie University or the present author.

HEADPHONE STIMULUS PRESENTATION

All experiments involving headphones (except Exp. 2.3) were conducted in a separate Eckel sound attenuating chamber. Stimuli were generated and presented by Matlab software running on a Power Macintosh 8600. Digitized signals were converted to analog through the built-in DA converter (sound output) of the Macintosh, and transduced through AudioTechnica ATH-M40fs Precision Studiophones or Sennheiser HD-25 headphones. Sound levels were determined by manipulating the signal intensity through Matlab, and were presented between 60-70 dB. The exact value used in a given experiment is specified in the methods section of each experiment. Absolute sound levels in SPL (re: 20 μPa; A-weighted) were determined by recording the experimental stimulus via a Brüel & Kjaer condenser microphone (model #4145) through a Brüel & Kjaer artificial ear (model #4152). The microphone output was analyzed with a Brüel & Kjaer sound level meter (model #2203).

Threshold determination procedures

Thresholds were typically obtained using two-interval, two-alternative forced-choice procedure (2I-2AFC) (except Chapter 4, Exp.1 and Exp.3A; and Chapter 5, Exp. 2). Listeners were presented with two stimuli – a 'standard' and a 'test' - in random order separated by an inter-stimulus interval of between 300-600ms. The task of the listener was to indicate the interval that contained the 'test' stimulus by pressing the '1' or '2' key on a computer keyboard, or a button box.

Thresholds for detection of the signal were determined in a block of trials using an adaptive tracking paradigm (after Levitt, 1971). There were two versions of this paradigm used in this research.

1. “2 Down, 1 Up” adaptive tracking procedure.

After every two successive correct responses, the size of the adaptive variable (e.g., the duration of a silent period) was decreased by a factor of 1.2, and after every incorrect response, the size of the adaptive variable was increased by the same factor. Each adaptive staircase continued for 14 reversals in the direction of change of the adaptive variable. The threshold for that block of trials was taken as the mean value across the last 10 trials that produced reversals in the adaptive step. This procedure tracks the 70.4% correct point on the psychometric function (Levitt, 1971).

2. “3 Down, 1 Up” adaptive tracking procedure

Until the first incorrect response, the size of the adaptive variable was decreased by a factor of 1.2 for each step. Subsequently, the size of the adaptive variable was increased by a factor of 1.2 after each incorrect response, or decreased by a factor of 1.2 after three consecutive correct responses. Each adaptive staircase continued for eight reversals and the mean of values for the adaptive variable for the last six reversals was defined as the detection threshold. This procedure tracks the 79.4% correct point on the psychometric function (Levitt, 1971).

Unless otherwise specified, at least three threshold determinations were obtained for each stimulus condition, and the listeners reported threshold was the mean of their last three thresholds measured for that condition. An exception to this was Experiment 1, where only two thresholds were obtained for each condition because of the large number of stimulus conditions.

METHODS SPECIFIC TO EXPERIMENT 1

Subjects

Results with ear-blocked listening were obtained from one untrained and three trained listeners. The trained listeners had participated in a variety of psychophysical experiments and in the earlier binaural study (Boehnke & Phillips, 1999). Their data from the earlier study are used for comparison with the monaural data obtained here. The untrained listener was recruited to complete the experiment first under binaural, then under monaural conditions. The listeners ranged in age from 22-33 years. All listeners had free-field noise detection thresholds within laboratory norms, and the three trained listeners were tested for audiometric sensitivity and were shown to have audiologically normal pure tone thresholds to at least 6.0 kHz (GSI 17 Audiometer). All participants knew the purpose of the experiment, and all except the author were paid for their participation.

Gap Stimuli and Apparatus

The experiment was conducted *prior* to implementation of the MATSS system (see Common Methods), and thus the method will be described here in detail. A custom-programmed Macintosh IIfx computer controlled stimulus presentation and data acquisition. All stimuli were wideband (20 kHz) noise bursts generated by a white noise source (Colbourn Instruments White Noise Generator, model S81-02), linearly ramped with 0.5 ms rise-fall times, including those defining the gap. The gap duration itself was defined as the length of the silent interval excluding rise/fall times. The gated stimuli were passed through a passive attenuator (Hewlett Packard 350D attenuator set), and a power amplifier (Realistic SA-155 integrated stereo amplifier) and transduced by two Optimus speakers (XTS 3640-1994) with a frequency response of 0.14-20.0 kHz. Each stimulus consisted of a leading marker that was 10 ms in duration and a trailing marker that was 300 ms in duration. The short leading marker was used because it was demonstrated in Phillips et al. (1997, 1998) that in between-channel designs, there was a greater elevation of gap thresholds for shorter leading marker durations, including the specific case of conditions where the markers were spatially separated. That is, a shorter marker was used to optimize stimulus conditions for visualizing the between-channel effect. The two speakers were suspended at approximately ear level by inflexible metal rods from a 70-inch diameter circular track mounted on the ceiling of an Eckel sound-attenuating room (95-in wide, 83-in deep, and 80-in high internal measurements), that was covered with 2-in. thick acoustic foam wedges (Auralex) (see Fig. 2.2A). In an attempt to control for head movement in the monaural condition, the subject's head rested on a chin rest such that it was centered in the room and by definition faced 0° azimuth. The listeners could see the two speakers being positioned for each block of trials and were therefore always aware of the location of the two markers.

Procedure

Sensation Level Determination

With the listener seated in the center of the room, noise detection thresholds were obtained by the method of limits using 300-ms noise bursts from a speaker located at 0° azimuth. This was repeated for both speakers used in the experiment. There were no differences between thresholds for stimuli delivered through the two speakers. For each subject, all experimental trials were presented at 35 dB sensation level, based on thresholds obtained at the 0° azimuth location.

Achieving Monaural Conditions

During the experiment, the less-preferred ear was plugged. The blocked ear will always be referred to in Figures as the left ear, although one listener preferred to block the right ear. These data for this listener were re-coded and presented in each Figure as if he had the left ear blocked. This is justified since there was no left-right asymmetry in the pattern or absolute value of thresholds obtained in the study of Boehnke and Phillips (1999). To achieve ear-blocking, the listener inserted a 3M 1110 Foam insert ear plug, with an NRR rating of 29 dB, as deeply into the ear canal as was safely possible. The listener also wore an industrial sound attenuating muff over the plugged ear (American Allsafe Company, Model HS-2600) with an Noise Reduction Rating (NRR) of 26 dB. In order to determine the amount of attenuation caused by this arrangement, detection thresholds were obtained at speaker locations of 0°, 180°, ±30°, and ± 90° under three conditions: both ears plugged simultaneously, less-preferred ear plugged only, and both ears unplugged. The

EXPERIMENT 1: EXPERIMENTAL SETUP AND STIMULUS

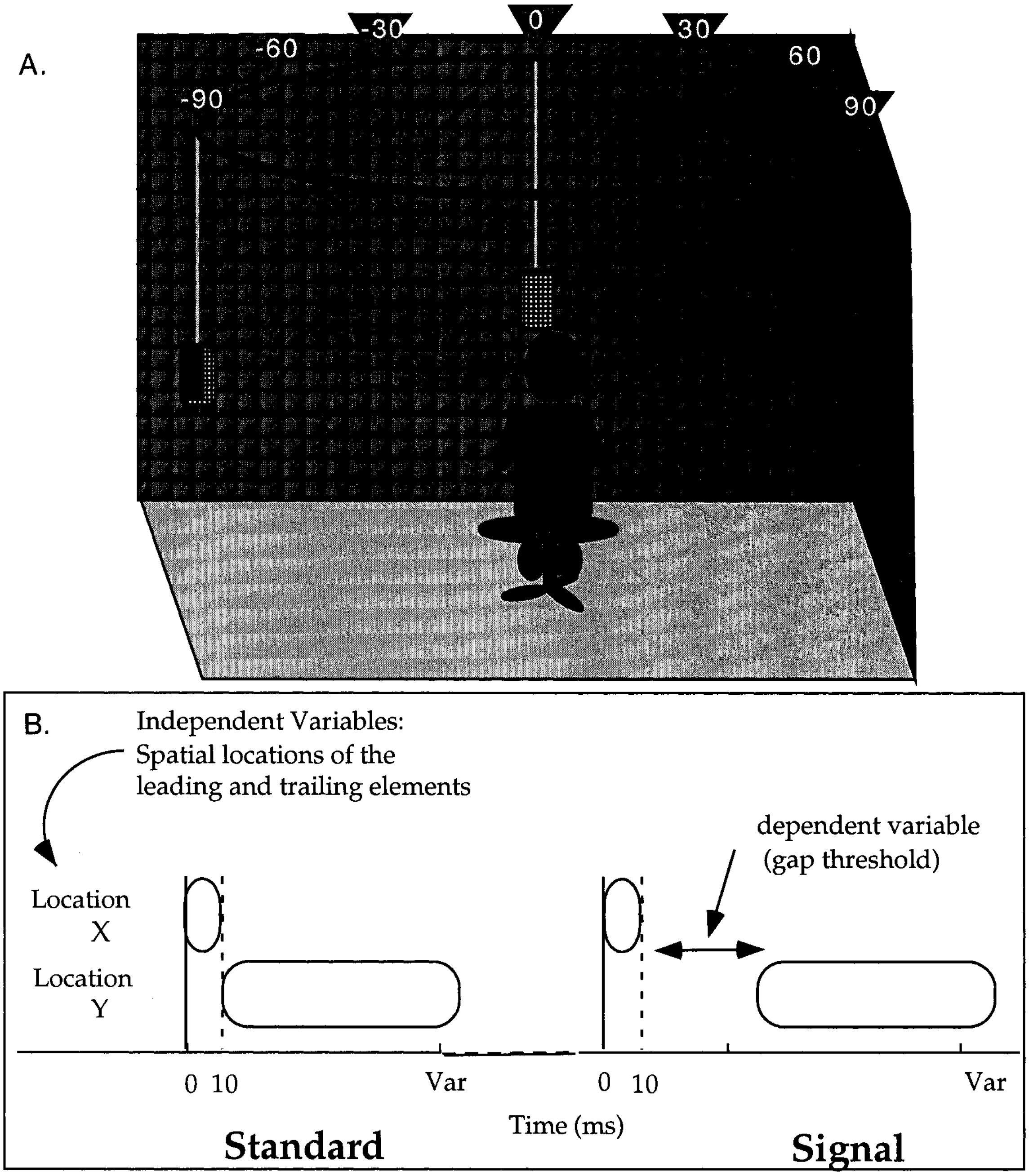


Figure 2.2

A. *Schematic of the free-field room used in Experiment 1 and 3a, see Fig. 2.2 for MATSS setup used in Experiments 2, 4, and 5.*
B. *Schematic of gap detection stimuli used in Experiment. 1 and typical of gap detection stimuli generally used throughout this chapter.*

amount of attenuation caused by the plug and muff at a given location was measured as Unplugged – Both Plugged, and, collapsed across location, the mean attenuation (±SEM) was 44 ±1.7 dB for the 4 listeners. In the experiment that follows, the sensation level at which stimuli were presented to the listeners was a few dB less than their threshold when both ears were plugged. With one ear blocked, this corresponded to a sensation level of between approximately 30 and 40 dB at the open ear, and was direction dependent, i.e. the sensation level was lowest for speaker locations lateral to the blocked ear. However, in the open ear's acoustic hemifield, these direction-dependent level differences were not likely a concern because gap detection has been shown to be insensitive to stimulus level for sensation levels above 30 dB (see Eddins & Green, 1995).

The Gap Detection Task

Each stimulus trial consisted of two sequences of noise bursts, separated by an interval of 300 ms (see Fig. 2.2B). One of the two sequences of noise - *the signal*, contained a detectable silent period (gap) between the leading and trailing marker. The other sequence was *the standard*, which included an undetectable 1.0-ms silent period between the markers (to control for gating transients present in the signal sequence). Gap thresholds for every stimulus condition were obtained in blocks of two-interval, two-alternative forced-choice (2I-2AFC) trials without feedback, using a two-down/one-up adaptive paradigm (after Levitt, 1971).

There were 10 locations of the leading marker (0°, ±30°, ±60°, ±90°, ±150°, and 180°), and for each leading marker condition, gap thresholds were obtained with trailing markers at each of 18 locations (0°, ±15°, ±30°, ±45°, ±60°, ±90°, ±120°, ±150°, ±165°, and 180°). Gap detection thresholds were obtained for each condition by each subject (fully repeated measures). In the binaural experiment, the conditions with frontally-located leading markers (0° to ±90°) were completed first and collected in a pseudo-random order. The rear leading-marker conditions (±120°, ±150°, and ±180°) were completed later, also in random order. In the monaural repetition, data for the leading marker conditions (front or back) were collected in a random order. For each leading-marker condition, the order of trailing conditions tested began at a random location and progressed, in serial order, around the azimuth. The randomization procedure across all conditions was repeated 2-4 times until subject's gap thresholds had stabilized. The number of repetitions required usually depended on the condition, in that, conditions with similarly-located markers ("within-channel") usually required the fewest repetitions to reach stability. The subject's final gap threshold for each condition was the mean of those for the last three estimations in the binaural experiment. Subjects' thresholds were quite stable after completing the binaural experiments; thus, only two threshold determinations were obtained for most of the 180 different stimulus conditions completed in the monaural repetition.

RESULTS

In Figure 2.3, gap detection thresholds obtained from individual listeners for the 18 trailing marker locations across the full azimuth are plotted in separate panels for 6 representative leading marker locations (0°, 180°, -30°, 30°, -90°, 90°). The grand mean data are shown in Figure 2.4. In each panel of Figure 2.3, the mean of the 3 threshold determinations obtained under binaural conditions (except in the new listener JH for whom only 2 thresholds were obtained) are plotted, with standard errors, as open square symbols. Plotted with these are the two threshold determinations obtained under monaural conditions (filled and open circles respectively) with a line drawn through the mean of the two (x-symbols). For convenience, the description that follows refers only to the grand mean data shown in Figure 2.4. A comparison of the appropriate mean and grand-mean data sets usually reveals that listeners followed the same general patterns of behavior, although absolute thresholds differ widely. Figure 2.4 shows the functions relating the mean gap threshold (all listeners) to trailing marker location obtained for each leading marker condition. The functions obtained under binaural and monaural conditions for a given leading marker location are always plotted together for comparison. The functions for leading markers on the midsaggital plane (center; panel B); in the acoustic hemifield of the occluded ear (left; panels A&D); and the hemifield of the open ear (right; panels C&E) are presented separately.

The effect of trailing marker location on gap thresholds differed for the various leading marker conditions, and between some binaural and monaural conditions. Under monaural conditions, the functions relating gap threshold to trailing marker location had the same general shape for all leading marker locations: gap thresholds were low for

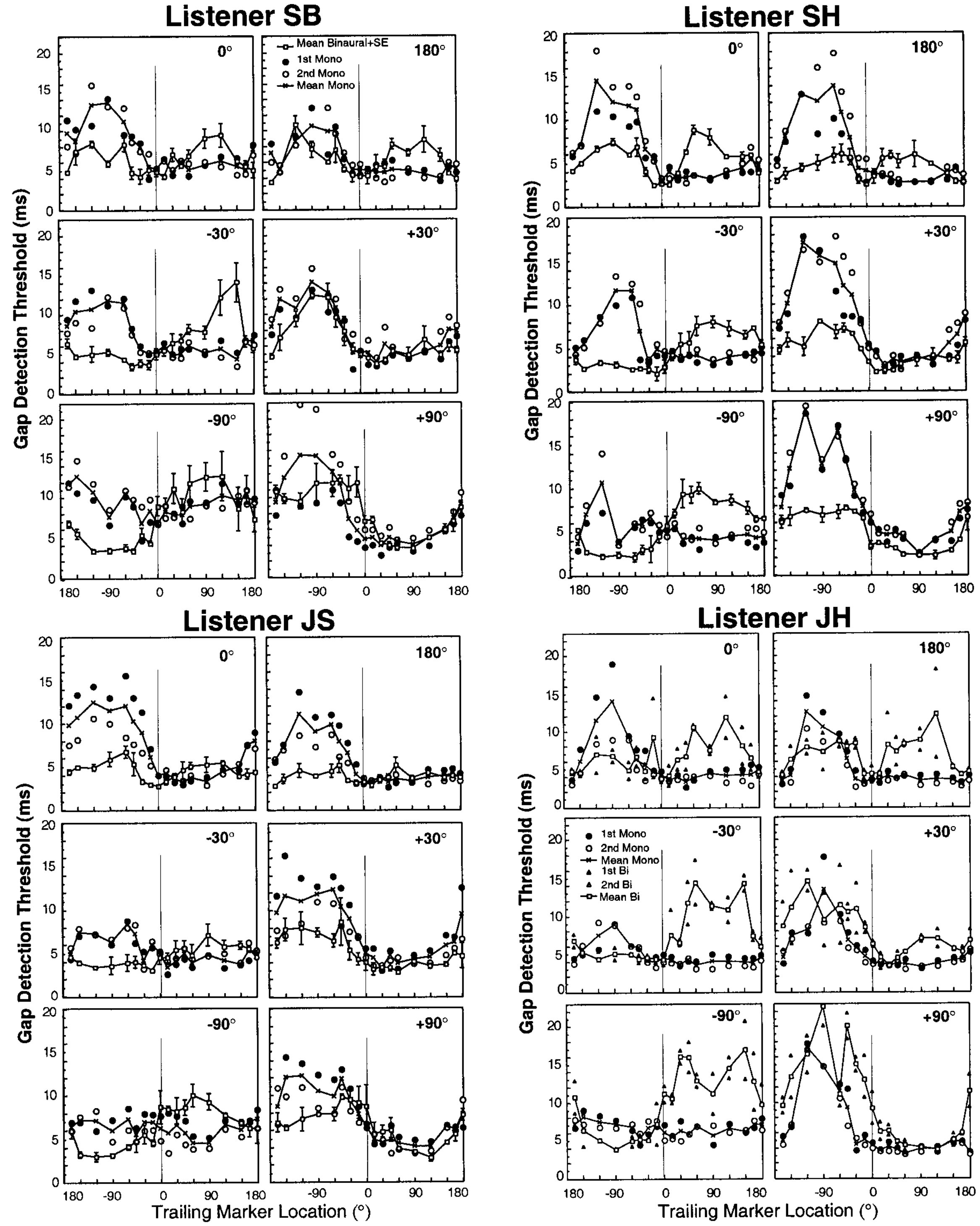
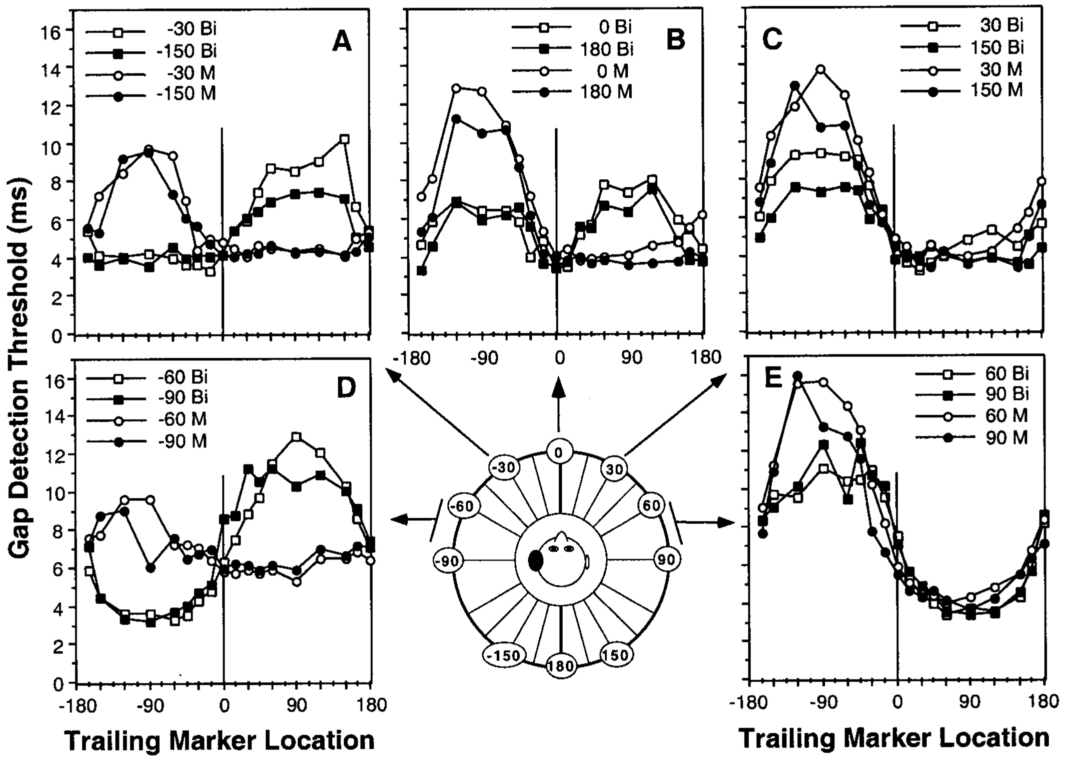


Figure 2.3

Gap thresholds obtained for individual listeners are plotted separately in the 4 sections of the figure. For each listener, the first and second gap thresholds obtained under monaural conditions are plotted as solid and open circles respectively and a line is plotted through the mean of those two estimates. The mean threshold of 3 estimates obtained under binaural listening (except in listener JH) are plotted as open squares with the standard error of the estimate for comparison. Each graph plots gap detection thresholds as a function of trailing markers location for 6 representative leading marker locations (0, 180, +30, -30, +90 and -90).



AZIMUTHAL FUNCTIONS FOR DIFFERENT LEADING MARKER LOCATIONS UNDER NORMAL (BI) AND EAR-BLOCKED(M) LISTENING

<u>Figure 2.4</u>

This figure compares mean gap thresholds obtained under binaural and monaural (ear-blocked) conditions for the same 4 listeners. In the lower center is a diagram of the azimuthal locations used for leading marker locations, with arrows pointing to the data for a given location. The functions show the dependence of gap threshold on trailing marker location under different listening conditions and for different leading marker locations. In all figures, circles and squares give the data obtained under binaural and monaural listening conditions, respectively. Frontal leading-marker locations are coded by open symbols, while rearmost leading marker locations are coded by filled symbols. See text for details.

A. *Functions for leading markers located 30° from the midsaggital plane in the hemifield of the blocked ear (left).*
B. *Functions for leading markers located on the midsaggital plane (0°, 180°).*
C. *Functions for leading markers located 30° from the midsaggital plane in the hemifield of the open ear (right).*
D. *Functions for leading markers located deep in the hemifield of the blocked (left) ear (60° and 90°).*
E. *Functions for leading markers located deep in the hemifield of the open (right) ear (60° and 90°).*

trailing markers throughout the open ear's hemifield and increased when trailing markers were in the blocked ear's hemifield. The functions for monaural conditions differed in systematic ways from those obtained for the same leading marker location under binaural conditions. Under binaural conditions, gap thresholds were low when trailing markers were in the same hemifield as the leading marker, and were raised when the trailing marker was in the hemifield contralateral to the leading marker. The single exception was for the midline leading marker condition, in which thresholds were raised when the trailing marker was at locations deep in either hemifield. Direct comparisons of each leading marker function obtained under binaural and monaural conditions are described individually below. Statistical comparisons were made with a repeated measures analysis of variance (ANOVA), using the Huynh-Feldt correction for violations of sphericity where appropriate (Huynh & Feldt, 1976).

Figure 2.4B shows that for binaural listening, the function of gap thresholds with trailing marker location obtained for leading markers on the midline had a characteristic "double-humped" shape. Gap thresholds were low when trailing marker locations were within 30° of the midsaggital plane and rose to asymptotically high values for trailing marker locations around the lateral poles. The elevation in gap thresholds that occurred for trailing markers at the lateral poles under binaural conditions was eliminated by ear blocking for locations in the hemifield of the good ear, and enhanced on the side of the blocked ear. This change in pattern was verified statistically by a significant interaction between listening condition (binaural, unilateral block) and trailing marker location for both midline leading marker locations, 0° [$F(17,51)=15.81$, $p<0.001$] and 180° [$F(17,51)=8.68$, $p<0.001$].

Figure 2.4A and 2.4D show the effect on gap threshold of trailing marker location, for leading marker locations in the hemifield of the *blocked* ear at 30° and 150°, and 60° and 90° degrees, respectively. In the functions obtained under binaural conditions, gap thresholds were low for all trailing marker locations in the same hemifield as the leading marker and thresholds rose to asymptotic levels for trailing markers deep in the opposite hemifield. Occluding the ear resulted in somewhat reversed functions. Thresholds were low for trailing markers spanning the entire hemifield of the open ear. These differences between the monaural and binaural functions were verified statistically by a significant interaction between listening condition and trailing marker location for leading marker locations at 30° [F(17,51)=15.74, p<0.001], 150° [F(17,51)=10.16, p<0.001], 60° [F(17,51)=20.69, p<0.001] and 90° [F(17,51)=16.28, p<0.001] in the hemifield of the occluded ear.

Figure 2.4C and 2.4E show the effect on gap threshold of trailing marker location, for leading marker locations in the hemifield of the *open* ear at 30° and 150°, and 60° and 90° degrees, respectively. These functions for binaural and monaural conditions for a given leading marker location were very similar. There was no interaction between the listening condition and trailing marker location for either the 30° [F(17,51)=1.76, ns], 60° [F(17,51)=2.42, ns], or 150° [F(17,51)=2.34, ns] leading marker. There was, however, a significant interaction between the functions obtained under monaural and binaural listening for the leading marker at 90° [F(17,51)=3.3, p=0.05]. This interaction could be accounted for by lower thresholds under monaural listening for locations across the midline region (note the downward displacement of the curve for 90° M in Fig. 2.4E). The thresholds were significantly lower for trailing markers at -15° [F(1,51)=6.4, p=0.05] and -30° [F(1,51)=8.6, p=0.04], indicating that under monaural listening, thresholds remained low across a broader region of space compared with the binaural listening for that condition.

DISCUSSION

The main results can be summarized as follows. In configurations where the leading marker was located laterally, monaural listening with the ear nearest to that leading marker resulted in gap detection threshold being a function of trailing marker location in a fashion similar to that seen with binaural listening. Monaural listening with the ear in the hemifield opposite to the leading marker location resulted in a somewhat reversed function of gap detection threshold with trailing marker location compared with that obtained for binaural listening. That is, thresholds were worse when both markers were in the blocked ear hemifield, likely because they were quiet and low-pass in nature, but better when the leading marker was in opposite hemifield. Finally, when the leading marker was located on the midline, thresholds were low for all trailing marker locations in the hemifield of the available ear, and elevated for trailing marker locations in the hemifield of the blocked ear.

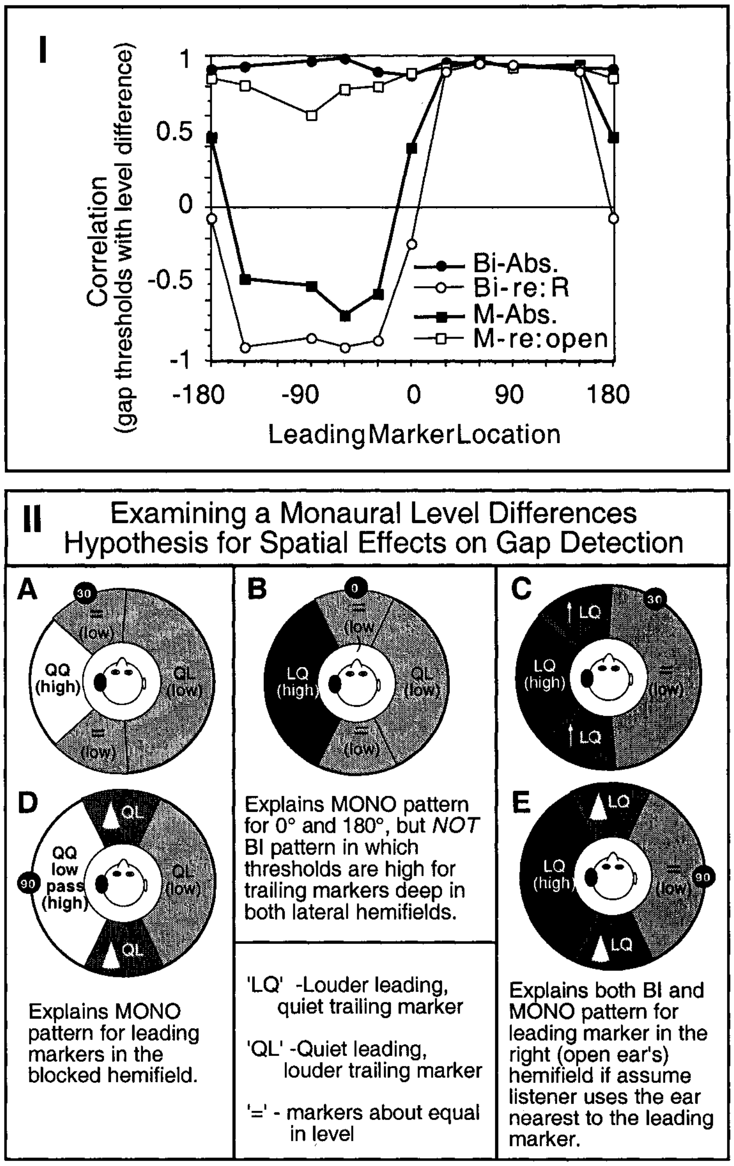
Boehnke and Phillips (1999) argued that gap thresholds became elevated with spatial separation to asymptotically high levels for marker pairs which activated non-overlapping “spatial channels”. These channels were proposed to be central in origin largely because of the functions of gap thresholds with trailing marker location using leading markers deep in a hemifield were so similar to the spatial tuning of cortical neurons. Oxenham (2000) proposed an alternative, within-channel explanation of the elevation in thresholds with free-field separation of markers. The increased gap detection thresholds obtained for free-field markers in opposite hemifields could be result of level differences between the markers at a single ear. In his experiment, thresholds were increased when dichotic leading and trailing markers differed in their interaural level. However, this increase could be accounted for by the observation that an equivalent monaural level difference between otherwise identical markers raised thresholds by a similar amount. Accordingly, without disputing the sensitivity of gap thresholds to marker locations, the mechanism mediating that sensitivity might have some or all of its basis in monaural effects rather than in binaural spatial processing. The implication is that gap detection between spatially separated markers may not necessarily be based on the ‘between-channel’ process (a relative timing operation) exploiting the output of centrally-generated representations of sound location.

Correlational analysis of binaural and monaural spatial gap detection thresholds with monaural level differences

Can the monaural and binaural data presented in Experiment 1 be accounted for by a monaural level differences hypothesis (after Oxenham, 2000)? The pattern of thresholds for leading markers located deep in a lateral hemifield was the same as that obtained for the ear nearest the leading marker. This suggests that thresholds under all conditions (except for a leading marker on the midline) might be accounted for by the level differences at the ear nearest the leading marker. To test this hypothesis, the absolute (unsigned) level difference between two sounds coming from any two azimuthal locations, and the signed level difference (as would be observed at the right ear) for any two azimuthal locations, was estimated from published data (Wightman & Kistler, 1993, pg. 162), for every pair of marker locations that were used to obtain gap thresholds in Experiment 1. These were then correlated with the gap thresholds obtained under both binaural and monaural listening conditions (see Fig. 2.5I). Interestingly, with an analysis including all leading and trailing marker pairs, the gap thresholds obtained under binaural conditions were highly correlated with the absolute level differences between the markers (r=.92), but the gap thresholds obtained under monaural listening only correlated about half as well (r=.47).

Examination of the pattern of correlations across leading marker location, presented in Fig.2.5, explicates this pattern. Under binaural listening, gap thresholds for all leading marker conditions were highly correlated with an absolute level difference (function "Bi-Abs."). Under monaural listening (function "M-Abs"), gap thresholds are positively and strongly correlated with absolute level difference for the open ear leading conditions, but are negatively correlated for leading marker locations in the occluded ear's hemifield. However, if we use the level difference re: the right (open) ear, gap thresholds for all leading marker locations under monaural listening are highly correlated with level differences at the open ear (function "M-re:open"), except for leading marker conditions deep in the blocked hemifield which are moderately well correlated. Under binaural listening, gap thresholds are positively correlated with right ear level differences only for right hemifield leading locations, negatively correlated with right ear level differences for left hemifield leading marker locations, and only partially correlated with right ear level differences for midline leading marker locations (function "Bi-re:R"). The points to take from this analysis are that, under these stimulus configurations (short leading marker duration, broadband markers), 1) thresholds are only elevated in configurations where the leading marker is louder than the trailing marker, and 2) if a level difference at one ear is limiting performance on gap detection thresholds it must be the difference found at the ear closest to the leading marker.

The two conditions that do not fit this analysis are the midline leading-marker conditions. In these cases, the leading marker does not direct by which ear the listener will be affected by a monaural level difference since both ears are equally activated by the leading marker. When only one ear is available, the level difference between a



Examining a Monaural Level Differences Hypothesis for Spatial Effects on Gap Detection

Explains MONO pattern for leading markers in the blocked hemifield.

Explains MONO pattern for 0° and 180°, but *NOT* BI pattern in which thresholds are high for trailing markers deep in both lateral hemifields.

'LQ' -Louder leading, quiet trailing marker

'QL' -Quiet leading, louder trailing marker

'=' - markers about equal in level

Explains both BI and MONO pattern for leading marker in the right (open ear's) hemifield if assume listener uses the ear nearest to the leading marker.

Figure 2.5 Level Difference analysis of Exp. 1
I. Plotted for each leading marker location is the correlation of gap thresholds obtained at all trailing marker locations with the level difference between that leading and all trailing marker locations. Circles and squares represent the data for binaural and monaural listening respectively. Filled and open symbols show correlations of gap thresholds with absolute (unsigned) level differences, and level differences re: the open ear.
II. A-E. Schematic diagrams show the regions where there is a level disparity between the leading and trailing markers, for leading marker located in the blocked ear (left) hemifield (A&D), the open ear (right) hemifield (C&E), and on the midline (B). The underlying text explains how the schematic could explain the pattern of results obtained. L=loud, Q=quiet. Areas with an arrow indicate a region of change in relative level that should correspond to a slope in the function. A level difference explanation of the data works well for all monaural conditions, but not all binaural conditions. See text for explanation.

leading marker at 0° and a trailing marker at the side of the good ear does not increase gap thresholds, but when two ears are available, the same level difference does. This is despite the fact that the listeners knew exactly where the sounds were located and could simply elect to focus their attention on one ear. The fact that this does not occur is important because it means that the pattern of data obtained (Fig. 2.2B) can not be explained by appealing to the listener's decision to use selectively the information from a single (most informative) ear.

A remaining question is why a "Loud-Quiet" marker level difference at one ear increases gap thresholds, but a "Quiet-Loud" level difference, as occurs when the leading marker is in the blocked ear hemifield and the trailing marker is in the open ear hemifield, does not. This is consistent with a rather small difference between Loud-Quiet and Quiet-Loud conditions found in a study of decrement detection, in which the markers on either side of the decrement differed in level (Plack & Moore, 1991). Such an effect might also be expected based on adaptive responses of the auditory nerve to such stimulus configurations (Smith *et al.*, 1985). Note, however, that such a disparity between "Loud-Quiet" and "Quiet-Loud" marker configurations has not been observed under headphone listening with broadband markers (Oxenham, 2000). The two configurations resulted in equally elevated gap thresholds relative to thresholds obtained with equal-level markers. Furthermore, the spectral characteristics of the free-field gap stimuli with spatially separated markers – somewhat low-pass quiet leading marker followed by a wideband trailing marker are consistent with elevated thresholds due to the presence of additional frequency information in the trailing marker (after Formby *et al.*, 1998b).

Putting this paradox aside for now (it will be addressed in Exp. 5), by using the distinction between Loud-Quiet and Quiet-Loud configurations, a monaural level-differences hypothesis can model both the binaural and monaural gap thresholds obtained for markers located at free-field locations quite well, as is illustrated schematically in Figure 2.5II (panels A-E) for monaural conditions. The regions of azimuth where the levels of the markers at the right ear were either roughly equal(=), Loud-Quiet (LQ), Quiet-Loud (QL), or changing towards QL or LQ (as indicated by an arrow), are outlined schematically for each leading marker location. Gap thresholds should be low for regions that are either equal or QL and should be raised as the stimulus configuration moves to

LQ. Assuming that under binaural conditions the level difference at the ear nearest the leading marker limits gap thresholds, this pattern predicts gap thresholds well for every binaural and monaural condition except for leading markers located at 0° or 180° under binaural listening. In these midline leading marker conditions, it is unclear why listeners could not achieve low thresholds for trailing markers at the lateral poles, particularly given that the stimulus is "Quiet-Loud" at the ear nearest both markers. Furthermore, in all other conditions, it is unclear why the listener can not simply use the information at the ear near the trailing marker to perform the task. At that ear, the stimulus is Quiet-Loud and so gap detection thresholds should not be limited by the periphery.

Modeling the binaural data from the monaural data: Are the binaural data limited by the 'best ear', 'worst ear', 'leading marker ear', or some 'average' of the two ears?

One way to consolidate all the results might be to assume that the limitation on gap detection thresholds under binaural listening is the result of some combination of the limits on gap detection thresholds achievable with either ear alone. The data-set from Experiment 1, in combination with the previously published data from Boehnke and Phillips (1999) for the same listeners provides for such an analysis. For example, one might postulate that the system always uses the 'best ear' under binaural listening. On the other hand, information processing might be such that the 'worst ear' limits the system under binaural listening. Alternatively, the system might average the information available at each ear, either through binaural integration or by separately analyzing, and then averaging the results from two independent functionally monaural processors. In the following analysis, these possibilities were tested by using the gap detection thresholds obtained with each ear alone for a given leading and trailing marker location pair to predict the gap detection threshold that might be achieved using both ears. The resultant predicted functions of gap threshold with trailing marker location, for a given leading marker location, were then statistically compared with the actual functions obtained under binaural listening for each leading marker location. The statistical details of this analysis are presented in Appendix A, and the results of this analysis are presented in passing here. In the first four columns of Figure 2.6B, the sums of the squared deviations of the predicted thresholds from the actual thresholds obtained under binaural listening are displayed for each leading marker location and for each model. This figure illustrates

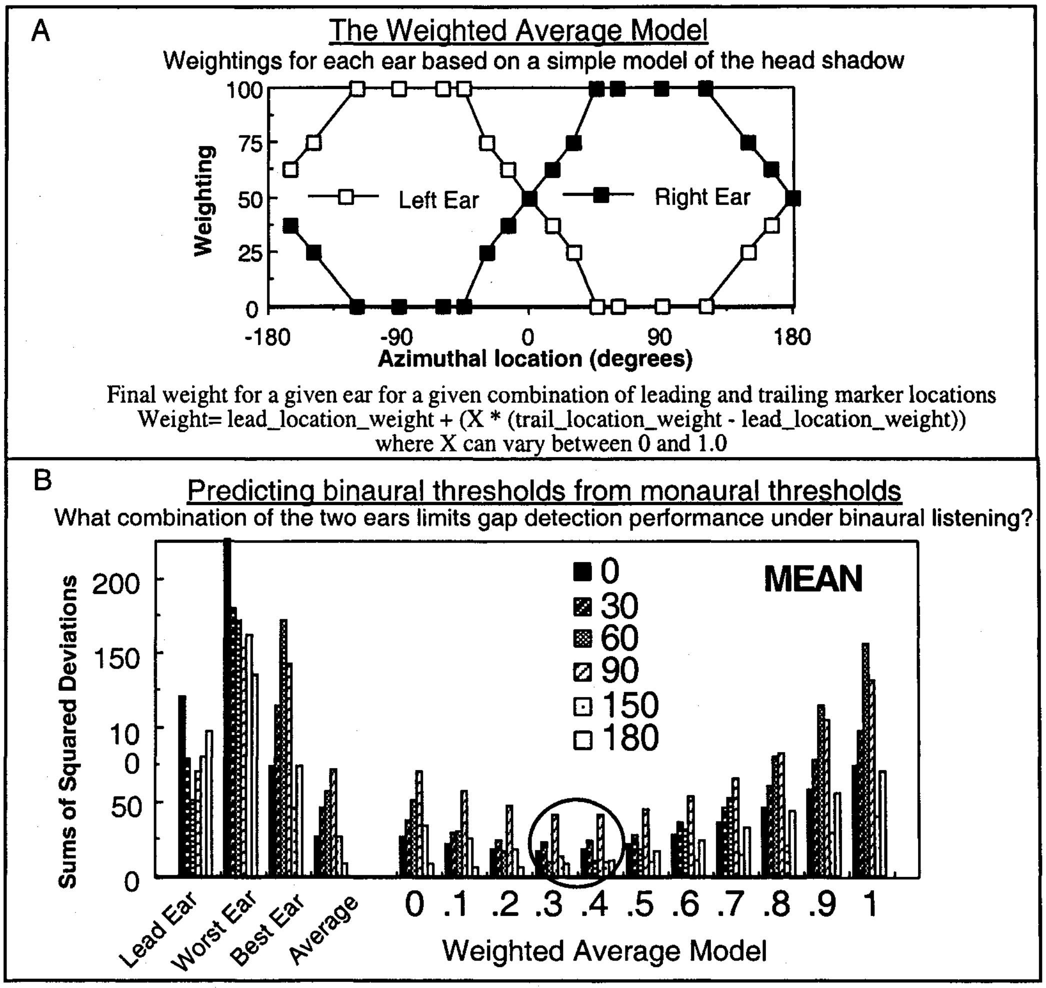


Figure 2.6

A. The Weighted Average Model.

Thresholds from each ear are weighted based on their expected intensity at each ear given estimates of the shadowing by the head as a function of location (see plot). For an X value of 0, only weights for the two ears specified by the leading marker location are used in the average of the two ears. For increased values of X, the weights for the trailing marker location are allowed to influence the final weightings for each ear as per equation.

B. Experiment 1. Predicting binaural thresholds from monaural thresholds

Experiment 1 generated gap detection thresholds as a function of spatial separation of gap markers for each ear separately and under binaural listening. This analysis examined whether thresholds obtained under binaural listening in the same listeners could be predicted based on performance under monaural listening. Predictions of binaural performance were computed using the monaural data in several ways. The sums of the squared deviations of that prediction from the actual binaural data were calculated for gap thresholds at all trailing marker locations for a given leading marker location. The model which best minimized the sums of squares across all leading marker functions best predicts both the pattern and absolute values of the thresholds. The LEAD EAR model examined whether thresholds under binaural listening were limited by the ear nearest to the leading marker. The WORST and BEST EAR models examined whether binaural performance was limited by the ear with the worst or best performance respectively. The AVERAGE model compared binaural performance to a linear average of performance at each ear. Finally, the Weighted Average Model, approximated both the shape and absolute value of the gap thresholds obtained under binaural listening for all leading marker locations when relatively more weight was given to the leading marker in the calculation of the weighting function for the two ears (see Text & Appendix).

the leading marker functions for which each model is effective and where it fails in its prediction. A good model will provide predictions that result in low sums of squared deviations. The WORST EAR model clearly fails to predict any of the data. The BEST EAR model predicts the MIDLINE leading marker functions well, but poorly predicts the hemifield-located leading marker functions. The LEAD EAR model predicts the hemifield-located leading marker functions well, but fails to predict the midline-located leading marker functions. The direct arithmetic AVERAGE model generally makes a fairly good prediction of all the data, particularly the midline-located leading marker functions; however, it is relatively poorer at predicting the pattern of thresholds in the hemifield-located leading marker functions (though still better than the other models).

The failure of any of these models to predict the pattern of thresholds for all leading marker locations indicated that they were likely too simple in their assumptions. When both markers are located in a given hemifield, a direct arithmetic average of the performance supported at the two ears doesn't really make sense since the majority of the stimulus energy is reaching the near ear, and only a minority is reaching the far ear. Thus, the expected sound intensity of the markers at each ear (relative to the other) that results from the shadowing of the head might be used as a weighting factor by which the information from the two ears is averaged. For example, for leading and trailing markers both located in the same hemifield, the nearest ear should receive a greater weight in the averaging process. For markers on the midline, the two ears should receive equal weighting in the average. The complication in such an analysis lies in determining what the relative weight given to the leading and trailing marker intensities at a given ear should be. For example, for markers in opposite hemifields, the leading marker is more intense at one ear, but the trailing marker is more intense at the other – so how should the ears be weighted?

A model was created in which the level of each marker at a given ear across azimuthal locations was roughly estimated by a function of the head-shadow (see Fig.2.6A), obtained from published measurements for wideband noise (Wightman & Kistler, 1993, pg. 162). The model starts out with the assumption that the location of the leading marker dictates the weighting to be given to each ear. For example, if the leading marker location is +30 deg., the right ear threshold contributes 0.75 and the left ear 0.25

to the binaural prediction, regardless of trailing marker location (see Fig.2.6A). This is like a *graded* version of the 'LEAD-EAR' model. Increasing the factor 'X' (see equation in Fig. 2.6A) allows the trailing marker location to influence the final weightings used in the average. For example, with a weighting of 0.5 and a leading marker at 30 deg., the weights of 0.75/0.25 will change as a function of trailing marker location. When the trailing marker is on the right, the weightings will be even more biased to the right ear – 0.85R/0.15L. When the trailing marker is on the left, the weightings will reduce in bias to the right ear – 0.625R, 0.375L. When the value of X is 1, the weighting functions are determined entirely by the weights for the location of the trailing marker. The two ears were then averaged using these weights in order to generate predicted thresholds under binaural listening.

The WEIGHTED AVERAGE model provided a good prediction of both the pattern and absolute values of the gap thresholds obtained under binaural listening (see Appendix A). Weighting the average of the two ears based solely on estimates of the level of the leading marker at the two ears (x=0) provided a better prediction of the actual binaural data than did the more simple models. As the levels of the trailing marker at the two ears were allowed to influence the weighting of the two ears in the average (x>0), both the patterns of the functions across trailing marker location and the absolute values of the predicted thresholds approached those actually obtained by Boehnke and Phillips (1999) for all leading marker locations. The sums of the squared deviations between prediction and the binaural data were best minimized when the trailing marker contribution was set to between .3 and .4, i.e. when there was a slight bias towards the leading marker location.

This bias toward the leading sound, particularly in the temporal configuration used here where the leading marker is relatively short, is not entirely surprising. One example of such an onset-sound dominance is found in the precedence effect and related phenomena. As described briefly in Chapter 1, when two brief, spectrally similar sounds are presented in rapid succession from different spatial locations the sounds are perceived as fused for temporal separations under about 5ms. The location of the perceived fused sound is biased to the location of the leading sound; hence, it is referred to as the precedence effect (see review by Litovsky *et al.*, 1999). It is as if the information about the second sound is either suppressed, as part of an echo suppression mechanism; or is simply not available due to temporal interactions in the transduction process (Hartung & Trahiotis, 2001). It might be that such temporal factors are also at work here, i.e. some feature (adaptation or inhibition) of the response to the leading marker reduces the size of the response to the trailing marker, thus reducing the contribution of its spatial information to the final perception.

Plomp (1964) and Penner (1977) related the decrease in gap threshold performance as the leading marker is made louder than the trailing marker to mechanisms of forward masking. The gap detection threshold reflects the duration required for the leading marker to decay in sensation to the same level as the trailing marker. In this regard, it might be instructive to compare the data from Experiment 1 with forward masking studies where the signal and masker are spatially separated in the free-field. No such experiments seem to have been reported. However, careful studies examining the release from simultaneous masking with spatial separation have been performed [wideband stimuli (Saberi *et al.*, 1991); low-, midfrequency-, and high-pass stimuli (Gilkey & Good, 1995)]. In Figure 2.7, the functions relating gap detection threshold to trailing marker location are plotted for leading markers at 0, 30 and 90 degrees in separate panels (right hand axis). Also plotted are the free-field masked thresholds (dB) as a function of signal location for noise maskers located at 0, 30, and 90 degrees. The patterns across azimuth are remarkably similar for the two tasks. The study by Saberi *et al.* (1991) sought to determine the relative contribution of monaural and binaural processes to this pattern of release. They determined that the pattern of detection under binaural conditions was very similar to that obtained under monaural conditions. Thus, they concluded that much of their data could be accounted for by the *improved signal to noise ratio* at the ear nearest the signal (due to the shadowing by the head and pinna), with a minor additional benefit that could be attributed to strictly binaural processes.

The general conclusion to be drawn from the previous analyses is that gap detection thresholds for free-field wideband noise markers can be conceived of as limited by monaural processes, without need to resort to mechanisms of binaural interaction at higher stations in the auditory system which produce different spatial location percepts. This is not to say that the output of these 'binaural' nuclei are not contributing in any way

COMPARISON OF FREE FIELD GAP DETECTION THRESHOLDS WITH FREE-FIELD MASKED THRESHOLDS

(Adapted from Saberi et al, 1991)

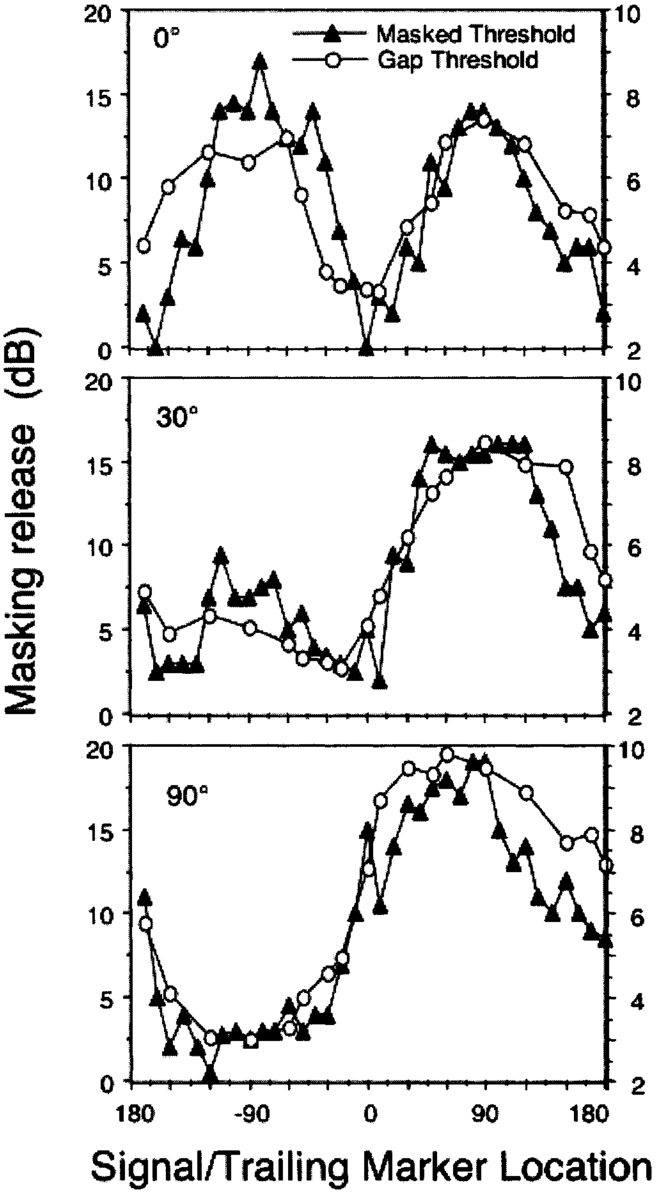


Figure 2.7 Comparison with Saberi et al. (1991)
A comparison of the dependence of gap detection threshold on trailing marker location obtained under binaural listening from experiment 1(Boehnke and Phillips,1999) for leading marker locations at 0°, 30°, and 90°. with the dependence of masked thresholds on signal location for masker locations at 0°, 30° and 90°. obtained by Saberi et al. (1991). In the gap experiment the question was "do you hear a gap between two sequential sounds" in the masking experiment "can you detect the signal in the presence of simultaneous masking noise". Interestingly, the pattern of results is the same.

to the temporal processing mechanism that is detecting the presence of the gap; only that all nuclei might receive the same degraded representation of the cues relevant to detection of the gap from the monaural processes. The level differences at a given ear due to spatial separation of the sound markers only occurs in frequency channels greater than about 1500 Hz, because the frequency-dependent filtering by the head and pinna occurs largely for frequencies greater than 1500 Hz. There would be little or no difference in marker level at lower frequencies with spatial separation, although there would be an ITD difference. Examining the effect of spatial separation with low-pass or high- passed markers might assess the relative importance of purely binaural cues (ITD and ILD) and the monaural spectral and level differences. Free-field release from masking studies have also been done for such band-passed stimuli (Gilkey & Good, 1995). Those results show a much-attenuated effect of spatial separation of a signal and masker when the sounds are low-pass compared with when they are high-pass. It is expected that the pattern of gap thresholds using low-pass and high-pass noise markers should show the same interaction of marker frequency with spatial separation.

EXPERIMENT 2:

THE EFFECT OF NOISE BAND-PASS ON GAP DETECTION PERFORMANCE FOR MARKERS AT DIFFERENT FREE-FIELD LOCATIONS

A simple way to test the relative contribution of level differences and ITD differences between markers to increased gap thresholds with spatial separation of markers in the free-field is by using free-field stimuli of different band-passes. This provides for a spatial location percept that is fully externalized, unlike conventional headphone experiments where the spatial quality is limited to intracranial locations. As described in the Introduction, there is little difference *in the level* of low frequency components at the two ears as a function of spatial location. As the stimulus component increases in frequency, its interaural level difference increases. Thus, presenting low-pass markers from different speakers distributed in azimuth provides for minimal differences in ILDs (and monaural level differences) but significant differences in ITDs (purely central coding) between the markers. High-pass noise stimuli result in large interaural level differences. For high-pass gap markers at locations in opposite hemifields, there would be a large relative level difference imposed between the markers at each ear (Loud-Quiet at the leading marker ear; Quiet-Loud at the trailing marker ear), and ILD differences between the markers. If level differences (or ILD differences) are the dominant feature contributing to increased gap thresholds in the free-field, then there should be a large effect of spatial separation for high-pass stimuli and a relatively small (or no) effect for low-pass stimuli.

METHODS

Subjects

Three subjects (1 male) participated. All were experienced in psychoacoustic tasks, and two were highly practiced in gap detection tasks.

Stimuli

Wideband (0-22050 Hz) leading and trailing markers were generated independently in MATLAB software (The Mathworks) in the frequency domain. Both markers in the standard and 'test' interval were of a random duration between 100 and 300 ms for each trial. This prevented the total duration, or the onset-onset duration, or any other interval other than the offset-onset interval from being used to aid performance (see Penner, 1977, for a discussion). To generate low-pass markers, energy above 1000 Hz was discarded and then the digital stimulus was transferred into the time domain. To generate high-pass markers, energy below 3000 Hz was discarded, and the digital stimulus was transferred into the time domain. The markers were then shaped with rise and fall times that were dependent on the pass band. Low-pass stimuli were shaped with 4 ms rise/fall times and high-pass stimuli were shaped with 2 ms rise/fall times. These were required to prevent spectral splatter at the onset and offset that would result in clicks which could provide a non-temporal cue for detection. Pilot testing indicated that these rise/fall times were sufficient to eliminate such effects. The standard stimuli were created by concatenating the shaped leading and trailing markers. A 1 ms gap was placed between the markers in order to closely replicate the method used in Experiment 1. The 'test' stimulus was created by concatenating the shaped leading marker, a period of silence equal to the gap duration for that trial, and the shaped trailing marker. The gap duration was defined as the absolute period of silence in the test stimulus, excluding the rise and fall times. This process was repeated for each trial. The stimuli were presented to free-field speakers through the MATSS presentation system (see General Methods).

Design and Procedure

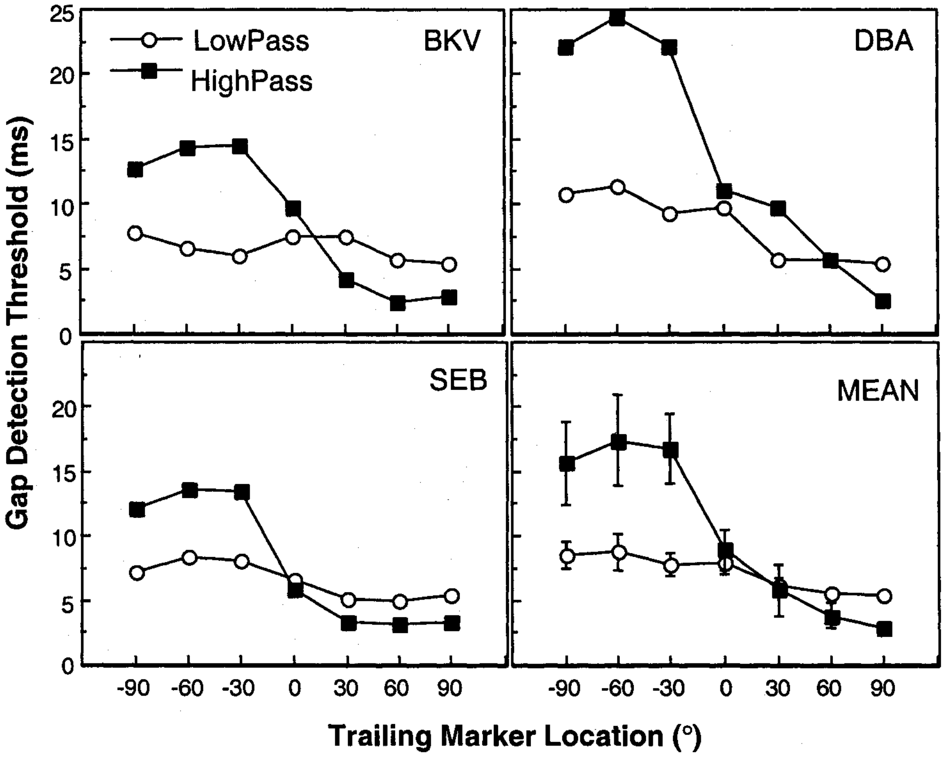
In order to determine the sensation level at which to present the gap detection stimuli, detection thresholds for the low- and high-pass noises were first obtained for each subject using a 2AFC 3-down 1-up adaptive tracking procedure (see common methods). Listeners were presented with two intervals congruent with illumination of a red LED and a green LED. The signal to be detected was presented in one of the intervals. The task of the listener was to indicate which interval contained the signal. The spectrum level of the signal was reduced adaptively and the detection threshold in spectrum level was determined. The stimuli for the gap detection task were always presented at 40 dB above the detection threshold, i.e., 40 dB sensation level (SL).

There were two variables: the pass-band of the markers (low, high) and the location of the trailing marker (+90, +60, +30, 0, -30, -60 and –90 degrees). For a given threshold determination the pass-band and trailing marker location were fixed. The leading marker was always presented to the speaker at +90 degrees. Gap thresholds were determined using the 2I-2AFC, '3 down, 1 up' adaptive procedure (see General Methods). Gap thresholds for the 21 conditions were obtained in a random order, usually within one session. This was repeated until at least three thresholds were obtained for each condition.

RESULTS AND DISCUSSION

Individual and mean gap-detection thresholds are plotted in Figure 2.8A as a function of trailing marker location, with the band pass of the marker noise as the parameter. Error bars on the means are standard errors. Gap thresholds for low frequency markers were generally higher than those obtained for high frequency markers when the markers were in the same hemifield. This is not surprising, because low frequency noise has a relatively narrower bandwidth, a factor known to increase gap thresholds (Shailer & Moore, 1983; Eddins *et al.*, 1992). When the markers were in different hemifields, thresholds were much higher for high-frequency markers. The interaction thus reflects the strong effect of spatial separation across the midline for high-frequency markers on one

A. SPATIAL SEPARATION EFFECTS ON GAP DETECTION WITH LOW- & HIGH-PASS MARKERS



B. SPATIAL SEPARATION EFFECTS ON MASKING RELEASE & GAP DETECTION: COMPARING LOW- AND HIGH-PASSED STIMULI

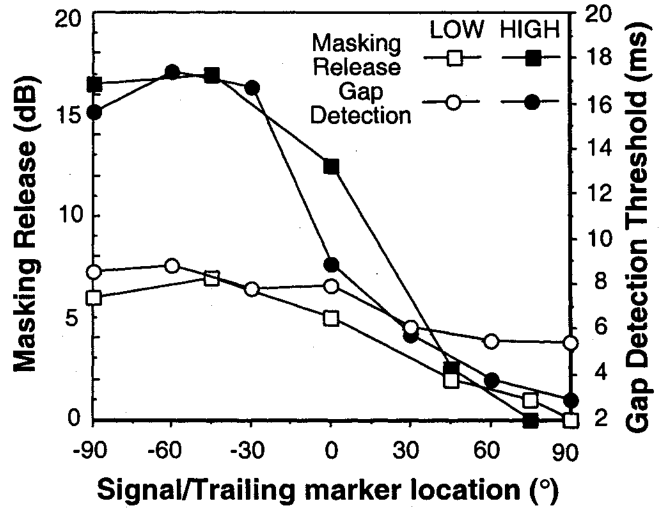


Figure 2.8 Experiment 2.

A. Gap detection thresholds as a function of bandpass and spatial separation. Low pass stimuli had an upper cutoff frequency of 1000 Hz, while high pass stimuli had a lower cutoff of 3000 Hz. The leading marker was always located at 90° azimuth.

B. A comparison of the data from A., with free-field release from masking data for the same spatial separation of a signal and masker, using similarly filtered stimuli (adapted from Gilkey and Good, 1995).

hand, and a lack of a spatial separation effect on thresholds for low frequency markers on the other. The data were statistically analyzed with a 2 factor (band pass and spatial separation) repeated measures ANOVA. The main effect of the noise band-pass did not achieve statistical significance [$F(1,3)=4.97$, $p=0.16$]. This is probably due to the relatively elevated thresholds with low-pass noise stimuli with markers in the same hemifield. There was a main effect of spatial separation [$F(1,3)=13.74$, $p<0.01$], and also a significant interaction, as expected, reflecting the greater effect of separation on the high- passed markers compared with the low-pass markers [$F(6,12)=18.16$, $p<0.001$]. Analysis of the low-frequency data separately by an ANOVA is technically not justified since there was no significant effect of frequency in the main analysis. However, such an analysis did show a significant effect of separation [$F(6,12)=3.8$, $p=0.05$]. This small (2-3 ms) increase with spatial separation using low-pass noise markers was observed in only two of the three listeners (DBA, SEB). It may have been due small residual level differences at either ear between the low-pass noise markers. It is also possible that the ramping of the low-frequency markers with rise/fall times of 4ms was insufficient to eliminate all spectral splatter. This possibility would result in the presence of low-amplitude, high-frequency energy, which would be subject to the head shadow and result in marker level differences in those frequency channels. Because of such possibilities, it can not be concluded that differences in ITD between the low-pass markers contributed to this small increase in gap detection thresholds with spatial separation.

These results lend support to the idea that level differences between the markers at either ear, or ILD differences in the high frequency components of the two markers, underlie the elevation in thresholds observed with spatial separation in Boehnke & Phillips (1999). As observed by Oxenham (2000) and here with low-frequency free-field noise markers, there is little to no effect of separation when the major difference between the markers lies in their relative ITD. This experiment is an important verification of Oxenham's results, as one might argue that the simple dichotic stimuli used in Oxenham's study were not comparable with 'real' fully externalized sound sources. A related implication resides with how one conceives of the 'between-channel' effect in gap detection. Low frequency sounds from speakers in opposite hemifield clearly differ perceptually; however, they apparently do not fall into different perceptual channels of the sort that have been conceived to underlie the elevation of gap thresholds with perceptually dissimilar markers (Phillips *et al.*, 1997; Phillips *et al.*, 1998; Boehnke & Phillips, 1999; Taylor *et al.*, 1999). The fact that low frequency sounds in opposite hemifields are also selectively represented by spatially sensitive neurons in different brain hemispheres raises questions about which channels are involved in this instance of temporal processing. It may be that the low-frequency channels that are sensitive to different ITDs are reserved as a code for sound location, and processing of non-ITD temporal differences occurs in a parallel monaural stream. This is not inconsistent with the fact that detection of temporal changes is better in high frequency channels than in low frequency channels (Shailer & Moore, 1983).

In Experiment 1, a compelling correspondence was shown between the pattern of gap detection thresholds obtained with spatial separation of wideband noise markers, and the pattern of masking release obtained with spatial separation of a wideband signal and noise masker (see Fig. 2.7). This correspondence can now be extended to low- and high-pass noise markers. In Figure 2.8B, the function relating gap detection threshold to spatial separation for low- and high-pass noise markers are plotted together with the functions of masking release with spatial separation of a low- or high-pass signal and noise masker (adapted from Gilkey & Good, 1995). What might this correspondence tell us about the factors influencing gap detection? Masking release has traditionally been discussed in terms of binaural processing. A wealth of data has demonstrated the improvement in detection when the interaural phase (and thus the perceived location) of the signal is made different from the masking noise (see Moore, 1997 for a review). Separation of a low-pass signal from a low-pass masking noise in the free-field only results in about a 6 dB release from masking for separation in opposite hemifields. This benefit is likely attributably to differences in ITD of the signal and masker. However, there is clearly a greater improvement in detection of a high-pass signal with spatial separation from a high-pass masker. This benefit can not be attributed to the difference in ITD, since ITDs are not coded for such frequencies. There are two possibilities - the benefit may arise from the more favorable signal to noise ratio (SNR) at the ear near the signal resulting from shadowing of the masker by the head, or as a result of ILD differences. Clearly, in the free-field with a wideband signal the listener would take advantage of the improved

SNR in the higher frequency channels rather than use the smaller benefit in the low-frequency channels that is available through the binaural system.

Now, it is difficult to know if the listener can (or does) exploit the difference in ILD (and attendant perceptual differences) between the signal and masker to perform the task or simply uses the monaural information at the ear nearest to the signal. Gilkey and Good (1995) reasoned if the ILD difference was being used, equivalent performance for middle-frequency band-passed and high-pass noise would be obtained because the function of masking level difference with ILD asymptotes for ILDs beyond 20 dB (Egan, 1965). Since 20 dB ILDs are already available in the mid-frequencies, then there should be no further benefit with the greater ILDs available for higher frequency components. Their data show considerable improvement in detection for the higher frequencies, a result that suggests that the monaural information at the ear nearest the signal is being accessed. Furthermore, they describe unpublished results, for which the free-field task was simulated over headphones using stimuli convolved by head-related transfer functions (virtual auditory space), that confirmed binaural performance to be comparable to monaural performance with the ear nearest the signal. That is, the dominant route by which detection of a signal in spatially separated noise is accomplished is monaural.

Given the correspondence of the gap detection data presented here with free-field masking release data for all frequencies, it is perhaps unlikely that ILD was playing a strong role in performance of the gap detection task either. However, it is not yet understood whether the elevation in thresholds observed by imposing ILD differences on dichotic markers is a result of the same mechanism that underlies the elevation in gap thresholds with free-field spatial separation. It might be that the poverty of spatial information provided by simple ILDs presented over headphones does not permit the markers to activate the spatial perceptual channels presumed to be activated by markers located at different free-field locations, i.e. real externalized spatial percepts might be required to activate a 'spatial perceptual channel'. Oxenham (2000) did not observe an effect of leading marker duration on thresholds obtained for markers with opposite ILDs presented over headphones. This is in contrast to the leading marker effect obtained by Phillips *et al* (1998) for markers at free-field locations in opposite hemifields, in which thresholds increased as the leading marker was shortened. This effect has been argued to be a characteristic of the ‘between-channel’ relative timing process (Phillips, 1999), which was described as a mechanism for timing the offset of activation in one perceptual channel with the onset of activation of another. If, indeed, the leading marker duration effect is a ‘marker’ for activation of different perceptual channels, then the failure to observe such an effect for markers lateralized with ILDs and presented over headphones, suggests that these stimuli were insufficient to activate the ‘spatial channels’ described by Phillips *et al* (1998) and Boehnke and Phillips (1999). In the next experiment, we compare the leading-marker duration effect for markers presented over headphones at different ILDs or presented from free-field locations in opposite hemifields.

EXPERIMENT 3A:

THE EFFECT OF LEADING MARKER DURATION ON GAP DETECTION FOR MARKERS DISTINGUISHED BY SPATIAL LOCATION, BY ILD OR BY LEVEL.

The preceding two experiments were conducted using short leading marker durations. This was because an earlier study examining the effect of leading marker duration on gap detection thresholds between markers presented from opposite hemifields indicated that the effect of spatial separation (the ‘between-channel’ effect) increased as leading marker duration decreased (Phillips *et al.*, 1998) This leading marker effect has been argued to be characteristic of ‘between-channel’ configurations, as the leading marker duration often has no effect (Forrest & Green, 1987; Phillips *et al.*, 1997; Phillips *et al.*, 1998) or a relatively small (Crawley, 1999; Schneider & Hamstra, 1999; Snell & Hu, 1999) effect on thresholds for gaps between identical markers, i.e., markers activating the same perceptual channel (frequency, ear or spatial location). In order to maximize the spatial-separation effect in order to measure the tuning of a ‘spatial channel’, the experiment by Boehnke and Phillips (1999) employed short leading markers. Oxenham (2000) observed no effect of leading marker duration when he used ILDs to lateralize the markers to opposite sides of the head over headphones. If we assume that a leading marker duration effect is characteristic of markers activating different perceptual channels, then it is possible that the elevation in thresholds in marker configurations with an ILD disparity is caused by a different mechanism – specifically, a within-channel effect of level differences between the markers at a given ear (after

Oxenham, 2000). Accordingly, experiment 3 compared the gap detection performance in a single set of listeners when leading and trailing markers were played from the same (RR) or different (RL) locations, under free field (speaker locations) or dichotic (intracranial locations) ILD conditions. These conditions were also compared with a monaural headphone presentation in which the leading marker was either 12 dB louder or 12 dB quieter than the trailing marker (same values used by Oxenham, 2000). Each condition was repeated for four leading marker durations (5, 10, 50 and 300 ms).

If a leading marker effect is found for gap thresholds obtained for different speaker locations, but is not observed for monaural level differences, then the threshold increases in the two cases might be due to different or incompletely overlapping mechanisms. It is possible that the level difference mechanism is affecting thresholds in both conditions, and an additional between-channel mechanism is causing an additive effect in the free-field. Such a mechanism might be a result of the enhanced perceptual differences between spatially separated free-field markers (compared with ILD-separated markers) due to the differential filtering of the sound at each ear by the head-related transfer function.

METHODS

Subjects

Data were obtained from four highly trained listeners, all of whom had participated in Experiment 1, and from an additional four listeners who had no experience in psychoacoustic experiments. Audiograms (GSI 17 Audiometer, Electro-medical Instruments) performed on all listeners were in the normal range. Free field detection thresholds for noise were similar across all listeners.

Gap Stimuli and Apparatus

Broadband noise stimuli were generated digitally on a custom-programmed Macintosh G3 computer. They were 400 ms in length with 1.0 ms raised cosine rise-fall times. To create different leading marker durations, silence was inserted destructively, beginning 5, 10, 50, or 300 ms into the 400 ms-noise, and thus the total duration of the stimulus remained about the same. Pilot testing with the 300-ms leading marker stimuli indicated the short *trailing* marker made the task very difficult (for a different reason than the one we were interested in examining - see also Snell & Hu, 1999). Accordingly, we used 600 ms noise samples and the gap was inserted destructively at the midpoint for the 300-ms leading marker conditions. Gap duration was specified as the duration of the silent period plus the fall time (i.e., the "equivalent rectangular" period of silence in ms). In this experiment, the two markers in the standard were shaped, but not separated by any silent period. This is in contrast to our previous experiments in which the markers in the standard were shaped and separated by a 1.0 ms gap. To partially address problems attendant to the use of "frozen" noise (i.e., the same sample of noise in all trials), each standard and signal stimulus was drawn randomly from 3 different saved noise samples for each of a series of gap durations decreasing as a factor of 1.2. The generated stimuli were saved as "resource files" by SOUNDEDIT™ software on a Power Macintosh 8600 computer. The experiment was then programmed in HYPERCARD™, which called up the resource files as needed (after Phillips & Hall, 2000).

In the headphone conditions, the stimuli were transduced by Koss TD/64 headphones. Stimulation was dichotic or monaural, as specified below. Absolute stimulus levels were in the range from 70 to 75 dB SPL (sound-pressure level: dB re 20 μPa, A-weighted; Bruel & Kjaer Instruments). In the free-field condition, the same stimuli were passed through a DA converter, amplified (Realistic SA-155 integrated stereo amplifier), and were transduced at a comfortable listening level (65-70 dB SPL) by two Optimus speakers (XTS 36 40-1994) with a frequency response of 0.14-20.0 kHz in the sound-deadened room described in Exp. 1.

Design and procedure

All listeners were tested individually. There were seven stimulus conditions and a minimum of three gap thresholds was measured for each. Stimulus conditions are illustrated schematically in left panel of Fig. 2.9. There were three "within-channel" control conditions: both the leading and trailing marker were from the same free-field speaker location (FF-RR); presented dichotically with the stimulus at the right channel 12 dB louder than the stimulus at the left (ILD-RR); or presented to the same ear (M-RR). All these conditions were expected to generate low gap thresholds. There were two "between-channel" conditions: the leading and trailing markers were from speaker locations at 90° in opposite hemifields (FF-RL) or were from opposite intracranial

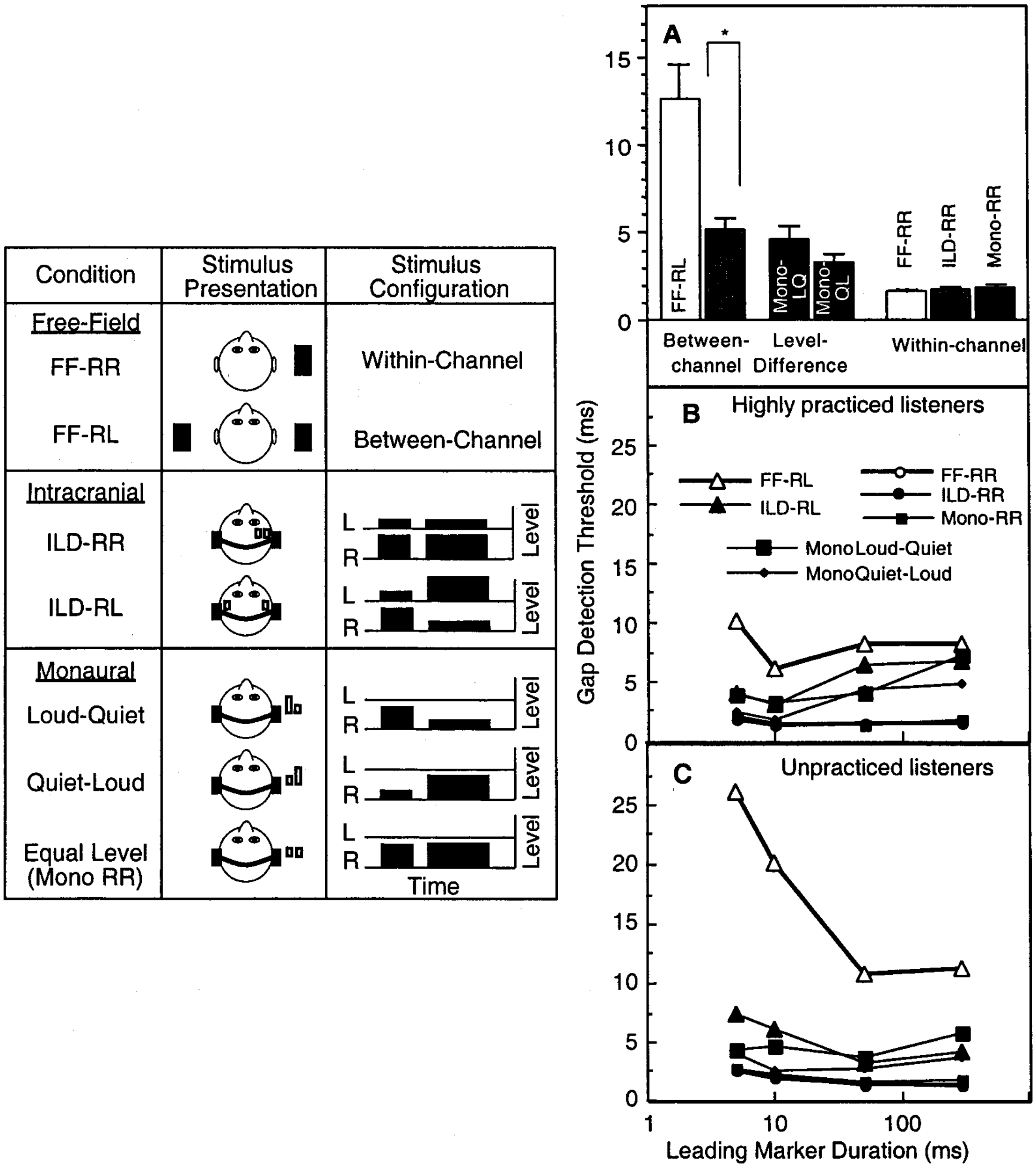


Figure 2.9 Experiment 3A.

Top Left: Schematic diagram showing the seven conditions. FF=freefield; ILD=interaural level difference; RR=both markers from right side (external or intracranial); RL=markers from opposite sides(external or intracranial); Mono=Monaural listening.

A. Data from the seven conditions. Mean gap thresholds averaged across leading marker durations are shown as a fucntion of stimulus condition with standard error bars (n=8).

B. Data from subset of highly practiced listeners (n=4).

C. Data from relatively naïve listeners (n=4).

In B and C, mean gap thresholds are plotted as a function of leading marker duration, with condition as the parameter. Thresholds obtained for markers at different freefield locations are always higher than those for markers presented at different intracranial locations, or markers presnented monaurally with a level disparity.

locations determined by imposing a 12 dB interaural level differences of opposite sign on the two markers (ILD-RL). Finally, there were two monaural conditions in which the leading marker was 12 dB louder than the trailing marker (M-LQ) or the trailing marker was 12 dB louder than the leading marker (M-QL). These two conditions were achieved simply by shutting off the left or right channel respectively of the ILD-RL condition. When the left channel was being used, the headphones were turned around so that all monaural stimuli were played to the right ear.

Each gap threshold was determined in a single block of trials. In each trial, the listener was presented with a standard and a signal in random order, separated in time by 750 ms. The task of the listener was to indicate by mouse-click which stimulus was the signal (i.e., a two-interval, two-alternative, forced choice). A two-down/one-up adaptive tracking procedure (Levitt, 1971) with a step factor of 1.2, was used to determine the gap threshold. Starting gap durations were in the range from 15-53 ms (depending on task difficulty). The block of trials was terminated after ten reversals in the direction of change of the adaptive step, and the gap threshold was defined as the mean gap duration associated with the last six reversals. Listeners were tested on each stimulus condition until their thresholds stabilized (3-6 thresholds). A listener's final gap threshold for any specified stimulus condition was defined as the mean of those obtained in the last three blocks for that condition. Trials were self-paced, and no feedback was provided.

RESULTS AND DISCUSSION

The general purpose of Experiment 3A was to compare within- and between-channel gap detection performance under free-field and dichotic stimulus conditions. In Fig. 2.9A, the mean threshold collapsed across leading marker duration are plotted for each free-field and dichotic condition. Apparent differences in performance were observed between practiced and unpracticed listeners in terms of both the absolute value and the pattern of thresholds across leading marker duration. Therefore, results for practiced and unpracticed listeners are plotted separately as a function of leading marker duration in Figure 2.9B and 2.9C respectively.

As expected, the means for the three "within-channel" conditions (FF-RR, ILD-RR, M-RR) at different leading marker durations were similar and low (1.2 - 2.4 ms

Fig.2.9A, far right). A repeated measures ANOVA showed a significant main effect of leading marker duration [F(3,21)=11.18, p<0.01)], attributable to slightly higher thresholds for the 5 ms leading marker condition (2.2, 2.38, 2.39 ms respectively for the three within-channel conditions). This increase in "within-channel" gap thresholds with a short leading marker has previously been seen only infrequently by some laboratories (Forrest & Green, 1987; Phillips *et al.*, 1997; Phillips *et al.*, 1998), and more commonly by others (Crawley, 1999; Schneider & Hamstra, 1999; Snell & Hu, 1999; Oxenham, 2000). The absolute values of these thresholds are lower than in the previous experiments of this thesis, but are in the range reported from other labs (e.g., Forrest & Green, 1987). In this experiment, the two markers in the standard were shaped, but not separated by any silent period. This is in contrast to Experiments 1 and 2, in which the markers in the standard were shaped and separated by a 1.0 ms gap. This may account for the 1 ms or so lower thresholds observed in this experiment compared to Experiment 1. Interestingly, there was a significant, albeit small, effect of condition, with thresholds for the monaural condition being slightly higher than the binaural and free-field conditions [F(2,14)=6.0, p=0.013]. This is consistent with the notion that an increase in the number of neural channels carrying information about the presence of the temporal gap (in this case, more neural channels when both ears are stimulated) facilitates gap detection. Such an explanation is arguably similar to that which describes how increased frequency bandwidth (which corresponds to an increase the total amount of neural activation) improves gap detection, i.e. better gap detection performance for gaps in wideband noise compared with narrow-band noise (Eddins *et al.*, 1992).

Effect of Condition

When the leading and trailing marker were presented from speakers in opposite hemifields, the mean gap threshold, collapsed across leading marker duration was 13.53 ms (Fig. 2.9A). This was 2.5 times greater than the 5.36 ms mean threshold observed when the markers were lateralized to opposite sides of the head with ILDs through headphones. Under the monaural condition, in which the leading marker was 12 dB louder than the trailing marker (Loud-Quiet), thresholds were also relatively low at 4.84 ms. Increasing the relative level of the trailing marker by 12 dB (Quiet-Loud) generated a mean threshold of 3.37 ms. A two factor, fully repeated measures ANOVA indicated there was a significant main effect of condition [F(3,21)=6.61, p<0.003]. Planned comparisons demonstrated that the gap thresholds obtained for the free-field, between-channel conditions were significantly greater than those obtained binaurally or monaurally when the free-field stimulus was simulated dichotically over headphones with an ILD (F(1,7)=19.15, p<0.001). The gap thresholds for binaural (ILD-RL) and monaural level difference (M-LQ, M-QL) conditions did not significantly differ (F(1,7)<1, ns). This is consistent with the pattern observed by Oxenham (2000). However, the thresholds obtained here are markedly lower than those obtained in his study (4-5 ms on average vs. 15-20 ms). The reason for this discrepancy is difficult to determine. The discrepancy should not be due to practice, as even our unpracticed listeners obtained relatively low thresholds on the ILD and level difference conditions. It is possible that our method of using a limited number of noise samples (a technical limitation of the method used here) somehow provided inadvertant cues in the headphone task which were less available in the free-field presentation. This might be due to the enhanced trial by trial variability in the noise spectrum due to small head movements slightly modifying the HRTF in free-field presentation conditions. Such issues are addressed in Experiment 3B.

Effect of Leading Marker Duration

In the FF-RL condition there was a tendency, albeit only marginally significant [F(3,21)=2.43, p=0.09], for increased thresholds when short leading markers were used. This varied across listeners, with some unpracticed listeners showing a major effect, selectively in the free-field between-channel case (Fig. 2.9C), and some practiced listeners showing either no effect of leading marker duration, or exhibiting a specific notch (better thresholds) in their function at the 10 ms duration (Fig. 2.9B). This notch is possibly a carryover effect due to selective experience in Experiment 1 with short (10 ms) leading marker durations, which were used exclusively for many hundreds of trials per listener. There was no effect of leading marker duration in the binaural level difference condition [F(3,21)<1, ns]. There was a significant effect of leading marker duration for both of the monaural level difference conditions; however, it was in the reverse direction from that observed in the free-field (FF-RL). Thresholds were worse for longest leading marker (300ms) whether they were in a Loud-Quiet configuration [F(3,21)=6.47, p<0.01] or the Quiet-Loud configuration [F(3,21)=4.7, p=0.01]. In this case, the increase was driven by practiced listeners who performed poorly with long leading markers, perhaps because of their extensive practice on short leading markers. The implication is that the difference in thresholds between the headphone and free-field conditions is larger for shorter markers than at the 300 ms leading marker, as evidenced by a significant interaction between condition and leading marker duration [$F(9,63)=2.74$, $p<0.01$]. However, a comparison between the headphone and the free-field conditions indicated that despite this, the free field thresholds were still significantly larger at the long leading marker durations [$F(1,7)=7.72$, $p<0.01$].

There are two possible explanations for this effect. The first is that spatial separation in the free-field leads the gap markers to activate different perceptual channels, following the arguments of Phillips *et al* (1997; 1998), but separation by ILD or level does not. The existence of the leading marker effect with free-field spatial separation (at least in unpracticed listeners) but not with spatial separation by ILD over headphones, lends support to this idea. However, an alternative possibility is that the higher thresholds in the free-field separation condition might be accounted for by the presence of extremely large ILDs (>25 dB) in the higher frequency channels, resulting in larger level differences between the markers in those channels at either ear. Such large ILD differences between the two markers may have made the task functionally 'between-ear', at least for the higher frequency components. This is because the level of the highest frequency components in a given marker at the far ear (25-30 dB attenuation due to the head shadow) would be only about 5 or 10 decibels above detection threshold. The leading marker effect in the free-field was not observed in practiced listeners. The existence of the effect in unpracticed listeners could, in theory, be related to the spectral differences at each ear between noise markers separated in the free-field due to the HRTF, which perhaps leads to a more reliable leading marker duration effect. However, unpublished results from the Phillips laboratory have shown that the leading marker duration effect for conditions where there are spectral differences between the markers can also be reduced or eliminated with practice, and is non-existent in some unpracticed listeners. In general, there is some controversy with the use of leading marker duration effects as a 'marker' for differences on a stimulus dimension leading to activation of different perceptual channels, both within our laboratory and others who have not observed such an effect (e.g., Grose et al, 2001; Oxenham, 2000).

EXPERIMENT 3B:

SECOND STUDY OF THE EFFECT OF INDIVIDUAL CUES TO SPATIAL LOCATION ON GAP DETECTION PERFORMANCE FOR NOISE MARKERS UNDER HEADPHONE LISTENING

The thresholds obtained in Experiment 3A for conditions with markers separated in level and ILD were considerably lower than those observed by Oxenham (2000).There were a number of methodological differences between the two studies. A shortcoming of Experiment 3A was the use of a limited set of 'frozen' samples of noise for the markers. Furthermore, both the leading and trailing markers were fixed in duration, and thus it is conceivable that total duration of the stimuli might be used as a cue. Experiment 3B was conducted as a better replication of the conditions tested by Oxenham (2000), and also included an extreme condition of spatial separation – presenting the markers to different ears (as in Phillips *et al.*, 1997). The stimuli were generated digitally through MATLAB software (The Mathworks) using a method that allowed for a new piece of random noise to be generated for each marker in the standard and test stimuli on each trial. The trailing marker duration in both the standard and test stimuli was randomly jittered in duration for each trial. Two fixed leading marker durations (10, 150 ms) were tested. This replication itself was conducted for two reasons: 1) to ensure Oxenham's results were repeatable given similar methods and 2) to compare thresholds obtained with markers differing in ITD, ILD and monaural level, with those obtained when the markers are presented to different ears. This also was done in order to provide baseline data for comparison with the effect of similar stimulus configurations on auditory stream segregation tasks to be described in Chapter 3.

METHOD

Subjects

Data presented here are from 2 extremely practiced (1 male), and 2 moderately practiced female listeners. All were familiar with gap detection tasks.

Stimuli

There were six stimulus configurations for which gap detection thresholds were obtained. For each condition, gap thresholds were obtained using two different leading marker durations – 10ms or 150ms. The trailing marker duration was always variable between 100-300 in order to reduce the use of duration cues other than the silent interval. These are illustrated schematically in Figure 2.10A. Stimuli were created using MATLAB software (The Mathworks) and presented at a sampling frequency of 44100 Hz at 16-bit amplitude quantization on the audio processor of an Apple PowerMacintosh 8600. The stimuli were constructed by concatenating two independent noise bursts (0-22050 Hz) separated by an interval of silence (the gap). Each noise burst was shaped with 1 ms rise and fall times, including the standard stimulus. There was no silent period imposed between the fall of the leading marker and the rise of the trailing marker in the standard.

In the CONTROL condition, the leading and trailing noise bursts were both presented diotically at 67dB SPL. In the EAR condition, the leading and trailing noise bursts were presented to left and right headphone channels respectively at 67 dB SPL. In the ILD condition, the noise bursts were presented dichotically. The leading marker had an ILD of 15 dB, which favored the left ear (left ear: 75dB SPL; right ear: 60dB SPL), and the trailing marker had an ILD of 15 dB, which favored the right ear (left ear: 60dB SPL; right ear: 75dB SPL). This resulted in the leading marker being lateralized to the left side of the head and the trailing marker being lateralized to the right side of the head. In the ITD condition, the noise bursts were also presented dichotically, at an equal level of 67 dB to each ear. The markers were presented at equal loudness to each ear, but the leading marker had an ITD of –500 μs and the trailing marker had an ITD of 500 μs. This was accomplished by starting the leading marker at the left ear .5 ms earlier than the right ear; and starting the trailing marker .5ms earlier at the left ear. This resulted in the leading marker being lateralized to the left side of the head and the trailing marker being lateralized to the right side of the head.

Note that in all these conditions (except the control) the leading marker was lateralized left and the trailing marker was lateralized right. For efficiency, the reverse

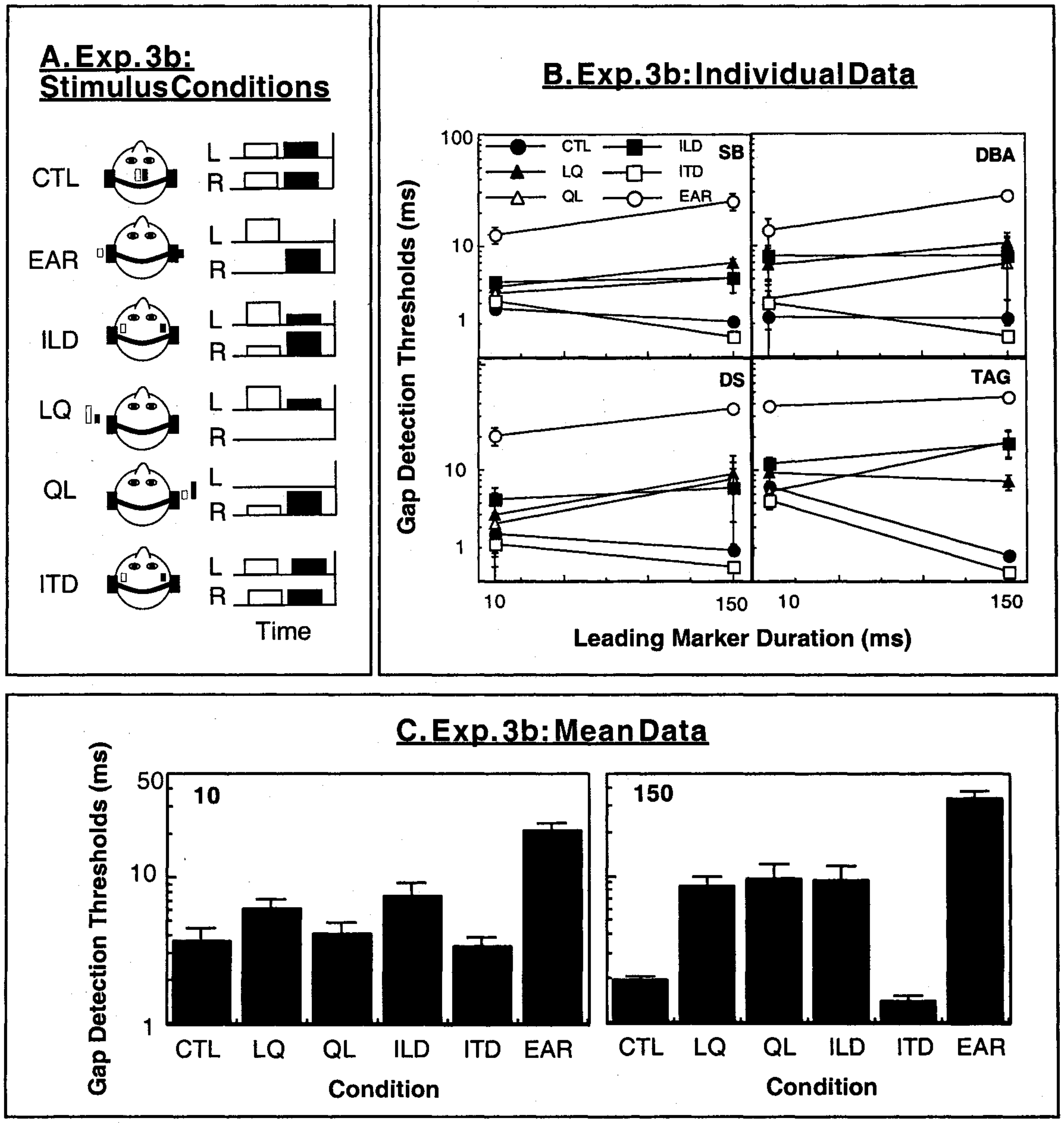


Figure 2.10 Experiment 3B

A. *Schematic of gap detection stimuli used for the 6 conditions.*

B. *Individual means of four listeners. Gap thresholds are plotted as a function of leading marker duration with the six conditions as parameters. The pattern was fairly consistent across listeners. Marker separation by EAR always resulted in highest thresholds, while separation by ITD always resulted in thresholds as low as the control conditions.*

C. *Mean data plotted separately for each leading marker duration (10 and 150 ms).*

conditions were not completed. This seems justified as no difference has been reported for the two directions (Phillips *et al.*, 1997; Phillips *et al.*, 1998). In the monaural level-difference condition in which the leading marker was louder, only the left channel of stimulus in the ILD condition was presented. In the monaural level-difference condition in which the trailing marker was louder, only the right channel of the stimulus in the ILD condition was presented.

For each stimulus presentation, new noises were generated for both markers in both the standard and test stimuli. The total stimulus duration varied, since the trailing marker varied in duration from 100-300ms. Stimuli were presented to subjects over AudioTechnica ATH-M40fs Precision Studiophones or Sennheiser HD 25 headphones, with listeners seated in a sound-attenuating booth.

Procedure

Thresholds for detecting the silent interval were measured using a two-interval, two-alternative forced-choice task, and a three-down, one-up adaptive method (see General Methods). At least three threshold determinations were made for each stimulus condition. The mean of the last three thresholds was taken as the subject's thresholds for that condition.

RESULTS AND DISCUSSION

Means from each of the four individuals are plotted in Figure 2.10B, with error bars representing the standard error of the 3 thresholds determinations which comprised the means. Thresholds are plotted as a function of the leading marker duration (10 or 150ms), with the condition (Diotic, 'loud-quiet'(LQ), 'quiet-loud'(QL), ILD, ITD, and EAR) as parameters. In the panel below, Fig. 2.10C, the mean thresholds across listeners are plotted as bars, separately for each leading marker duration. A 2-factor fully repeated measures ANOVA was performed using the Huyndt-Feldt correction for violations of sphericity where possible. The ANOVA revealed significant main effects for both leading marker duration [$F(1,3)=47.39$, $p<0.01$], and condition [$F(5,15)=24.08$, $p<0.01$], and a significant interaction [$F(5,15)=11.5$, $p<0.03$]. The data for each leading marker duration were analyzed separately in 1-way repeated ANOVAs. There was a significant effect of condition for both the short [F(5,15)=11.2, p<0.04] and long leading marker [F(5,15)=32.7, p<0.002].

Thresholds obtained when the markers were characterized by opposite ITDs never differed from the control condition with diotic markers (CTL). As can be seen in the mean data in panel C, for the short leading marker the only condition that is significantly elevated is the EAR condition. For the longer leading marker duration, separation of the gap markers by level (monaural or interaural) resulted in elevated thresholds, although not as elevated as when the markers were separated by EAR. These findings basically replicate those of Oxenham (2000), and extend them by demonstrating the relative ineffectiveness of separation by level compared with separation by ear. The thresholds obtained here for the ILD, LQ, and QL conditions were significantly lower than those obtained by Oxenham. The reason for this is not clear, though probably reflects that our listeners were more practiced than his were (Oxenham, personal communication; Feb, 2002). Most notably, the patterns are all similar. The pattern of thresholds across leading marker duration is inconsistent with that found by Phillips et al (1997). One of the main messages in that paper was that thresholds obtained in 'between-channel' conditions, where markers were separated by either ear or frequency, were sensitive to the brevity of the leading marker. In this study, the opposite effect was found: between-ear thresholds were significantly lower with shorter leading markers. This might be expected if listeners were using the onset-onset duration as a cue. With shorter leading marker, that base duration would be shorter, and following Weber's law acuity should be greater. In this study, the diotic conditions thresholds were slightly higher with short leading markers. This is consistent with data from several other labs (i.e. Snell, 1999; Oxenham, 2000).

The important point to take from this experiment is that gap detection was dependent on the similarity with which the markers activated the auditory periphery. If the markers were spectrally and intensively identical at each ear (as in the CTL and ITD conditions), then thresholds were low. If the markers differed in level at one or both ears, then thresholds were somewhat elevated. If the markers were separated completely at the periphery (segregation by ear), then thresholds were always elevated.

EXPERIMENT 4:

THE EFFECT OF SPATIAL SEPARATION, MONAURAL AND BINAURAL LISTENING, AND LEADING MARKER DURATION ON GAP DETECTION USING VIRTUAL AUDITORY SPACE METHODS.

A nagging, unexplained result from Experiment 1 is the benefit in free-field gap detection performance with spatially separated markers achieved when the ear near the leading marker ear is blocked (see Figure 2.4 A,B,D). One hypothesis was that the blocking of that ear caused both the leading and trailing markers to be misperceived as coming from the same hemifield, making the task 'within-channel'. However, in the case of blocking the ear near the trailing marker, both markers were also perceived as emanating from the open ear hemifield, yet thresholds were not improved due to the task becoming 'within-channel'. This could be attributed to a distinction between loud-quiet and quiet-loud configurations (as was argued in Experiment 1), yet Experiment 3B indicated that when presented over headphones there was no difference in thresholds under such LQ and QL configurations. There was also no leading marker effect for ILD or level difference conditions presented over headphones in Experiment 3, but there was a leading marker duration effect for markers separated in the free-field in unpracticed listeners. The monaural free-field conditions that are congruent with the headphone level-difference conditions in Experiment 3 were not completed. That is, we do not know if there is a leading marker effect for spatially separated free-field markers presented under monaural listening.

The design of Experiment 4 examined this issue, and addressed some arguable methodological shortcomings of the Ear Blocking method in Experiment 1. The first was that the binaural conditions were collected prior to the ear-blocking conditions. Thus, some of the threshold differences may have been caused by practice effects. The second is the problem attendant to achieving 100% ear blocking, particularly of low frequency components (see Wightman & Kistler, 1997b for a discussion). Most importantly, in Experiment 1 the sounds presented under binaural listening were presented at slightly different sound pressure levels (SPL) than those presented under monaural listening. Under binaural listening stimuli were presented at 35 dB above sensation level, but, under monaural listening, stimuli were presented at a level which corresponded to 2-3 dB below the listeners detection threshold when both ears were blocked. This corresponded to a sensation level of about 35 dB at the open ear for a sound played from 0 degrees azimuth, but was probably a slightly higher SPL than was used under binaural listening, since detection thresholds are higher under monaural listening.

In an effort to control for these issues, a method inspired from 'virtual auditory space' techniques was employed. While dichotic manipulation of sounds over headphones can shift the perceived laterality of a sound, this is limited to 'intracranial' space. As such, the spatial quality of the sound is not very realistic – i.e. the sound is not externalized. This occurs because the dichotic presentation does not take into account the HRTF – or filtering of the spectrum of the sound at each of the ears. Recent 'virtual auditory space' techniques have used recordings from the ear canals of clicks or noise bursts presented in the free field in order to obtain the head-related transfer function for a given spatial location. The transfer function taken at each ear is then mathematically imposed on a digital sound waveform. When the two filtered waveforms are presented to the two headphone channels, the resultant perception is of a spatially realistic, 'virtual' sound source corresponding to the location of the sound for which the HRTF was obtained (Carlile, 1996).

In order to re-examine the effect of monaural listening and leading marker duration on free-field gap detection performance, bilateral ear canal-entrance recordings were made of gap stimuli of various gap durations and two leading marker durations (10 and 100 ms) for 4 conditions from Experiment I. These conditions included both markers located at –90, leading and trailing markers located at –90 and 90 respectively, both markers located at 0, and leading and trailing markers located at 0 and 90 respectively. Those captured sounds were then replayed to the listener through headphones, one ear at a time (or, of course, binaurally) to obtain gap detection thresholds.

METHODS

Subjects

Participants included two of the female listeners from Experiment 1 (aged 29 and 36) and two relatively naïve listeners (a male and female, aged 25 and 21 respectively).

Gap Stimuli and Apparatus

Stimuli were generated and presented using the MATSS stimulus presentation system (see General Methods). An FFT of these waveforms showed little variation in the response of each speaker across frequency. Stimulus amplitude was set at 70 dB SPL(A-weighted) as measured with a sound level meter (Bruel and Kjaer) at the center of the room. All stimuli were wideband noise bursts, linearly ramped with 0.5 ms rise-fall times, including those defining the gap. Each stimulus consisted of a leading marker that was either 10 ms or 100ms in duration and a trailing marker of 300 ms duration. There were four configurations of leading and trailing marker locations (leading location/trailing location): 0°/0°, +90°/+90°, 0°/+90°, +90°/-90°.

Procedure

Binaural recordings

To generate sound files that could be presented over headphones, stimuli to be used in a 2AFC staircase procedure were presented over the speakers in the free-field and recorded through SoundMan Studio 'binaural microphones' (matched to within 0.5 dB) placed at the entrance of each ear canal. The output of these microphones was passed through the input channels of the ECHO Darla sound card and saved digitally in MATLAB. For a given leading marker duration and a given leading and trailing marker speaker pair, five recordings were obtained for each of 21 gap durations, decreasing from 30 ms to .9 ms by a factor of 1.2 (i.e. 30, 25, 20.83, 17.36…). In each recording, new noise was generated for each marker.

The Gap Detection Task

Stimuli were presented to subjects at a comfortable listening level, near that at which the sound was recorded, over AudioTechnica ATH-M40fs Precision Studiophones or Sennheiser HD 25 headphones, while they were seated in a sound-attenuating booth. For the +90/+90 condition, sound levels recorded from the headphone channels were 75dB SPL for the right ear channel and 63 dB SPL for the left headphone channel. For the 0/0 condition, sound levels were 72 dB in each ear. These values are internally consistent with expectations of the size of the head shadow, validating the recording technique. Each stimulus trial consisted of two of the above recorded sounds, separated by an interval of 300 ms. One of the two was *the standard*, and was randomly chosen on each trial from the set of recorded gap stimuli which included an undetectable .9 ms silent period between the markers (to control for gating transients present in the signal sequence). The other was *the signal*, and was randomly chosen from the set of five stimuli with the appropriate gap duration for the point in the adaptive staircase. For every trial, the duration of the trailing marker of both the standard and test stimuli was shortened by removing a random value between 0 and 100 ms from the end and re-imposing a fall time of 1ms. Gap thresholds for every stimulus condition were obtained using the 3 down, 1 up adaptive tracking procedure described in the General Methods (p.56). Thresholds were obtained for three listening conditions: stimuli presented to both headphone channels (binaural configuration which simulates binaural listening in free field) or presented only to the left or right headphone channel (simulated blocking of the right and left ear in the free-field respectively).

RESULTS AND DISCUSSION

The mean thresholds obtained for four individual subjects, with standard errors of the three threshold estimates contributing to that mean, are displayed in the first four rows of Figure 2.11. The mean thresholds, with standard errors, are presented in the bottom row. The figures plot thresholds obtained with each ear and with both ears for leading markers of either 10 or 100 ms. Each column presents the data obtained for a given spatial configuration – 90/90, 90/-90, 0/0, 0/90. The results were consistent across individuals (first four rows, Fig 2.11) so the mean results will be discussed (bottom row, Fig. 2.11).

Leading Markers at +90

The first two columns of Figure 2.11 show the results for leading markers at +90 degrees and trailing markers at +90 and –90 degrees. These data were analyzed by a 3-way repeated measures ANOVA (listening condition, leading marker duration, and spatial separation). The only significant main effect was of 'spatial separation' [$F(1,3)=146.43$, $p<0.001$]. There was an interaction of 'listening condition' with 'leading marker duration' [$F(2,6)=21.32$, $p<0.003$], which reflected that left-ear thresholds (for which the leading marker shadowed) were insensitive to leading marker duration, while

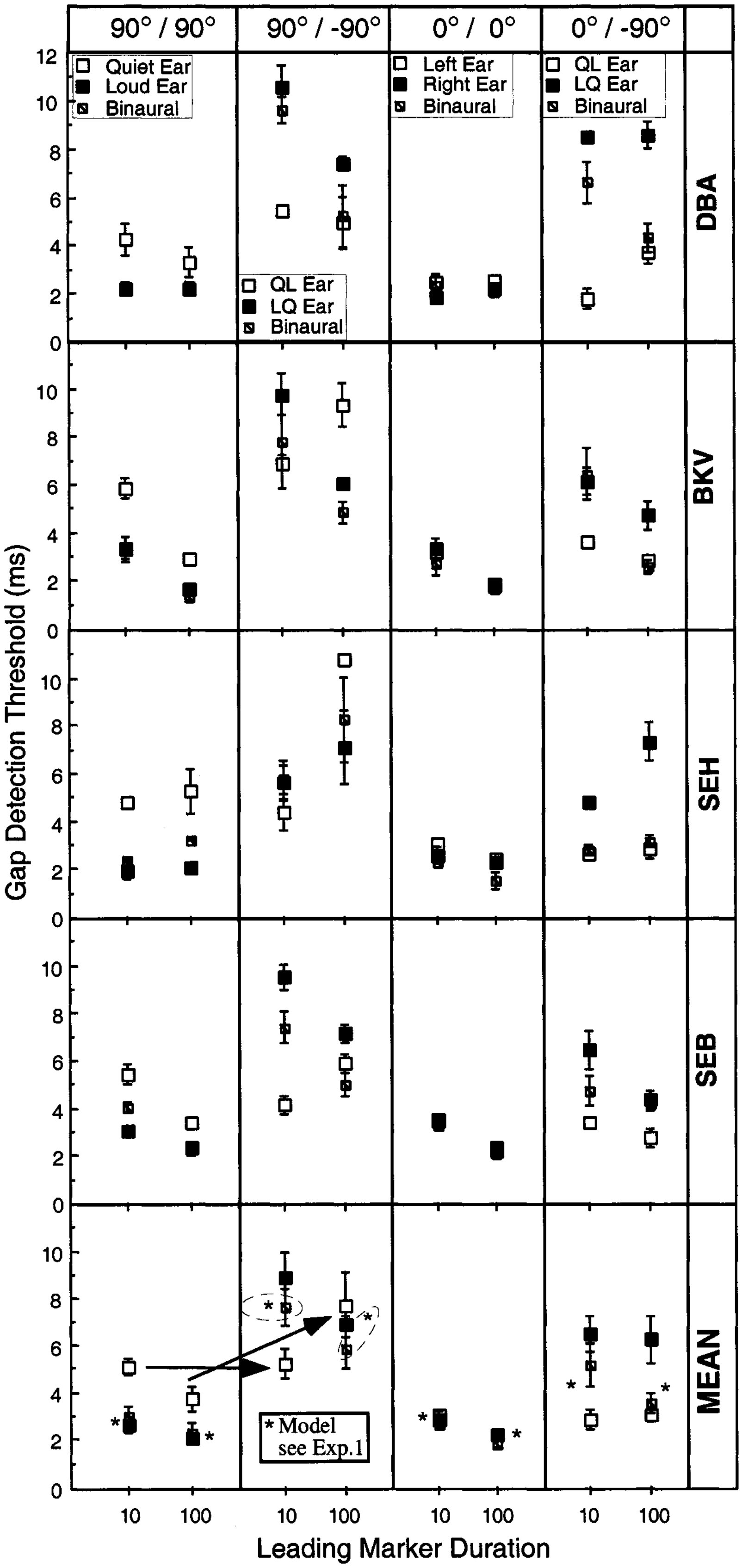


Figure 2.11 Experiment 4. Spatial separation effects on gap thresholds obtained over headphones using virtual auditory spatial stimuli. Gap stimuli were played from a subset of marker locations used in Exp.1 (leading/trailing marker locations specified at the top of each column) and two leading marker durations (10, 100 ms). Recordings of the gap stimuli were made at the entrance of each ear canal of each listener. Thresholds were obtained by playing the recorded stimuli back to the listener over headphones, under binaural listening or through each headphone channel alone. Predicted thresholds from the weighted average model from the analysis of Exp. 1 (weighting factor X=0.3), are plotted as stars (*). Arrows show the interaction between the trailing marker ear performance advantage only holds for the short leading marker.

right-ear and binaural thresholds were worse for the short leading-marker stimuli. There was also an interaction of 'listening configuration' with 'spatial separation' [F(2,6)=10.15, p=0.01] which reflected a lack of spatial separation effect for the left-ear thresholds compared with a strong effect of separation on binaural and right-ear thresholds. Finally, there was a 3-way interaction [F(1,3)=45.0, p=0.0002]. A 2x3 (listening condition x spatial separation) repeated measures ANOVA was performed on the data for each leading marker duration to elucidate it. This revealed a prominent interaction of listening configuration with spatial separation with a 10 ms leading marker [F(2,6)=40.1, p<.001], but not the 100 ms leading marker (F<1). This reflects the lack of spatial separation effect with left-ear listening at 10 ms (see arrow), compared with the large effect at 100 ms. This suggests that the benefit achieved with ear-blocking in Experiment 1 with spatially separated markers ('quiet-loud' configuration), and generally the pattern of data obtained with spatial separation with ear-blocking, might not generalize to longer leading marker durations.

In the +90/+90 marker configuration, low thresholds (2-3 ms) were obtained both under binaural listening, and under listening with the right ear (the ear of the hemifield in which the markers were located). Consistent with the reduced level and bandwidth of the markers at the shadowed ear, thresholds are elevated when listening with the left ear (4-5ms). This pattern was the same for both short and long leading markers. A 2x2 (lead duration x listening condition) ANOVA revealed a significant main effect of listening condition [F(1,3)=37.2, p<0.01], reflecting the higher thresholds at the shadowed ear, but no significant main effect of marker duration (p>0.2). While there is some evidence suggesting that gap thresholds become elevated as the leading marker duration is shortened, this effect is typically not as salient in practiced listeners such as those in this study (Snell and Hu, 1999). This replication of both the absolute size (2-3 ms) and effect of level and bandwidth on 'within-channel' gap thresholds suggests that this method of recording free-field gap stimuli and playing them back over headphones was successful.

The critical condition is that of 90/-90 where the markers are located in opposite hemifields. For the short leading marker (10ms), as used in Experiment 1 and Boehnke and Phillips (1999), thresholds obtained at the 'Quiet-Loud' ear (ear in hemifield opposite to leading marker) were improved relative to binaural listening and thresholds at the 'Loud-Quiet' ear were elevated relative to binaural listening. This is precisely the pattern of thresholds predicted by the 'WEIGHTED AVERAGE' model in the analysis of Experiment 1 (also see Appendix A). For comparison, the data from the left and right ears were averaged using the weighted average model that best predicted mean gap detection thresholds in Exp.1. The predicted threshold for binaural listening is plotted next to the actual value obtained (* symbol). In that regard, this experiment replicated the results for these conditions in Experiment 1. The most important result obtained in this experiment was that the 'benefit' of monaural listening in the 'QUIET-LOUD' configuration was abolished for longer leading marker durations (100ms). For the 100ms leading marker, thresholds under 'quiet-loud' conditions were as elevated as thresholds obtained under binaural listening and right-ear listening (as shown by the arrow). This suggests the data and analyses from Experiment 1 are limited to configurations with short leading markers.

Leading Markers at 0

In the last two columns of Figure 2.11, the results are plotted for conditions in which the leading marker was located at 0 degrees and trailing marker was located at 0 or +90 degrees. These data were analyzed by a 3-way repeated measures ANOVA (listening condition, leading marker duration, and spatial separation). The only significant effect was a main effect of 'spatial separation' [$F(1,3)=23.04$, $p<0.02$], no other effects reached even marginal significance. The pattern of the mean thresholds for the 0/+90 condition for the 10 ms leading marker do follow the pattern obtained in Experiment 1 (low thresholds for the 'quiet-loud' ear, elevated thresholds for the 'loud-quiet' ear, and intermediate thresholds with binaural listening). However, the magnitude of difference between left-ear, right-ear, and binaural listening is attenuated. Thresholds were generally lower for all conditions in this study – as if the more practiced subjects became, the more attenuated was the effect of spatially separating the markers.

Why then does the 'benefit' with ear blocking, observed in Experiment 1, disappear at long leading markers for markers separated in different hemifields? The fact that Oxenham did not observe any difference in gap thresholds for 'Loud-Quiet' and 'Quiet-Loud' configurations with spectrally identical markers presented over headphones suggests that differences observed here might result from spectral differences between the markers imposed by the location-dependent spectral filtering by the head and pinna. In the case where the QL benefit occurs, the stimulus is a quiet, brief, low-pass sound followed by a louder wideband sound. Both the existence of a level difference between the markers, and a spectral difference between the markers due to head-shadowing, would suggest that gap thresholds would be elevated for left ear listening. This elevation was observed for the100 ms leading marker, and thus it is puzzling why it was not observed with the shorter leading marker duration (10 ms). The improved thresholds obtained near the trailing marker ear for short leading marker durations require another explanation. Introspection about the cues used to perform the task under that configuration suggested that the task was different. In the 2AFC task, the short, low-pass leading marker of the standard (no gap) might be rendered inaudible when followed immediately by the loud trailing marker. This is logically consistent with a backward masking effect (Elliot, 1962). Excellent performance can be achieved by listening not for the GAP per se, but for the interval in which a leading marker is audible. When the gap duration in the 'test' stimuli reaches about 4-6 ms, the leading marker in the 'test' begins to also become inaudible and the detection strategy fails. In the same condition when the other ear is blocked, such a cue is unavailable as the leading marker is louder than the leading marker and always audible. Such a strategy is impossible in the case with a 100 ms leading marker, as that marker is always audible in both the 'standard' and 'test' intervals.

If there are 'temporal masking' factors contributing to the pattern of gap detection thresholds across various spatial separations, then it is less surprising that the results from Boehnke & Phillips (1999) map so precisely onto the free-field release from masking results of Saberi *et al.* (1991) because both patterns are, at least in part, the result of the spectral and level changes at either ear caused by the shadowing of the head. Consistently, the results of Experiment 2 map well onto the results of Gilkey and Good (1995), who measured release from masking with spatial separation as a function of the frequency pass-band of the stimuli (see Fig. 2.8B). In both experiments, there was minimal effect of spatial separation for low-pass stimuli, for which there is minimal shadowing by the head. A remaining question then is whether the pattern of results obtained in Boehnke & Phillips (1999), and by extension the hemifield model they proposed based on that data, generalizes to longer leading markers. The previous four experiments examined the contributions of pass-band, monaural listening and leading marker duration to the effect of spatial separation of markers on gap detection thresholds. However, they did not always use the same listeners or the same equipment and methodology. Listeners that participated in multiple experiments were subject to practice effects and order effects. As such, these stimulus factors have not be examined together in a balanced design using similarly experienced listeners. Such an experiment is described next.

EXPERIMENT 5:

FULLY FACTORIAL STUDY OF EFFECTS OF SPATIAL SEPARATION OF MARKERS, BAND-PASS OF MARKERS, LEADING MARKER DURATION AND LISTENING CONDITION (MONAURAL/BINAURAL) ON GAP DETECTION PERFORMANCE BETWEEN MARKERS AT DIFFERENT FREE-FIELD LOCATIONS.

This experiment takes advantage of the knowledge acquired from the previous experiments regarding the relevant variables to examine and the various aspects of stimulus control. The story emerging from the previous four experiments is as follows. Spatial separation of wideband markers in the free-field elevates gap detection thresholds (Exp. 1). This elevation is not due to perceived lateralization, as differences in the ITD of the markers, or in the spatial location of low-pass markers, did not elevate gap detection thresholds (Exp.2 and 3B). Thresholds obtained with high-pass noise markers are strongly affected by spatial separation (Exp. 2). What is unclear is whether the effects of marker spatial separation are comparable for high-pass and wideband noise markers. This matters, because it addresses the question of the extent to which marker level differences influence gap detection thresholds. Because Experiment 1 and 2 used different listeners and a different methodologies, we can not know for certain if thresholds obtained with high-pass noise were more strongly influenced by spatial separation than those obtained with wideband noise. If true, the results would support the proposed mechanism that marker level differences are the major contributor to threshold increase with separation.

In a fully factorial design (see schematic in Fig. 2.12), this experiment examines all the variables that have been shown to be relevant in elucidating the mechanisms underlying marker spatial separation effects on gap detection in the same practiced listeners. The effect of marker band-pass (low, high, wide), leading marker duration (10, 50, 100 ms), listening condition (right ear only, left ear only, binaural) and spatial separation of markers in the free field (90/90, 90/0, 90/-90) on gap detection thresholds was determined. Thus, it was a four factor, fully-repeated measures design, with three levels of each factor (3x3x3x3). This results in 27 functions of gap detection threshold with spatial separation: One for each marker band-pass, listening condition, and leading marker duration. This allows two types of comparisons. First, it allows visualization of the degree to which binaural performance is dictated by that available at the left and right ears alone. That is, are we doing the task with our right or left ear? Second, it allows determination of the degree to which the wideband results can be explained by those for high- and low-pass noise. That is, are we listening to the high- or low-frequency channels when we do the task with wideband noise?

METHODS

Listeners

The four most experienced listeners in our subject pool (SEB, SEH, BKV, and DBA) participated in this experiment. Each had participated in at least two of the studies previously described studies in this chapter, in addition to many other auditory temporal processing tasks.

Gap Stimuli and Apparatus

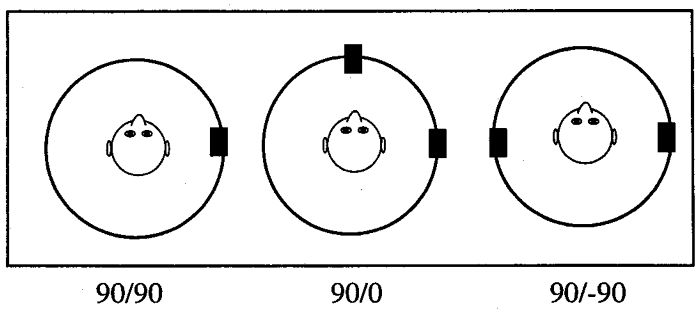
The stimulus construction and Methods were identical to Experiment 2. To reiterate, the low-pass noise stimuli contained frequency components of equal amplitude between 50 Hz and 1 kHz; the high-pass stimuli ranged from 3 kHz to 22.05 kHz. The wideband noise ranged from 50 Hz to 22.05 kHz. The main difference was that the leading marker duration took values of 10, 50, or 100 ms.

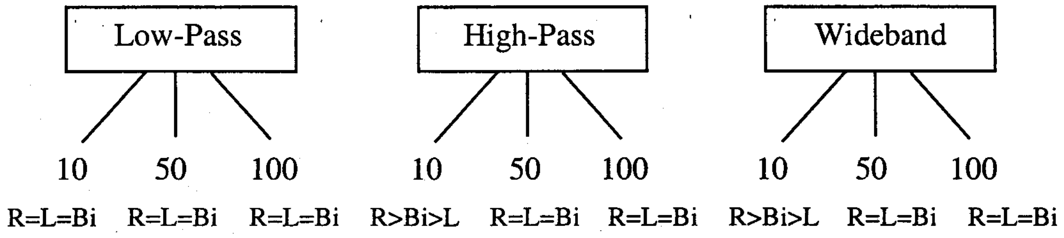
Procedure

The procedure was similar to that of Experiment 2. Detection thresholds for 500 ms noise bursts, of each spectral pass-band used in the experiment, were obtained for

EXPERIMENT 5: DESIGN SCHEMATIC

3 spatial separations (90/90,90/0,90/-90)
X
3 frequency pass-bands (Low,High,Wide)
X
3 leading marker durations (10,50,100)
X
3 listening conditions (L,R,Binaural)





No effect of spatial separation

Strong effect of spatial separation

Thresholds highest overall due to
greater level differences

Benefit due to Ear-blocking only
with short leading marker

Figure 2.12 Schematic of the design of Experiment 5, and predicted results.

each listener using a 2AFC 3-down, 1-up procedure. Stimuli were presented at 35 dB above the detection threshold level for that stimulus.

In the gap detection task, listeners received feedback via a light emitting diode (LED) at zero azimuth which flashed green for correct responses and red for incorrect responses. In order to control for order and practice effects, each listener provided one threshold for each of the 27 stimulus conditions for a given listening condition (Binaural, Left or Right Ear blocked) in a random order. One set of 27 conditions was completed for each of the three listening conditions (complete set of 81 conditions) before going on to complete a second threshold determination. Stimulus generation, presentation, collection of responses and randomization of the conditions were controlled entirely by a program running through MATLAB. Listeners typically completed one block of 27 thresholds in a 2 -3 hour session. Listeners took forced breaks after every 10 threshold determinations, but could pause the program to take a break at any time. Ear blocking was achieved as in Experiment 1.

RESULTS AND DISCUSSION

There were four variables, each with three levels (81 conditions). An 'omnibus' 3x3x3x3 repeated measures ANOVA was first performed in order to establish the main effects and interactions. Significance threshold was always determined by the Huynh & Feldt adjustment of the degrees of freedom in order to correct for violations of sphericity (Huynh & Feldt, 1976). There was a main effect of marker separation [F(2,6)=32.32, p<.004], but only weak, statistically insignificant effects of stimulus pass-band (p=.08) and listening condition (p=.07). None of these are surprising, because the interesting effects are in the interactions. There were two statistically significant 2-way interactions. The first was of pass-band with leading marker duration [F(4,12)=6.12, p=0.02]. This reflected that thresholds obtained with low-pass filtered markers on average improved with leading marker duration, while thresholds obtained with high-pass and wideband markers were relatively unaffected by leading marker duration. The selective effect of marker duration on responses to low-pass noise is probably explicable in terms of the extreme trial by trial variability in quality of a brief low-pass noise. Examination of the waveform of a low-pass noise reveals that very few cycles of stimulus frequency occur in

10 ms of low-pass noise (e.g. usually less than three waveform fluctuations). This is reflected in the perception of randomly generated 10 ms low-pass markers, which vary widely in pitch and loudness. The second interaction was of pass-band with separation [F(4,12)=33.46, $p<0.0001$]. On average, there was no effect of spatial separation on thresholds obtained with low-pass leading markers, and a strong effect of separation on high-pass and wideband markers. This replicates the results of Experiment 2. Finally, there was a 3-way interaction between pass-band, listening condition and separation [F(8,24)=3.52, $p< 0.02$]. Additional 3x3x3 ANOVAs were performed for each level of pass-band, and for each listening condition in order to elucidate this interaction.

The analysis is complex, so the data are presented in two ways to illustrate the story. The spatial separation variable is always on the x-axis; the frequency pass-band, listening condition and leading marker duration are the parameters varied across graphs. The best strategy is to compare the shape of the function across these parameters. The three columns of Fig. 2.13A-C present the data for each frequency pass-band of the markers - low-pass, high-pass, and wideband, separately. In each column, there are three plots – one for each leading marker duration (10, 50, 100 ms). In each of these panels three functions of gap thresholds with trailing marker location are plotted – one for each listening condition - Binaural (black line), Left ear (blue line), Right ear (red line). In essence, this first set of graphs was designed to facilitate comparison of the spatial separation effects on the three listening conditions. Examination of the three functions, elucidates the contribution of the thresholds obtained with each ear to those obtained under binaural listening.

Another question we might ask is whether the wideband data are best accounted for by the high-pass data, the low-pass data, or some combination. In other words, are we limited by our high-frequency channels or low-frequency channels during the task? To facilitate such an analysis, in a second set of Figures (Fig. 2.14A-C), the same data were plotted in another way. In each column (A-C) a set of 3 graphs (one for each leading marker duration) are presented for a given listening condition. Three functions of trailing marker location with gap detection threshold are plotted, one for each frequency pass-band used. The line colors follow the color of the lines for each listening condition in Fig. 2.13A-C (black for binaural, blue for left-ear, red for right-ear).

Figure 2.13 ***Experiment 5:***

Marker passband X Listening condition X Leading marker duration X Spatial separation

This figure illustrates the effect of spatial separation of markers on gap detection thresholds obtained with either ear alone and using binaural listening. This effect is examined for leading marker durations of 10, 50 and 100 ms in the three panels of each column. The results for each marker frequency condition are presented in separate columns. This facilitates evaluation of the contribution of thresholds obtained with each ear to the results obtained under binaural listening.

A. LOW-PASS NOISE

There were no significant effects. In particular, there was no effect of spatial separation, consistent with data from Experiment 2.

B. HIGH-PASS NOISE

The only significant main effect was spatial separation ($p<.001$). There was also a significant interaction between listening condition and spatial separation ($p<0.03$). The elevated thresholds observed for left-ear listening for a trailing marker at 0 might reflect the presence of irrelevant frequency information in the trailing marker due to shadowing of the leading marker by the head (see text; Formby, 1998b).

C. WIDEBAND NOISE

There are main effects of listening condition ($p=0.03$) and spatial separation ($p=0.02$). There is also a 3-way interaction between listening condition, leading marker duration and spatial separation. This interaction reflects the fact that listening with the left ear for short leading markers shows no effect of spatial separation (examine relative placement of the blue line across the 3 panels). This interaction is important. It demonstrates that the benefit of earblocking observed in Experiment 1 for configurations in which the leading marker was in the blocked ear hemifield is a product of the short leading marker used (see text).

Figure 2.14 ***Experiment 5***

This figure presents the same data as those in Fig. 2.13, but plots functions of gap thresholds with spatial separation for different marker frequencies together, in separate columns for each listening condition. This facilitates evaluation of the contribution of data obtained with low- and high-passed stimuli to that obtained with wideband stimuli.

A. LEFT-EAR LISTENING

There was a main effect of spatial separation ($p<0.04$), an interaction of marker frequency and separation (as in A & B) and additionally, a 3-way interaction between leading marker duration, frequency and separation ($p<0.04$). This reflected the lack of spatial separation effect for wideband markers with a 10ms leading marker (the so-called quiet-loud effect from Exp. 1) compared with a strong separation effect at longer leading markers.

B. RIGHT-EAR LISTENING

There was a main effect of spatial separation ($p<0.01$) and an interaction between marker frequency and separation (as in A).

C. BINAURAL LISTENING

There was a significant main effect of spatial separation ($p<0.01$) and an interaction between marker frequency and separation ($p<0.01$). This interaction indicates that gap thresholds obtained with low-passed markers showed a far weaker spatial dependence than that of high-passed and wideband markers.

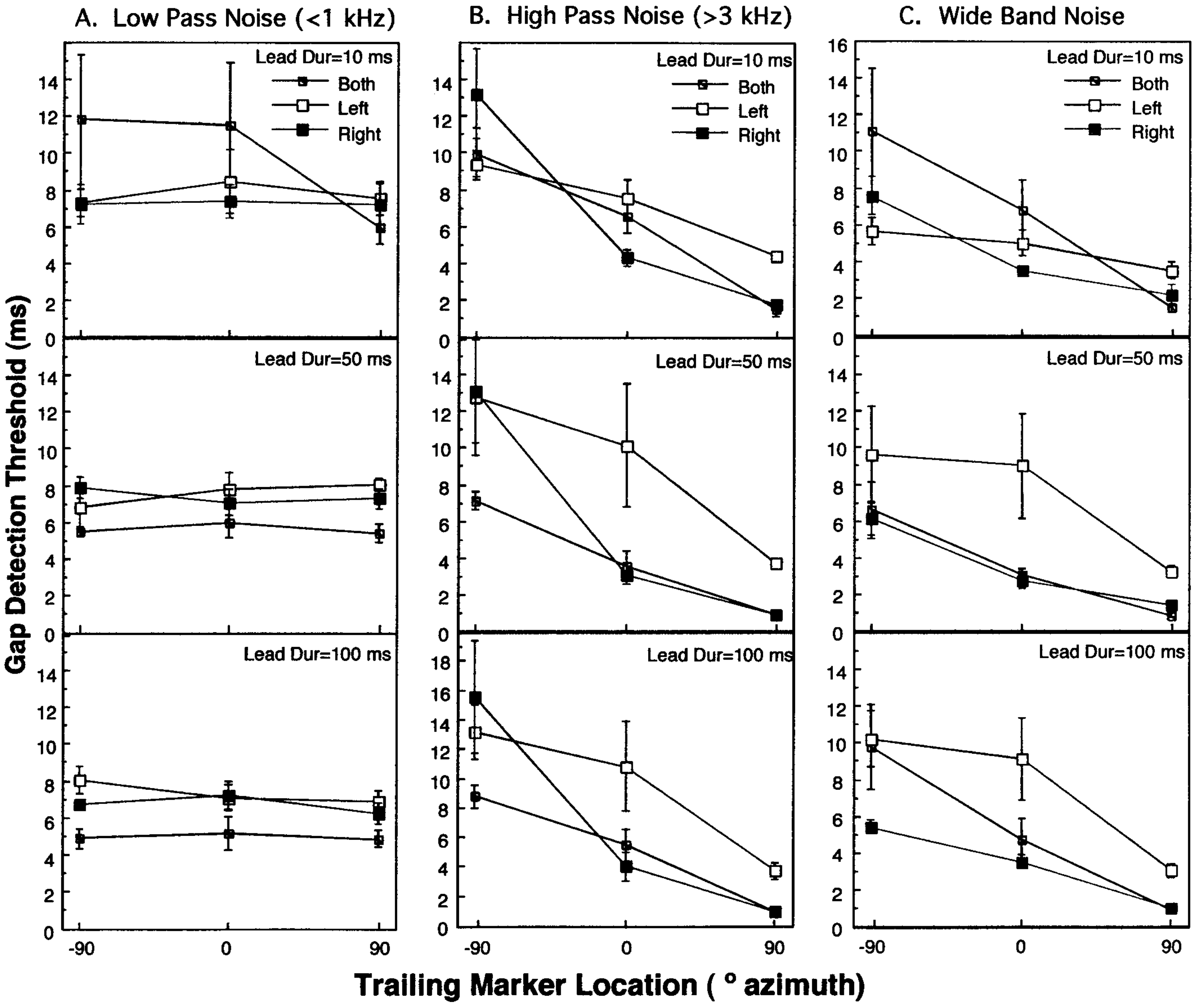

A. Low Pass Noise (<1 kHz)
B. High Pass Noise (>3 kHz)
C. Wide Band Noise
Lead Dur=10 ms
Lead Dur=50 ms
Lead Dur=100 ms
Both
Left
Right
Gap Detection Threshold (ms)
Trailing Marker Location (° azimuth)
0
2
4
6
8
10
12
14
16
18
-90
0
90

Figure 2.13

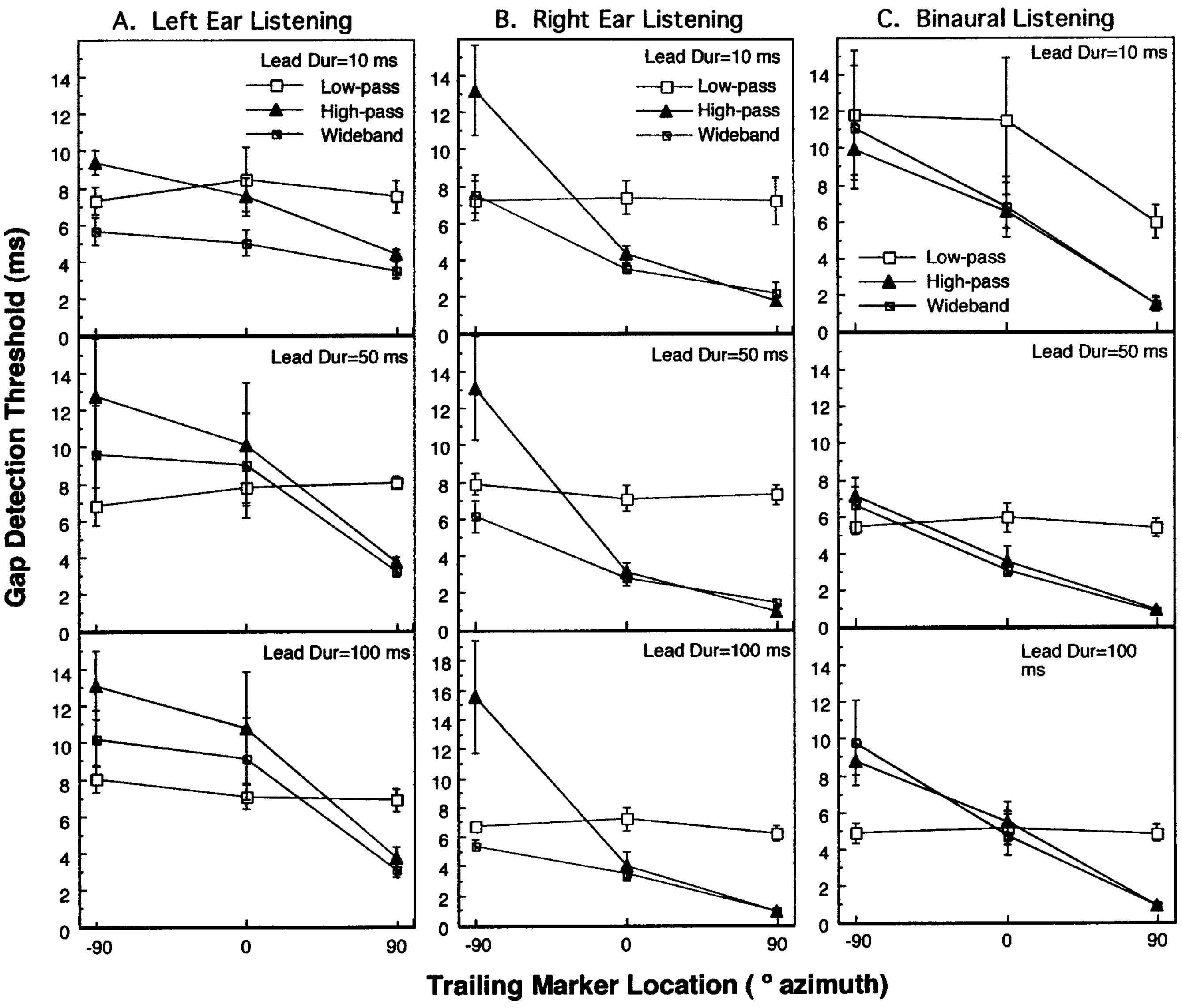

A. Left Ear Listening
B. Right Ear Listening
C. Binaural Listening
Lead Dur=10 ms
Lead Dur=50 ms
Lead Dur=100 ms
Low-pass
High-pass
Wideband
Gap Detection Threshold (ms)
0
2
4
6
8
10
12
14
16
18
Trailing Marker Location (° azimuth)
-90
0
90

Figure 2.14

Low-pass Noise

The data obtained using low-pass noise markers (<1kHz) are plotted in Fig. 2.13A. The functions are generally flat. That is, gap thresholds were unaffected by spatial separation. With two exceptions, all mean thresholds fell between 5 and 9 ms, regardless of listening-condition and leading marker duration. Note also that the absolute values of the thresholds are higher for co-located markers than observed with high-pass or wideband markers. This is likely due to bandwidth effects on gap detection. A 3X3X3 ANOVA revealed there to be no significant main effects or interactions. The apparent exception in the Figure is that of binaural listening with a short leading marker (see black line, Fig. 2.13A, top) for which thresholds appeared to be affected by spatial separation. This effect was largely driven by a single listener, and was not statistically significant. Brief low-pass markers tend to vary dramatically in quality from trial to trial, since only a few cycles of stimulus frequency may occur during the time interval. This ambiguity is probably the source of the elevated and variable thresholds for that condition.

High-pass Noise

The data obtained using high-pass noise markers (>3kHz) are plotted in Fig. 2.13B. Note that the absolute values of the thresholds for the largest marker separation were the highest in the experiment, while the thresholds for similarly located markers were among the lowest in the experiment (note the scale change in the y-axis for the bottom panel). There was a strong effect of spatial separation [$F(2,6)=53.0$, $p<0.0002$] and an interaction between listening condition and separation [$F(4,12)=4.1$, $p<0.03$]. This interaction reflects the different patterns of sensitivity to marker separation for each listening condition. Under binaural listening and left-ear listening, thresholds increased roughly linearly with separation. Under right-ear listening (for which the stimulus becomes 'loud-quiet' with spatial separation) thresholds increased more exponentially with spatial separation of the markers. A dramatic increase was observed when the markers were separated across the midline, consistent with the decrease in level of the trailing marker at that ear by the shadow of the head.

Of note are the generally elevated thresholds for the left ear, particularly at long leading marker durations. This is the ear near the trailing marker location, for which thresholds in Experiment 1 were low with separated markers (the so-called Quiet-Loud benefit). This elevation, particularly for the trailing marker at 0 degrees, was puzzling at first, particularly since thresholds for the other ear (at which the markers were 'loud-quiet') were not much elevated at that separation. It suggests two things. First, the apparent distinction between 'QL' and 'LQ' proposed in experiment 1 was less a product the intensity of the stimuli as it was a result of some other factor relating to the short leading marker duration. As mentioned in Exp. 4, backward masking of the short leading marker is one possibility. Second, the results here suggest that spectral factors might play a role when the leading marker is long enough to be 'sampled' for its frequency content. The difference in the magnitude of thresholds for the left and right ears for the trailing marker at zero might be explained by the different spectral shape imposed on the markers by the head shadow. At the left ear (blue symbols), the gap stimulus is a low-pass marker followed by a broadband marker and thresholds are elevated. At the right ear (red symbols), the gap stimulus is a broadband noise marker followed by a low-pass noise marker, and thresholds are low. The main difference is that there is additional spectral information in the trailing marker at the left ear, and less spectral information in the trailing marker at the right ear. Formby (1998b) showed that additional spectral information in the trailing marker (but not in the leading marker), elevated thresholds. Thus, the difference in thresholds obtained through the two ears can be explained well by known effects of level and spectral differences between the markers, and the change in thresholds across spatial location reflects the change in level and spectrum of the two markers at each ear as a function of their spatial location, as described in the Chapter 1.

Wideband Noise

The data obtained using wideband noise markers are plotted in Fig. 2.13C. This panel shows data for stimulus configurations which are analogous to those used in Exp. 1 and Exp. 4. There were main effects of listening condition [F(2,6)=6.84, $p<0.03$] and separation [F(2,6)=19.27, $p<0.02$]. Furthermore, there was a significant 3-way interaction between listening-condition, duration, and separation. An analysis of that interaction indicates that the interaction of listening condition with separation was different for the three leading marker durations. Specifically, there was no effect of separation for the left ear at the 10ms leading marker, but a large effect of separation at longer leading marker durations (consistent with Exp. 1 and 4). The results for the short leading marker generally replicate the results from Experiment 1 – that thresholds obtained with the ear nearest the trailing marker are low with spatially separated markers (the so-called 'Quiet-Loud' benefit). This benefit is clearly related to the brevity of the leading marker, and may be best explained by backward masking of the leading marker as described in the discussion of Exp. 4. The duration of the gap threshold (4-6ms) is consistent with the temporal distance over which backward masking effects tend to have a strong effect (Elliot, 1962). The effects of separation for right-ear and binaural listening were generally similar across all leading marker durations, generally showing a linear increase in gap thresholds with increases in spatial separation across the midline. However, the binaural function tended to have a steeper slope, particularly at the long and short leading marker durations. This result is not consistent with Exp.'s 1 and 4, for which thresholds obtained with the ear near the leading marker were elevated relative to binaural listening. The reason for this is not clear. Because binaural thresholds were not intermediate between those obtained with the two ears, predicting binaural thresholds based on any averaging of the thresholds from the two ears, as was done in the analysis of Experiment 1, would be unsuccessful with this data set.

There seems to be no simple correspondence of the binaural data to the data from either ear alone. For short leading markers, the relatively larger elevation in binaural thresholds with spatial separation compared to those with either ear alone was not consistent across listeners. For the 50ms leading marker, the binaural thresholds are clearly determined by the ear near the leading marker, because the thresholds for the two conditions are basically indistinguishable in absolute pattern and values across spatial separation of the markers. For the longer leading marker duration, the binaural listening condition appears determined by the right ear for the same location, and small spatial separation, but then is similar to the left ear for the large separation.

What is most surprising, is the large elevation in gap thresholds with spatial separation obtained with the left ear for longer leading markers, as was observed for the high-pass markers described above. In this case, the leading marker is always shadowed (i.e., low-pass). Thus, it appears that the addition of energy in the high frequency components at that ear as the trailing marker was moved towards it was detrimental to gap detection. Again, this seems to be consistent with the observations of Formby et al. (1999), suggesting it is a spectral effect.

We now turn to comparing the functions relating gap threshold and spatial separation obtained using different frequency bands. This facilitates analysis of the contribution of the low and high frequency components to the function obtained with wideband noise.

Binaural Listening

The data obtained under binaural listening are plotted in Fig. 2.14C. Regardless of leading marker duration, the thresholds obtained with wideband noise perfectly map onto the thresholds obtained with high-pass noise markers. Clearly, there is little contribution of the low frequency channels in binaural listening. This is consistent with the lack of an effect of imposing ITD differences between the markers (Oxenham, 2000 and Experiment 3), a cue which is coded by low frequency neurons (Goldberg & Brown, 1969). There was a main effect of separation [F(2,6)=18.55, p<0.02] and an interaction of frequency with separation [F(4,12)=6.87, p<0.04] that reflects the fact that there was less effect of spatial separation on thresholds obtained with low-pass noise markers.

Right Ear Listening (Leading marker ear)

The data obtained using right-ear listening (leading-marker ear) are plotted in Fig. 2.14B. There was a main effect of separation [F(2,6)=29.11, p<0.003] and an interaction of frequency with separation [F(4,12)=16.7, p<0.013]. The latter reflected the strong dependence on spatial separation of thresholds obtained with high-pass filtered markers, and the lack of effect of spatial separation on thresholds obtained with low-pass filtered markers. Note that the wideband data maps well onto the high-pass data, except at the largest separation (90/-90), at which wideband thresholds are similar to the low-pass thresholds. For that separation, the wideband stimulus at the right ear is a loud, wideband noise followed by a noise which is of low-pass character due to the headshadow. For the high-frequency stimulus, there is no low frequency stimulus information so the stimulus at the right ear is a loud, high-frequency noise followed by low-intensity high frequency-components due to the head shadow. The filtering by the head for frequencies above 3 kHz is greater than 15 dB. It is well established that such a level difference will elevate thresholds. Now, at the wide separation, the thresholds with low-pass markers are similar to those obtained with wideband markers, i.e., thresholds for wideband markers approximates that supported by the low-pass noise only. Thus, it appears that listeners can resort to using the low-frequency channels to improve performance when performance in high-frequency channels is marred by the presence of marker level differences.

Left Ear Listening (Trailing marker ear)

The data obtained using left-ear listening (trailing-marker ear) are plotted in Fig. 2.14A. There was a main effect of separation [$F(2,6)=11.18$, $p<0.04$], an interaction between frequency and separation [$F(4,12)=8.94$, $p<0.02$]. Note that the pattern of the functions obtained for the three frequency conditions was almost identical for the 50 and 100ms leading marker durations. This pattern differed markedly from that obtained with the 10ms leading marker. As was discussed previously, the lack of an effect of spatial separation on thresholds with wideband markers most likely reflects a task change to a backward masking task when the markers are located in opposite hemifields. If listeners are performing a different task in the free-field spatial separation experiment with short markers, then one may ask why subjects did not use this strategy with the high-pass markers. Because of the greater head shadow effect for higher frequencies, one would expect the backward masking effect to be even more obvious. Again, introspection with the stimuli revealed that the brief, high frequency leading marker was audible in the standard. This might be accounted for by the greater 'spectral level' of those high frequency components in high-pass noise compared with wideband noise presented at the same sensation level. That is, reducing the bandwidth while maintaining the same absolute level results in a greater intensity of the components in the narrower band.

GENERAL DISCUSSION: DETECTION OF TEMPORAL GAPS BETWEEN SPECTRALLY IDENTICAL MARKERS DIFFERING IN SPATIAL LOCATION

Five main experiments were performed in an attempt to elucidate which of the stimulus features that vary as a function of spatial location underlie the increase in thresholds observed with spatial separation of spectrally identical gap markers first described by Phillips *et al,* (1998), and further investigated by Boehnke & Phillips (1999). Boehnke and Phillips argued that the pattern of gap detection thresholds obtained with spatial separation of gap markers throughout azimuth, reflected that auditory azimuth could be conceived of as being represented through 2 broadly tuned-hemifield channels. Specifically, separation of gap markers anywhere within a hemifield did not elevate thresholds, but separation across the midline did. The experiments completed here attempted to identify which stimulus features that arise from spatially separating the gap markers led to this pattern of results. A summary of the results from the experiments was presented in the introduction to the chapter, and thus will not be reiterated here.

In the introduction to this chapter (see Table 2.1), several predictions were made for the effect of spatial separation of gap markers by Ear, ITD, ILD, and monaural level, over headphones or in the free-field. In general, the predictions were supported by the experimental evidence, although we also determined through the course of these experiments that the leading marker duration and previous practice also considerably complicated the effects. These will each be discussed in turn. The effects of marker spatial separation in opposite hemifields (or sides of the head) on gap detection obtained across experiments are summarized in Table 2.2. The top half of the Table summarizes the results obtained under headphone listening with manipulation of different interaural cues, while the bottom half examines similar conditions probing those cues in the free-field. In both, results are separated in terms of whether the leading marker was short (10ms) or long (>50ms). Results are only presented for spatial separation of the markers across the midline (e.g. –90/90). Summarizing across these seemingly disparate experiments, a distinct pattern of gap threshold dependence on stimulus variables emerges.

1. Listening Condition.

This could be binaural, or monaural using either the ear near the leading marker or the ear near the trailing marker.

2. Lateralization Cue

This was probed in two ways – directly over headphones by manipulation of the ITD or ILD of the markers (Exp. 3), or in the free-field with spatially separated markers of either high or low frequency content (Exp.2). In the free-field case, spatially separated low-pass noise predominantly probes marker differences in ITD (minimal ILD influence), while spatially separated high-pass noise predominantly probes marker

Headphone TASK		**Binaural ILD (+/-12 dB ILD)**	**Monaural (OL)**	**Monaural (LO)**	**Binaural ITD (+/-600µs)**	**Diotic-within**
Short Lead Marker		Elevated (4-8 ms) ↑ Exp 3A&B	Elevated (4-8ms) ↑ Exp 3	Elevated (4-8ms) ↑ Exp 3	Good (1-3ms) − Exp 3B	*Control* *(1-3 ms)*
Long Lead Marker		Elevated (4-8 ms) ↑ Exp 3A&B	Elevated (4-8ms) ↑ Exp 3	Elevated (4-8ms) ↑ Exp 3	Good (1-3ms) − Exp 3B	*Control* *(1-3 ms)*
Free-Field TASK	**90/-90 (BI) Wideband Noise**	**90/-90 (BI) High-pass noise (ILD only)**	**90/-90 (Mono) Lead Ear blocked**	**90/-90 (MONO) Trail Ear blocked**	**90/-90 (BI) Low-pass noise (ITD only)**	**Diotic: Same Loc**
Short Lead Marker	Elevated ↑ (10-12 ms) Exp 1, 4, 5	Elevated ↑ (10-12ms) Exp 2 & Exp 5	Good Wide:4-6 ms − High:9-10 ms ↑ Low: 6-7ms − Exp 1, 4, 5	Elevated Wide:5-7 ms ↑ High:9-10 ms ↑ Low: 6-8 ms −	Elevated ** (8-14 ms) Exp 2 & Exp 5	*Control* Wide:1-3 ms; High:1-2 ms Low: 6-8
Long Lead Marker	Elevated ↑ (7-10ms) Exp 1, 4, 5	Elevated ↑ (7-10 ms) Exp 2 & Exp 5	Elevated Wide:9-10 ms ↑ High:12-14 ms Low: 6-7ms − Exp 1, 4, 5	Elevated Wide: 5-6ms; ↑ High: 12-16ms ↑ Low: 6-8 ms −	Good − (4-5 ms) Exp 2 & Exp 5	*Control* Wide:1-3 ms; High:1-2 ms Low: 4-8 ms

Table 2.2.

Summary of the results from Chapter 2. Illustrates the size of the increase in gap thresholds with spatial separation between the hemifields for each listening conditions and each cue exploited.

differences in ILD and monaural level (minimal ITD influence), and spatially separated wideband noise probes marker differences in both cues.

3. Duration of the leading marker

Recall that this variable had previously been proposed as a 'marker' for stimulus differences which led the markers to activate different perceptual channels (Phillips *et al.*, 1997; Phillips *et al.*, 1998).

4. Magnitude of spatial separation

This was also examined only under free-field listening. In Exp. 1 a full range of spatial separations across azimuth were examined. In Exp. 2, separation across the frontal plane for a fixed, laterally-located leading marker was examined. In Exp. 5, a subset of separations (90/90; 90/0; 90/-90) used in Exp. 2 was examined.

Discussion of these variables in isolation is difficult, as the effect of each one depends on the effects of the others, i.e. the important effects were in the interactions. Thus, the results will be discussed in terms of the main messages.

Spatially separating gap markers by presenting them to different ears (complete peripheral segregation) impairs gap thresholds significantly more than spatial separation by any other cue. Free-field separation in opposite hemifields may effectively approximate this.

In experiment 3B, it was demonstrated that separation of gap markers by ear of entry was more detrimental to gap thresholds than any other marker manipulation. The only other conditions across the five experiments that resulted in double-digit thresholds (>10ms) were those in which spatially separated high-pass or wideband noise were used (Exp. 2 and 5). It might be argued that this stimulus configuration effectively mimics a 'between-ear' case. The ILD for noise components high-pass at 3 kHz is at minimum about 15 dB and up to 40 dB for the highest frequencies. Consider a high-pass leading marker at -90 degrees and a trailing marker at +90 degrees. The leading marker will maximally activate the left ear and any energy reaching the right ear will be very low in intensity; and the trailing marker will maximally activate the right ear and be shadowed at the left. Given that the stimuli in the present studies were usually presented around 35 dB

SL, the energy in some frequency bands could have been below detection thresholds at the shadowed ear. This set of circumstances might have led to an effectively 'between-ear' condition for the high-pass noise. The observation that the effect of spatial separation of wideband noise markers followed that of high-pass markers precisely in pattern and absolute values under binaural listening (see Fig. 2.14), suggests the same may be true for wideband noise. This is also consistent with the fact that the pattern of thresholds for the high-pass and wideband noise differed only for the large spatial separation when listener had only one ear with which to do the task (see Fig. 2.14 A and B). It is as if the listener could access low-frequency information to improve thresholds under monaural listening, but not binaural listening. If the listeners were functionally reverting to a between-ear, relative timing operation under binaural listening with spatially separated high-pass or wideband markers, it is not clear why the absolute values of the thresholds are so low (<10ms). It is possible that these values simply reflect the improvement in between-channel thresholds with practice.

Spatially separating markers using different ITDs (for which there is minimal peripheral differences) does NOT appreciably increase thresholds for any magnitude of spatial separation, under binaural or monaural listening with either ear, under headphone listening or in the free-field

In experiment 3B, gap markers were lateralized to opposite sides of 'intracranial' space with ITDs of opposite signs imposed on wideband dichotic markers. This had the effect that the leading marker was lateralized to the right side and was followed, after the gap, by a trailing marker lateralized to the left. This was a replication of experiment 2 by Oxenham (2000). In both his experiment, and the replication reported here, gap detection thresholds for markers differently lateralized by ITD, did not differ from thresholds obtained when the markers were both presented diotically, i.e. presented from the same intracranial location. Listeners reported that the marker lateralization shifted through the gap stimulus, but that shift was irrelevant to task performance. In the standard interval there was a smooth shift with no discontinuity in the stimulus, while in the 'test' interval, there was a marked discontinuity that could be detected using the same cues as a within channel task. This discontinuity is presumably the brief change in stimulus level that occurs during the gap in the stimulus that is present in the noise representation at either ear.

One argument may be that such differences in ITDs of identical noise markers presented over headphones do not activate grossly different neural channels, and so support only weak location percepts. There are two counterpoints to this. One is the fact that considerable evidence points to the ITD cue available in the low frequency channels as behaviorally the most heavily weighted spatial cue to sound lateralization (Wightman & Kistler, 1992). Furthermore, there is evidence that listeners perceive the spatial quality of stimuli lateralized by an ITD and an ILD as similar and in fact under some conditions the two perceptions of lateralization are tradable (see review by Grantham, 1995). In addition, clicks lateralized by ITD or ILD can be used interchangeably in tasks of perceived motion, such as the saltation effect (Phillips *et al.*, 2002b).

In the case of wideband noise markers differing in ITD, it could be said that the neural bandwidth actually sensitive to the ITDs was small compared with the stimulus bandwidth. Excellent performance was achieved by detecting the gap in the high frequency channels. However, this was addressed in Exp. 2 and 5, where low-frequency (<1000Hz) gap markers were played from speakers in opposite free-field hemifields. With low frequency markers, there is a minimal ILD difference in the stimuli since low-frequency components are not much shadowed by the head and pinna, and the perception of lateralization emerges from the coding of ITD. There was only a minimal effect of spatial separation of the markers, which was dramatically smaller than that observed with wideband or high-pass markers. Furthermore, that small effect of spatial separation was only present for configurations where the leading marker duration was short (10 ms). In Experiment 5, it was demonstrated that there was no effect of spatial separation for low-pass noise markers when a longer leading marker duration was used.

Spatially separating markers using different ILDs increases thresholds relatively equally under binaural and monaural listening

Presented over headphones, the imposition of different interaural level differences on wideband markers elevates thresholds somewhat. In experiment 3A and 3B, this elevation is clearly no greater than that obtained when the task is done monaurally and the markers differ in level. The story was different in the free-field. The effect of spatial separation of high-pass noise markers, for which ILD was the dominant cue to lateralization, the pattern of thresholds obtained differed for each ear, and while neither ear predicted binaural performance particularly well, the leading marker ear (Right) was generally more predictive of thresholds under most conditions. The thresholds for each ear alone can be well explained by spectral and intensive differences between the markers at that ear. Under binaural conditions, it is unclear how to account for the data. What does this suggest for the role of the ILD on its own? Is the elevation in thresholds with markers differentiated by ILD a 'between-channel' effect, or the product of the limitations of the two monaural processors. It is impossible to generate an ILD without generating differences in level at the two ears, so it would appear to be impossible to know. However, Phillips (1999) has argued that the 'between-channel' marker configurations might be differentiated by their sensitivity to leading marker duration. Thus, if gap thresholds for markers that differed in ILD exhibited sensitivity to leading marker duration, but thresholds for monaural marker level differences did not, ILD might be considered to be having an effect beyond the monaural effects. This was tested in Exp. 3A, and no consistent effect of leading marker was observed. Moreover, in experiment 3B, gap thresholds were actually better with a shorter leading marker when the markers were presented to different ears, in contrast to the original report by Phillips et al.(1997). In conclusion, the leading marker duration effect may be a capricious effect, and it turned out to be an ineffective means by which to tease apart the use of monaural level differences from interaural level differences under binaural listening conditions.

There is another way in which the influence of monaural and binaural level differences might be disentangled. An experiment could be conducted in order to determine the 'tuning' of gap detection thresholds across a range of both monaural and binaural level differences. If the gap detection thresholds plateau (reach an asymptote) with increases in ILD, but not with increases in monaural level differences, that might be suggestive of a shift in strategy to a relative timing operation in the ILD case.

Increases in gap thresholds are related to decreases in masking with free-field separation: Related processes or coincidence?

As was argued in the general introduction, the processing mechanisms underlying detection of a signal (i.e. detection of a tone in a noise masker, or in temporal proximity to a masker) may be the same as those that underlie detection of a gap. The similarity in the pattern of gap detection thresholds obtained as a function of spatial separation is strikingly similar to the pattern of detection thresholds for a signal in the presence of simultaneous noise as a function of spatial separation (see Fig. 2.7). This similarity in pattern also holds for low-pass and high-pass noise (see Fig. 2.8) where the effect of spatial separation is attenuated or enhanced respectively. This similarity in pattern suggests that both tasks are affected by the same variable that changes as a function of spatial separation. In the detection task, the signal becomes more detectable as the masker is moved into the other hemifield. This effect can almost entirely be attributed to the greater signal to noise ratio at the ear near the signal that comes about from shadowing of the masking noise by the head. That is, the relative levels of the signal and noise at a single ear. Somewhat surprisingly, research has shown that the binaural interaction cues (ITD and ILD) play only minor roles in the benefit of spatial separation in the free-field. This is unsurprising when the physics of sound filtering are considered. Similarly to Experiment 1 here, this has been determined by comparing the benefit of spatial separation on signal detection under both binaural and monaural conditions. With one ear blocked, there is a nearly an equal improvement in signal detection with spatial separation. There is also an improvement in signal detection with certain spatial separation on the mid-sagittal plane where the binaural cues do not change (Saberi et al, 1991).

A qualitative description of how gap thresholds with dissimilar markers might be predicted via within-channel mechanisms

As outlined in Chapter 1, most effects in classic gap detection experiments in which broadband noise markers presented monaurally can be accounted for by assuming the gap is detected by the decrease in level in a given frequency channel after the stimulus is averaged over time through a sliding temporal integration window. We will elaborate on this further now.

It is known that gap detection thresholds are strongly influenced by stimulus bandwidth. Gap detection is relatively poor when performed within a narrow band of frequencies. Increasing stimulus bandwidth well beyond a critical band improves gap detection thresholds. This suggests that information regarding the presence of a gap is combined across frequency channels to improve detection (Hall *et al.*, 1996). A simple way to conceive of this is to assume the mechanism detecting the gap has access to the output of an array of frequency channels with overlapping filters. Such a channel responds maximally to frequencies near the center of its filter. Its filter has a tuning such that it will be activated by other frequencies, but their input will be progressively more attenuated with distance from the filter's center frequency. Each channel receives information about each marker within the range of frequency to which it is sensitive. Imagine the response of such an array to a 1000 Hz tone. The channel with a center frequency of 1000 Hz will be activated maximally. Channels with center frequencies away from 1000 Hz will also be activated, but to a progressively lesser extent. The presence of the gap is determined independently in each channel, and each channel provides some probability about the presence of the gap to the detector. If very few channels are activated (as occurs in response to a tone), the representation of the gap in their output is uncertain because there are natural fluctuations in the output of a narrowband filter. The fewer the number of channels activated, the larger the gap would need to be in order to generate conclusive evidence that a gap was truly present. However, if the bandwidth of the stimulus was increased so that many channels were activated, a level fluctuation present simultaneously in all channels would statistically provide for greater certainty that the gap's presence was real, and not due to random fluctuations. Thus, a statistical analysis of the output of a multi-channel model might explain the improvement of gap detection with increases in stimulus bandwidth. For visualization purposes, we will examine one implementation of this kind of cochlear filter bank - Seneff Ear Model (Seneff, 1988).

> "This model is based on properties of the human auditory system. A bank of critical-band filters defined the initial spectral analysis. Filter outputs are processed by a model of the nonlinear transduction stages in the cochlea, which accounts for such features as saturation, adaptation and forward masking. The parameters of the model were adjusted to match existing experimental results of the physiology of the auditory periphery." This is a model of the activity across the auditory nerve array which takes into account realistic filtering of sound

> frequency by the cochlea, the half-rectification that occurs when the hair cell response is transduced into output on the auditory nerve, and finally includes an adaptive gain control mechanism which reflects adaptive processes in the auditory nerve." (Slaney, 1998)

The details of the model are less important than the visualization of the stimulus that it provides. The code for the main stages has been implemented by Malcolm Slaney in his Auditory Toolbox for Matlab (Slaney, 1998). Fig. 2.15 shows the output of the Seneff model to a short (10ms) wideband leading marker and a longer wideband trailing marker presented at the same level, or in the loud-quiet or quiet-loud configurations as used in Exp. 3B. Note that the representation of the 3 ms gap becomes clearly evident across all the channels in the 'Equal Level' condition, consistent with a gap detection threshold of 2-3 ms observed psychophysically in wideband noise. When the leading marker is 15dB louder than the trailing marker, adaptation prevents much of a response in any of the channels to the quieter trailing marker until a gap duration of at least 10 ms. In the condition where the trailing marker is louder than the leading marker, there is evidence of relatively small gaps. This suggests that an adaptation model such as the Seneff Model, may not be able to account for the psychophysically determined thresholds for that condition. Persistence models, like the temporal window model, which integrate intensity for times before and after its center, might be more successful in this case. Note that the model used for illustration here only shows the expected output of the auditory nerve, and does not provide us with any further information on transformations occur at higher stations.

Now let us consider the case where the markers are bands of noise with different center frequencies (1 kHz and 4 kHz), each with a bandwidth of 5 equivalent rectangular bandwidths (ERBs). Many channels that will be activated only by energy from one marker, as the other marker has no frequency content that will pass through that channel's filter. Such channels can provide no evidence on the presence of the gap. However, there are subsets of channels, those with center frequencies in between that of the two noise bands, which will be activated equally (although with some overall attenuation) by both markers. These channels will provide accurate estimates of the gap's presence. Furthermore, there will be another subset of channels which have center

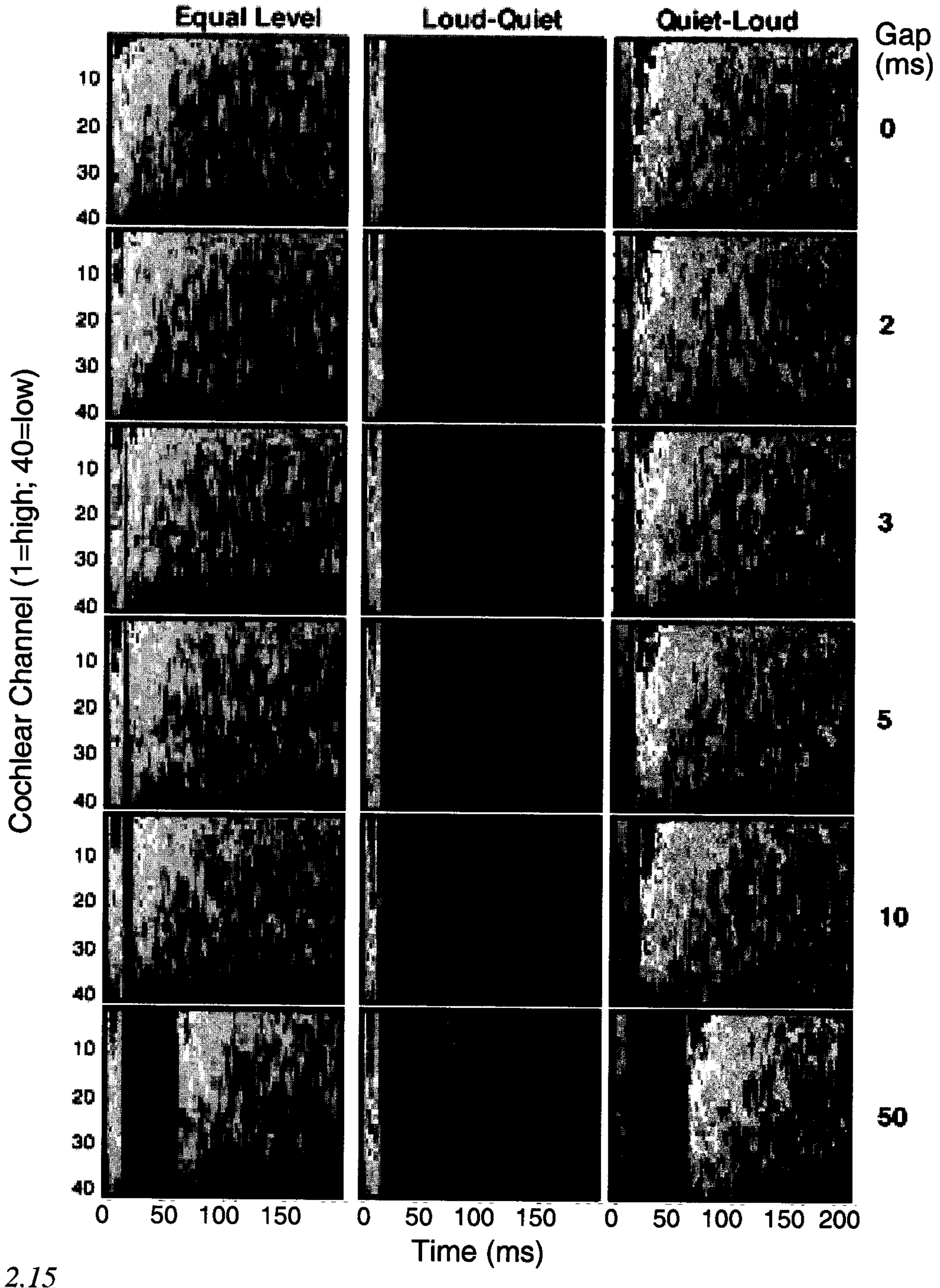


Figure 2.15

The output of the Seneff auditory nerve model for each ear in response to a gap detection stimulus such as used in Exp. 3B. Specifically, wideband noise markers (10 ms leading marker duration) of equal level (1st column), louder leading marker (2nd column), or louder trailing marker (3rd column. Each row shows the response to the same stimuli with increasing gap duration (0-50 ms). The y-axis shows the 40 Seneff channels – increasing channel numbers represent auditory filters with low centre frequenies. The x-axis shows the intensity of the response (colour-coded) of each Seneff channel over time. Darkest blue shows no response; while red is the strongest response.

frequencies closer to that of one noise burst that will respond to one marker more strongly than the other. Gap detection can be accomplished based on the output of these channels (after Heinz *et al.*, 1996), which will report on the presence of the gap within the constraints of the level difference. In such a model, gap detection threshold will rise with spectral separation of noise bands for several reasons. 1) There are fewer channels activated equally by both markers. 2) Other channels that are activated by both, are activated unequally and so their gap representation is marred by level differences/ 3) There are a number of channels not activated by both markers which provide no evidence for the gap. Finally, in a case where each marker activates a completely different set of channels, then the stimuli are completely segregated and the task must be performed by some other means. One possibility is a relative timing operation. This would compare the interval between the offset of activity in one set of channels and the onset of activity in another set.

Now let's consider the case that is relevant to this chapter: wideband noise markers presented from opposite free-field locations. Fig. 2.16 shows the output of the Seneff model in response to a right and left ear canal recording of the gap detection stimulus from Experiment 1, with leading marker at 90 degrees in the left hemifield and the trailing marker at 90 degrees in the right hemifield. The effect of the HRTF imposed on the two markers is immediately apparent. The uppermost plots are the output in response with no gap (0) as was used in the standard interval of the gap detection task. Subsequent plots show the response to larger gap durations. On the y-axis in each plot are the 40 Seneff channels which respond to sound energy within their pass band. Channels higher on the axis respond to higher frequencies. The x-axis represents the stimulus across time. Evidence of the frequency dependent effect of the head shadow is readily apparent in the output. The left-located leading marker activates a broad range of left ear channels with relatively high intensity, but only activates lower frequency channels of the right ear with any significant intensity. The reverse is true for the representation of the trailing marker at the two ears. As can be visualized here, the gap stimulus is a broadband loud sound followed by a low-pass quieter sound at the left ear. Clearly, gap detection thresholds will be affected by the existence of fewer channels reporting

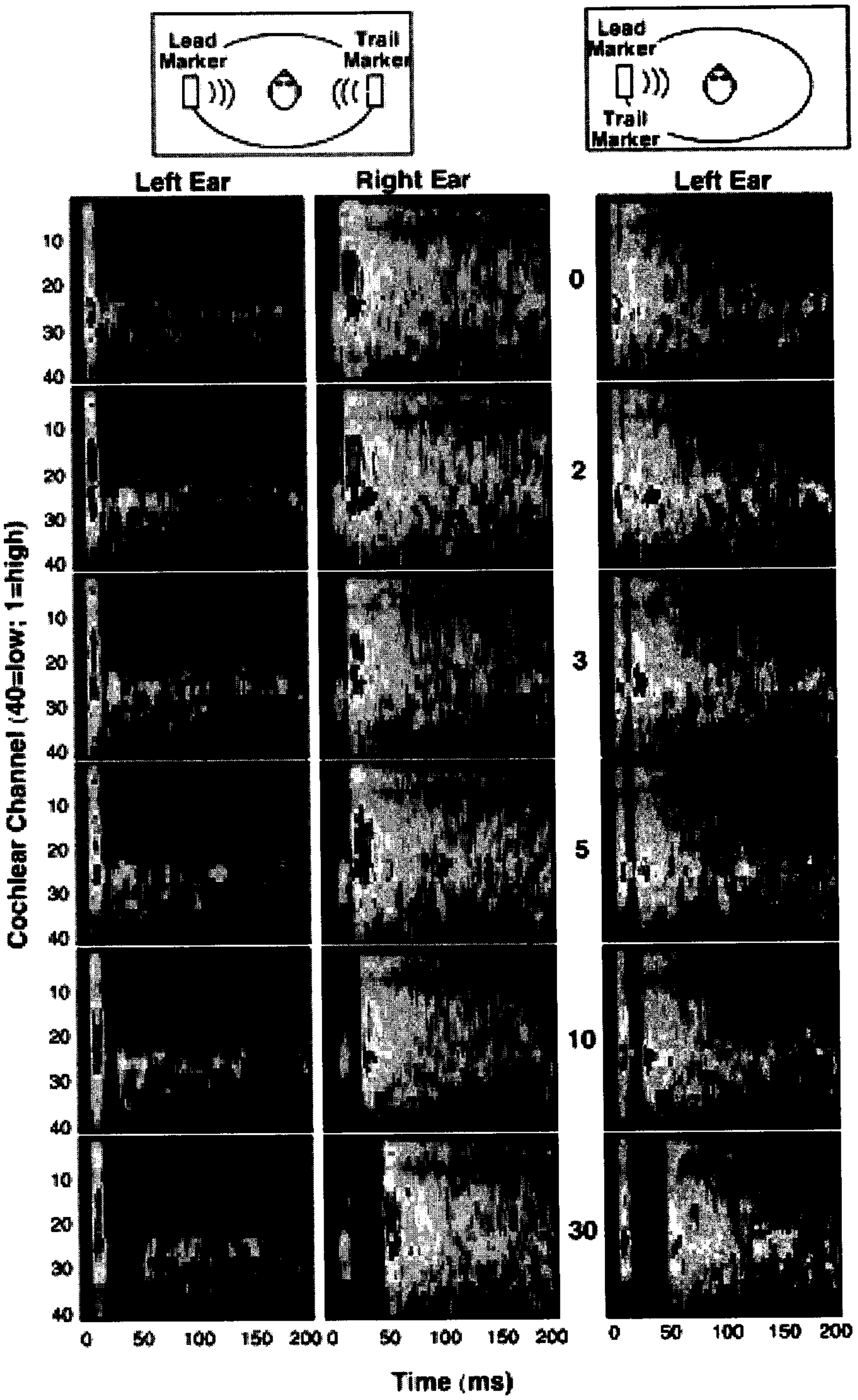


Figure 2.16
The output of the Seneff auditory nerve model for each ear in response to a gap detection stimulus such as used in Exp. 1. Specifically, wideband noise markers (10 ms lead duration) positioned in opposite hemifields (first two columns) or the same hemifield (3rd column). Other details as in Fig. 2.15 caption.

evidence on the presence of a gap. Furthermore, a range of channels receive relatively quieter information from the trailing marker and thus the presence of the gap in their output will be marred by persistence of the leading marker. This is in contrast to a gap stimulus in which both markers come from the left speaker. In this case the representation of the gap emerges quite clearly for gap durations as low as 2-3 ms (akin to Fig. 2.16A).

It could be argued that based on this paucity of information about the presence of the gap at either ear alone, the system might simply revert to another mechanism for gap detection – a relative timing operation between the offset of the leading marker in the left ear and the onset of the trailing marker in the right ear (after Phillips et al., 1997). Such a strategy might have better acuity in binaural listening with widely separated markers than performing the task based on the impoverished information available at a single ear. This paucity might make the gap detection task with spatially separated markers functionally a 'between-ear' task.

Implications for the 2-Channel Model described by Boehnke and Phillips (1999)

The 2-channel model was less of a model, than a *functional description* of the pattern of data obtained for spatial separation of gap markers throughout azimuth. It remains as such. Based on the similarity of the pattern of the data with free-field release from masking of a signal, and the failure to demonstrate a dependence on binaural cues in the experiments reported here, it appears both tasks track the similarity of the HRTFs for various pairs of locations. Spatial separation across the midline releases a signal from a laterally located masker by allowing it to activate channels in the other ear that are less activated by the masker noise. Similarly, spatial separation of the trailing marker across the midline from a laterally located leading marker *reduces* the number of channels commonly activated by the two markers, thereby reducing the information available for detecting the gap. The results here revealed that the sensitivity of gap detection to spatial separation of gap markers was probably not resulting from processes stemming from binaural interaction, but more likely due to monaural spectral and level differences emerging from HRTFs of markers at different locations. This result provides a mechanism for the spatial separation effect on gap detection that can be firmly couched in known within-channel limitations, and known characteristics of spatial filtering of sounds by the head and pinnae. The conclusions from the study of Boehnke and Phillips (1999) – that events within a given hemifield will not be differentially processed for selective attention remains completely intact – sound sources with similar spectral components within a given hemifield will be poorly segregated (Cherry & Bowles, 1960). The correspondence between spatial separation effects on masking and spatial separation effects on gap detection revealed in this chapter is also strongly suggestive of common underlying processes for masking and gap detection. Further study of this correspondence should reveal that a single model might account for simultaneous masking, temporal masking and temporal processing.

The preceding experiments also provide some evidence to suggest that the pattern obtained by Boehnke and Phillips (1999) was not limited to short leading marker durations. In Experiment 2, gap detection thresholds obtained with spatial separation of high-pass gap markers followed roughly the same pattern with spatial separation. In that experiment the leading marker duration was randomized between 100-300ms, thus by no means were the results limited by leading marker duration. The results from Experiments 4 and 5, suggest that spatial separation effects obtained with longer leading markers might be somewhat attenuated, but still existent whether the leading marker is at the midline (Exp. 4) or laterally located (Exp. 2, 4, 5).

CHAPTER 3

THE EFFECT OF SPATIAL CODING ON THE GENERATION OF SEQUENTIAL STREAM SEGREGATION

When a flutist moves through a musical phrase, she will sometimes end it with a trill – a wavering, fluttering note. This is done by rapidly pressing a tiny key, which slightly increases the pitch of the note being 'trilled', by increasing the frequency of the sound produced. Done correctly it sounds like a single warbling note, but never like separate notes. The trill is a single voice, always a singular perceptual event distributed in time. This effect works because the trill never wavers too far from the note being played. If it were to waver too far in pitch, then the trill would no longer be a single voice, but two separate voices. This point marks the 'trill threshold', which was first reported by Miller and Heise (1950), but had long been understood by instrument makers and musicians, and is the most simple example of a major perceptual effect relevant to our understanding of the source segregation problem. Miller and Heise pointed out that beyond the trill threshold "the high tones group to give one melody, and low tones another, and the listener does not notice that in fact successive tones from the 2 melodies alternate [i.e. the listener has no sense of their temporal order]". This conceptualizes one of the main topics of interest in this thesis: is there some measurable perceptual distance between two sounds beyond which the listener has difficulty determining temporal relations between those two sounds. In the trill example, the perceptual dimension is frequency (spectral pitch), while the dimension of interest in the thesis is rather the spatial location of the two sounds. Can we think of a location separation threshold, such as was arguably measured in Chapter 2 (experiment 1) that is analogous to the trill threshold? That is, is there some degree of separation of the locations of two, otherwise similar, sounds beyond which we are impaired at making temporal judgments between them? In a real world case, imagine a conductor who has two violinists playing the identical melody – how can he know which player missed a note? Can he use their relative spatial location in order to segregate the musicians and monitor each of their performances. Relevant to this, it has recently been shown that conductors perform well above average on tests of auditory spatial discrimination, and indeed have greater neural sensitivity to spatial location changes (Munte *et al.*, 2001).

This general line of enquiry has been termed sequential auditory stream segregation ('streaming'). Streaming refers to the change in perceptual organization that occurs when the components of a temporal sequence of sounds become organized to form two or more individual perceptual entities or streams. An auditory stream is "the perceptual unit that represents a single happening" as opposed to a 'sound' which is a "physical happening" (Bregman, 1990, pg.10). The 'stream' is the perceptual representation of an auditory object that can be a combination of related sounds distributed in time, e.g. a series of footsteps. Bregman views the stream as "a computational stage on the way to the full description of an auditory event" which serves the purpose of "clustering related qualities". Bregman differentiates a 'stream' from a perceptual channel, such as proposed by Broadbent (1958) to explain loss of intelligibility when speech is alternated between the ears.

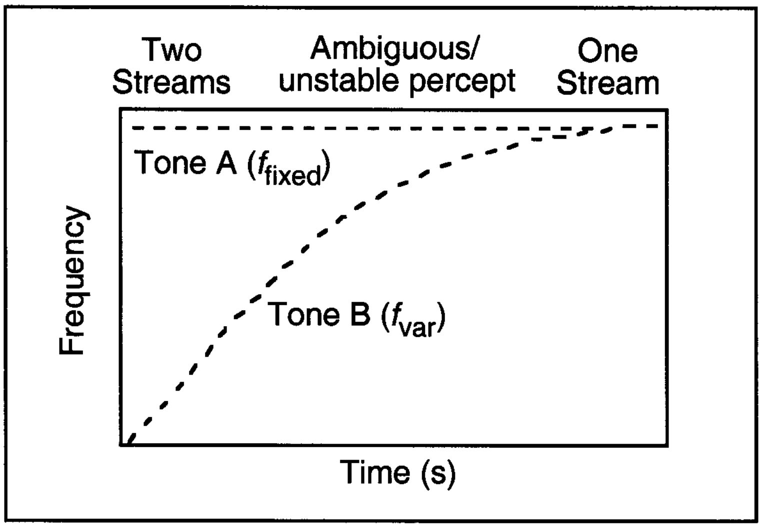
> "A perceptual channel was understood as being like an input channel on a computer. The events that it contained were thought to be linked together by the fact that they had a common input device, like an ear. In contrast, the events in a stream are conceived of as having been linked together by the fact that the perceptual system has concluded that they must have come from a common source. It is reasonable, therefore, to say that the concept of channel is focused on the mechanical structure of the listener while the concept of stream is focused on the listener's apprehension of the structure of the external world. Auditory stream theory would argue that the listener is not built to listen to ears but to events, and it is among these events that attention must be switched." (Bregman, 1990, pg. 80).

One standard stimulus paradigm used to study auditory streaming was developed by Van Noorden (1975). A sequence of A and B tones that differ in frequency are presented in a repeating ABA_ABA_ configuration. The 'A' tone is fixed in frequency and the 'B' tone differs from 'A' in frequency. If the frequency difference is large enough, i.e. beyond the trill threshold, separate A and B streams are perceived. Over repetitions of the sequence, if the 'B' tone if shifted in frequency towards that of the 'A' tone, then the two-stream perception begins to fall apart and 'B' becomes grouped with 'A' by into a single integrated stream (see Fig. 3.1A). This state referred to as 'coherence' or 'integration', and sounds like a galloping rhythm. Increases in frequency separation and/or rate result in the 'segregation' of the A and B tones into separate streams, a state also referred to as 'fission'. Van Noorden (1975) mapped the boundaries within which segregated and integrated perceptions were possible as a function of frequency disparity and repetition time (the period of the alternation of the tones) (see Fig. 3.1B). The integration boundary is fixed. Beyond a range of given frequency separations and rates, the formation of separation streams is obligatory, i.e. it becomes almost impossible to maintain a perception of a single integrated stream. Bregman refers to this as 'primitive' or obligatory stream segregation, in order to specify that segregation will occur even if the listener is attempting to maintain an integrated perception. There is a range of frequency separation and rate conjunctions within which segregation is strongly influenced by intention. In this range, either an integrated or a segregated perceptual organization can be experienced. For these stimuli, auditory stream segregation is not automatic. The system is geared to begin hearing the sequence as a single integrated stream, and segregation may build up over time as if the perceptual system is accruing evidence that two events are present (Bregman, 1978). This leads to a powerful perceptual effect. A sequence may begin as a single coherent stream (galloping rhythm), and then dramatically split into two separate streams.

PERIPHERAL CHANNELING AND STREAM SEGREGATION

Bregman's stream theory focuses on the ecological importance of grouping sound components that are likely to come from the same source in order to form a stable mental representation of the source, and that those components are not easily defined by the physical properties by which the sounds are channeled. It does not ignore century old knowledge that the ear channels information at the periphery by ear and frequency: it simply sees the solution to the problem at a different level of analysis. The reductionist would argue that one must determine how much of the relevant data on segregation can be explained by a peripheral channeling before resorting to a more cognitive explanation involving grouping.

A. Perception of alternating tones (ABABAB) at a fixed rate, as the frequency of 'B' approaches that of 'A'



B. Integration (or temporal coherence) and Segregation (fission) boundaries revealed for various frequency separations across tone repetition rate

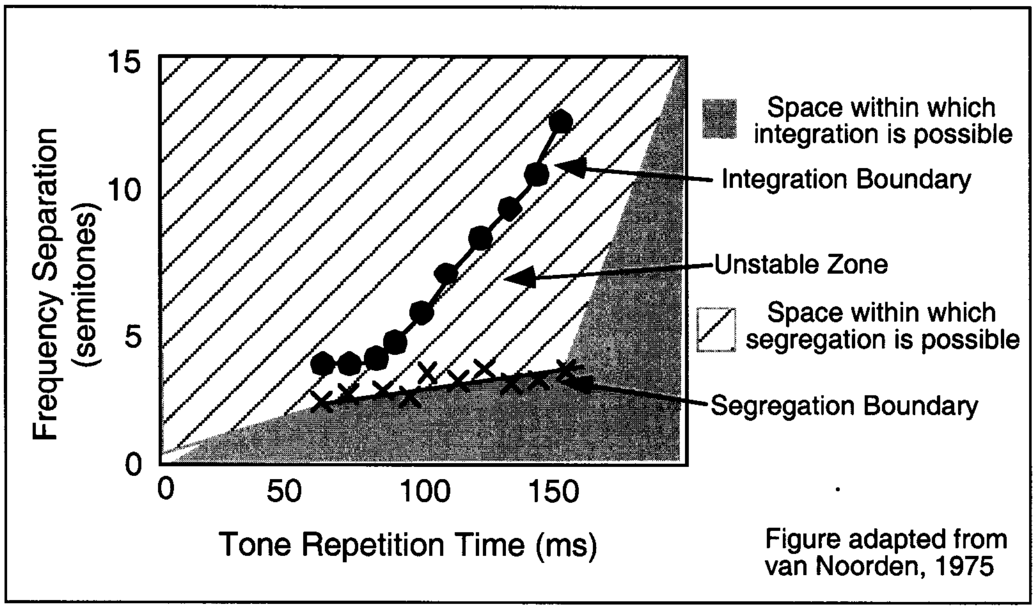


Figure adapted from van Noorden, 1975

Figure 3.1

Integrated (one stream) and segregated (two stream) perception of rapidly alternating tones. A. Tone A is of fixed frequency. Tone 'B' begins at a distant frequency and the perception is of two separate 'A' and 'B' streams. As 'B' approaches the frequency of 'A', the two-stream perception becomes unstable and eventually a single stream is perceieved.

B. The integration and segregation boundaries as defined by van Noorden (1975). If the alternation rate is increased, the sequence segregates at lower frequency separations. The integration boundary is primitive. Segregation is automatic to the left of this point, and the listener can not force their perceptual system to experience the sequence as integrated. The segregation boundary is more flexible, but segregation is impossible to the right of this point.

Frequency separation has a powerful influence on stream segregation. Based on his observations regarding this relationship, Van Noorden (1975) concluded that overlap in activation of A and B at the level of the cochlea was a necessary (although not sufficient) condition in order to maintain an integrated percept. Hartmann and Johnson (1991) performed a critical study which examined the relative importance of peripheral channeling on segregation. They examined various perceptual differences that could lead two overlapping melodies to segregate, and thus be identified. Simple, recognizable melodies were played in an alternating fashion. For example, the notes for 'Yankee Doodle' were played alternately with the notes for "Ode to Joy", both melodies being played using the same range of notes. Played in this fashion, it is almost impossible to identify either of the melodies – they are a jumbled set of seemingly unrelated notes. However, if the notes of one of the melodies were to be shifted upwards an octave, then the two melodies segregate and can easily be identified. They examined how separating the notes comprising the two melodies in different perceptual dimensions (sharpness, loudness, timbre, ear, location, and others) improved this recognition performance. Their results indicated that conditions in which the melodies were changed so that they activated different *peripheral* channels (an octave shift, played to different ears) were most effective in improving recognition of the melodies. Separation by ITD, a location change perceptually similar to that of playing the melodies to different ears, improved recognition, as did separation by pitch, intensity or duration, but the improvement was less dramatic. They concluded that peripheral channeling was "of paramount importance" to segregation.

Such observations led other authors (Beauvois & Meddis, 1996; McCabe & Denham, 1997) to propose models of the stream segregation process based on the concept of peripheral channeling by frequency. The peripheral channels found on the cochlea have a logarithmic frequency spacing - the filters become wider (in Hertz) with increasing frequency. This is usually expressed as the equivalent rectangular bandwidth (ERB) of the auditory filter. The model of Beauvois and Meddis (1996) predicted that tones that were separated in frequency by a constant difference in ERBs (rather than Hertz) would result in the same amount of streaming (i.e., the strength of streaming for an

A of 100Hz and a B of 130Hz would be about the same as for an A of 1000 Hz and a B of 1130 Hz). This relationship was confirmed by Rose and Moore (2000) using repeating ABA sequences. The frequency of the 'B' tone at which the sequence became integrated (the Segregation Boundary) was determined as a function of the frequency of 'A'. When the frequency difference at which this boundary occurred was expressed in terms of ERBs, it was relatively independent of the frequency of 'A'. The boundary occurred for frequency differences that were about half an ERB apart.

CENTRAL INFLUENCES ON STREAM SEGREGATION

While the peripheral channeling hypothesis has received some support, and may be the basis for obligatory segregation, a number of lines of evidence have recently accrued which suggest that the stream segregation is highly influenced by central factors and can occur with complete peripheral overlap in the activation patterns of the cochlea.

In the ambiguous zone, where either integration or segregation of tones of different frequency can occur, the sequence always begins with an integrated perception. Exposure to the sequence over time will lead to an increased tendency to segregate. This "build-up time" to segregation of a sequence of A and B tones differing in frequency can be influenced by several factors. One factor is attention. Carlyon *et al* (2001) presented an ABA sequence to one ear and measured the build-up time to segregation while the subjects attended the sequence. In a second condition, they had the listeners perform a noise discrimination task in one ear while the sequence was played to the other ear, that is, the peripheral auditory system was being activated by the sequence, but the listener was not attending it. After some time the listeners were told to switch tasks and report their perception of the sequence in the other ear. The idea was that if stream segregation resulted from peripheral channeling, than when the listeners switched their attention to the sequence that had been playing, it should have already built-up to a 'segregated' state so they should have a much shortened build-up time to report segregation. Instead, they found that the build-up time after switching tasks was roughly equivalent to the build up time which occurred when there was no preceding task to attend to. Thus, the build-up time to segregation appears to begin when the listener begins attending to the sequence.

Other observations also suggest that the tendency to hear the segregated state is influenced by central factors. Sudden perceptual changes occurring mid-sequence will cause perception of the sequence that has built up to a segregated state to revert back to an integrated state. For example, sudden switching of the sequence to the other ear (Anstis & Saida, 1985) will cause a segregated percept to shift back to an integrated state. This 'resetting' capacity of the perceptual stream segregation system has been examined by asking listeners to rate their perception of a monaurally presented ABA test sequence that followed a long 'induction' sequence. High segregation ratings of the test sequence were most likely, and occurred at smaller frequency separations, if the induction sequence was identical (presented to the right ear). Segregation was far less likely if the induction sequence differed in perceived location (ear, interaural time, interaural intensity or speaker location) or if the induction sequence was quieter than the test sequence, but not if it was louder (Rogers & Bregman, 1993, 1998). The increased tendency toward segregation after exposure to an inducer is also subject to decay over time in the absence of perceptual information. That is, if there is a pause between the end of the inducer and the presentation of the test sequence, the likelihood that that sequence will be perceived as segregated reduces exponentially with temporal distance from the inducing sequence (Beauvois & Meddis, 1997).

The most compelling evidence that the stream segregation is not entirely dependent on peripheral channeling is that sounds that result in the same peripheral activation patterns can segregate based on perceptual differences that are generated centrally. If A is a tone and B is a narrowband noise (less than an ERB wide) centered on the frequency of 'A', the peripheral activation pattern is very much the same. Despite this a tone and noise will segregate, presumably due to timbre differences arising from their envelope differences. Bands of noise that overlap considerably in spectral content, and thus in peripheral activation patterns, will form separate streams depending on the distance between the center frequencies of A and B.

If A and B are both wideband noises (same peripheral activation pattern) that have different amplitude modulation rates (AM-rate) segregation occurs as a function of the disparity in AM-rate (Grimault *et al.*, 2002). Increases in AM-rate generate increases in perceived pitch, and so segregation appeared to occur based on this centrally generated percept of pitch. Segregation can also occur with harmonic tone complexes that have different fundamental frequencies, but are filtered to include the same spectral region (Vliegen & Oxenham, 1999). Such stimuli have similar peripheral activation patterns, but different centrally generated perceptions based on the temporal pattern of their envelope fluctuations.

THE INFLUENCE OF SPATIAL LOCATION (EAR, ILD, ITD) ON STREAM SEGREGATION

It would seem intuitive that separation of the A and B sounds by spatial location would promote streaming. Indeed, Bregman (1990) describes a role for spatial separation in the formation of auditory streams and effectively demonstrates that spatial separation of two interleaved melodies will cause them to perceptually segregate (Track 41 on compact disc associated with Bregman, 1990). In this demonstration, spatial separation was generated by imposing opposite interaural level differences (ILDs) on the two melodies, with the size of the ILDs increasing until each melody was effectively played to a different ear (Bregman, 1990). The following will review evidence suggesting that it may not be the auditory spatial *location* differences per se that lead to segregation, but rather the peripheral representational differences that are the result from sounds at different locations.

A number of studies have demonstrated that segregation by spatial location can occur, and most of these have used ear of entry as the spatial cue. Van Noorden (1975) first described streaming of tones sequences that alternated between the ears. He also anecdotally reported that he had difficulty in perceiving the temporal relations between the tones presented to different ears. However, he only used himself as a subject, and never quantified this degraded perception of the temporal relations between the tones. Segregation of interleaved melodies by presentation to different ears is described by Bregman (1990) and is effectively demonstrated on the CD associated with his book. This demonstration is consistent with the results of Hartmann and Johnson (1991), who found that identification of two interleaved melodies was most improved when they were presented to different ears compared with segregation based on stimulus differences subserving other perceptual dimensions. This is consistent with the peripheral channeling hypothesis (van Noorden, 1975; Hartmann & Johnson, 1991), and the models based on it

(Beauvois & Meddis, 1996), suggesting that segregation based on different ear of entry could be considered the result of peripheral 'channeling', rather than perceptual differences in spatial location *per se*. To refer back to the overview of auditory processing of spatial location, recall that a real-world sound normally activates both ears with relative intensities that vary depending on the sound's frequency components. Activation of a single ear could only under under rare circumstances, e.g. a relatively quiet, lateralized high frequency tone might not activate the shadowed ear with any detectable intensity, nor would a quiet sound presented very close to one ear. The peripheral channeling hypothesis might suggest that segregation would be automatic when sounds are alternated between the ears. If this were the case, there would be no reported 'build-up' to segregation, because the sounds inherently could never be perceived as 'integrated'. The published literature reports no explicit test of this hypothesis.

The only examination of the effect of different values of ILD on stream segregation comes from Bregman's demonstration of segregation of interleaved melodies on the CD associated with his book (Bregman, 1990). In the demonstration, each of the melodies begins with equal intensities at the two ears. They are both located in the middle of the head (diotic) and only a single complex series of notes are perceived. Over time the relative intensity of each melody is at the two ears is changed, such that melody A shifts to the left and melody B shifts to the right. There appears to be some degree of ILD difference required before segregation can occur and allow the listener to recognize that there are two instruments playing different melodies. There is no report on the minimum ILD difference required to produce segregation. Nor does this demonstration inform us of any impairment in temporal judgments between these streams.

In a study of concurrent sound segregation, that has sequential aspects, Gockel and Carlyon (1998) found that the ability to detect a mistuning in a target harmonic complex was impaired if a distracter was presented to the same ear, but not the contralateral ear. This alleviation of the distracters effects on detection of the mistuning when it was presented contralaterally can be attributed to its effective segregation from the target. However, presentation of the the distracter with an ILD of 10 dB, so that it was 'perceived' as contralateral, was far less effective in alleviating the effects of the distracter on detection of the mistuning. Thus, segregation by 'ear' probably does not have its effects because the components are made to differ in 'perceived location'.

Spatial segregation by ITD is an interesting case for the peripheral channeling hypothesis. For A and B noises that differ only in their ITD, the peripheral channels activated are identical, and they are differentiated only through their activation of different populations of central binaural neurons. Now, a pure peripheral channeling hypothesis, which looks only at the pattern of activity across the auditory nerve array, would predict no segregation by ITD. However, if the periphery were to be expanded to include the output of the brainstem, then ITD might contribute to segregation. In fact, in studies of concurrent sound segregation, ITD has been shown to be a weak segregation cue. Culling and Summerfield (1995) conducted a study of vowel identification in which listeners were required to identify a vowel from a variety of formants present. If the relevant vowel formants were given the same ILD, or presented to the same ear, listeners could group together vowel formants in order to accurately report the vowel identity. However, presenting the relevant formants at the same ITD, did not facilitate identification. Thus, ITD appears to be a poor *simultaneous* grouping cue.

In contrast, ITD has been reported to be more useful in studies of sequential grouping. In the identification of interleaved melodies, a task in which the listener is biased to try to perceive segregated streams, segregation of the melodies by ITD provides some benefit, although not as much as separation by ear of entry or frequency (purely peripheral channels: Hartmann & Johnson, 1991). This suggests that while streams might segregate by ITD alone in a repeating ABA sequence, the ITD cue is likely relatively less effective than presentation of A and B to separate ears or different ILDs. What is not known is if the listener can maintain an 'integrated' perception in the face of sounds differentiated by ITD. This is the latter of two basic questions relevant to the debate on the relative importance of peripheral channeling and central influences on stream segregation.

1. To what degree is segregation possible in the absence of peripheral channeling?
2. To what degree is integration possible in the face of large differences in the peripheral channels activated.

To answer the first, a task that biases the subject to form separate streams in order to do the task is required. The only such task developed to date that fits this requirement is probably the melody identification task (Hartmann & Johnson, 1991) or speech sound identification tasks (e.g., Culling & Summerfield, 1995). Evidence from that task suggested that segregation by ITD was possible, but less effective than segregation by ear of entry. To evaluate the second question, a task that biases the subject to perceive a single integrated stream is required. The best task for this exploits the fact that temporal judgments between sounds occupying separate streams are impaired.

TEMPORAL PROCESSES ARE ACUTE WITHIN A STREAM, BUT IMPAIRED BETWEEN STREAMS

Bregman and Campbell (1971) first showed that perception of the temporal order of the A and B tones became difficult as they became segregated into different streams (Bregman & Campbell, 1971; Dannenbring & Bregman, 1976). More recent studies have used detection of a temporal asymmetry between the A and B sounds in a repeating ABA sequence to determine the influence of stream segregation on temporal judgments. Detection of such an asymmetry is more difficult for sequences which form separate streams based on frequency differences, and less so for sequences that differ in centrally generated pitches (Vliegen et al., 1999; Vliegen & Oxenham, 1999). It is an empirical question as to whether this is analogous to the degradation in both gap detection and gap duration discrimination observed with markers of similar frequency disparities. Separation of gap markers by presentation to different ears (Phillips et al., 1997) or ILDs (Oxenham, 2000, Exp.3 in Chapter 2) or different spatial locations (Phillips *et al.*, 1998; Boehnke & Phillips, 1999) also impairs thresholds, while separation by ITD alone does not (Oxenham, 2000, Exp.3 in Chapter 2). In the latter case, the peripheral representations of the two sounds at each ear are identical; the different perceptions arising from a central computation of the interaural time difference. If the impairment in temporal processing is the result of the stream segregation process or if both processes are affected by a common limitation (Neff *et al.*, 1982), then we would expect a similar pattern in the strength of auditory stream formation when the streams to be segregated are based on these spatial cues.

The following experiments examined the relative effectiveness of spatial separation in the promotion of separate A and B streams during a repeating ABA sequence. In Experiment 1, listeners were asked to track their perception of a long sequence of A and B noises which were differentiated by ear of entry or spatial position specified by ILD or ITD, or monaural levels. In Experiment 2, duration thresholds for detection of a temporal asymmetry between the A and B noise bursts were obtained for sequences comprised of the similar stimulus configurations. In Experiment 3, identical noise bursts were played from speakers at different locations in the free field (same speaker, opposite hemifields, two locations in the same hemifield), and stream segregation was measured using the temporal asymmetry task. The noise bands were either wideband, in which all spatial cues contribute to the spatial perception, low-pass (<1000Hz) in which the ITD cue dominates, and high-pass (>3000Hz) in which the interaural differences in spectrum and level dominate according to the duplex theory of sound localization (Stevens & Newman, 1936).

The goal of the set of experiments described here was to evaluate whether the pattern of data obtained in the earlier gap detection experiments corresponded with the degree to which two sounds characterized by different values of a given spatial cue would segregate noise stimuli into two streams. To review, lateralization of the gap markers to opposite intracranial locations degrades gap detection performance most if the markers are presented to different ears, less so if the markers are presented at different ILDs, and not at all if the markers are presented at different ITDs. Note that in the EAR case, the markers are completely segregated at the periphery, in the ILD condition, the markers differ in level at each ear and interaurally, while in the ITD case the markers are spectrally and intensively identical at each ear, but activate different central neurons sensitive to ITD. Thus, the pattern of thresholds reflects the degree of peripheral similarity in the markers.

EXPERIMENT 1:

RATINGS OF SEQUENTIAL STREAM SEGREGATION BY EAR, ITD, ILD, OR LEVEL DIFFERENCES

Bregman (1990) describes a number of methods that can be used to measure the subject's perception of streams. Typically, a rating scale is used in studies of stream segregation in which subjects may rate the ease with which the sequence segregated or remained integrated (depending on task instructions), simply state whether it was integrated or segregated (e.g., Beauvois & Meddis, 1997; Beauvois, 1998; Vliegen & Oxenham, 1999; Bregman *et al.*, 2000; Rose & Moore, 2000; Bregman *et al.*, 2001). Bregman also refers to a method of perceptual tracking of the two states, where listeners report their changes in perceptual state as they occur (by button press for example) (Roberts *et al.*, 2002). In this experiment, the latter method was implemented for two reasons. First, pilot testing indicated that stream segregation by spatial location was often ambiguous and was subject to perceptual state switches. Second, by using a tracking procedure, it would be possible to measure a) the time to build up to segregation and b) the proportion of segregation reported over time.

METHODS

Subjects

Results presented are from 9 listeners (7 female). Three were very experienced in psychoacoustic tasks; the rest were experimentally naive. One subject's data (TAG) had to be discarded because the data set was incomplete.

Design

There were three independent variables by which sequences were characterized: the spatial stimulus attribute leading to perceived lateralization of the noise bursts, the rate of the noise burst alternation, and the noise burst quality.

There were 6 lateralization conditions – DIOTIC (both A & B centered), EAR (A & B presented to different ears), ILD (A and B at ILDs of +/-15dB), LQ (A is 15 dB louder than B), QL (B is 15 dB louder than A), and ITD (A and B presented at ITDs of +/-600μs). These ILD and ITD values were chosen because they tend to result in roughly equal intracranial lateralization. The level difference conditions (LQ,QL) represent the stimulus at each ear present in the ILD condition. These were included in order to dissociate monaural level differences on streaming from the central ILD computation (after Oxenham, 2000).

Each condition was presented at two different rates – an ISI of 30ms or 90ms. For the 'A' and 'B' durations of 90ms, this corresponded to a 'ABA repetition time' of 480 ms and 720 ms respectively. These were chosen because the same values had been used to generate good streaming in a similar task examining stream segregation based on the pitch of tone complexes (Vliegen & Oxenham, 1999).

The noise burst that characterized A and B could be identical ('A=B') or independently generated ('A≠B'). This variable was included for two reasons. One goal was to examine segregation by spatial cues where A and B activation at the periphery was held constant. This was achieved in the "A=B" condition where the same piece of noise characterized 'A' and 'B' throughout the sequence. Independently generated random pieces of noise can have different 'qualities', although still activate the same cochlear filters roughly equally. It was hypothesized that characterizing 'A' and 'B' by independent noise bursts "A≠ B" (yet maintaining the same 'A' and 'B' throughout the sequence) might influence segregation 'additively' in some conditions. A sequence in which 'A' that could be discriminated from a 'B' by such quality differences would not lead to segregation, but an additional spatial difference between independent 'A' and 'B' noises might enhance their segregation relative to condition where 'A' and 'B' were identical ('A=B'). Such additive effects of spatial location differences have been described for rhythmic masking release (Turgeon *et al.*, 2002). In total, this resulted in 24 different sequences (6 x 2 x 2).

In addition to the 6 'spatial' conditions, 2 frequency conditions were completed as controls at each of the two rates. They were inherently 'frozen' because they were tones. In both conditions, 'A' was a tone of 300 Hz. In the 'Near' condition, the 'B' tone was 357 Hz; while in the 'Far' condition, the 'B' tone was 600 Hz. The near frequency region lies in the ambiguous zone while the 'far' condition lies in the segregated zone (see Fig. 3.1B). These were completed in order to have baseline measures to validate the procedure by comparison with other results from stream segregation experiments, which have predominantly used tone bursts as stimuli.

There were three dependent measures. The first was the proportion of time listeners reported hearing two streams, referred to as the 'Proportion Segregation'. The second measure was the time from sequence onset until the listener pressed a key to indicate the sequence had split into two streams, referred to as the "Build-up Time" to stream segregation. The final measure was the number of perceptual switches reported by the listener, referred to as "# Perceptual Switches". It was postulated that this might be considered a measure of the ambiguity of the perceptual states for a given condition.

Stimulus Construction

Stimuli were created using Matlab software (The Mathworks) and presented at a sampling frequency of 44100 Hz at 16-bit amplitude quantization on the audio processor of an Apple PowerMacintosh 8600. Each sequence was in the form A-B-A--- A-B-A--- ...A-B-A (as per Fig. 3.2). A and B were newly generated noises (0-22050 Hz) of 90 ms duration that included 10 ms rise and fall times. In the 'A=B' condition, A and B were identical noises (frozen) for all iterations; while in the 'A≠B' condition, A and B were independent noises, but A and B remained constant across all iterations. For the fast repetition rate, the AB and BA intervals (represented by '-' above) were 30 ms each, while in the slow repetition rate condition these silent intervals were each 90 ms in duration. The AA interval ('---') between triplets was 2 times the AB interval plus the length of the B noise. The A and B noises were concatenated with a segment of zeros the length of the AB, BA, and AA intervals. The A-B-A--- sequence had a total duration of 480ms for the fast rate and 720ms for the slow rate. The A-B-A--- stimulus was replicated 30 times for the fast rate and 20 times for the slow rate, and concatenated in sequence for a total sequence duration of 14.4 seconds for any given sequence.

Procedure

Listeners received instruction on the concept of stream segregation and were presented with several examples of stream segregation involving A and B tones of various frequency separations (as per Fig 3.1B). In some of these examples, segregation was automatic. In others, segregation did not occur or would require some time in order

EXP.1 STREAM SEGREGATION SEQUENCES SCHEMATICS (TO SCALE)

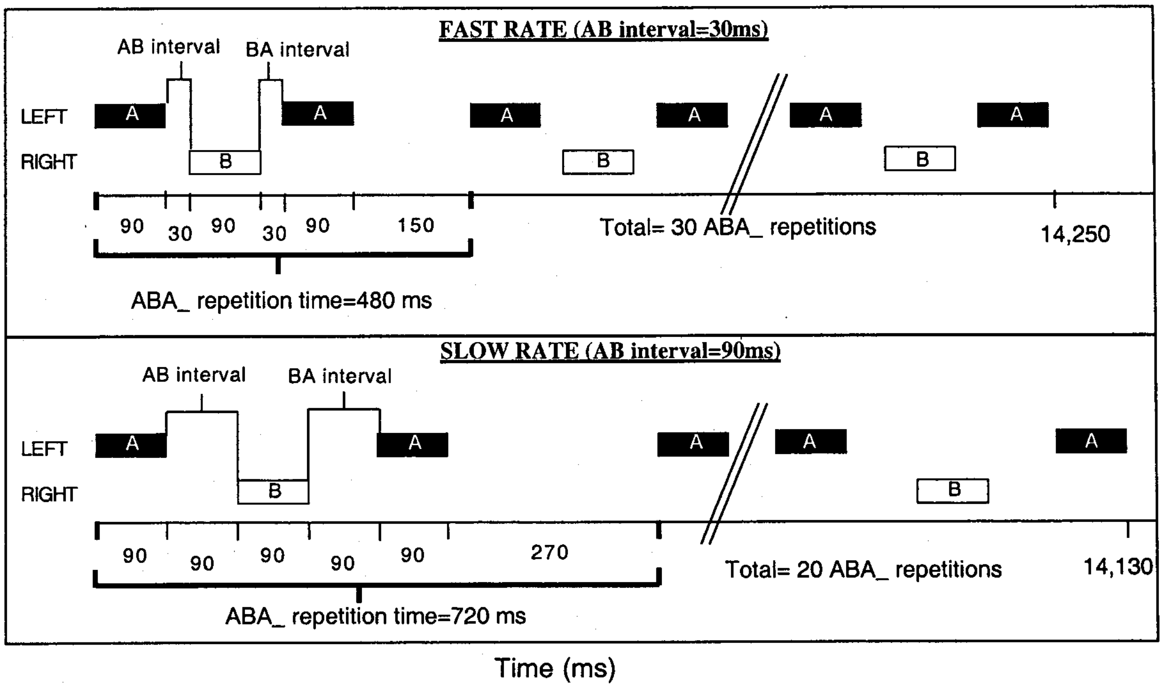


Figure 3.2

This figure shows the temporal configuration of the ABA sequence. All noise bursts were 90 ms in duration. In the fast and slow sequences the AB and BA intervals were short (30ms) and long (90ms) respectively. Total sequence length was the same (14.4 seconds) if the silent period following the last ABA repetition is included. In the lateralization conditions (EAR, ILD, ITD), the A sounds were always presented to the left, while the B sounds were lateralized right. In the FREQ and DIOTIC control conditions, both A and B were presented at equal levels to each ear. In the LQ and QL conditions, both A and B were presented to only the left ear and right ear respectively.

to build up to segregation. After the listeners were able to correctly label these states, they were given two practice sessions of the sequences used in the present experiment. Subjects were told to try to hear out the B tones as a separate stream. That is, they were instructed to attempt to achieve the perceptually segregated state. This is important because the task instructions are known to influence reports of streaming in ambiguous stimulus configurations (van Noorden, 1977). A diagram (see Fig. 3.3A) was presented before each trial that gave a schematic representation of the integrated and segregated states. Listeners always began the trial holding down the '1!' key on the keyboard. This is because there is a bias towards hearing an integrated perception at the beginning of a sequence (van Noorden, 1977), i.e. one can not start hearing a sequence immediately because 'segregated' as segregation is a function of the perceptual system's decision that there are two sources.

The Task

The sequence was presented to subjects at a comfortable listening level (70 dB SPL) over Sennheiser HD25 headphones, while they were seated in a sound-attenuating booth. Listeners were presented in a random order with one of the 28 sequences. The instruction screen (Fig. 3.3A) was presented prior to the onset of each sequence . Listeners held down the '1' key until they experienced the sequence split into two streams. At this point, the listener was instructed to press and hold down the '2' key for the time during which they perceived 2 streams. If the sequence became integrated again, they were to switch keys and hold down the '1' key again. Listeners were instructed to 'track' their perception by pressing one or the other key. In the event that there was ambiguity as to the presence of one or two streams, they were to press the '1' key. The '2' key was reserved for the time in which they were certain that two streams were heard.

After the sequence ended, listeners were given 5-10 seconds during which they were to return to pressing the '1' key and prepare for presentation of the next sequence, which was presented automatically. The last 3 seconds before sequence onset was indicated by an onscreen countdown (3...2...1..). There were 28 sequences presented in a single block, for a total duration of 8-10 minutes. Listeners completed 10 such blocks separated by breaks, usually in a single 2-hour session.

A. ON-SCREEN INSTRUCTIONS PRIOR TO AND DURING SEQUENCE PLAY

GALLOPPING? PRESS 1!

```
A_      _A ___ A_      _A
   -_ _-         -_ _-
     B             B
```

STREAMING? PRESS 2@

```
A_ _ _A_ _ _A_ _ _A
----------------------------
_ B_ _ _ _ _ _ B_ _
```

HOLD THE KEY DOWN FOR THE TIME YOU EXPERIENCE
THE PERCEPTUAL STATE

***START NOW BY HOLDING DOWN THE 1 KEY FOR GALLOPING

Ready? 1.....2......3

B. KEY RESPONSES DURING A DIOTIC, EAR & ITD CONDITION

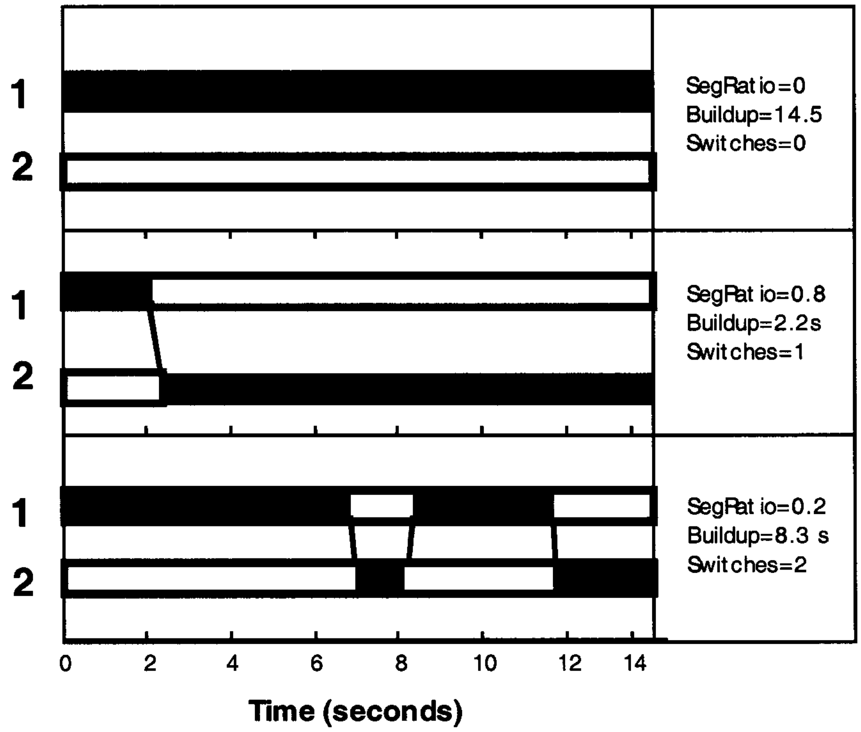


Figure 3.3
Experiment 1: The perceptual rating task instructions and response examples

Data Collection

The stimulus was generated in a MATLAB script. The keypresses were monitored using functions from the Psychophysics toolbox. The start time (t=0) was determined by polling the computer clock (using "GetSecs") immediately after sequence play was initiated. The sequence was played asynchronously (using the SND('play') command). This type of playback released MATLAB during sequence play in order to monitor the keyboard and clock and log the results. In a loop, MATLAB checked the system clock and the status of the keyboard (KbCheck). For efficiency, only information regarding the occurrence and time of *changes* in keyboard state were recorded to a data file. Precision was in the ms range, as MATLAB could check the keyboard at least every 2ms. It is likely that operating system interrupts affected this checking rate occasionally. However, given the coarse nature of this measure (subjective reporting of a perceptual state), this level of precision is more than sufficient for the present purposes.

Data Analysis

There were three measures obtained from the perceptual tracking data:

1. Proportion of time spent in a segregated state
   "Prop.Seg" =(segregation time)/(integration time+segregation time)
2. Time from sequence onset until the first segregation
   "Build-up time" = time until the first 2@ keypress
3. Number of times the subjects reported a change in perceptual state
   "State switches".

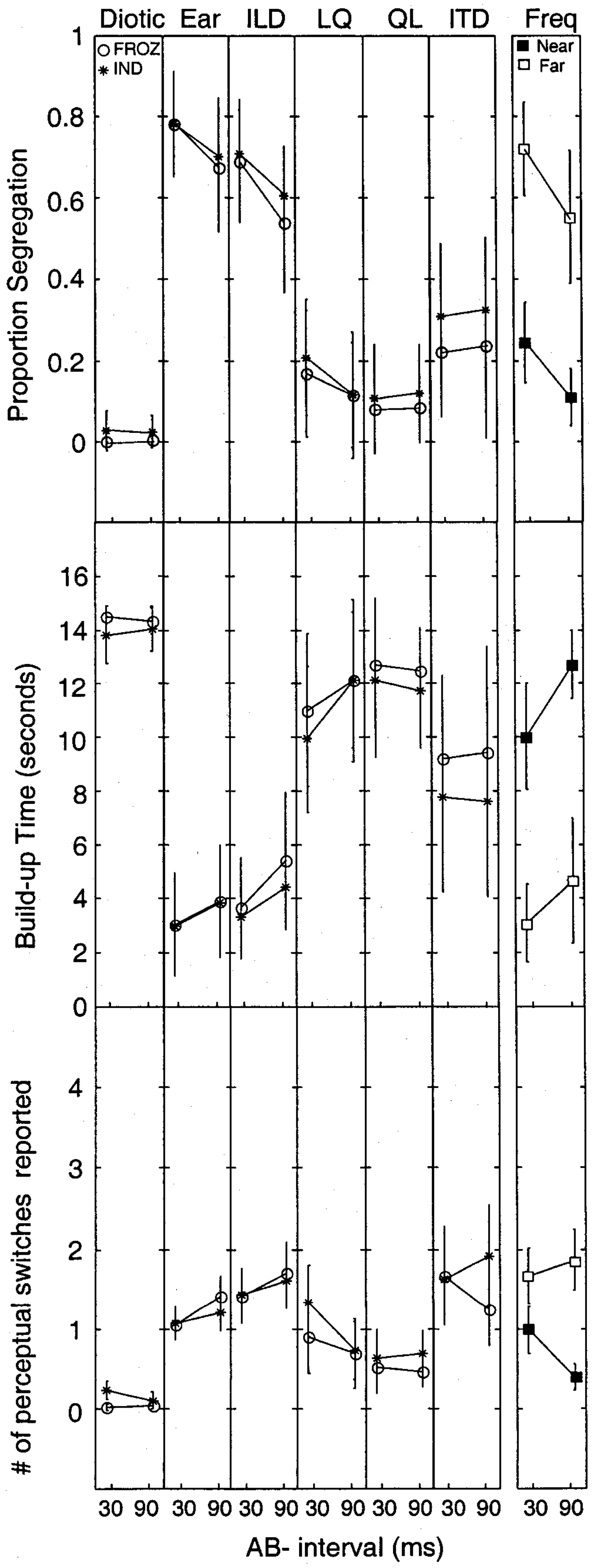
Note that in the measure of state switches there are only two states -'integrated' or 'segregated' – indicated by listeners pressing key 1 or key 2 respectively. During state switching listeners might press both keys, or neither key for a few ms. These key states were logged, but not classified as a state change, nor did transition to the first integration state count as a state switch. If the listener continued pressing the '1@' key from sequence onset to offset, no state switch was logged. If the listener switched from the initial integration to segregation and remained pressing the 2@ key, then one transition and one segregated state was logged, but only one state switch occurred.

A post hoc analysis of the data measures was completed to eliminate error trials. If the listener was not pressing the '1' key within 300 ms of sequence onset, the trial was an error and removed from the analysis (listener did not start in 'integrated' state). If the listener made a transition to the '2' key in less than 700 ms after sequence onset (approximate time to hear one ABA iteration and press the key), the trial was classified as an anticipatory response and removed from the analysis. Such errors were fairly rare or non-existent in most subjects' data. After elimination of these trials, a count of the total trials left was completed. In all listeners, at least 8 error-free trials were available in each condition. To retain a consistent number of trials across the conditions, additional trials beyond 8 were discarded. Thus, the results described below are based on 8 repetitions of the perceptual tracking procedure for each condition.

RESULTS

In Fig. 3.3B, examples of the time course of key presses across the duration of the sequence are shown. In the first panel, a representative time course for the control condition is shown: the listener simply continued to press the '1!' key for the duration of the sequence. In the second panel, a representative example of an 'EAR' condition is shown: the subject rather rapidly logged a perceptual switch to a segregated state. Some listeners made the key press less than 1 second after sequence onset. In the third panel is a representative example of a 'ITD' condition. There is some "buildup time" before the listener indicated segregation had occurred, and then they switched back and forth between integration and segregation one more time. Only the mean results and those of two representative listeners are presented here.

The mean results for each condition are presented in Figure 3.4, and two individual examples are presented in Figure 3.5. Error bars are standard errors of the mean. The three panels of Fig. 3.4 show the mean "Proportion Segregation", "Build-Up Time" to first segregation, and "# of Perceptual Switches". In each panel, the results for each condition are plotted in separate columns as a function of the AB-interval (see Fig. 3.2). This corresponds to the rate of A and B alternation, with 30 ms referring to the rapid rate and 90 ms referring to the slower rate. Results for conditions in which the A and B



<u>Figure 3.4.</u>
Mean results for Exp 1. with standard error bars.
The condition is listed across the top. In columns 1-6, open circles represent "FROZ" condition - where 'A' and 'B' were identical. Asterisks represented the "IND" condition where 'A' and 'B' were independent noises.
In the last column results for the two frequency disparities tested are presented. Closed squares are the near in frequency condition (A=300, B=357); while open squares are the far apart in frequency condition (A=300Hz,B=600Hz)

<u>Top:</u> proportion of time spent in a segregated perceptual state.
<u>Middle:</u> the time until the first key press indicating segregation (the 'build-up' to segregation).
NB: for trials in which the subject never experienced segregation, the build up time was set to the duration of the sequence.
<u>Bottom:</u> The number of perceptual state changes reported.

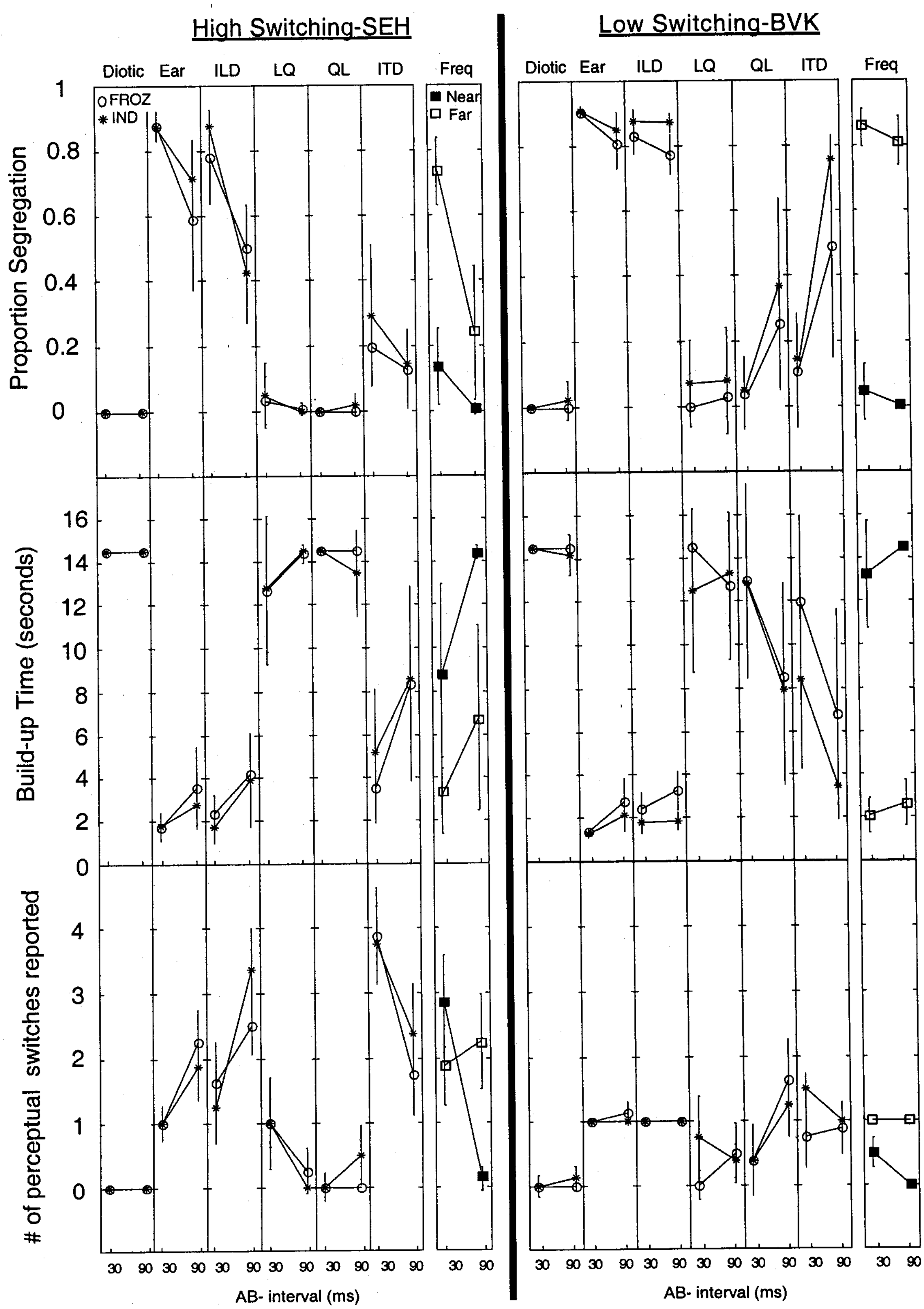


Figure 3.5
Data from 2 individual subjects performing the rating task. The subject on the left showed greater perceptual instability (greater number of perceptual state changes), while the subject exhibited more perceptual stability. See caption of Fig.3.4 for plot details.

noise bursts were identical pieces of noise (“A=B”) are plotted as open circles, while results for conditions in which A and B were independently generated noise bursts (“A≠B”) are plotted as asterisks. The first panel shows the results for conditions in which both A and B were presented diotically (equally to both ears). The second column shows the results for the condition where A and B are presented to different ears (EAR). The third column shows the results from the ILD condition, where A was lateralized to the left with a –15 dB ILD and B was lateralized to the right with a 15 dB ILD. The fourth and fifth columns show the results obtained for the ILD sequence when the listener heard only the left headphone channel (stimulus was ‘Loud-Quiet-Loud’) or the right headphone channel (stimulus was ‘Quiet-Loud-Quiet’) respectively. The sixth column displays the results obtained for a sequence in which A was presented at an ITD of –600 μs and B was presented at an ITD of +600 μs. In the final column (dissociated from the first six columns), the results from the frequency control conditions are plotted. The conditions where ‘B’ (357 Hz) was near in frequency to ‘A’ (300 Hz) are plotted in closed squares, while the conditions where ‘B’ was far in frequency (600Hz) are plotted as open squares.

PROPORTION OF TIME SPENT IN A SEGREGATED STATE (“PROPORTION SEGREGATION”)

The segregation ratio indicates the proportion of time the listener heard the sequence in a segregated state, i.e., the proportion of time they heard two separate A and B streams. Segregation ratios from all conditions using broadband noise (see Fig. 3.6, top panels) were subjected to a three-way ANOVA (6 conditions x 2 rates x 2 noise-types). This analysis revealed significant main effects of condition [F(5,40)=48.73, p<.001] and noise-type [F(1,9)=22.18, p<0.001]. There was no main effect of rate [F(1,9)=2.16, ns], or interaction of condition with rate [F(5,45)=1.7, ns]. There was, however, an interaction of condition with noise-type [F(5,45)=3.03, p<0.05]. Simple effects were used to elucidate this interaction. Significant differences in segregation ratios obtained using different noise types were found when A and B were differentiated by ILD [F(1,45)=7.42, p<0.01] and ITD [F(1,45)=32.75, p<0.001] conditions only. In both conditions, segregation was greater when A and B were independent noise bursts (A≠B).

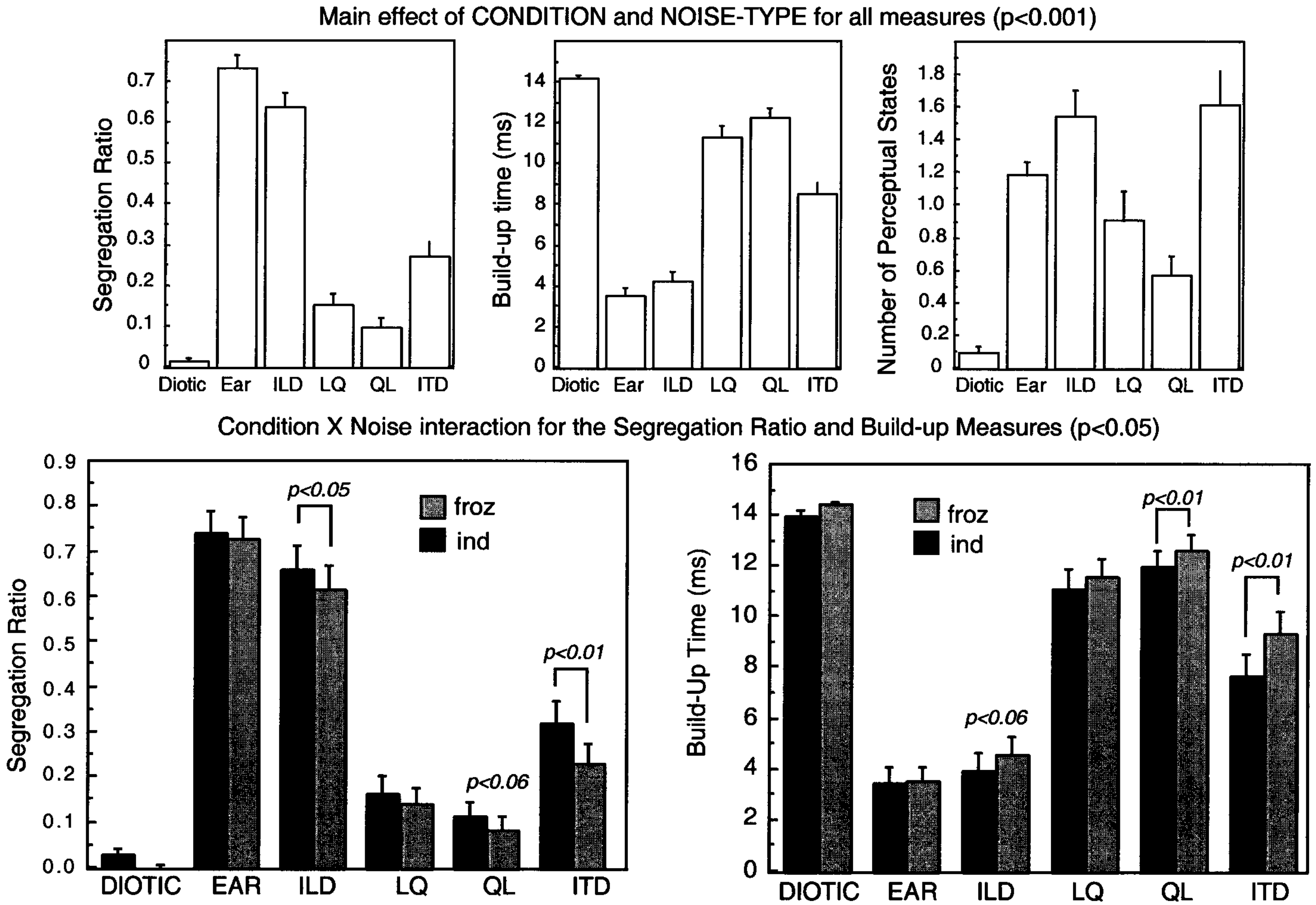


Figure 3.6

Summary of the ANOVA results for Experiment 1. There were main effects of condition and noise type for all three measures. Ear and ILD resulted in high Segregation Ratios and short Build-up times. The level difference conditions generated similarly low segregation values. Segregation was reorted when the markers differed in ITD, but took significant time to build up and was also subject to perceptual ambiguity. There were significant interactions of condition with noise for the Segregation Ratio and the Build-up measures. The interaction suggested that the use of independent noise bursts for A and B, with their attendent quality differences, enhanced segregation when the markers were also characterized by different ITDs, and somewhat when they differed in ILD and level (QL).

Examination of the main effect of condition indicated that separation of A and B by EAR and ILD resulted in similarly high reports of segregation [F(1,45)=2.6, ns]. Imposing opposite ITDs on A and B resulted in segregation that was significantly greater than DIOTIC [F(1,45)=18.4, p<0.01], but less than EAR or ILD [F(1,45)=62.0, p<.001]. Differentiation of A and B by level monaurally (LQ and QL) resulted in equally poor segregation on average, that was only marginally greater than that seen in the DIOTIC condition [F(1,45)=4.5, p=0.06]. However, the effect of a monaural level difference between A and B on reports of segregation was highly idiosyncratic across listeners. Listeners were split into those who reported more segregation with LQ than QL (n=4), those who reported no segregation with either (n=4), and a single listener reported more segregation with QL than LQ.

The two tone conditions (see Fig. 3.4, 7th panel,) were analyzed separately, as they could not be differentiated by noise type. A two-way repeated measures ANOVA on condition (near or far) and rate (fast or slow) revealed main effects of condition [F(1,8)=32.34, p<0.001] and rate [F(1,8)=10.24, p=0.001], but no interaction.

BUILDUP TO SEGREGATION

The time to the first segregation (the 'buildup time') should be inversely related to the segregation ratio data, and in general, this was the case. A three-way ANOVA revealed significant main effects of condition [F(5,40)=37.6, p<.001] and noise-type [F(1,9)=29.07, p<0.001], but no main effect of rate [F(1,9)=2.16, ns], or interaction of condition with rate [F(5,45)=1.7, ns]. Again, there was an interaction of condition with noise-type [F(5,45)=2.68, p<0.05]. Simple effects were used to elucidate this interaction. A greater build-up time in the 'A=B' condition was found when A and B were differentiated by ILD [F(1,45)=4.07, p=0.056] or ITD [F(1,45)=26.75, p<0.001] or level (QL condition) [F(1,45)=4.08, p<0.05]. This is correlated with the smaller 'Proportion Segregation' obtained for these conditions. Thus, there was less segregation because it took longer for the integrated perception to switch to a segregated one, i.e. a longer build-up to segregation.

Examination of the main effect of condition indicated that separation of A and B by EAR and ILD resulted in comparably fast build-up times to segregation [F(1,45)=2.6, ns]. Segregation on the basis of ITD was slower to build up compared with that for segregation based on EAR or ILD [F(1,45)=28.09, p<0.01], but faster than the DIOTIC control condition where segregation did not occur [F(1,45)=31.46, p<0.001]. The comparison with DIOTIC is not strictly justified, however, because DIOTIC conditions rarely resulted in segregated percepts, and thus the buildup time for them is infinite. Note that for trials in which no segregation occurred, the build up time was artificially set to the duration of the sequence (14.5ms). Because of this, and because of the variability among listeners in perceptual segregation of A and B by stimulus level, a the comparison of the buildup times for the LQ and QL conditions is problematic. This consideration aside, the buildup time was equally long for both the LQ and QL conditions, and less than the maximum length of the signal, which characterized the DIOTIC conditions [F(1,45)=7.43, p>0.02].

NUMBER OF STATE SWITCHES

The number of perceptual switches in this paradigm is related to perceptual state instability. It can be considered to be related to switching rate, only because the duration of the sequences was constant. Listeners may hear the sequence segregate into two streams for a few seconds, then revert back to a single stream. This would be an example of 2 switches – the first switch from an integrated percept to a segregated one, the second from segregated back to an integrated percept. A three-way ANOVA revealed significant main effects of condition [F(5,40)= 8.32, 24.88, p<.001] and noise-type [F(1,9)=29.07, p<0.001]. No other effects approached significance. Perception was more unstable when independent noise bursts characterized A and B, compared with identical (frozen) noises. The conditions were ordered as follows from highest to lowest numbers of state switches: ITD, ILD, EAR, LQ, QL, and DIOTIC. An examination of the simple effects revealed that the binaural conditions (EAR, ILD, ITD) were similarly higher than the monaural conditions (DIOTIC, LQ, QL) [F(1,40)=14.46, p<0.01]. That is, switch rates for the EAR, ILD and ITD conditions were all equally elevated, while the DIOTIC, LQ and QL condition were all equally low.

Qualitative analysis of the individual data suggested that they could be split into two groups: five listeners who typically switched states only once (or not at all) from integrated to segregated; and four listeners whose perception was reported to be more unstable. The results from a representative individual from each group (listener SEH, a high switcher; and listener BVK, a low switcher) are presented in Fig. 3.5. Interestingly, these groups had no obvious correspondence to listener musical ability or experience in psychoacoustic tasks. In order to determine the effect of this ‘switching status’ on the three measures, it was treated as a between-subjects variable and tested in a mixed ANOVA with three within-subjects variables and one between-subjects variable. There was a main effect of switching status for the “# of Perceptual Switches” data, confirming the groups did indeed differ quantitatively in the number of reported perceptual switches [$F(1,7)=8.8$, $p=0.02$]. Group status also interacted with condition [$F(5,35)=7.0$, $p<0.01$]. Most notably, the groups differed most in the number of state switches reported in the ITD and ILD conditions. Using the “Proportion Segregation” data, there was no effect of ‘switching status” ($F<1$), and it did not interact with any of the repeated measures variables ($p>.2$ for all interactions). Using the “Build-up” data, no main effect of switching status was found, but again, status interacted with condition [$F(5,35)=3.16$, $p=0.03$]. The unstable group were generally faster to report segregation in the ITD and ILD conditions, but more likely to switch back to an integrated state. The stable group took longer to make a switch to segregation (if they did at all) but rarely switched back. This explains why the two groups did not differ in their segregation ratios.

DISCUSSION

The goal of this experiment was to evaluate whether the pattern of data obtained in the earlier gap detection experiments corresponded with the degree to which two sounds oppositely lateralized by a given spatial cue would segregate into separate perceptual streams. Qualitatively such a pattern was found – separation of the A and B noises by ear of entry (EAR) was greater than, or equal to, that obtained with different ILDs, which in turn, was greater than that obtained by different ITDs. However, there were important differences between the gap detection and stream segregation results.

Characterization of A and B by ear of entry (EAR) resulted in the highest segregation ratio and the fastest buildup time. Almost uniformly across subjects, as soon as the listener had heard one or two ABA repetitions they pressed the ‘2@’ key. The mean build-up time was 3-4 seconds, but was as low at 1-2 seconds in 3 listeners. Given the ABA period (480 or 720 ms), plus some time to make the decision to respond, plus the reaction time to make the response (likely about 300ms), segregation appeared to be nearly automatic. This suggests that stream segregation by ear of entry is likely the obligatory or 'primitive' type of stream segregation described by Bregman (1990). The ILD condition also resulted in similarly high segregation ratios, and rapid build-up times. As such, the lack of a difference in the EAR and ILD condition does not follow the results from gap detection data where marker differentiation by EAR elevated thresholds more than markers differentiated by ILD. It is possible that there is a ceiling effect on the stream segregation data, i.e., the variability in decision and reaction time masks any small difference in the buildup to segregation in the ILD condition.

In gap detection, thresholds obtained using markers of opposite ITDs are as low as those obtained with identical diotic markers. A complete correspondence might suggest that stream segregation would not occur for A and B noises of opposite ITDs. Indeed, peripheral channeling theories of the stream segregation process (Beauvois & Meddis, 1996; McCabe & Denham, 1997) might suggest this to be the case, since 'A' and 'B' activate similar or identical peripheral patterns of activity. However, on average the listeners studied here reported segregation on the bases of ITD alone using spectrally and intensively identical (frozen) A and B noise bursts, albeit such segregation usually had a long buildup time and was less stable. However, not all listeners reported segregation by ITD, particularly when A and B were identical frozen noise bursts (A=B condition). The significant effect of noise-type obtained was more notable for ITD than any other condition.

In the DIOTIC condition, thresholds were slightly, but not significantly greater for A and B streams comprised of different random noise bursts. No listener reported segregation on any trial for identical diotic frozen noise-bursts. However, on the rare diotic trial using independent random noise bursts, five of the listeners reported segregation for some period of time. This result in itself is compelling. Randomly generated random noises, in theory, have identical long-term spectra. However, this similarity reduces with noise duration, and short noise bursts can differ in a quality somewhat like timbre. They can be 'buzzy' or have a 'pitch'-like quality. It appears that some listeners might have exploited such quality differences to briefly segregate such bursts into different streams. This result in itself is not surprising as many experiments have now shown an effect of temporally generated pitch (Vliegen & Oxenham, 1999), amplitude modulation rate (Grimault *et al.*, 2002) using otherwise spectrally similar stimuli. An interesting result from this study is that an additional difference in ITD or ILD on such independent stimuli enhanced segregation in most listeners, quite dramatically in some. This approach of using independent noise bursts is likely not the ideal method by which to probe such additive effects in the stream segregation process. A more controlled method might be to add an ITD or ILD to A and B tones that differ by a few Hz, but are below the segregation boundary. Such an interaction suggests that ITD can be used to enhance segregation if the A and B tones can be differentiated by some quality difference, even if that quality difference on its own is insufficient to generate streaming. The main reason that noise was used in this experiment was in order to provide a comparison to gap detection studies using noise.

The main limitation of this study is its subjective nature. It is not possible to know if listeners were accurately reporting their perception, although if it was not the case one would not expect such a similar pattern in all subjects across conditions. This has always been a criticism of research into stream segregation (Hartmann & Johnson, 1991), and most studies involving subjective rating tasks such as this Experiment are supplemented with a less subjective, more psychophysically rigorous task. One class of such tasks exploits the fact that under streaming conditions (segregation) the acuity of temporal judgments made of an event occurring between sounds occupying separate streams is impaired, while such acuity between sounds occurring within the same stream is not. One such task is to have listeners discriminate the temporal intervals between A and B elements in a sequence. Experiment 2 examines the effect of the spatial differences (Ear, ITD, and ILD) between A and B, on the detection of a temporal asymmetry (AB interval ≠ BA interval) in the A and B sequence.

EXPERIMENT 2:

SEQUENTIAL STREAM SEGREGATION BY EAR, ITD, ILD, OR LEVEL DIFFERENCES – DETECTION OF A TEMPORAL ASYMMETRY BETWEEN STREAMS

In Experiment 1, a perceptual tracking procedure was used in which subjects rated their current perceptual state (integrated or segregated) throughout the presentation of long repeating ABA sequence at two repetition rates (fast or slow). The A and B noises differed in either their ear of presentation (EAR), ILD, the left and right channels of the ILD stimulus presented individually (loud-quiet (LQ) or quiet-loud(QL)), or ITD. As a control, listeners also rated sequences in which A and B differed in frequency (Freq-Near and Freq-Far) or were noises at the same location (DIOTIC). In Experiment 2, we examine these A and B stimulus differences in a non-subjective temporal discrimination task. Listeners were presented with two A and B sequences. In one sequence the AB and BA silent intervals were identical while in the other sequence, the AB interval was longer than the BA interval. Using a 2AFC adaptive tracking task, the threshold asymmetry a listener could detect was determined.

METHODS

Listeners

Seven of the listeners who participated in Experiment 1 of this chapter (5 female) and the author participated in Experiment 2 (n=8). Four were experienced in auditory temporal processing tasks, and the rest had only participated in Experiment 1. All participants completed Experiment 1 prior to participation in the present Experiment 2. The author was excluded from the analysis of Experiment 1 due to the subjective nature of the task (the author was biased). All had at least some musical training, and three (HA, RJS, BKV) played music regularly.

Design

The same 8 stimulus conditions (EAR, ILD, LQ, QL, ITD and two frequency controls) from Experiment 1 were examined for a single AB interval (30 ms) and using the same frozen noise bursts for all A and B repetitions.

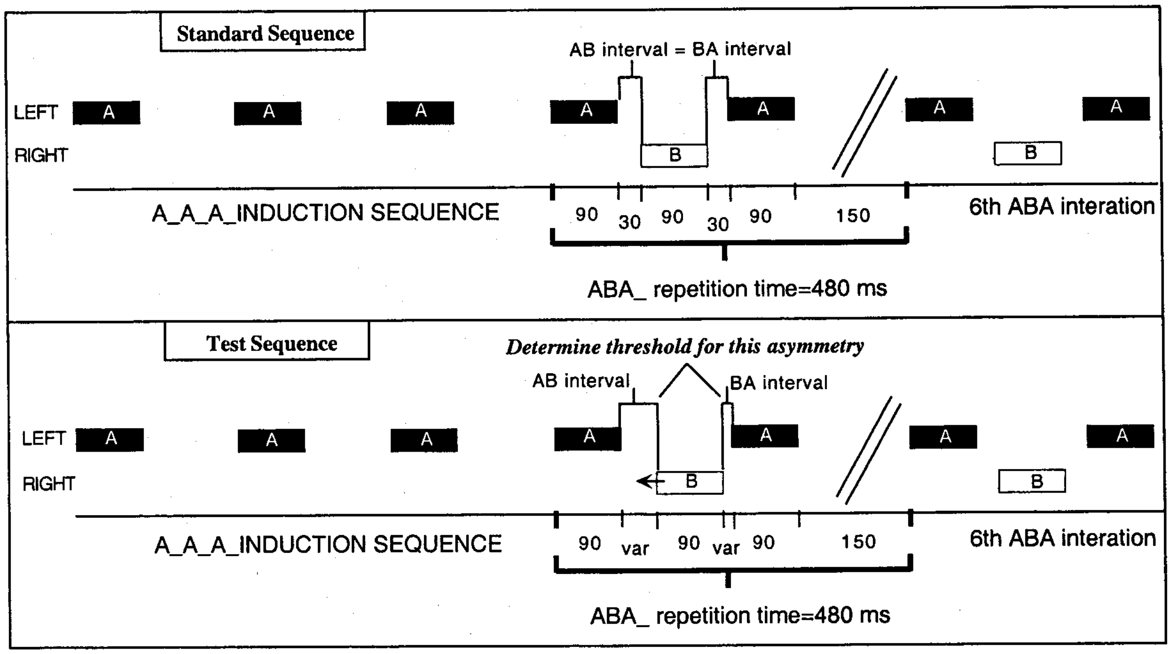
Stimuli

The standard sequence was in the form A---A---A---A-B-A--- A-B-A---...A-B-A (as per Fig. 3.7). The three 'A' iterations prior to the first ABA formed an induction sequence which was included to attempt to bias the perceptual system towards segregation, that is, to push the A's into their own stream before the first 'B' was presented (after Vliegen *et al.*, 1999). The induction sequence was 720 ms in duration. For each sequence of each trial, A and B were comprised of a newly-generated, identical 90 ms noise burst (0-22050 Hz), shaped with 10 ms rise and fall times. In the frequency condition, A was always comprised of a 300Hz tone-burst, and B was either a 357 or 600 Hz tone burst, each 90 ms in duration and shaped with 10 ms rise and fall times. AB and BA intervals (represented by '-') were 30 ms. The AA interval ('---') between triplets was equal to the AB interval + BA interval + B noise duration. The A and B noises were concatenated with a segment of zeros the length of the AB, BA, and AA intervals. Five copies of this A-B-A---stimulus were concatenated for a total sequence duration of about 3 seconds (480*5 + 720). The test stimulus was identical except that the AB interval was greater than the BA interval, while the A-A interval remained constant. As such, the A stream on its own remained temporally regular, as did the B-stream, and the asymmetry existed only in the temporal relation of A and B. That is, the B stream started later in the test stimulus, so that each instance of B was always closer in time to the second A in each ABA sequence.

PROCEDURES

Listeners were seated in a sound attenuating room in front of a Macintosh 8600 running MATLAB (The Mathworks). All stimulus generation, presentation, and collection of responses were co-ordinated through a script written in MATLAB. The sequences were presented to subjects at a comfortable listening level (70 dB SPL) over

EXP.2 TEMPORAL ASYMMETRY TASK SCHEMATIC (TO SCALE)



<u>Figure 3.7</u>
This figure shows the configuration of the ABA sequences using in the temporal asymmetry detection task of Exp. 2. Listeners were presented with two sequences in a random order, one that was perfectly rhythmic (the standard) and one in which there was a temporal asymmetry in the spacing between the A and B tones. The task was to report the interval with the asymmetry, and a threshold was determined using an adaptive tracking procedure. Each sequence began with an inducing sequence of A-sounds in order to bias the listener to hear segregation when the first 'B' sound was introduced. All noise bursts were 90 ms in duration. The AB and BA-intervals were always short (30ms), and the ABA repetition rate was 480ms. In the lateralization conditions (EAR, ILD, ITD), the A sounds were always presented lateralized to the left, while the B sounds were lateralized to the right. In the FREQ and DIOTIC control conditions, both A and B were presented equally to both ears. In the LQ and QL condition, both A and B were played only to the left ear and only to the right ear respectively.

Sennheiser HD25 headphones. Thresholds for detecting the temporal asymmetry in the repeating A_B_A stream were obtained using a two-interval, two-alternative forced-choice task and a modified two-down, one-up adaptive method (Levitt, 1971), as described in the General Methods section. The adaptive variable was the difference in the AB and BA intervals. On each trial, listeners were presented, in random order, with a standard stimulus sequence (A---A---A-B-A---A-B-A...), and a test stimulus in which the AB interval was larger than the BA interval, i.e. the sequence had a temporal asymmetry (A---A---A---A--BA---A--BA---A--BA...). The inter-sequence interval was set at 500 ms. The task was to indicate whether the test stimulus was in the first or second position. Listener response was not timed, and so the inter-trial interval varied. The initial asymmetry duration was above detection threshold and depended on the stimulus conditions (20 ms to a maximum of 30 ms). The maximum asymmetry size was 30 ms. If the listener did not make a correct response at this asymmetry, the sequence was repeated. Each adaptive staircase continued for only eight reversals and the mean temporal asymmetry size for the last six reversals was defined as the detection threshold. The reduced number of reversals was used in order to make a threshold determination more manageable. Due to the length of each sequence, determination of a single threshold took about 6 minutes.

RESULTS

The mean thresholds for each listener are plotted separately in the top eight panels of Figure 3.8 with standard error bars from the 3 thresholds estimates used to generate the mean. The grand mean thresholds are plotted in the bottom panel with standard error bars. In each plot, data for the six conditions in which A and B are comprised of wideband noise bursts are plotted on the left. At the far right, the means for the two frequency separation conditions are plotted. Temporal asymmetry threshold (ms) is on the y-axis.

Baseline performance in the DIOTIC condition, for which only a single stream could be perceived was 9.9 ms. The ITD condition was not different from baseline (11.03 ms). Thresholds were highest (23.25 ms) when A and B were presented to different ears

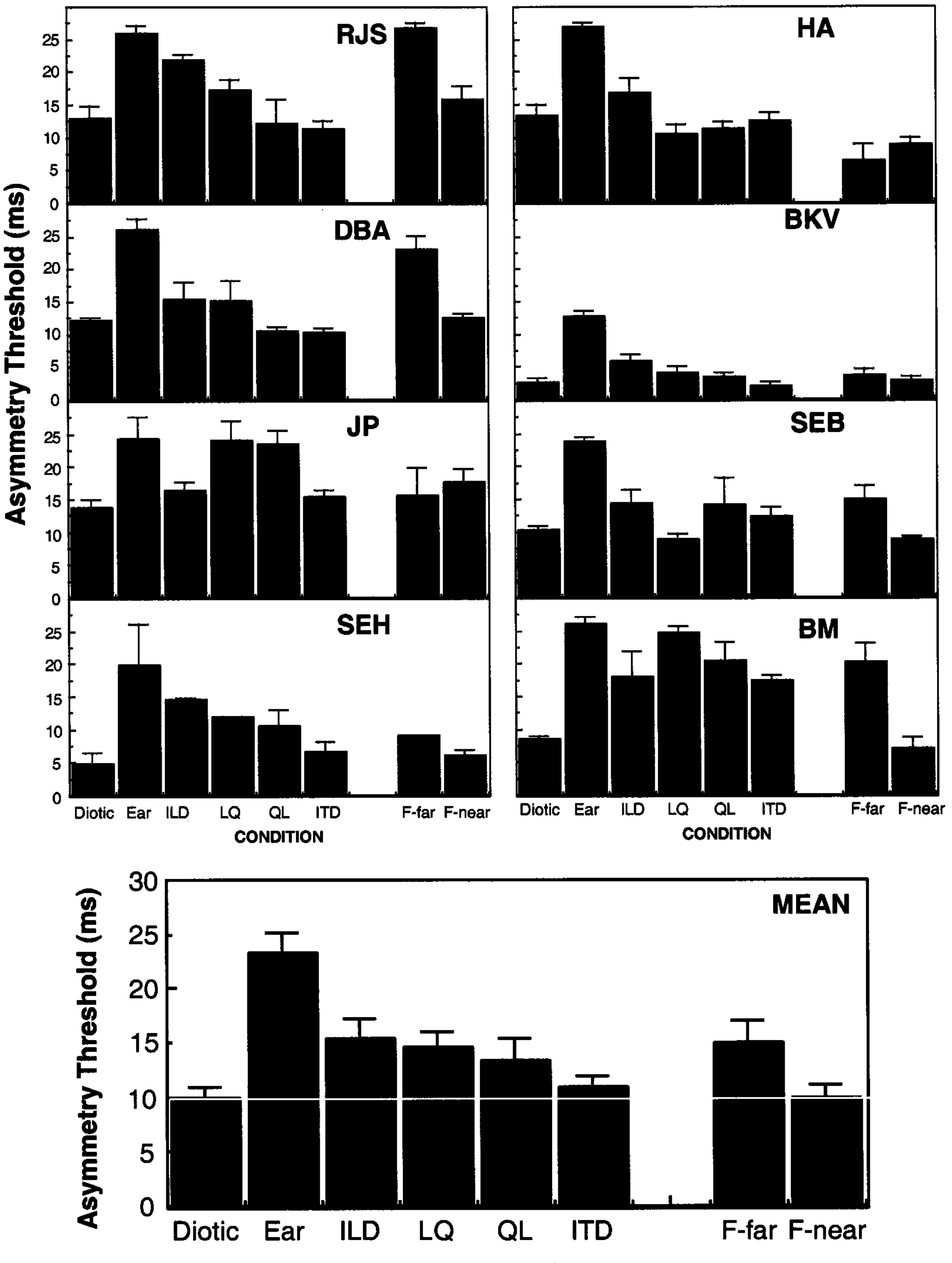


Figure 3.8.

Individual and mean temporal asymmetry detection thresholds from Experiment 2. Error bars on the plots for individuals are the standard error from the 3 threshold determinations which comprised that mean. The error bars on the mean data (bottom) are the mean of each listener's standard error. This estimate of error was used because there were large individual differences in absolute threshold which are not relevant to the pattern of data obtained across conditions.

(EAR). The mean asymmetry threshold for the 'EAR' condition was 23.2 ms. Since the maximum possible asymmetry is 30 ms, listeners obtaining thresholds above 25 ms (RJS, HA, DBA, BM) were not achieving any level of performance on the task, as the first adaptive step is from 30 is 25 ms. Visual inspection of the staircases indicated that subjects could often do the task early in the staircase, but lost whatever cue was permitting performance, and the stimulus then returned towards the maximum asymmetry at later reversals.

The data were subjected to a 1-way repeated measures ANOVA, which revealed a significant effect of condition [F(7,49)=11.87, p<0.001]. Specific hypotheses were tested using simple effects. Experiment 1 suggested that stream segregation when the A and B noises were differentiated by EAR or different ILDs was equivalent. A contrast of EAR and ILD, however revealed temporal asymmetry thresholds to be significantly higher in the EAR condition [F(1,49)=19.28, p<0.001]. In the previous gap detection experiments, thresholds obtained with markers differing in ILD were the same as that obtained for monaurally presented markers that differed in level (see Chapter 2, Exp. 3.). A contrast of the thresholds obtained for the ILD with those for the two level conditions revealed that there was no difference in temporal asymmetry thresholds (F<1). However, mean thresholds for those conditions were significantly elevated relative to those for the DIOTIC and ITD conditions [F(1,49)=9.96, p<0.01]. In the gap detection experiments, thresholds were unaffected by imposing different ITDs on the gap markers, and consistent with this, the present experiment revealed no effect of ITD on temporal asymmetry thresholds relative to the DIOTIC condition (F<1). Thresholds for the A and B tone conditions, where B was disparate or near to A in frequency, generally followed the pattern that was expected. Thresholds were significantly higher with the larger frequency separation [F(1,49)=8.00, p=0.01]. The thresholds obtained, however, were lower than those obtained with segregation by EAR [F(1,49)=21.5, p<0.001].

DISCUSSION

The temporal asymmetry detection task used in this experiment inherently biases the listener to attempt to perceive the sequence as integrated. That is, in order to achieve good performance, it was in the listener's best interest to try to force the A and B noises into the same perceptual stream. Thus, the task is measuring the segregation boundary – the point beyond which segregation is obligatory and integration fails. The main results were as follows. Thresholds to detect a temporal asymmetry were significantly elevated relative to DIOTIC when the A and B markers differed by EAR of presentation, ILD, level or the large frequency disparity. Thresholds were unaffected by imposing different ITDs or a small frequency difference on the A and B sounds. Separation of A and B by ear was clearly the most detrimental to performance, consistent with the obligatory stream segregation observed in listener's ratings of stream segregation in Experiment 1. Interestingly, while separation of A and B by ILD resulted in perceived stream segregation *ratings* similar to EAR, in the temporal task where the listener was biased to perceive an 'integrated' sequence, listeners appeared to be able to maintain integration in the ILD condition in order to better perform the task.

While imposing different sound levels on the A and B noises did not result in high ratings of stream segregation (although they were significantly greater than the DIOTIC condition), it did affect performance on the temporal asymmetry task. This is in contrast to imposing different ITDs on the A and B noises, which caused significant reports of stream segregation, but did not elevate temporal asymmetry thresholds. The temporal asymmetry thresholds for the two monaural level conditions were not different from those obtained when A and B were characterized by different ILDs. This pattern is precisely that observed in gap detection studies using markers that differ in level or ILD (Experiment 3 in Chapter 2, Oxenham, 2000), but differs from that observed in ratings of stream segregation, for which ILD was much more effective. This suggests that the rating task and the temporal task might be measuring different aspects of streaming. The rating task might have encouraged listeners to try to 'hear out' segregated streams, while performance on the temporal asymmetry task measured the ability of the listeners to maintain an *integrated* perception.

Consistent with this, all listeners reported the temporal task to be cognitively and attentively demanding. The cue typically used to perform the task was a rhythmic difference between the two sequences. The standard sequence was temporally rigid, like the snare drum in a military march. The asymmetric sequence had a 'lazy lilt' in its rhythm, akin to some jazz rhythms. This distinction was facilitated if an integrated perception could be maintained. Another strategy, reported by the best listener (BKV) involved attempting to force the first ABA iteration after the induction sequence to remain integrated. The task then became a gap discrimination task in which she determined in which sequence the first AB interval was longer. In a sense, the task was still measuring some aspect of the integration boundary, which clearly was difficult to overcome in the 'EAR' condition, for which even the thresholds of BVK were elevated. This strategy was also used in the 'F-Far' condition where the A and B tones were separated by 300 Hz. In three listeners, thresholds were nearly as elevated as those obtained in the EAR condition. In three other listeners, however, thresholds were as low as control conditions. There is reason to believe that this difference might not necessarily be related to differences in streaming strength in those conditions. The unfortunate fact that the choice of tone frequencies (300 and 600 Hz) were harmonically related may have facilitated this to some degree through some within channel mechanism.

The temporal asymmetry task was modelled after that used by Vliegen *et al. (1999)*, who used pure tones differing in frequency, similar tone complexes which differed in fundamental frequency (VarFo), or tone complexes with the same fundamental, but differing in the frequency range in which they were filtered (VarSpec). In general, the absolute values of the thresholds obtained in the present study are consistent with those of Vliegen *et al. (1999)*. Similar to the present results (see data for listeners BKV and HA), they found that some listeners, particularly those with musical training, achieved asymmetry thresholds of 5 ms or lower for the conditions with no frequency difference. Other listeners had baseline thresholds closer to 10 ms. Their maximal thresholds approached 25 ms for the largest frequency separation, similar to that obtained in the present study in the EAR condition. Thus, the results for the two studies are generally consistent. Furthermore, the greatest elevation in temporal asymmetry thresholds was for pure tones of disparate frequency or tone complexes filtered at different spectral regions – both cases where the stimuli were differently channeled at the periphery. Thresholds were less affected by differences in the fundamental frequency (a centrally generated difference in pitch) of the tone complexes which activated similar spectral regions (similar peripheral channeling).

Importantly, the pattern of results of the present Experiment 2 mapped on precisely to the pattern of results obtained in the gap detection study (Chapter 2, Exp. 3B) using markers that differed in the same spatial cues (EAR, ILD, monaural level, ITD). The gap detection task is somewhat akin to the first A-B transition in the asymmetry detection task. Oxenham (2000) also observed that the pattern of gap detection thresholds obtained using stimuli similar to those used by Vliegen *et al. (1999),* mapped onto the pattern of temporal asymmetry performance. That is, gap detection thresholds were higher for tone complex markers of similar Fo but filtered at different spectral regions (VarSpec) than for tone complexes of similar spectrum that differed in Fo. The similarity in the pattern of thresholds for gap detection and temporal asymmetry thresholds observed for both spatial stimulus attributes (reported here) and for pitch attributes (Vliegen *et al.*, 1999; Oxenham, 2000), is consistent with a dominant effect of peripheral similarity of the gap markers (or A and B sounds) on the acuity of the temporal judgement. Veridical judgement of the gap duration in both tasks might depend on the number of frequency channels commonly activated by both sounds. The overlap in peripheral activation in the gap detection stimulus, and in the first components of the asymmetry detection stimulus, might dictate performance. In terms of streaming, the degree of common activation of channels by both sounds might determine the immediacy with which the sounds will be segregated. In the ITD condition, both sounds equally activate all channels, and so integration is maintained early in the sequence. It may be that ITD differences between sounds can not be used for segregation until the system has accrued enough evidence that there is a consistent location difference between the two identical noises. In essence, grouping by frequency similarity is being pitted against segregation by location. This is not entirely surprising as it is known that grouping by frequency similarity, at least under some conditions, can even overide segregation by ear of entry (Deutsch & Roll, 1976). In this case, however, in the EAR condition segregation appeared to be complete from the outset of the sequence, and was, in effect, obligatory and lasting. In the ILD condition, all channels were activated by both noises but at different levels. Thus, early in the sequence the stimulus may have been integrated. With repetition, however, the common activation of the channels by both noises might be reduced, perhaps through adaptation. The response to the louder sound might gradually come to dominate the response to the quieter sound, so that the ILD condition builds towards being effectively similar to the EAR condition. This is of course speculative, but could be tested physiologically. Candidate neural correlates of the stream segregation process will be addressed in the General Dscussion to this Chapter.

EXPERIMENT 3:

SEQUENTIAL STREAM SEGREGATION BY FREE-FIELD LOCATION – DETECTION OF A TEMPORAL ASYMMETRY BETWEEN STREAMS

Experiments 1 and 2 demonstrated that segregation by Ear of entry was obligatory from the onset of the stimulus, segregation by ITD was not immediate – listeners could maintain integration, at least early in the sequence, and segregation by ILD was effective, though perhaps only after the some iterations, since listeners showed better performance on the asymmetry task compared with the EAR condition. These results were consistent with the pattern of results obtained with gap detection between markers that differed by ear of entry, ITD or ILD (Chapter 2, Exp.3B). In Chapter 2, the effect of free-field separation also affected gap detection, but primarily under conditions when the markers differed in level at either ear. If temporal asymmetry detection is subject to the same limitations as gap detection, then it should show a similar dependence on stimulus parameters in the free-field. Recall in Chapter 2 (Exp. 5) that spatial separation of gap markers did not affect thresholds when the markers were low-pass noises which were not subject to the shadowing by the head (which introduces level disparities between the markers). Spatial separation effects were strongest when the markers were high-pass noises which thus *were* subject to such shadowing. In Experiment 3, the temporal asymmetry task was taken into the free-field in order to extend the correspondence. A and B were both low-pass, high-pass or wideband in spectral content, and were systematically separated in spatial location. It was expected that thresholds would follow the same pattern as observed in the gap detection task. Specifically, temporal asymmetry thresholds would not be affected by spatial separation when the markers were low-pass noise, and would be highly affected when the markers were high-pass noise. It was expected that noise bursts at different locations would segregate into two streams more easily with high-pass noises, consistent with the difference in segregation observed with ILD and ITD in Exp. 1.

METHODS

Subjects

Five of the listeners from Experiment 2, including the author, participated (SEH, BKV, SEB, BM, HA). All were highly practiced in auditory temporal processing tasks. Three of these had participated in gap detection task in the free-field described in Chapter 2.

Design

In a completely repeated measures design, two variables were manipulated. These are outlined in Figure 3.9. The first was the spectral content of the noise bands. The noises were either low-pass at 1000 Hz, high-pass at 3000Hz or wideband. These values follow those used in Experiment 5 described in Chapter 2. The second variable was the spatial location of the 'B' noise burst. While 'A' was always located at –90 (deep in the left hemifield), 'B' could be located at either at –90, -45, 0, 45, or 90 degrees azimuth (see Fig. 3.9). In addition, there were 4 tone conditions. A and B could be either 500 Hz or 4000 Hz tone bursts co-located at -90 degrees azimuth, or located in opposite hemifields (-90/90).

Stimuli

Sequence construction was identical to that described for Experiment 2, with the exception that the noises were processed in the frequency domain in order to generate the different spectral conditions. This process was described in Experiments 2 and 5 of Chapter 2.

Procedures

All stimulus generation, presentation, and collection of responses were co-ordinated through a script written in MATLAB, which called upon the MATSS stimulus presentation system, described in the General Methods. Initially, it was intended that the stimulus level be set at 40 dB above sensation level for the different noise types (low-pass, high-pass, wideband) following the protocol of experiment 5 in Chapter 2.

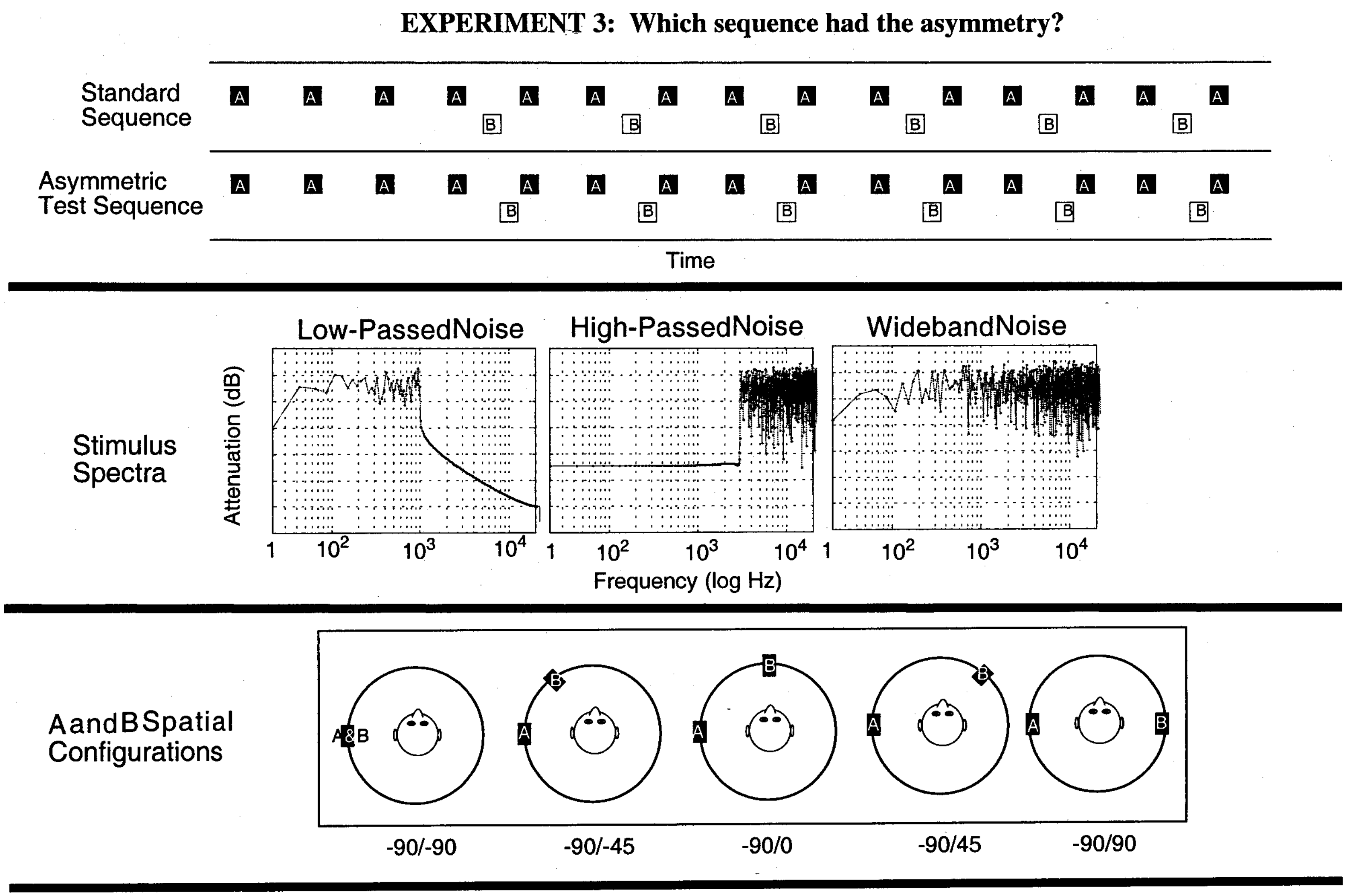


Figure 3.9 Experiment 3
Top: Schematic showing the ABA sequence provided in the 2AFC task. Listener chooses which stimulus had the asymmetry. Details as in Figure 3.7
Middle: Spectra of the A and B noisebursts. Note the spectra of A and B were always the same in a given threshold determination.
Bottom: Locations of A and B in the five spatial conditions.

However, while the low-pass noise was comfortable at this level, the high-pass noise and 4kHz tone were extremely uncomfortable to listen to at this level, due to the percussive nature of the ABA sequence. Thus, a comfortably low stimulus level for the high-frequency conditions was jointly determined by the listeners (about 50 dB SPL) and the stimulus was fixed at that level for all listeners (about 30-35 dB SL depending on the listener). Thresholds for detecting the temporal asymmetry in the repeating A_B_A stream were obtained as described in Experiment 2 (this chapter). The only differences were in the method of response collection (button box instead of keyboard), and the method of feedback, which was given via illumination of a green or red LED located at 0 degrees azimuth instead of on a computer screen.

RESULTS

The results from each of the 5 listeners, and the grand mean data are presented in Figure 3.10. Mean temporal asymmetry detection thresholds (ms) are plotted as a function of the location of 'B'. One function is plotted for each of the three spectral conditions. The mean thresholds for the 500Hz and 4kHz tone conditions are plotted as isolated symbols. In the individual subject plots, error bars are the standard deviations. In the mean plot, error bars are the average of each of the subject's standard errors. This type of error bar was used in order to avoid the large between-subjects differences in absolute thresholds values.

Contrary to hypothesis, and as is clearly shown in the Figure, there was no main effect of spectral condition ($F<1$), and no expected interaction of spectral condition with spatial separation ($F=1.0$). There was, however, an effect of spatial separation [$F(4,16)=5.13$, $p<0.01$]. A trend analysis indicated that effect of spatial separation was completely accounted for by a linear [$F(1,16)=5.3$, $p<.03$] and a quadratic trend $F(1,16)=15.06$, $p=0.01$]. Together they accounted for 99% of the sums of squared deviations attributed to the separation variable in the ANOVA. A separate 2x2 repeated measures ANOVA performed on the data from the four tone conditions led to a similar conclusion. There was a marginal main effect of frequency [$F(1,4)=2.16$, $p=0.07$], which reflected that thresholds were better with the 500 Hz tone; and a main effect of separation [$F(1,4)=10.97$, $p<0.03$], but no expected interaction ($p=0.37$).

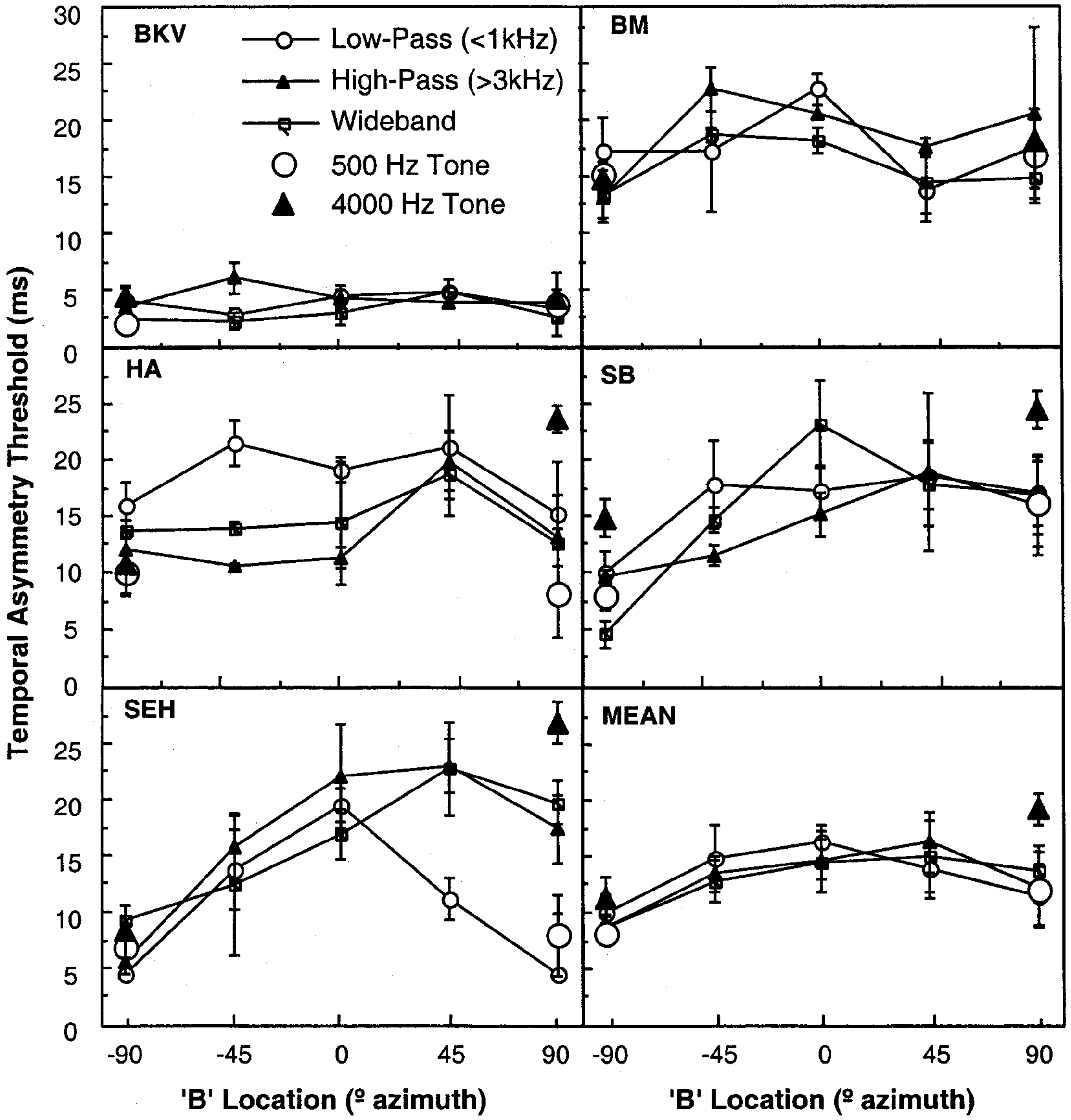


Figure 3.10.
Temporal asymmetry threshsolds are plotted as a function of the spatial location of the 'B' sound. The 'A' sound was always fixed in location at -90 deg. A separate function for each noise band used (low-pass, high-pass, and wideband) is plotted, in addition to thresholds for 500 or 4000 Hz tones at the same location or in opposite hemifields (large circles and triangles). Results from each of the five listeners are presented with error bars showing the standard error of the three threshold determinations which comprised the mean. The mean data from those five listeners is presented in the bottom right panel with error bars being the average of the individidual subject standard error values.

The significant variability across listeners warrants that they each be examined independently. Listener BKV, the best listener from Experiment 2, was also the best listener in Experiment 3. She had no effect of separation or spectral condition. As in Experiment 2, she found she could perform the task, regardless of condition by maintaining integration for the first repetition and doing a gap duration discrimination between the first AB interval in the test and standard stimuli. This alone suggests that no spatial separation conditions result in 'obligatory' stream segregation. Listener BM also showed no modulation with spatial separation or spectral condition; her thresholds were generally poor. This is consistent with her performance in Experiment 2, where her performance was generally poor, and only improved for the diotic and near-frequency conditions. Why she was unable to achieve similar thresholds for the same location condition in the present experiment is a little worrying. However, it may simply speak to the general difficulty all subjects found when this task was taken into the free-field (with the exception of BKV). Listener HA, showed the expected pattern for the high-pass and wideband stimuli and for the 500 Hz tone. Her performance using low-frequency noise was poor for all conditions, suggesting a possible difficulty with that stimulus. Recall that gap detection thresholds were also elevated in general with the low-pass noise. The author's data (SB) were the least consistent with the hypotheses. Thresholds were elevated with spatial separation for the low-pass noise and the 500 Hz tone. The elevation of thresholds with spatial separation with the 4000 Hz tone followed expectations, as did the effect of spatial separation for high-pass noise. The pattern of data for high-pass noise did not follow expectation, showing significant elevation at mid-separations. Finally the pattern of data obtained for SEH, the single most practiced subject in the laboratory, are most interesting. She showed the expected interaction between frequency and separation for the tone conditions. She also showed the expected pattern of data for the wideband and high-pass noise conditions. However, her pattern of data for the low-pass noise condition is inexplicable, having definite tuning with a peak at 0, and being lowest for the no separation and the largest separation conditions.

Discussion

There was significant variability across listeners on this task and it was uniformly agreed that the task was more difficult in the free-field than under the headphone conditions. One possibility was that timbre differences resulting from the speaker transfer functions influenced results by additively contributing to segregation. However, there are two arguments against this. The speakers were independently calibrated online via the MATSS system so to be roughly matched in level for each stimulus type. That is, the SPL of each speaker was determined for each type of noise, and the speakers were equalized by varying the potentiometers, which controlled the amplifiers. This procedure would not entirely eliminate timbre differences, as intensity differences in different frequency ranges may have been present. However, the fact that there was no interaction between frequency and spatial separation for the tones is stronger evidence against an explanation based on speaker transfer functions. The speakers were independently calibrated to be matched in output level for each of the tones. Tones are obviously not subject to speaker transfer functions in the same way that noise is.

Why then might free-field spatial separation of 500 Hz A and B tones affect temporal asymmetry thresholds, while A and B noises differing in ITD over headphones do not? In theory, the only difference between A and B tones at the same free-field location and A and B tones at different spatial locations is in their ITD. A 500 Hz tone is not subject to shadowing by the head and can not be filtered in any way as a function of location as it is a tone. One possibility is that the poverty with which noise bursts of opposite ITDs led to segregation was due to the use of wideband noise. The higher frequency channels would have been less sensitive to the ITD, yet higher frequencies components (>2kHz) comprise much of the stimulus. The stimulus intensity in a given low frequency channel might have been too low to sufficiently activate channels sensitive to ITDs. This was tested (n=1) as a possibility, but the author was easily able to achieve temporal asymmetry thresholds ranging between 4 and 5 ms for A and B tones of 500 Hz when both were presented diotically OR when A and B were lateralized to opposite sides of the head with ITDs of +/- 500μs. Thus, while stream segregation seemed to be fairly strong under that condition, the temporal asymmetry could always be determined by listening 'underneath' the lateralized stimulus. It is as if there is a dual perception of integration and segregation simultaneously.

One possiblity for the various results stems from the multitude of stimulus differences that arise in the free-field. There are spectral and intensive factors that differ as a function of location resulting in a myriad of cues. Perhaps this, combined with the task difficulty simply made detection variable. Furthermore, because the conditions were randomized, an effective strategy used in a previous condition might have been completely ineffective for the following condition. That is, if listener worked on one stimulus condition at a time, thereholds may have been more stable. Due to the variability in the patterns of data across listeners, it is very difficult to draw definite conclusions from this data set.

GENERAL DISCUSSION: SEQUENTIAL STREAM SEGREGATION BY SPATIAL LOCATION

The goal of this chapter was to determine the spatial stimulus attributes that might lead to sequential stream segregation of spectrally similar noise bursts. In the first experiment subjects tracked their perception of a repeating ABA sequence, under conditions where A and B were identical (DIOTIC), differed in ear of presentation (EAR), differered in interaural level (ILD) or monuaral level (LQL or QLQ), or differed in their interaural time difference (ITD). In general, listeners perceived sequences in the EAR condition to segregate nearly immediately (after the first ABA iteration) and remain segregated throughout the sequence. The ILD sequences were also perceived to segregate quickly, although occasionally reverted back to an integrated perception. The monaural level difference conditions (LQL and QLQ) were generally not perceived to segregate, although this was variable among listeners. The ITD condition led to segregation in some listeners, but the perception of these sequences tended to be unstable, in that listeners tended to report a return to an integrated perception during the course of the sequence. It is arguable that in the rating task the listener were biased to try to hear out the sequence as *segregated*, in order to utilize the response medium. The basic pattern of results – EAR>=ILD>ITD>LQL=QLQ>DIOTIC – generally reflected the similarity in peripheral activation patterns of A and B: complete segregation in the peripheral channels activated in the EAR condition and no segregation in those activated in the ITD and DIOTIC conditions. The fact that segregation was possible based on differences in ITD alone is important. This demonstrates that sequential stream segregation can occur in the absence of peripheral segregation, consistent with the results of Hartmann and Johnson (1991), even when A and B are spectrally identical. This is consistent with recent results demonstrating segregation with A and B tone complexes that are spectrally similar, but differed in fundamental frequency (Vliegen & Oxenham, 1999), and A and B tones with the same carrier but different amplitude modulation rates (Grimault *et al.*, 2002). In both cases the peripheral channels activated are similar, but the stimuli differ in centrally generated mechanisms underlying the extraction of different pitches. The power of the present study was to demonstrate that ITD was *less effective* than other spatial stimulus attributes that produce similarly lateralized perceptions (the ILD condition). This suggests that similiarity in peripheral channel activation provides a bottom-up limit on the segregation process. This bias towards an 'integrated' state may then be overidden as other evidence for A and B differences are deemed to be consistent over time. The immediacy of segregation may be determined by bottom-up processes, but more centrally generated stimulus differences can be used for segregation upon a 'second look' by the perceptual processor. This is consistent with the fact that when segregation occurred with ITD sequences it took somewhat longer to 'build-up'.

The second experiment determined the effect of segregation by the same stimulus attributes on temporal judgements (detection of a temporal asymmetry) made between the A and B stimuli. This task biases the listener to attempt to hear the sequences as *integrated* in order to optimally perform the task. The results indicated that relative to the DIOTIC control condition (no streaming), detection of a temporal asymmetry was extremely difficult when A and B differed by ear of entry (EAR), moderately difficult if A and B differed in either their interaural (ILD) or monaural level (LQL QLQ), and not at all impaired if the stimuli differed in ITD. In this experiment, in which listeners were biased towards hearing the sequence as integrated, the basic pattern of results was EAR>ILD=LQL=QLQ>ITD=DIOTIC. Performance was completely dictated by the similarlity of peripheral channels activated by A and B. This suggests that listeners were able to maintain an integrated perception in order to do the task. This is most notably in a condition like ITD, where segregation is not automatic, but takes time to build up. The similarity of this pattern of data with that obtained for gap detection under the same stimulus configurations (Chapter 2, Exp. 3B) is striking. This raises the possibility that the same constraints limiting gap detection, may be those limiting the immediacy of sequential stream segregation (i.e., the ability to maintain an integrated perception). Gap detection thresholds obtained between markers that differ perceptually on some dimension may thus be predictive of the build-up to stream segregation for those conditions. For those stimulus configuration for which segregation is automatic (between-ear, disparate frequencies), gap detection thresholds should be elevated. For those which require some time to build-up to segregation, gap detection thresholds should be less impaired. The dominant feature in determining performance appears to be the similarity in peripheral activation patterns.

THE EFFECT OF SPATIAL STIMULUS ATTRIBUTES ON STREAM SEGREGATION

From the results of experiments 1 and 2 the following can be concluded about the effect of spatial stimulus attributes on sequential stream segregation of noise bursts:

1. Segregation by ear is obligatory and immediate

This is evident from the immediate transition to 2@ key (indicating segregation) in the rating task, and the difficulty listeners had in detecting a temporal asymmetry in the EAR condition – a task that was impossible for some listeners.

2. Segregation by ITD is weak

Some listeners did not report segregated percepts in the ITD condition, and in those who did, segregation often took some time to build up, and was more unstable. In the temporal asymmetry task, thresholds obtained in the ITD condition were equal or less than those obtained in the control condition (DIOTIC). This indicated that listeners could easily suppress segregation in order to hear enough ABA iterations to perform the task. Furthermore, in 3 out of 4 listeners there was no effect of spatial separation on temporal asymmetry threshold for a 500 Hz tone.

3. Segregation by ILD is obligatory, but not immediate

In the rating task, listeners reported segregation as quickly as in the EAR condition. However, there was a greater tendency for the perception of the ILD sequence to revert back to an integrated state compared with the ear condition, suggesting the ILD cue was not as strong as segregation by EAR. Furthermore, in the asymmetry detection task, which biased the listeners towards maintaining an integrated perception, thresholds were significantly lower in the ILD condition than the EAR condition. This suggests that segregation by ILD can be delayed/suppressed for some time in order to successfully perform the asymmetry detection task, in a fashion that is not possible in the between-ear condition.

4. Segregation by monaural level differences is rarely reported, but impairs asymmetry detection nonetheless

The fact that temporal asymmetry detection was equally impaired with either binaural (ILD) or monaural (LQ,QL) level differences between A and B is intriguing, since listeners rarely reported segregation in the monaural conditions. This raises the possibility that the limit for temporal asymmetry detection might not necessarily be related simply to the ability to maintain integration, but perhaps may be limited by temporal integration of sound level such as may degrade veridical perception of a silent interval in gap detection (Plomp, 1964; Penner, 1977; Oxenham, 2000, various experiments in Chapter 2) and possibly gap duration discrimination (Divenyi & Danner, 1977). One possibility is that listeners did experience segregation in the monaural level difference conditions but had difficulting in labelling the perceptual state as such. When the stimuli differ in location or pitch, it is much easier to label the two streams. In the monaural case, for those who did report segregation, the streams were roughly related to distance – a quiet stream beating in the background and a louder stream beating at its own rhythm near the ear. When perceived as *integrated*, the stimulus was more simply perceived as a single event modulating in intensity. Distance is not a particularly strong perceptual dimension, as it is inherently ambiguous, requires previous knowledge of the source and possibly contrast with other sources in the environment. For example, a quiet low-pass sound can be veridically so, and located nearby, or it could be a broadband loud source that is some distance away. Resolution of such ambiguity would be resolved by source familiarity, or by experiencing the sound moving away, and thus reducing in intensity over time (particularly in the higher frequencies). Intensity differences can be used to segregate melodies to some extent (Hartmann & Johnson, 1991). This might be partly accounted for by narrower auditory filter widths at quieter intensities, so that the louder sound activates more peripheral channels than the quieter sound.

5. Ability to maintain 'integration' related to degree of common peripheral activation

In general, the degree to which common peripheral channels are activated by A and B noises is correlated with the ability to maintain an *integrated perception.* That is, the duration over which an integrated percept can be maintained is determined by the degree of common peripheral activation – EAR<ILD<ITD. The acuity with which temporal judgements can be made between A and B noises (e.g. the temporal asymmetry task) depends on the degree to which an integrated percept can be maintained. For example, conditions for which integration can be maintained for at least the first few iterations of the sequence were those with greatest peripheral similarity (ITD, CTR, FREQ near); conditions for which segregation was automatic (EAR), integration was almost impossible and so thresholds were most elevated.

6. Segregation by free-field spatial location is difficult to measure and requires further study

In experiment 3, temporal asymmetry thresholds for ABA sequences were determined for A and B noises and tones at various degrees of spatial separation. The basic hypothesis was that spatial separation would affect detection of a temporal asymmetry between A and B noises or tones with high frequency content but not A and B noises or tones with exclusively low-frequency content. Each of the four subjects had a unique pattern of results. One subject had low, relatively undifferentiated thresholds for all conditions. Another had high, relatively undifferentiated thresholds for all conditions. The other two showed results that were affected by spatial separation, but not in the same way. The fact that three of these were equally practiced at this (and other) psychoacoustic tasks suggests that segregation of streams based on free-field spatial location is a complex task that is solved differently by different listeners. Little can be concluded from Experiment 3, except that the issue deserved more study.

NEUROPHYSIOLGICAL CORRELATES OF STREAM SEGREGATION

Although beyond the scope of this thesis to address empirically, predictions might be made regarding the response of auditory neurons to the sequences presented here. In the cortex for instance, it is known that a response to a second equally intense stimulus can not be generated with an onset asynchrony (SOA) shorter than about 40-50 ms (Eggermont, 2000). While the shortest SOA of A and B stimuli in this experiment were 120 ms, it is possible that the ability of a cortical neuron to respond to a second, less intense stimulus might require a longer SOA, particularly over the course of a long repeating sequence. There is some evidence for this in auditory cortical cells of awake monkeys in response to sequences of A and B tones of different frequencies in the A_B_A_B...A_B configuration (Fishman *et al.*, 2001). Cells with a best frequency centred on the A tone will also respond with a lower firing rate to the B tone, since the neuron will exhibit an attenuated response to tones on the skirts of its frequency response area. With increases in the ABA repetition rate, the cortical neuron can no longer respond to the B tone, presumably due to the reduced time the neuron has to recover after having responding to the A tone. The source of this suppression of responses after the offset of another stimulus is not known with certainty, but appears to be the same one which underlies forward masking (Phillips, 1985) and release from forward masking with frequency separation (Brosch & Schreiner, 1997; Fishman *et al.*, 2001). It could be a reflection of increasing adaptation of neurons at the various synaptic relays through the ascending auditory system. It could also be active inhibition from neighboring cortical neurons, or from inhibitory inputs to cortex from other cortical areas. In reality, it is probably a reflection of many of such processes, and regardless of origin, the resultant supression provides a possible neural bases for the increase in sequential stream segregation with increasing rate for tones of different frequencies (van Noorden, 1977, see Fig. 3.1). This view is further bolstered by behavioural evidence in humans which suggests that the effect of rate on stream segregation is largely determined by the duration of the silent interval between the 'A' and 'B' components of the sequence (Bregman *et al.*, 2000). For example, the same repetition rate can be generated by having either short duration sounds with longer silent intervals or long duration sounds with shorter silent intervals. For a given rate, the ability to maintain an integrated percept was more difficult when the silent interval is short, that is when there is less time for the neuron to recover after the offset of 'A' before the onset of 'B'. Finally, recent unpublished observations made of the response of cortical cells to ABA_ABA_ sequences ('A' presented at cortical BF) indicate that the reduction in the response to the 'B' tone due to such suppressive mechanisms 'builds up' over the duration of the sequence (Micheyl *et al.*, 2002). For the frequency separations and presentation rates used, the build-up time until complete suppression of the 'B' response is in line with that reported by human listeners.

Can the pattern of stream segregation obtained with the various spatial stimulus attributes examined in the present experiments be considered in terms of such processes? While neurophysiological recordings in response to the analogous stimulus conditions have not been reported, the known spatial sensitivity of cortical neurons might be used to evaluate this question. For sequences in which A and B differ in level, the response of neurons which have monotonic rate/level functions would likely be similar to that observed with A and B of different frequencies, but the same level. It might be expected that the response to the weaker stimulus would be gradually suppressed over time. Most auditory cortical neurons in A1 are binaurally sensitive, but will usually respond better to stimulation of the contralateral ear than to stimulation of the ipsilateral ear, consistent with sensitivity to contralateral spatial positions. In fact, some cells are inhibited by activation of the ipsilateral ear. Thus, the configuration where A and B were presented to different ears (EAR) would tend to result in a strong 'A' response in one A1 (contralateral to the ear to which A was presented) and a strong 'B' response in the other A1. Weaker responses to A and B in their respectively ipslateral cortex might be suppressed in the same fashion as 'off-BF' responses are suppressed in the condition where A and B differ in frequency. For A and B differing in ILD, a similar response pattern would likely be observed. For neurons sensitive to both ears, there would likely be one ear which would be dominant, and thus there would be a strong response to the stimulus presented to that ear and a weaker response to the stimulus presented to the other ear. In that case, through the suppressive mechanisms described earlier, it is likely that the responses to the weaker ear would be suppressed with increases in repetition rate, akin to the condition where weaker off-BF responses are suppressed. This would lead to considerable segregation of A and B into different neural populations. The exceptions would be neurons activated equally by both ears, or neurons which require simultaneous activation of both ears to respond (i.e., PB cell; see Chapter 1). However, as descibed in the introduction, the largest proportion of neurons have contralateral hemifield-tuning.

The response of cortical neurons to A and B stimuli differing in ITD would likely vary according to the BF of the neuron. Neurons with higher BFs would likely respond equally to both A and B, as they are less sensitive to ITD. Neurons with low BFs that are sensitive to ITDs would likely respond best to the A or B with an ITD favouring the contralateral ear. Responses to A and B noises differing in ITD might be segregated in the lowest frequency neurons. The fact that less perceptual segregation occurs in the ITD condition might simply reflect that relatively fewer neurons are responding only to only A or B. It is unclear how differentiated responses of low frequency neurons would be to A and B of different ITDs, because there is a paucity of recordings of low best-frequency ITD-sensitive neurons in the cortex (but see Wallace *et al.*, 2000; Wallace *et al.*, 2002).

COMPARISION WITH OTHER EXPERIMENTS

The pattern of results obtained in the rating task, which arguably biased the subjects to hear segregation, was similar to the pattern of benefit in melody recognition with separation by EAR and by ITD (Hartmann & Johnson, 1991). Presentation of A and B to different ears (EAR) resulted in the highest segregation ratings (Exp 1 here) and separation of melodies to the two ears resulted in the greatest improvement in melody recogition (Hartmann & Johnson, 1991). Giving A and B (or the two melodies) different ITD values resulted in intermediate ratings of segregation and an intermediate improvement in melody recognition. ITD has also been shown to be useful in segregating distracting (interfering) spectral information presented before and after target spectral information on which subjects performed a fundamental frequency discrimination task (Gockel *et al.*, 1999). This is basically a concurrent segregation task with sequential attributes. When it is a purely concurrent task (fringes presented at the same time as the target), then only separation of the target and distracting spectral information to different ears (but not different ILDs or ITDs) improved performance (Gockel & Carlyon, 1998).

The pattern of results obtained here in Experiment 2 using the temporal asymmetry task better mimics the effect of spatial stimulus attributes on such concurrent stream segregation tasks, i.e. a strong effect of ear of entry and no effect of ITD.

This distinction, that ITD is effective in sequential stream segregation, but not concurrent stream segregation, has been demonstrated by other authors (Culling & Summerfield, 1995; Darwin & Hukin, 1997, 1998, 1999). Giving different frequency components the same ITD is also not useful in grouping frequency components to solve a speech-sound recognition task (Culling & Summerfield, 1995). To explain such a pattern of results, Darwin and Hukin (1999) suggested that ITD might be used to direct attention between auditory objects that had been formed based on other grouping cues. This is consistent with the fact that ITD was more effective when the A and B noises were independent and so differed in timbre slightly. In that case there was already a basis on which A and B could be differentiated, and the additional ITD difference between them could lead them to form separate streams. That our listeners could occasionally segregate identical noise bursts that differed only in ITD suggests that ITD on its own can sometimes lead to segregation. More importantly, the fact that some listeners never experienced segregation based on ITD differences alone speaks to its weakness as a segregation cue.

LIMITATIONS AND FUTURE DIRECTIONS

The major limitation of this study is the lack of an objective task in the ABA format that biases listeners towards stream segregation rather than integration. The temporal asymmetry task, and arguably gap detection with dissimilar markers, are good tasks for which to measure the listeners ability to maintain integration. A task needs to be invented during which it is to the listener's advantage to segregate the A and B streams to perform the task. One possibility is to make listeners count the number of occurrences of some event in a single stream. In an ABA_ task, randomly a low-depth amplitude modulation might be applied to either the A or B. Listeners might be asked to report, by button press, when two such events occurred in a row in the B channel. Performance on such a task would benefit from segregation of the A and B streams and selective attention to the B stream. Another possibility is to alter the temporal asymmetry task to take advantage of the build-up to segregation. In Exp. 2, listeners needed only to maintain segregation through the first few ABA iterations in order to perform the task. An alternative would be to randomly place an asymmetry in one of the last few ABA iterations. Listeners are less likely to be able to maintain integration throughout the sequence in order to perform the task, if segregation is possible. Finally, the asymmetry could be introduced gradually into the 'test' sequence – increasing by some proportion with every ABA iteration. It would only then be detectable if subjects could maintain integration until the end of the sequence when the asymmetry was large enough to be detected. This strategy was recently employed with good success to demonstrate segregation of tone complexes in which A and B had identical power spectra, but the components of the complexes could differ in phase (leading to a kind of timbre difference: Roberts *et al.*, 2002).

The possibility that timbre differences between the A and B noise bursts (independently generated A and B bursts) which do not result in segregation on their own might be enhanced by spatial segregation by ITD or ILD was examined in Exp. 1. The initial inclusion of that noise-type variable was not intended for that purpose, and is a rather uncontrolled method by which to test such a hypothesis. A more controlled test might be to apply different ITDs to tone complexes that differ in component phase (after Roberts *et al.*, 2002). They demonstrated that such timbre differences induced by changing the phase spectra could enhance ratings of segregation with differences in frequency passband (which results in differential peripheral channel activation). It might be interesting to evaluate whether such phase-spectra mediated timbre differences might enhance segregation based on different ITD. In this case, both variables (phase spectra and ITD differences) are centralled mediated. This would determine the additive effect of ITD in a far more quantitative fashion than achieved in Experiment 1. And such a result would demonstrate additive effects of two types of centrally-generated stimulus differences, in the absence of differences in peripheral channel activation. This would provide a powerful demonstration that peripheral channel differences are not required for segregation.

Most importantly, a basic set of experiments are required to characterize the segregation and integration boundaries for ILD, ITD, EAR as a function of repetition rate, for a different AB silent interval durations (see Bregman *et al.*, 2000) and different frequency composition of the A and B noises. The fact that no strong effect of repetition rate was found in Exp. 1 suggests that the repetition rates, for which frequency differences modulate segregation, may not be the same as those for which spatial stimulus attributes modulate ratings of segregation.

CHAPTER 4

TEMPORAL RESOLUTION OF PURELY BINAURAL CHANNELS

Specific contributions of Torsten Marquardt and Susan Hall to the experiments in this chapter are gratefully aknowledged. Susan Hall created the program which presented the stimuli at different interaural correlations, and Torsten Marquardt implemented the binaural temporal window model used here. Many of the ideas emerged from lively discussions of binaural processing between the three of us from March-June, 2001.

PREFACE

One of the interesting results from Chapters 2 and 3 was the relative ineffectiveness of ITD as a segregation cue. This was surprising because ITD is the strongest cue for sound localization (Wightman & Kistler, 1992). Furthermore, the strong correspondence between the data of Experiments 1 and 2 with studies of free-field release from masking demonstrate that binaural interaction effects seem to play less of a role in the free-field than predicted from data obtained in headphone experiments. This is likely because the head shadow is such an overwhelmingly dominant aid to sound segregation in the free-field. The reason this is surprising is because the most studied effect in binaural hearing is the *benefit* accrued in detecting a signal in noise presented over headphones, when its interaural properties are made to differ from those of the noise (the binaural masking level difference – BMLD). This benefit can be up to 15 dB – an improvement that can be attributed solely to processes underlying sensitivity to interaural correlation – specifically, coincidence detection in the MSO.

The reason that separation of gap markers, or A & B noises, by ITD was relatively ineffective in modifying task performance relative to diotic conditions may be because temporal tasks basically measure a change in the stimulus level, and unless there is primitive segregation occurring in the representation of the markers, the task is probably limited by within-channel mechanisms. In the ITD case, the dominance of frequency and intensity similarity at the two ears biases the system to maintain an integrated state despite the change in ITD. This allows the tasks of Chapter 2 and 3 to be performed as easily as the diotic control condition when the markers are similar. Thus, we never examined pure changes within the binaural channels - changes in the outputs of binaural channels may have always been overridden by the similarity of the markers in the monaural channels. This chapter is a bit of a sidestep from the preceding chapters, but stemmed from a desire to understand temporal processing mechanisms in purely binaural channels, uncluttered by concomitant changes in monaural channels. This required a deeper understanding of the mechanisms underlying the ITD cue to spatial location. Recall from the discussion in the Chapter 1, that ITD is processed by MSO neurons, which detect the interaural correlation of the sound components entering the neuron's frequency pass-band from the two ears. Manipulating the ITD of sine waves presented dichotically modulates the firing rate of the MSO neuron such that the rate increases and decreases as the excitatory inputs from each ear arrive in phase and out of phase respectively. This kind of a neuron is usually referred to as an EE cell – to specify it has two excitatory inputs (although it is more accurately an EIEI cell as there is additionally an inhibitory input arriving from each ear, see Grothe, 2000); or as a "peak-type" neuron – to specify that the firing rate peaks at the same interaural delay, regardless of the frequency of the input; or as a coincidence detector – to specify that the ability of that neuron to generate an action potential is dependent on the coincident arrival of at least two excitatory spikes (one from each input). The MSO neuron is sensitive to ITDs, and this sensitivity arises from the neuron's sensitivity to interaural correlation. This chapter describes experiments that examine the sensitivity of the binaural system to rapid temporal changes in interaural correlation, under a variety of interaural correlation configurations. Determining the shortest detectable duration of change in interaural correlation has been deemed a 'binaural analog' of monaural temporal gap detection, in which the shortest duration of change in intensity (decrease) is determined (Akeroyd & Summerfield, 1999). As in the monaural case, there is a limit on the duration of change which is detectable. This limit is presumed to occur as a result of smearing of the representation of interaural correlation over time due to temporal integration (persistence). The main goal of the present experiments was to evaluate this analogy between monaural and binaural 'gap' detection. The second goal was to examine if the pattern of sensitivity of binaural gap thresholds to marker and 'gap' interaural configurations was related to pattern of sensitivity of binaural signal detection (BMLD) to the interaural configuration of the signal and masking noise.

INTRODUCTION

The interaural correlation ('ρ') of a dichotic noise stimulus is a measure of the similarity of the waveforms presented to the left and right ears. If the waveforms presented to each ear are identical (diotic or 'correlated', $\rho = +1$), the resulting perception is of a compact sound source whose location can be specified precisely at the center of "intracranial" space. Reducing the interaural correlation, for example by adding independent noise to either channel, widens the perceptual image and makes it more diffuse. When noises from fully independent sources are presented to each ear ('uncorrelated', $\rho = 0$), the resulting perception is that of a diffuse sound that fills the head, but may have two dominant sources near the ears (see Blauert & Lindemann, 1986). When identical noises are presented to each ear and the waveform at one ear is inverted, the stimulus becomes interaurally 'negatively-correlated'($\rho = -1$). There are no formal reports on the perception of negatively-correlated noise. It has been described as two relatively compact sources near the ears (see Hirsh, 1948), although a few of our listeners reported one dominant blurred image displaced laterally. It is occasionally described by listeners as unpleasant or irritating. In this paper, the terms 'correlated', 'negatively-correlated' and 'uncorrelated' refer to interaural correlations of +1, -1 and 0 respectively. Following mathematical convention, any change towards –1 will be considered a decrease and any change towards +1 will be considered an increase in interaural correlation. Note however, that while shifts towards –1 will be referred to as a decrease, the noises presented to each ear have maximal independence at zero correlation.

There has been a renewed interest in examining human sensitivity to interaural correlation, particularly with respect to its relation to the binaural masking level difference (e.g. Durlach *et al.*, 1986) and dichotic pitches (Culling *et al.*, 1998a; Culling *et al.*, 1998b). Recently, Culling et al. (2001) reported a study on the discriminability of different positive interaural correlations of a sub-band of noise within a correlated broadband noise. Their results were similar in form to those first reported by Pollack & Trittipoe (1959a) who manipulated the correlation of the entire spectrum. Specifically, discrimination was very fine for decreases in interaural correlation from a reference of +1 ($\Delta\rho = 0.02$-0.04), progressively degraded with decreases in the reference value, and was worst ($\Delta\rho = 0.3 - 0.5$) for an uncorrelated reference (Pollack & Trittipoe, 1959a, 1959b;

Gabriel & Colburn, 1981; Koehnke *et al.*, 1986; Jain *et al.*, 1991; Bernstein & Trahiotis, 1992; Culling *et al.*, 2001). The binaural system is unable to follow rapid changes in interaural correlation, as evidenced by the relative difficulty listeners have in detecting or discriminating dynamic changes in interaural disparities (Grantham & Wightman, 1978, 1979). This 'binaural sluggishness' has been characterized in terms of a 'binaural temporal window' - a moving-average filter that integrates binaural information over time according to some weighting function. Several studies have attempted to determine the shape and equivalent rectangular duration (ERD) of the window (Kollmeier & Gilkey, 1990; Culling & Summerfield, 1998; Bernstein *et al.*, 2001). In an attempt to measure the ERD but not the window shape, Akeroyd and Summerfield (1999) recently examined detection of dynamic changes in interaural correlation in an experiment described as a binaural analog of gap detection. In monaural temporal gap detection, the parameter of change is intensity (Moore *et al.*, 1988); in the binaural analog, the parameter of change is the interaural correlation. The "gap" was represented by uncorrelated noise embedded in correlated noise markers. This stimulus can be described as a sequence of three contiguous noises in which the first and last (markers) are of interaural correlation X, and the center (target) noise is of interaural correlation Y, i.e. X/Y/X. At either ear alone a continuous noise is heard through the duration of the stimulus. The analogy assumes that there is a decrease in activity in the binaural temporal window's output specifically during the uncorrelated Y-noise, similar to that effected in the output of a monaural temporal window by a decrease in intensity (a gap). It follows that as the Y-noise decreases in duration it contributes less weight to the smoothed output (Akeroyd & Summerfield, 1999). This means that the effective change in interaural correlation is less than that specified by the stimulus. The minimum detectable target duration is thus a measure of the ability to detect a short departure from the marker's interaural correlation. For stimulus conditions in which noises of different bandwidths and center frequencies were used, Akeroyd and Summerfield (1999) found that a binaural temporal window model could predict the binaural gap thresholds obtained in the +1/0/+1 configuration by knowing the subject's interaural correlation discrimination acuity for the same stimulus conditions, i.e. their JND ('just noticeable difference') from a reference of +1. Akeroyd and Summerfield (1999) estimated the equivalent rectangular duration

(ERD) of the binaural temporal window to be about 140-210 ms depending on the model employed. This is much longer than the analogously defined Gaussian-shaped monaural temporal window (e.g. Moore *et al.*, 1988, 27 ms [corrected value, see Akeroyd and Summerfield, 1999]). While useful as an analogy, the term "binaural gap detection" limits the implied possibilities of the paradigm. Since the 'gap' could in theory be an increase or decrease in interaural correlation from a reference noise of any given correlation value, the task might provide a more general way to probe the detectability of interaural correlation changes. In purely stimulus terms, the general paradigm is analogous to increment- and decrement-duration detection in the intensity domain, of which gap detection is a single case. For clarity, the "binaural gap" paradigm might be more generally construed, and will be referred to in this report, as 'interaural correlation change-interval' (ICCI) detection.

Another task that uses the full range of interaural correlation is the binaural masking level difference or BMLD. Like ICCI-detection, the BMLD has been linked to interaural correlation discrimination (Durlach *et al.*, 1986; Koehnke *et al.*, 1986; Jain *et al.*, 1991; Culling *et al.*, 2001). The BMLD is the binaural advantage in detecting a signal (S) in noise (N) when the interaural configuration of the signal [e.g., correlated (o), negatively-correlated (π), uncorrelated (u)] differs from that of the masking noise, as compared typically to the condition in which both the signal and noise are presented diotically. At present, the correspondence between the BMLD and correlation discrimination has only been made clear for cases of correlation discrimination from positive reference values, such as the NoSπ condition (Durlach *et al.*, 1986; Culling *et al.*, 2001). In light of a relation among the interaural correlation JND, and both ICCI-detection and the BMLD, one might make predictions for different ICCI-detection and correlation discrimination conditions based on what is known about the BMLD under analogous configurations. For instance, the BMLD is largest for NoSπ, and is several dB less for the NπSo condition (Hirsh, 1948); one would therefore predict better interaural correlation discrimination from a reference of +1 than from a reference of -1. See Table 4.1 for an overview of corresponding BMLD, ICCI-detection, and JND tasks. Note that these predictions would be expected to hold only to the degree that the BMLD can be conceived in terms of correlation discrimination, a correspondence that is good, but not complete (Durlach *et al.*, 1986).

The present study examined the relation between performance on the interaural correlation JND and the ICCI-detection tasks. Thresholds were obtained for all X/Y/X configurations of correlated, negatively-correlated, and uncorrelated wideband noise. JNDs for interaural correlation were obtained from corresponding reference interaural correlations (ρ_{ref}) of +1, -1, and 0 in the positive (0+) or negative (0-) direction (see Table 1). JNDs from $\rho_{ref} = +1$ and -1 should be related to performance in the 1/0/1 and -1/0/-1 conditions respectively. JNDs from $\rho_{ref} = 0$ in the positive and negative direction should be related to performance in the 0/1/0 and 0/-1/0 conditions respectively. In order to determine whether the two data sets were related by the same binaural temporal window, we used the "ρ_w" model to estimate the window length for each related pair of JND and ICCI-detection conditions. The model is described in detail by Akeroyd and Summerfield (1999), and follows the strategy of earlier authors (Grantham & Wightman, 1979; Kollmeier & Gilkey, 1990; Culling & Summerfield, 1998).

Table 4.1.

Relation among the binaural masking level difference, interaural correlation change interval (ICCI) detection, and interaural correlation discrimination (JND) for various interaural configurations. Higher scores on the BMLD task indicate better performance while lower scores on the dynamic (ICCI) and static (JND) interaural correlation tasks indicate better performance. Data presented are results from this study. Binaural temporal change detection and correlation JND values are from the present study.

	BMLD(dB) Np Sp	Interaural Correlation Change-Interval (ICCI) Detection	**Correlation JND**	
			Ref (ρ)	**Comp**
Signal causes decorrelation from +1	$N_{+1}S_{-1}$ (NoSπ) 14 dB	1/-1/1 1.48 ms	+1	Δρ= -0.045
	$N_{+1}S_0$ (NoS90°) 6 dB *	1/0/1 2.40 ms		
Signal causes decorrelation from –1	$N_{-1}S_{+1}$ (NπSo) 10 dB	-1/1/-1 3.98 ms	-1	Δρ=+0.086
	$N_{-1}S_0$ (NπS90°) 4 dB *	-1/0/-1 7.60 ms		
Signal causes a change in correlation from 0	N_0S_{+1} (NuSo) 4 dB	0/1/0 21.01 ms	0	Δρ= +0.32
	N_0S_{-1} (NuSπ) 2 dB	0/-1/0 42.88 ms	0	Δρ= -0.46

*The comparison of uncorrelated noise, as used in our experiments, with a tone with an interaural phase difference of 90°, as used in this BMLD condition, is not entirely correct. Two uncorrelated noises originate from independent sources, i.e. have a *coherence* of zero. The phase shift introduced to the tone does result in an interaural correlation of 0, but it retains of *coherence* of 1. However, for illustrative purposes only, the values for the NoS90° andNπS90°conditions reported in Durlach and Colburn (1978) are displayed.

EXPERIMENT 1: DETECTION OF STATIC AND DYNAMIC CHANGES IN INTERAURAL CORRELATION

METHODS

Listeners

Data for the ICCI-detection conditions were collected from eight subjects (5 female) ranging in age from 20 to 35 years. Five of those listeners also participated in the correlation discrimination (JND) experiment. Listeners SB, SH, and JH were highly experienced in a variety of auditory temporal processing tasks; TM, DR, and MC had participated in a few other psychoacoustical tasks, listeners BV and RO were relatively inexperienced. All had normal audiograms from 250 Hz to 8 kHz. In both the ICCI-detection and the JND tasks, listeners completed many practice runs of each condition until thresholds stabilized, and the mean threshold for each condition was taken from the last three threshold determinations. Subjects were always provided with visual feedback.

Interaural Correlation Change Interval Detection (ICCI)

There were 6 X/Y/X configurations tested for detecting a dynamic change in correlation. These are illustrated schematically in Figure 4.1. Stimuli were created using Matlab software (The Mathworks) and presented at a sampling frequency of 44100 Hz at 16-bit amplitude quantization on the audio processor of an Apple PowerMacintosh 8600. The stimuli were constructed by concatenating three independent noise bursts (0-22050 Hz), and the total stimulus was shaped with 10 ms rise and fall times. For each stimulus presentation, new noises were generated. The total stimulus duration was fixed at 500 ms; thus, the marker durations increased as the target noise decreased adaptively. Stimuli were presented to subjects at a comfortable listening level (70 dB SPL) over AudioTechnica ATH-M40fs Precision Studiophones, while they were seated in a sound-attenuating booth.

Thresholds for detecting the target (Y) noise were measured using a two-interval, two-alternative force-choice task and a three-down, one-up adaptive method, estimating the 79.4% -point on the psychometric function (Levitt, 1971). On each trial, listeners were presented, in random order, with a standard stimulus consisting of noise of correlation X and a test stimulus in the X/Y/X configuration. The task was to indicate

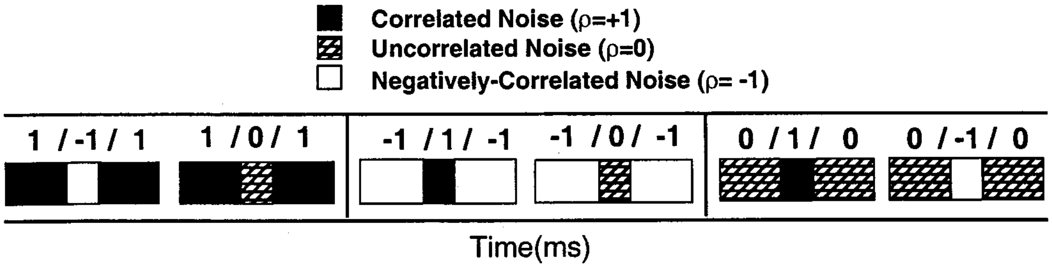


Figure 4.1
Schematic of the interaural correlation change interval (ICCI) detection conditions. The stimuli are a contiguous sequence of 3 broadband noises of X and Y correlation of equivalent intensity, in the configuration X/Y/X. The interaural-correlation is represented by the black (correlated, ρ = 1), hatched (uncorrelated, ρ = 0) and white (negatively-correlated, ρ = -1) sections. In the 2AFC task, detection is of the presence of Y noise.

whether the test stimulus was in the first or second interval. The initial Y-noise duration was set well above detection threshold for a given stimulus condition (10-70 ms). Until the first incorrect response, the center noise duration was decreased by a factor of 1.2 for each step. Subsequently, the target (Y) noise duration was increased by a factor of 1.2 after each incorrect response, or decreased by a factor of 1.2 after three consecutive correct responses. Each adaptive staircase continued for 8 reversals and the geometric mean of the target noise durations for the last 6 reversals was defined as the ICCI-detection threshold.

The just noticeable difference (JND) in Interaural Correlation

The general experimental details were the same as above, except for the following. The just noticeable difference (JND) for the reduction in interaural correlation from a reference of $\rho_{ref} = +1.0$ and $\rho_{ref} = 0$, and for an increase in interaural correlation from $\rho_{ref} = -1$ and $\rho_{ref} = 0$, was determined. The stimuli were wideband noises of 400 ms duration (10 ms rise/fall times) with a fixed value of interaural correlation. Reference stimuli (the standard) were noise bursts with an interaural correlation of 1, -1, or 0. The interaural correlation of the test stimuli was controlled by mixing two independent Gaussian noises, one interaurally correlated (N_{+1}, $\rho = 1$) and the other negatively-correlated (N_{-1}, $\rho = -1$), as described in equation (1) (van der Heijden & Trahiotis, 1997). In this equation, N_{+1} is the power of the interaurally correlated noise, N_{-1} is the power of the negatively correlated noise, and N_p is the power of the sum.

$$N_{\rho} = \frac{1}{2}(1 + \rho)N_{+1} + \frac{1}{2}(1 - \rho)N_{-1} \qquad (1)$$

JNDs were measured using a four-interval, two-alternative forced-choice task and a three-down, one-up adaptive method, estimating the 79.4% point on the psychometric function (Levitt, 1971). On each trial, the first and last intervals had the correlation of the reference stimulus. The test stimulus was randomly located in the second or third position. The task of the listener was to indicate which of the middle two noise bursts was of discrepant correlation. For JNDs from $\rho_{ref} = +1$ and -1, the initial interaural correlation of the test stimulus was +0.5 and -0.5 respectively, and the correlation was changed towards the reference correlation with correct responses. For JNDs from a $\rho_{ref} = 0$, the initial correlation of the test stimulus was either +1.0 or -1.0 and was changed with correct responses towards 0. The same three down, one up adaptive procedure (as described above) was used to determine the threshold using a step size factor of 1.2 based on the difference between the test interaural correlation and the reference interaural correlation.

RESULTS

Interaural Correlation Change Interval (ICCI) Detection Thresholds

Thresholds for duration of a noise of interaural correlation Y were obtained from 8 listeners for each of the X/Y/X conditions. These are plotted on a log scale in Figure 4.2. Individual subject means and standard errors (based on the final three observations collected) for each condition are presented in the main graph, and the group means for each condition are presented in the inset graph. There are two important trends in the data. Targets embedded in correlated or negatively-correlated markers (1/-1/1 and 1/0/1, or -1/1/-1 and -1/0/-1) were generally more easily detected than targets embedded in uncorrelated markers (0/1/0 or 0/-1/0). That is, it was easier to detect a brief change in correlation from markers of $\rho = +/- 1$ than from markers of $\rho = 0$.

Second, detection performance of listeners was better when detection of the change was executed in the positive range of interaural correlation than in the negative range. The mean thresholds for detecting uncorrelated noise amid correlated or negatively-correlated markers, 1/0/1 and -1/0/-1, were 2.34 ms and 7.43 ms respectively, a difference which was statistically significant [$F(1,7) = 13.42$, $p < 0.01$][1]. For larger changes in correlation, 1/-1/1 and -1/1/-1, the mean thresholds across listeners were 1.43 and 3.91 ms respectively. A repeated measures ANOVA indicated that this difference in thresholds was also statistically significant [$F(1,7) = 7.5$, $p<0.03$]. These conditions are analogous to the first pair (1/0/1 and –1/0/-1), but the change in correlation is twice the magnitude of change caused by an uncorrelated Y noise (e.g. a correlation change of 1.0 for 1/0/1 vs. 2.0 for 1/-1/1) for a given duration. As a consequence, a shorter duration of Y would be required for the smeared window output to reach the threshold for correlation discrimination. That is, the mixing of negatively-correlated noise with correlated noise within the temporal window has twice the decorrelating effect as the mixing uncorrelated

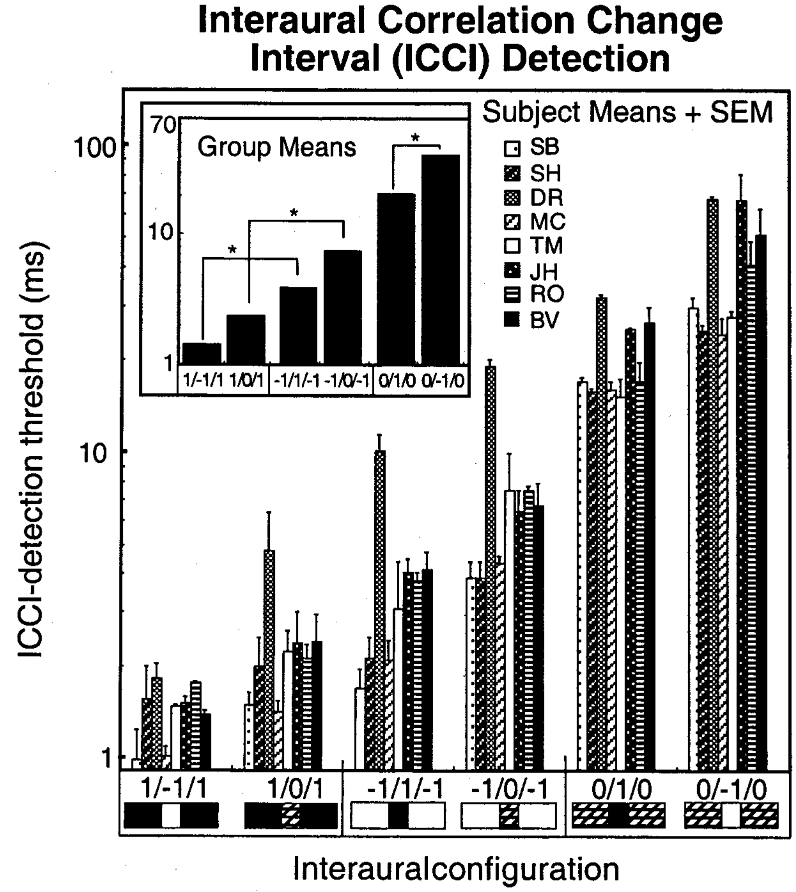


Figure 4.2
Duration thresholds (ms) for a dynamic measure of interaural correlation sensitivity - the interaural correlation change interval (ICCI) detection task (see figure 1 for schematic of stimuli). Individual subject means with standard errors for the six interaural configurations (see text) are presented in the main graph, and the group means are presented in the inset graph.

and correlated noise. The benefit in performance seen with a doubling of the correlation change of the target, however, was proportionately greater for negatively correlated markers (49% decrease in threshold) than for positively correlated markers (29% decrease). A repeated measures ANOVA with two factors – the interaural correlation of the markers ($\rho = +1$ or -1) and the size of correlation change caused by the target (correlation change of 1.0 or 2.0) – resulted in a significant interaction [$F(1,7)=22.35$, $p<0.01$]. This interaction indicates that the benefit of a larger correlation change was significantly greater for increases in correlation from -1 than for decreases from +1.

Finally, detection of a correlated noise amid uncorrelated markers was relatively poor, particularly if the Y-noise was negatively-correlated. The mean thresholds for the 0/1/0 and 0/-1/0 conditions were 20.57 ms and 41.45 ms respectively. These results show that the detection of changes in interaural correlation from zero was also significantly asymmetric in the positive and negative range of interaural correlations [$F(1,7) = 22.5$, $p<0.01$].

Interaural Correlation JNDs

The interaural-correlation JNDs are shown in Fig. 4.3. Results from five individual listeners are plotted in the main graph with standard errors based on their 3 final threshold determinations. Group mean data are presented in the inset graph. JNDs for every listener were consistently best for $\rho_{ref} = +1$ with a mean of 0.048. The mean JND for $\rho_{ref} = -1$ was 0.084. A repeated measures ANOVA indicated that this asymmetry was statistically significant [$F(1,4) = 9.5$, $p<0.05$]. JNDs for $\rho_{ref} = 0$ were poorer in general and more variable across subjects. The mean JND for an increase in correlation towards +1 was 0.31; while that for a decrease in correlation toward –1 was 0.45. Again, this asymmetry was statistically significant [$F(1,4) = 19.6$, $p=0.01$].

THE BINAURAL TEMPORAL WINDOW MODEL

We wondered whether the processing mechanism limiting performance on all these tasks could be characterized as sharing the same binaural temporal window ERD. If this were the case, the duration of the binaural temporal window measured would be independent of the interaural configuration used. In order to calculate an estimate of this

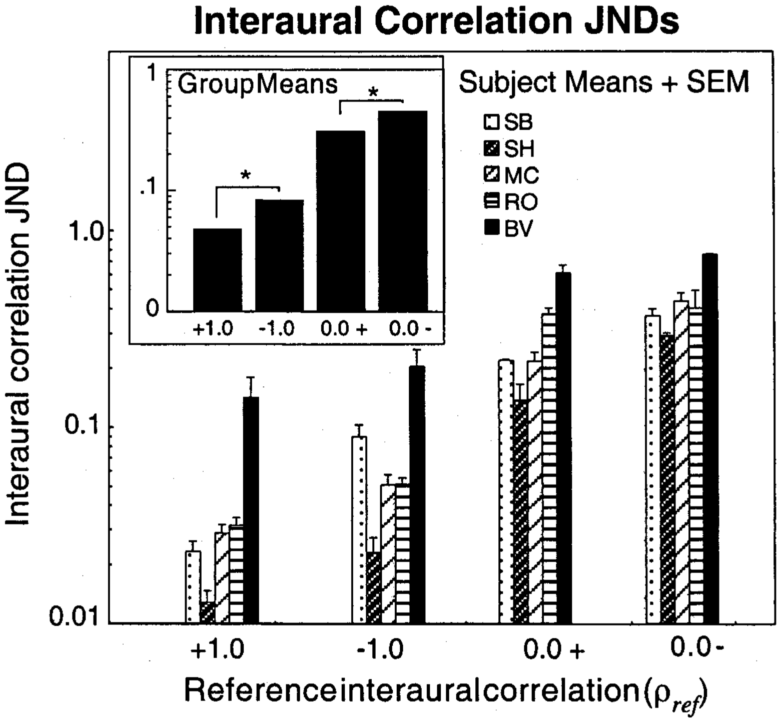


Figure 4.3

Interaural correlation JNDs. Individual subject means with standard errors are presented in the main graph for 4 conditions – the just noticeable decrease in interaural correlation from references of 1 and -1, and a just noticeable increase in correlation from $\rho_{ref} = 0$, *in either a positive or negative direction (see text). Group means are presented in the inset graph. In both A and B asterisks indicate that the difference between the specified bars was statistically significant (p<0.05).*

binaural temporal window ERD for each subject and configuration, we used the “ρ_w“ model (see Akeroyd & Summerfield, 1999: equations 3-6). Following the strategy employed by others (Grantham & Wightman, 1978; Kollmeier & Gilkey, 1990; Culling & Summerfield, 1998) this model measures the interaural correlation directly on the stimulus waveform. The temporal window used in the model is symmetric, and has a Gaussian shape. The model assumes that the ICCI is detected if the change in the output of the temporal window during the X/Y/X stimulus exceeds the JND for the corresponding change in interaural correlation.

In Figure 4.4, the ICCI-detection thresholds (durations of Y-noise in ms) are plotted against the corresponding correlation JND for each subject (see Akeroyd and Summerfield, 1999, Figure 10). Superimposed on these data are lines that represent predicted relations between ICCI-detection threshold and JNDs, based on different binaural temporal window ERDs. ICCI and JND pairs that fall along a given line can be explained by a binaural temporal window with a duration corresponding to the ERD represented by that line. The location of the data points moves upwards along these lines for the conditions with poorer thresholds, indicating that the performance for different interaural configurations is correlated for both the static and dynamic tasks, and thus, for a given listener the window ERD remains relatively constant across configurations. The lines representing the different ERDs in this figure are generated only for changes in correlation with a magnitude of 1.0 (e.g. +1 to 0). It is for this reason that the 1/-1/1 and –1/1/-1 conditions, for which the magnitude of change in correlation is twice the size (2.0), are not plotted in Figure 4.4. If the lines of equal ERD were plotted for a correlation change of 2.0, they would be shifted downward because the model predicts that the ICCI-detection thresholds would be approximately half the size for the same JND values.

Figure 4.5 shows the actual individual ERDs calculated for each condition, including the conditions where the X and Y noise are of opposite correlation, 1/-1/1 and -1/1/-1. These were also generated using the “ρ_ω“ model. Although individual listeners' ERDs vary somewhat across condition, there appears to be greater variability across listeners. There was no significant difference in the mean ERD, collapsed across

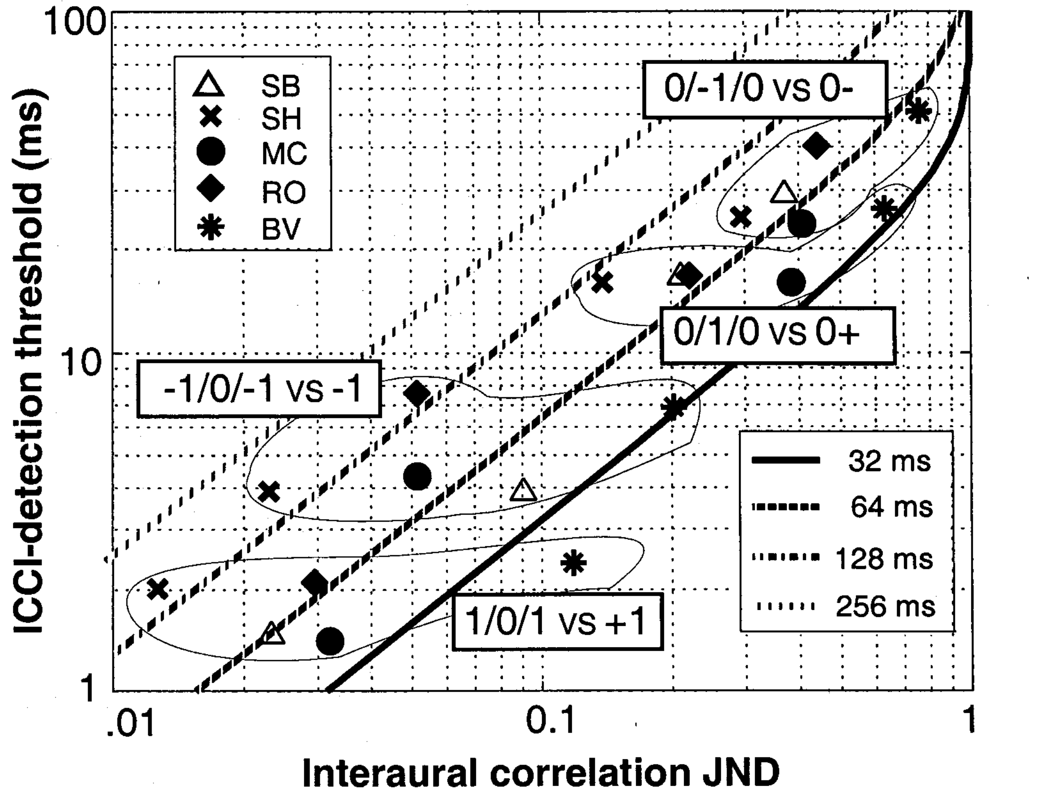


Figure 4.4

The relation between the dynamic measure of interaural correlation sensitivity (ICCI-detection) and the static measure (JND) can be explained via a temporal window that smears temporal changes in interaural correlation. The data points show the experimental measurements from each subject from Figs. 4.2 and 4.3. Interaural correlation JNDs are plotted against the ICCI-detection threshold (ms) for related configurations. Superimposed on these points are lines which represent a binaural temporal window length (ERD) calculated using the "ρw" model (see Akeroyd and Summerfield, 1999, Fig. 10). A binaural temporal window with a fixed duration (ERD in ms as per the legend) can explain all JND and ICCI threshold pairs that fall along a single line. The location of the data points for different conditions shifts upward along these lines for the conditions with poorer thresholds. This linear shift demonstrates that changes in acuity are correlated for the static and dynamic tasks, and thus the window ERD remains relatively constant across conditions. The lines calculated by the model are based on a change in correlation of 1 (e.g. +1 to 0). A different set of lines would be generated if the change in correlation was 2 (e.g. +1 to –1). It is for this reason that the data for the 1/-1/1 and –1/1/-1 conditions are not plotted here. The actual ERDs for all conditions are reported in Figure 4.5.

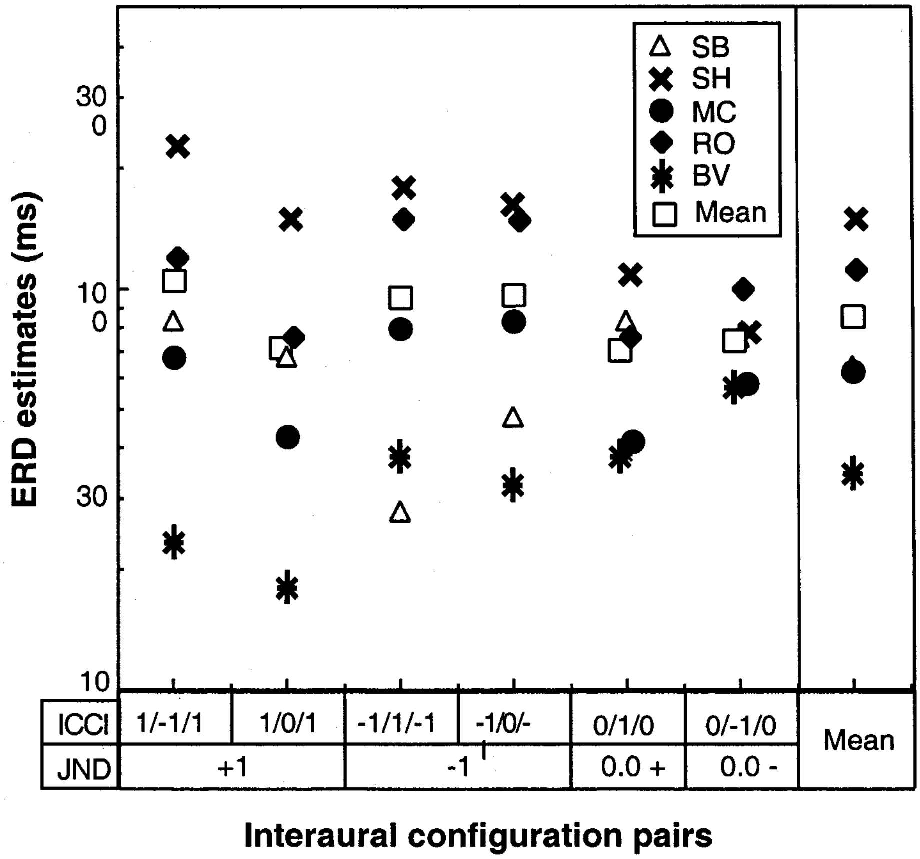


Figure 4.5

Binaural temporal window ERD estimations (ms) for individual subjects and the grand mean ERD for each interaural configuration (open squares). The ICCI-detection and JND conditions used to generate the window estimate are labeled on the x-axis. The "ρw" model was used to estimate the binaural window ERD required to explain the ICCI-detection threshold based on the subjects JND. While ERDs are variable across listeners, there is less variability across condition for a given listener.

listeners, obtained for different conditions [F(4,20)=1.4, p>0.26]. The grand mean ERD across listener and condition was 86 ms, with a standard deviation of 52.7 milliseconds.

EXPERIMENT 2: A RE-EVALUATION OF THE BMLD FROM NO, NPI OR NU

The binaural masking level difference (BMLD) is one of the most studied phenomena in hearing. It is also a common demonstration used for teaching. The values for most interaural conditions are well characterized and very reliable. The classic case, for which the BMLD is largest, is the 13-15 dB release from masking obtained when the signal is presented 180 degrees out of phase (Sπ) from diotic noise (So). The second largest BMLD, slightly lower at between 10-13 dB occurs for the NπSo condition. As outlined in Table 4.1, these values correspond with the temporal conditions +1/-1/+1 and –1/+1/-1 respectively. However, the appropriate comparison BMLD condition for the ICCI conditions of +1/0/+1 and –1/0/-1, are the rarely examined BMLD for uncorrelated noise- NuSo and NuSπ respectively. There are only a few reports for these values, and while most studies find BMLD around 4 dB and 3 dB for NuSo and NuSπ respectively, the difference has not always been significant (Robinson & Jeffress, 1963). In this experiment we obtained BMLDs for all four conditions in order to provide a 'within-lab' comparison for the ICCI and JND tasks.

METHODS

Subjects

Masking level differences were obtained from 4 relatively practised subjects, all of whom had previously participated in the ICCI and JND tasks of Experiment 1.

Stimuli and Procedure

The signal was a 500 Hz tone and the masker was a low-pass (cut-off = 4071Hz) random noise presented at 70 dB SPL. The initial signal level was about 60 dB and was clearly audible. The tone and noise could be set to different interaural parameters. The noise could be 100% correlated (No), 180 degrees (or π radians) out of phase (Nπ), or have an interaural correlation that approximated 0 (Nu: independent noise bursts presented to each ear). The signal was either presented diotically (same starting phase at each ear, 100% correlated) or interaurally 180 degrees out of phase (Sπ). This allowed for two diotic control conditions – NoSo and NπSπ; and 4 BMLD conditions – NoSπ, NπSo, NuSo and NuSπ. Stimuli were generated online in Matlab, with a new randomly generated noise used for each trial. Listeners were presented with two intervals – noise alone and noise plus signal – in a random order. Listeners indicated which interval contained the signal by pressing a key. The masked threshold level for detection of the signal was determined using a 2 down - 1 up adaptive procedure (see General Methods). Binaural masking level differences were determined by subtracting the masked threshold for the experimental condition from the masked threshold obtained for one of the control conditions: e.g. BMLD=NoSo – NoSπ; NπSπ-NπSo; NoSo-NuSo or NπSπ-NuSπ.

RESULTS AND DISCUSSION

Individual (n=4) and mean BMLDs for each of the four conditions tested are plotted in Figure 4.6. BMLDs obtained for the NoSπ and NπSo conditions ranged from 12- 16 dB and 7-11 dB respectively. This accords well with published data for the NoSπ condition, which fall between 13-15 dB (Moore, 1997). The NπSo condition, as predicted, resulted in a significantly lower BMLD of 9 dB. A planned comparison between NoSπ and NπSo thresholds (simple effects) indicated this difference was significant [$F(1,9)=65.5$, $p<0.001$]. The values of greater interest were those for NuSo and NuSπ, which resulted in mean values of 4 and 2 dB respectively. In 3 of the 4 subjects, the MLD was greater for the NuSo condition than the NuSπ condition. In the fourth subject (the author), the values were the same. Statistical comparison of these means indicated the difference was significant [$F(1,9)=14.8$, $p<0.02$]. This provides more evidence to support the existence of a true difference between these conditions. Recall from the introduction that the existing evidence showed a small (1 dB) difference in the same direction as that observed here, but across studies the difference was not always significant. Not only does the difference appear to be real, but also it is consistent with the idea that changes within positive-correlation space are more detectable compared with changes in negative-correlation space, consistent with the pattern of results in Experiment 1.

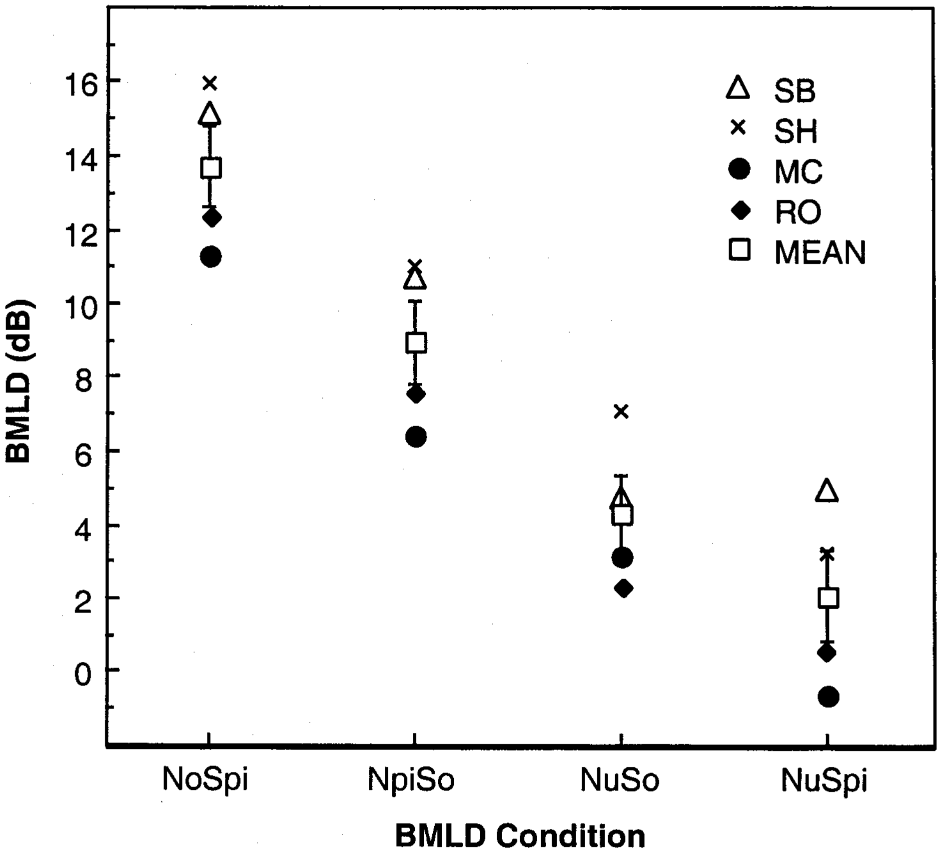


Figure 4.6
Binaural masking level differences (BMLD) for masking noises of either +1 (No), -1 (Nπ), or 0 (Nu) interaural correlation. The well-known asymmetry in binaural detection, such that negatively correlated signal in correlated noise is better detected than a correlated signal in negatively correlated noise, was replicated. This study sought to re-examine the size of the BMLD for a correlated or negativey correlated signal obtained in uncorrelated noise.

GENERAL DISCUSSION

The purpose of this chapter was to examine static (JND) and dynamic (ICCI - detection) interaural correlation sensitivity under a variety of interaural configurations using wideband noise. These were compared with results for the BMLD under similar configurations. Results for all three tasks followed the same dependence on interaural configuration – the lowest thresholds were for discrimination of a change in interaural correlation from +/-1, and the highest were for a change in correlation from 0. The results replicate those of other investigators for the positive range (Pollack & Trittipoe, 1959a, 1959b; Gabriel & Colburn, 1981; Koehnke *et al.*, 1986; Jain *et al.*, 1991; Culling *et al.*, 2001). However, they are the only data for the negative range of which we are aware other than an informal mention by Pollack and Trittipoe (1959a). The second major result is that interaural correlation sensitivity (dynamic or static) is more acute in the positive than in the negative range. Specifically, discrimination of a change in interaural correlation from a reference of +1 was significantly more acute than from -1, and discrimination of increases in correlation from a reference of 0 toward +1 were significantly more acute than a decrease towards -1. These results are consistent with those from the BMLD experiment. A signal is more detectable when the channel coding for the signal intensity changes within the positive range (NoSπ – signal causes decrease in interaural correlation of the correlated noise components near the signal frequency) compared with changes made in the negative range (NπSo – signal causes increase in interaural correlation of the negatively correlated noise components).

INTERAURAL CORRELATION SENSITIVITY FOR ρ_{REF} = +1

Interaural correlation sensitivity is most acute from a reference of +1, and this acuity is probably what underlies the maximal release from masking in the classic NoSπ BMLD condition (see Section D). The JNDs for a departure in correlation from a reference of +1 ($\Delta\rho = 0.045$) observed in this study were close to values previously reported for broadband noise (e.g. Pollock and Trittipoe, 1959a: $\Delta\rho = 0.04$; Gabriel and Colburn, 1981: $\Delta\rho = 0.03$; Koehnke et al, 1986: $\Delta\rho = 0.02$, Akeroyd and Summerfield, 1999: $\Delta\rho = 0.023$). Gabriel and Colburn (1981) also described a bandwidth dependence on the JND from a reference interaural correlation of +1, such that acuity degrades with increasing bandwidth. Consistent with this effect, our thresholds for wideband noise are slightly higher than those obtained by Gabriel and Colburn for their widest band of noise (4.5 kHz, $\Delta\rho = 0.02$-.03). Interestingly, bandwidth effects act in the opposite direction for interaural correlation JND from uncorrelated noise references (Gabriel & Colburn, 1981), i.e. performance improves with bandwidth.

The absolute values of thresholds in the 1/0/1 configuration were consistently lower (< 3ms in 7/8 listeners) than the 5 ms mean observed by Akeroyd and Summerfield (1999) for their widest stimulus bandwidth (100-500 Hz). This may indicate that performance remains stable, or continues to improve with increases in bandwidth in the ICCI-detection task, akin to bandwidth effects in monaural gap detection (e.g. Fitzgibbons, 1983). This is contrary to the bandwidth effect observed for the JND from +1, in which acuity tends to degrade with increasing bandwidth (Gabriel & Colburn, 1981).

In the dynamic task (ICCI-detection), performance improved with greater distance in correlation of the target noise from marker noise of +1 interaural correlation. A change to –1 (i.e. 1/-1/1) resulted in ~30% lower thresholds than a change to 0 (i.e. 1/0/1).

INTERAURAL CORRELATION SENSITIVITY FOR $\rho_{REF} = -1$

Sensitivity to interaural correlation changes from a reference of –1 were found to be relatively acute, but significantly worse than performance from a reference of +1. This is inconsistent with the study on interaural correlation sensitivity by Pollack and Trittipoe (1959a), in which they reported that “extensive informal tests” with negative reference correlations “yielded results that were indistinguishable from results with positive correlations” (pp. 1251). The reason for this discrepancy is not clear, though it may be related to the difference in methodology. Comparisons between the two studies are difficult because Pollack and Trittipoe did not describe their method. It may be that our use of a staircase procedure was more sensitive, which may be reflected in the lower mean thresholds we obtained in the positive range, at least from uncorrelated references. Similar to the results with reference correlation of +1, performance for the –1/1/-1 condition was better than for the –1/0/-1 condition. This benefit , with a doubling of the interaural correlation change from negatively-correlated markers (~50%), was significantly greater than that observed for correlated markers (~30%).

INTERAURAL CORRELATION SENSITIVITY FOR $\rho_{REF} = 0$

JNDs for an increase in correlation from $\rho_{ref} = 0$ ($\Delta\rho = 0.31$) are consistent with previous reports (range: $\Delta\rho$ =0.3 to 0.5, Pollack and Trittipoe, 1959a; Gabriel and Colburn, 1981; Koehnke et al., 1986; Jain et al, 1991; Culling et al, 2001). They match well with Gabriel and Colburn's value ($\Delta\rho$ =0.3) for their widest stimulus bandwidth (4.5 kHz), but are somewhat lower than that obtained by Pollack and Trittipoe (1959a) for wideband noise (0.44). A new result was that discrimination of an increase in correlation towards +1 ($\Delta\rho = 0.31$) was better than discrimination of a decrease towards –1 ($\Delta\rho$ = 0.45), a condition for which we have found no comparable data reported. This was consistent with the pattern of data we obtained for ICCI-detection, in which the duration required for detection of an increase in correlation from uncorrelated markers (0/1/0, 20.57 ms) was lower than that required for detection of a decrease in correlation (0/-1/0: 41.45 ms). Therefore, the correspondence in performance on the dynamic and static tasks holds for uncorrelated references, as observed for references of ρ= +/- 1.

Pollack and Trittipoe (1959a) made the reasonable suggestion that the poorer sensitivity at lower reference correlations was the result of the increased sampling variance for correlations near 0 (also see Gabriel and Colburn, 1981). That is, that performance would be degraded because of the increased trial-to-trial variation in correlation of the stimuli as the nominal correlation approaches 0. However, this explanation would predict equal performance in the positively- and negatively-correlated direction, as there is no difference in the sampling variance in the two configurations.

INTERAURAL CORRELATION SENSITIVITY AND MASKING LEVEL DIFFERENCES (BMLD)

Both dynamic and static measures of sensitivity to departures of correlation from a negatively-correlated reference were poorer than those observed from a positively correlated reference. The asymmetry in performance is consistent with the difference in binaural masking level difference (BMLD) (see Table 5.1, Exp. 2). The largest BMLD (~14 dB) occured for the NoSπ condition, which corresponds to the JND for ρ_{ref} = +1. In

The NπSo condition, which corresponds to the JND from $\rho_{ref} = -1$, provides less masking release (~9 dB). This asymmetry in correlation discrimination from positive and negative references is thus consistent with the prevailing theory that performance on BMLD and correlation discrimination tasks may be limited by the same mechanism (e.g. Durlach et al, 1986; Koehnke et al, 1986, Jain et al, 1991; Culling et al., 2001). Models which explain the difference between the NoSπ and NπSo BMLD conditions (and by extension the JND from ρ_{ref} =+1 and –1 respectively) traditionally require the postulation of coincidence detector neurons with a distribution of internal delays centered on zero interaural difference. In the NoSπ condition, the decorrelation of the noise by the antiphasic signal is assumed to be most salient in neurons with near-equal internal delays, the density of which is proposed to be maximal. Neurons most highly activated by the negatively-correlated noise in the NπSo condition have greater internal delay differences. There are fewer of these neurons available, which would make detection of the decorrelation (decrease in activity) in the NπSo condition worse than in the NoSπ condition (Jeffress *et al.*, 1956; Colburn, 1977). An alternative explanation is that the phase-delayed masker noise (Nπ) is not perfectly compensated, except at center frequency, by the optimal internal delay. Thus, the noise can not be cancelled entirely by a time delay, as proposed by, for example, the Equalization/Cancellation model of Durlach (1978). And in other models (e.g., Colburn, 1977), the decorrelation in the optimum ITD-channel effected by the signal must be detected in comparison to a representation of the phase-delayed masker noise that is already less than unity. All these explanations would similarly predict the observed difference in the correlation JND between references of +1 and –1. However, they assume that for a given frequency band there is an array of coincidence-detecting neurons with a wide range of different internal axonal delays that can compensate for large interaural time differences. The physiological validity of such a required distribution of internal delays has been called into question (McAlpine *et al.*, 2001; Brand *et al.*, 2002).

Given that the asymmetry in performance on the tasks using markers or references of +1 and –1 seemed to follow the patterning of data in BMLD conditions, we wondered whether the asymmetry from an uncorrelated reference might also follow BMLD patterns for uncorrelated noises. Reports usually show a small BMLD (~2-5 dB) for uncorrelated noise with either a correlated signal (NuSo) or a negatively-correlated signal (NuSπ) (Hirsh, 1948; Blodgett *et al.*, 1958; Robinson & Jeffress, 1963; Langford & Jeffress, 1964). However, reports of a *difference* in the BMLD for a negatively-correlated (Sπ) and correlated signal (So) in uncorrelated masking noise (Nu) have been conflicting. There is either no reliable difference, or the NuSo condition provides approximately 1 dB more masking release than the NuSπ condition (Blodgett *et al.*, 1958; Robinson & Jeffress, 1963; Langford & Jeffress, 1964). Current theory suggests that there may be no relation between the BMLD and correlation discrimination for conditions where the reference noise is uncorrelated (see Durlach *et al.*, 1986). In these BMLD conditions, monaural cues contribute to detection of the signal as the sound pressure level at each ear is significantly increased in the frequency band containing the signal. The observed BMLD in the Nu condition can be accounted for by having 2 independent chances to detect the signal monaurally in the uncorrelated noises at each ear (Langford & Jeffress, 1964; Durlach *et al.*, 1986; Jain *et al.*, 1991). The monaural account is further bolstered by the fact that NuSo thresholds show the same frequency dependence as monaural conditions (Durlach & Colburn, 1978). In contrast, the noise stimuli used in the correlation discrimination tasks do not contain any monaural cues (no change in level) and detection must be mediated solely by binaural processes. The asymmetry observed in the correlation JND predicts that the binaural cues would be stronger for the NuSo than the NuSπ conditions. The data presented here in Exp. 2 supports the presence of a real difference between these conditions. Although the difference was small (< 2 dB), it was evident in three of the four listeners.

EVIDENCE FOR AN ASYMMETRY IN POSITIVE AND NEGATIVE CORRELATION SPACE

In summary, there are a number of lines of evidence from this study which suggest the existence of an asymmetry in acuity for interaural correlation in the positive and negative range. First, while interaural correlation sensitivity was most acute from unity (positive or negative), sensitivity was significantly more acute from +1 compared with −1. Correlation sensitivity was generally poor from an uncorrelated reference. However, it was significantly poorer if the change is made towards more negative than towards more positive correlations. Finally, in the ICCI-detection task, there was a significantly greater benefit with the increasing size of correlation change, when that change was in the positive direction. Specifically, thresholds in the 1/-1/1 condition were ~30% more acute than in the 1/0/1 condition; while thresholds in the –1/1/-1 condition were ~50% more acute than in the –1/0/-1 condition. It is yet to be determined if this benefit for increases in correlation holds for reference correlations other than unity. The asymmetry in performance from references of +1 and –1 is consistent with the asymmetry observed NoSπ and NπSo BMLD conditions, thus providing further evidence that the BMLD might be explained in terms of interaural correlation discrimination.

The observed asymmetry might reflect an asymmetry in the internal neural representation of negatively- and positively-correlated noises. The internal representation of uncorrelated noise might not be half-way on some continuum between the representation of negatively-correlated noise and positively-correlated noise, but may be closer to the representation of negatively-correlated noise. That is, the gradient in internal representation between negatively-correlated and uncorrelated noise is smaller than the gradient between uncorrelated and positively-correlated noise.

THE REPRESENTATION OF INTERAURAL CORRELATION

An interaural correlation change is represented differently in different binaural coincidence detecting neurons, depending on their internal delay and type of binaural input. In the example of a change from correlated to uncorrelated noise (1/0/1), the activity in a binaural channel sensitive to correlated noise would briefly decrease. This decrease would represent the reduction in intensity of the positively correlated component of the uncorrelated noise (see Eq.1). In contrast, there would be a brief increase in activity in a channel that does not respond to correlated noise, but is strongly activated by negatively correlated noise. This increase would represent the intensity of just the negatively correlated component of the uncorrelated noise. Activity in these channels would change inversely when the interaural correlation is changed: they form an opponent channel pair. Thus, in contrast to a monaural intensity change for example, any correlation change regardless of its direction, results in both an increase and a decrease in activity as cues for detection. Although the absolute change in activity may be equal in both channels, the proportional change in activity would differ. That is, the percentage decrease in the activity of the channel maximally activated by the reference noise would be less than the percentage increase in activity in the channel minimally activated by the reference noise. This might suggest that the brief increase in activity in the minimally activated channel would be a more salient cue for detection of the interaural correlation change.

The neurophysiological basis for binaural channels selectively activated by positively or negatively-correlated stimuli might lie in binaurally-innervated neurons with small and large axonal delay differences between predominantly excitatory inputs from each ear, as postulated in the classical Jeffress model (Jeffress, 1948). A channel with equal axonal travel time from both ears would be most sensitive to correlated noise and far less responsive to negatively-correlated noise. However, if the difference in axonal travel time from the ears is equivalent to half the period of the characteristic frequency of the neuron, the response to correlated noise will be minimal and the neuron would be most sensitive to negatively-correlated noise. As an alternative to an implementation using axonal delays, these opponent channels might be comprised of the 'peak-type' and 'trough-type' neurons which are maximally sensitive to positive and negative interaural correlation of noise respectively (Fitzpatrick *et al.*, 2002; Shackleton *et al.*, 2002). ‘Peak-type’ neurons are coincidence detectors receiving predominantly excitatory, phase-locked inputs from each ear. ‘Trough-type’ neurons receive phase locked excitation from one ear and phase-locked inhibition from the other. For this reason, they are most responsive when the phase-locked inputs from each ear arrive out of phase, that is, when the signal is interaurally negatively-correlated.

BINAURAL TEMPORAL WINDOWS

Our mean binaural temporal window ERD (86 ms) is within the same general range as found previously (e.g. Grantham and Wightman, 1979: 44-243 ms; Culling and Summerfield, 1998: 97 ms; Akeroyd and Summerfield, 1999: 40-400 ms), and larger than the reported values for the analogous Gaussian monaural temporal window (e.g. Moore et al, 1988, 27 ms [corrected value, see Akeroyd and Summerfield, 1999]). Subjects differed in their mean temporal window ERDs. For example, the ERD of listener SH was always higher than that of BV. Subjects also differed in their *pattern of ERDs across* *configuration*, but not in any consistent way. There was no significant difference between the mean ERDs obtained for the different interaural configurations. The mean threshold reported here of 86 ms is significantly shorter than that reported by Akeroyd and Summerfield (1999) using the same methods and model (210ms). According to the "ρ_w" model, the reason for this could either be higher JND values or shorter ICCI-detection thresholds in our experiments The JNDs reported here for $\rho_{ref} = +1$ (with the exception of listener BKV), were not substantially larger than those reported by Akeroyd and Summerfield (1999) for their widest bandwidth (100-500 Hz). However, our ICCI-detection thresholds were far lower for (2.34 vs. 5.3 ms), despite similar methods. A major difference in both experiments is the bandwidth of the stimuli used. Akeroyd and Summerfield (1999) used exclusively low-pass filtered stimuli with a cut-off frequency of 500Hz. One possibility is that frequency components above 500 Hz contribute to ICCI-detection performance but not to static correlation discrimination. In fact, static correlation discrimination has a tendency to degrade with increased bandwidth (Gabriel and Colburn, 1981). The "ρ_w" model would then predict a dependency of the ERD on lower cutoff frequency. A detailed examination of the bandwidth and frequency dependence of the two tasks in the same listeners would be required to determine whether our lower temporal window ERDs were the result of the wider bandwidth used, or simply that our listeners fell in the lower end of the population ERD range.

The method employed here for estimating the binaural temporal window (after Akeroyd and Summerfield, 1999) provided an efficient way to compare the window ERD for the different binaural configurations tested. The method also has the advantage that it exclusively probes the binaural system. However, it does not provide for estimation of the window shape. Other methods, such as those which use BMLD-like stimuli distributed in time (i.e. No noise followed by a Sπ signal) have provided estimations of the shape of the window (e.g. Kollmeier & Gilkey, 1990; Culling & Summerfield, 1998; Holube *et al.*, 1998). These studies have revealed a symmetric double-sided exponential or rounded exponential weighting function with a duration between 40 and 200+ms depending on the listener and the task parameters (see Holube *et al.*, 1998). This ERD range is similar to that estimated in this study using the "ρ_w" model. More recently, Bernstein et al. (2001) estimated the binaural temporal window using a task similar in form to that used here; however, detection was of a change in lateralization by ITD of a brief correlated noise embedded in correlated or uncorrelated markers. Their data was best accounted for by a window with two time constants: a very short one that was heavily weighted at the temporal midpoint and a second one that was 14 ms. This might suggest that the binaural system exploits different temporal windows for tasks probing different aspects of binaural temporal processing (i.e. changes in ITD vs. changes in interaural correlation) (Kollmeier & Gilkey, 1990; Bernstein *et al.*, 2001).

LIMITATIONS AND FUTURE DIRECTIONS

Conclusions from this study are limited to broadband noise stimuli, and extreme references correlations (+/-1 and 0). It remains to be determined if the asymmetry that we observed can be observed for other reference values, although this might be a way to determine the temporal window shape. Future studies are also required to show whether the asymmetry holds for correlation discrimination at various noise bandwidths and center frequencies, and in fringed correlation discrimination (Jain et al, 1991). This latter configuration mimics the BMLD task for a narrow band signal. The change in correlation occurs only in a narrow frequency band fringed by noise of a reference correlation, but the energy in the frequency band used for detection remains constant. We expect that detection of a decorrelation in a band surrounded by a positively correlated fringe would be better than in a negatively-correlated fringe.

CONCLUSIONS

1. Static and dynamic measures of interaural correlation follow the same general pattern of dependence on interaural configuration.
2. There is an asymmetry in performance of both static and dynamic measures of interaural correlation sensitivity in the positive and the negative range in that correlation sensitivity is more acute in the positive range.
3. The mean binaural temporal window equivalent rectangular duration (ERD) was relatively constant across interaural configuration, although individual listeners ERD patterns varied in unique ways. Listeners also differed in their mean individual window length. The mean ERD was 86 ms.

CHAPTER 5

GENERAL DISCUSSION

This thesis began with the premise that one of the main functions of the auditory system was source segregation (de Cheveigné, 2001), and that spatial location was a possible cue by which segregation could be achieved. More accurately, within contraints of the auditory system architecture, the system segregates sounds when the task demands segregation, and integrates (groups) sounds under other listening conditions. The general purpose of the present work was to determine which of the many attributes of the proximal stimulus that change as a function of distal source location (outlined in Chapter 1) were important in generating source segregation. An outline of the various stimulus dimensions examined in the experiments presented in Chapters 2-4 is presented in Table 5.1.

The experimental tasks used in Chapters 2 and 3 exploited the observation that temporal judgements made between two sounds are impaired when those two sounds are perceived to emanate from different distal sources, i.e. they are segregated/occupy separate auditory streams (after Bregman, 1990). The two main stimulus paradigms used were auditory temporal gap detection and sequential stream segregation – related tasks which have rarely been examined in parallel. Both tasks are characterized by three elements: two sounds (A and B), which differed in various spatial stimulus attributes, and a silent interval (_) positioned between them. In gap detection, the ability to detect the silent period between the two sounds in the configuration A_B was determined. In the stream segregation experiments, a sequence was presented in the configuration A_B_A___A_B_A____...A_B_A. In one stream segregation task listeners rated the degree to which the stimulus was perceived as a single integrated stream, or two segregated A and B streams. In a second task, which biased the listener towards hearing the sequences as integrated, the relative durations of the temporal intervals between the A and B sounds was judged (temporal asymmetry detection task). In Chapters 2 and 3, the stimuli used were characterized by a silent interval between two sounds. In Chapter 4, an examination of the processing of changes in the purely binaural characteristics of a sound was examined (Table 5.1, lower). The relation of Chapter 4 to the studies on gap detection and stream segregation is somewhat indirect, so this general discussion will focus on integrating the results from Chapters 2 and 3.

Table 5.1

Thesis summary. Upper table presents a summary of results obtained from Chapters 2 (gap detection) and 3 (stream segregation) for congruent stimulus conditions. The break out table outlines the experiments in Chapter 4, which had their genesis in an attempt to understand the nature of temporal processing in the ITD channel.

HEADPHONE TASK	EAR	ILD	ITD	Mono LQ/QL	Diotic (within)
Gap Detection Chap.2, Exp.3	Impaired	Impaired	Not impaired	Impaired	Control
Stream Segregation Chap.3, Exp.1	80% (stable)	75% (stable)	25-30% (variable)	15-20% (variable)	0%
Temporal Asymmetry Detection Chap.3, Exp.2	Impaired ~23 ms	Impaired ~16 ms	Not Impaired ~10 ms	Mod. Impaired ~13 ms	Control ~10ms

FREEFIELD TASK	Wide Band 90/-90	Hi Pass (ILD only) 90/-90	Low Pass (ITD only) 90/-90	Ear Block (LQ or QL) 90/-90	Same Location 90/90
Gap Detection Chap.2,Exp.1,2,4,5	Impaired	Impaired	Generally Not impaired	Impaired	Control
Temporal Asymmetry Detection (ms) Chap.3, Exp.3	Variable	Variable	Variable	Variable	Variable

ITD

TASK	1/-1/1	1/0/1	-1/1/-1	-1/0/-1	0/1/0	0/-1/0
I-Corr Change Interval (ICCI) Chap.4, Exp.1	1	3	2	4	5	6
I-Corr Discrimination (JND) Chap.4, Exp.2	1	x	2	X	3	4
BMLD (dB) Chap.4, Exp.3	NoSπ 1	NoSu x	NπSo 2	NπSu X	NuSo 3	NuSπ 4

*performance is ordered from 1-6 (or 1-4)

In the general introduction, some goals were outlined for the thesis:

1) The first goal was to determine the relative contribution of various spatial stimulus attributes to the sensitivity of gap detection to spatial separation of wideband gap markers described by Boehnke and Phillips (1999).

As will be discussed in more detail below, the dominant effect of spatial separation on gap detection was well accounted-for by marker spectral and level differences at each ear. No independent effect of binaural interaction could be demonstrated using these methods, evidenced by the failure of large ITD differences (in the free-field with large separations of low-pass noise markers, or over headphones) to affect gap detection thresholds. In general, the pattern of results indicated that gap detection was most impaired for spatial separation conditions which led the markers to selectively activate different ears, and thus different peripheral channels (separation of markers to different ears, free-field separation of high frequency noise markers to opposite hemifields).

2) The second goal was to determine if the pattern of sensitivity of gap detection to spatial stimulus attributes was related to the degree to which those attributes support the perception of segregated streams in sequential stream segregation tasks. Specifically, it was hypothesized that those attributes leading to elevated gap thresholds would be those leading to the most segregated percepts. This hypothesis was generally supported. In the rating experiment, the greatest segregation was reported for A and B sounds presented to different ears or with opposing ILDs, while the least segregation was for markers differing in ITD. In the temporal asymmetry experiment, the pattern of threshold duration to detect a temporal asymmetry precisely mirrored the pattern observed in gap detection experiments.

FACTORS UNDERLYING SENSITIVITY OF GAP DETECTION THRESHOLDS TO MARKER SPATIAL SEPARATION

A sensitivity of gap detection thresholds to spatial separation of wideband noise markers was demonstrated previously (Phillips *et al.*, 1998; Boehnke & Phillips, 1999). This sensitivity was such that spatial separation of gap markers within an acoustic hemifield minimally affected gap detection thresholds, while spatial separation across the midline led to increases in thresholds that reached asympotically high levels when markers were in opposite hemifields at any position in the range from 30-150°. This azimuthal dependence on spatial separation effects, as opposed to an absolute dependence on spatial separation, was an important result. It demonstrated that stimulus locations within a hemifield effectively activated similar neural representations. The pattern of dependence was argued to reflect the existence of two hemifield-tuned, partially overlapping spatial channels. This channel tuning was consistent with animal behavioural-lesion studies (and some human neuropsychological studies) in which auditory cortex was ablated, which showed similarly-tuned contralateral deficits in sound localization acuity (see Chapter 1). Furthermore, auditory cortical neurons tend to have broad spatial tuning for the contralateral hemfield. This convergence of evidence suggested that the spatial channels of interest had their genesis in these binaurally-sensitive, contralaterally-tuned neurons in the auditory cortex.

This argument stemmed from a theoretical framework which saw gap detection as a tool by which to evaluate central stimulus representations (Phillips *et al.*, 1997; Phillips *et al.*, 1998; Phillips, 1999; Taylor *et al.*, 1999). The basic idea was that there were two processes in gap detection. When markers were similar, gap thresholds were determined by a discontinuity detection mechanism and were low. When the markers differed perceptually, gap thresholds were determined by a relative timing operation based on the offset of stimulation in one channel (subserving the perception of the leading marker) and the onset of stimulation in another (subserving perception of the trailing marker). Thresholds were elevated under such conditions because that mechanism had inherently poorer acuity. The evidence for such a mechanism stemmed from the fact that gap detection thresholds were similarly elevated for markers of disparate frequency content, markers presented to different ears, and markers presented from different spatial locations.

Attribution of the elevation of thresholds with marker spatial separation to a relative timing mechanism was contentious because stimulus differences which affect acuity of the monaural discontinuity detection mechanism are present at each ear in the spatial case. In addition to differences in ITDs and ILDs (binaural interaction cues), spatial separation results in marker level differences at each ear, a ‘within-channel’ effect which is known to impair gap acuity (Plomp, 1964; Penner, 1977; Oxenham, 2000). In Chapter 2, the hypothesis that spatial separation effects on gap detection were dependent on binaural interaction mechansims was evaluated, and could not be supported. In the first experiment, spatial separation sensitivity was determined under monaural conditions. The pattern of dependence of gap detection thresholds on trailing marker location for leading markers in the hemifield of the open ear were similar under both monaural and binaural conditions, suggesting much of the pattern of thresholds reported by Boehnke and Phillips (1999) could be accounted for by the effect of monaural marker level differences. An analysis was undertaken to determine whether the pattern of thresholds obtained under binaural listening might be accounted for by a weighted averaging of the thresholds obtained with either ear alone (an attempt to salvage the binaural hypothesis). Using the data set acquired in Experiment 1, such a model was successful. This analysis suggested that while some limits on gap detection acuity were likely monaural/peripheral in nature, in the free-field condition, where acuity differed at each ear depending on the condition, the system makes used of a weighted average of the acuity available at each ear. However, empirical results from Experiments 4 and 5 demonstrated that the success of this model was made possible by a third variable – the short duration of the leading marker used in Experiment 1. The model could not be generalized to longer marker durations.

One might hypothesize that the similar sensitivity of gap detection to spatial separation observed under binaural and monaural conditions was nonetheless determined by different mechanisms – a between-channel mechanism in the binaural case and poor within-channel acuity in the monaural case. This was discounted in Experiment 2. The sensitivity to spatial separation of gap markers under binaural listening was shown to be minimal when low-pass noise markers were used, which provide binaural (ITD) cues to source location but generate minimal level differences at each ear with separation. In contrast, a strong effect of spatial separation on thresholds was found when high-pass noise markers were used, which generate large monaural level differences with spatial separation. This general pattern was replicated with controlled stimuli over headphones using markers differing in ear of entry, ITD, or ILD (Exp. 3b). Marker differences in ITD alone were inconsequential to gap detection acuity (akin to the low-pass markers in the freefield). Marker differences in ILD (which result in monaural level differences at each ear, akin to high-pass noise markers in the freefield) impaired gap detection acuity equally under binaural and monaural listening. By far, the condition leading to the largest increase in gap detection thresholds was when markers were presented to different ears (complete peripheral segregation).

Further experiments were conducted to identify the interactions of marker spectral content, leading marker duration and available ear in the effect of spatial separation on gap detection. Although the results were complex, in general it can be concluded that marker level and spectral differences, rather than binaural interaction cues, dominated the effect of spatial separation. In conditions where the markers are in different hemifields the stimulus may effectively become a between-ear task. Thus, the pattern of gap thresholds as a function of separation of gap markers in azimuth described by Boehnke and Phillips (1999) probably reflects the similarity in the *spectral profile* of the proximal stimulus (that reaching each cochlea) from any two marker locations. Which ear dominates the contribution to gap thresholds for a given spatial separation condition was not entirely clear from Experiment 5. Under some conditions, binaural thresholds appeared to be determined by the ear near the leading marker, and under others configurations the far ear seemed to influence results. Such effects might be better evaluated and understood by developing a model of gap detection mechanisms that examines the cochlear output of both ears and uses the representation of the gap in all channels in some statistical decision process.

SEQUENTIAL STREAM SEGREGATION BY SPATIAL LOCATION ATTRIBUTES

The experiments in Chapter 3 provide the first direct comparison of different spatial stimulus attributes in a sequential stream segregation task in the A_B_A___ format. It also provides important data regarding the necessity of peripheral channelling cues for stream segregation. The main result is that presentation of A and B to different ears, or with different ILDs, promotes a rapid build-up to a segregated percept, which remains relatively stable through the duration of the sequence (see Table 5.1). The most important finding was that differences in ITD, which provide no peripheral channeling cues but differently lateralized percepts, *can* support stream segregation albeit relatively weakly. This suggests that differences in peripheral channelling of A and B noises is *not* a necessary requirement for stream segregation to occur, as has been suggested by advocates of the peripheral channelling hypothesis (Beauvois & Meddis, 1991; Hartmann & Johnson, 1991; McCabe & Denham, 1997; Vliegen & Oxenham, 1999). However, segregation by ITD alone took longer to build and was less stable than that obtained with EAR or ILD differences. Experiment 2 of this study also presents new data on the ability to make temporal judgements between sounds characterized by different spatial stimulus attributes (temporal asymmetry detection). The pattern of results from this study parallel the pattern of results observed with gap detection using the same spatial stimulus attributes (EAR > ILD = "mono level differences" > ITD = DIOTIC). This suggests that peripheral channeling might dictate the speed with which a segregated percept can emerge, but is not the limiting factor on which stimulus differences can result in stream segregation given some exposure time to the sequence. This is consistent with recent results from Robertson and Moore (2002) which demonstrate that stream segregation occurs with A and B with tone complexes whose components have random or coherent starting phases (resulting in different timbres), but otherwise activate identical peripheral channels. The weakness of ITD as a perceptual grouping cue has been shown previously in other paradigms (Culling & Summerfield, 1995; Hukin & Darwin, 1995).

Temporal asymmetry thresholds appeared to be related to the speed with which segregation occurs. In that sense, thresholds appeared to track the temporal coherence (integration) boundary – that point in stimulus space beyond which segregation is obligatory. For the EAR and ILD conditions, segregation occurred very quickly, while for the ITD condition, segregation sometimes did not occur at all (remained as a galloping rhythm); if separate A and B streams were formed, it usually took some time to build-up to that segregated state. The results for the temporal asymmetry task, which arguably forced the listeners to maintain an integrated state in order to do the task, followed the pattern of results for gap detection using the same stimulus attributes.
The pattern of temporal asymmetry thresholds obtained for free-field spatial separations varied dramatically across subjects in contrast to the rather stereotypical pattern obtained for the more simple headphone conditions. It is possible that spectral and timbre differences between sounds from different speakers made the task difficult for some

listeners, or that some listeners weighted different cues more strongly than others. For that reason, no absolutely definitive conclusions could be drawn about the free-field stream segregation case. Futher study of free-field location effects on stream segregation would be required to delineate the factors contributing to performance.

The present results also have possible implications for experimental psychological studies into cognitive processes such as attention and memory. Studies examining attention or cross-modal processing that use auditory spatial location as a variable might require re-interpretation. Different effects might be found depending on the cue used to indicate auditory 'location', and the frequency of the sound which acted as the stimulus. For example, ear of entry is often used to specify left and right auditory locations in such studies. Given the results reported here, it might not be surprising if ear of entry, but not actual free-field spatial location, would lead to an effect (or vice-versa).

DIMENSIONAL RELATIONS BETWEEN MASKED SIGNAL DETECTION, GAP DETECTION, AND STREAM SEGREGATION

Probably the most interesting result to emerge from Chapter 2 was a striking correspondence in the pattern of effect of spatial separation on gap detection and spatial separation on the release of a signal from a masking noise (Saberi et al, 1991). Spatial separation of a signal (a click train) from a simultaneous noise masker improves the signal detection threshold – the signal is released from the masker and can be heard at a lower intensity. This *improvement* in signal detection with spatial separation was congruent with the *impairment* in gap detection with spatial separation. For wideband stimuli, this correspondence held for a variety of leading marker/masker locations (see Fig. 2.7). Even more compelling, the same correspondence was found for low-pass and high-pass sounds. Gilkey and Good (1995) examined spatial release from masking using the same low and high passbands, and similar spatial separations, as were used in Experiment 2 of this thesis. Spatial separation of low-pass gap markers minimally elevated thresholds, and similarly, separation of a low-pass signal from a low-pass masker resulted in only a minor improvement in signal detection. Spatial separation of high-pass gap markers maximally elevated gap detection threhsolds, and similarly, separation of high-pass signal from a high-pass marker resulted in maximal improvement in signal detection. Finally, both Saberi (1991) and Gilkey and Good (1995) reported that a similar patterns of results were obtained under monaural listening with the ear with the best signal to noise ratio (SNR). This is consistent with the fact that a similar pattern of gap detection with spatial separation under monaural listening with the ear near the leading marker was observed in this thesis (Chap.2, Exp. 1).

What might this correspondence tell us about gap detection and signal detection mechanisms? Spatial separation allows for segregation of sound sources probably not because the sounds activate different binaural, spatially-tuned 'channels' coding for the contralateral hemifield (after Boehnke and Phillips, 1999), but because each sound is veridically coded in the ear which is shadowed from the other sound source (at least in the high frequencies). The relationship between masking and gap detection is confusing at first glance. Gap detection thresholds are best under conditions for which signal detection is worst. This is easily reconciled if one considers that gap detection performance is determined by the degree to which the sound markers equally activate channels; while signal detection performance depends on the degree to which the signal and masker differentially activate channels. The same general dependence of stimulus variables is found in the two tasks. Separation in frequency improves signal detection and degrades gap detection. Increasing the level of the signal improves signal detection, increasing the level of one of the gap markers degrades gap detection. The tasks measure two different aspects of perception, but are limited by the same basic processing mechanisms that originate in the cochlea. There is one condition that does not correspond: imposing different ITDs on the signal and masker improves detection (BMLD), while giving gap markers different ITDs has no effect on gap detection. This may simply reflect the differences in task strategy – in signal detection the goal is to use any stimulus difference available that provides for detection of the signal; while in gap detection, the goal is to ignore any stimulus differences in order to detect the gap in the available channels responding to both markers.

The relationship between masking and gap detection is more obvious when a temporal masking task (forward masking) is considered. In this case, the masker precedes a signal in time. For similar stimulus configurations, simultaneous masking, temporal masking, and gap detection have been shown to have the same dependence on stimulus parameters in a within-channel task (Smiarowski & Carhart, 1975). Another within-channel similarity between gap detection and signal detection is the existence of the overshoot effect in both tasks (Crawley, 1999). This is the effect in which detection of a signal or a gap degrades when the temporal position of the event (gap or signal) is placed close to the onset or offset of the ongoing noise. Analogous experiments have not been completed in parallel which compare signal detection (under simultaneous or forward masking conditions) with gap detection using stimulus configurations in which the markers differ in some dimension. However, such variables have been examined independently, though rarely have they been related to each other. In Table 5.2, the dimensional relationships between signal detection, gap detection and stream segregation are presented, all tasks which are characterized by the interaction between 2 sounds (signal and masker; leading and trailing marker; A&B sounds). In general, differences between the two sounds on a given perceptual dimension (i.e., ear, frequency, pitch, location), result in an inverse patterns of results between tasks encouraging segregation of the two sounds (signal detection, stream segregation) and tasks encouraging integration of the sounds (temporal resolution tasks). For example, it is well known that separation of a signal and masker in frequency systematically improves signal detection until the separation exceeds a critical bandwidth (the definition of a critical band actually). Similarly, separation of gap markers in frequency degrades gap detection until a critical separation, after which thresholds become asymptotic in many listeners (Formby & Forrest, 1991; Formby *et al.*, 1996a; Forrest & Formby, 1996; Heinz *et al.*, 1996; Phillips *et al.*, 1997; Formby *et al.*, 1998a). Separation of a signal and masker to different ears

Table 5.2 Dimensional relations between masking, gap detection and stream segregation type tasks, where the two sounds comprising each task differ in ear of entry or ILD, spectral range, periodic pitch, or ITD. Tasks are split into those which encourage segregation of the two sounds, and those which encourage grouping of the two sounds.

Tasks Encouraging Segregation ***-weight channel differences higher***	**Ear / ILD**	**Spectral**	**Pitch**	**ITD**
Simultaneous Masking	Release	Release	*CMR?	BMLD
Temporal Masking	Release	Release	*CMR?	BMLD
Stream Segregation: Rate: try to acheive segregation	Easy	Easy (beyond segregation boundary)	Possible, but requires time	Possible, but requires time
Stream Segregation: Task: interleaved melody recognition	Easy	Benefit increases with spectral distance	Some benefit	Some benefit
Tasks Encouraging Integration ***-weight channel similarity higher***				
Gap Detection Gap Duration Discrimination (with relatively short standard gaps)	Difficult	Difficulty increases with spectral separation	Some impairment	No impairment
Stream Segregation: Rate: try to maintain galloping rhythm	Difficult	Impossible beyond temporal coherence boundary	Difficulty increases over duration of sequence	Difficulty increases over duration of sequence
Stream Segregation: Task: temporal asymmetry detection	Difficult	Difficulty increases with spectral separation	Some impairment	No impairment

*CMR= comodulation masking release, a phenomena by which the comodulation of a signal with a spectrally distant flanking masker causes it to be released from masking by a noise with signal centre-frequency (Moore, 1997).

greatly improves signal detection, and similarly, gap detection thresholds are increased to about the same duration as obtained with large frequency separations (Phillips *et al.*, 1997; Formby *et al.*, 1998a). Signal detection may be improved if the signal and masker are given different modulation envelopes (with different correlated pitches) although they activate the same cochlear channels. However, this improvement only occurs if the signal modulation is correlated with that of a spectrally distant flanking masker. This effect occurs under both simultaneous (see Moore, 1997 for a review) or forward masking configurations (McFadden & Wright, 1987).Within-channel gap detection can also be affected under similar conditions, i.e., when the gap markers are co-modulated with spectrally distant flanking maskers which have no gap. However, there is no evidence that differential modulation of a signal and masker alone (so that they have different centrally-generated pitches) improves signal detection. Similarly, making spectrally similar markers dissimilar in centrally generated pitch by imposing different amplitude modulation rates, or using iterated rippled noise markers of different delays, does not appreciably impair gap detection (Grose *et al.*, 2001b). Gap thresholds are impaired as the markers are tonal complexes which differ as a function of fundamental frequency, but are filtered in the same spectral region, although far less than for markers filtered at different spectral regions (Oxenham, 2000). The main condition for which signal detection and gap detection diverge is when the signal and masker or the gap markers are given different interaural parameters. Significant improvement in signal detection occurs through the binaural system when the signal and masker are given different interaural properties (BMLD, see Chapter 4). As we have demonstrated here, there is no effect of marker differences in ITD on gap detection thresholds (see also Oxenham, 2000).

Do these dimensional relations between signal detection and gap detection hold when the task is explicitly one of stream segregation? Certainly, segregation increases as a function of the frequency difference between A and B, and if A and B are presented to different ears (Bregman, 1990). Segregation in these conditions is nearly immediate. Stimuli evoking pitch differences in the absence of spectral differences can be percieved as segregated, but maintaining segregation is more difficult and may take longer to build up (Vliegen & Oxenham, 1999), similar to the weaker effect of pitch differences between gap markers in elevating thresholds (Oxenham, 2000). Similarly, in this thesis it was demonstrated that stream segregation is relatively weak for sounds differing only in ITD. Note, however, that this is dependent on whether the task performance encourges segregation or integration. In tasks encouraging segregation, such as interleaved melody recognition, or masked signal detection, ITD differences can be used to some extent (Hartmann & Johnson, 1991). However, in tasks encouraging integration, these differences can be ignored to maximize task performance (gap detection, temporal asymmetry detection, maintenance of a galloping rhythm).

EXPLAINING MASKING, GAP DETECTION, AND STREAM SEGREGATION – A COMMON MULTICHANNEL MODEL OF AUDITORY PROCESSING (INCLUDING TEMPORAL INTEGRATION) WITH DIFFERENT DECISION VARIABLES

What is the mechanism by which these tasks are related? Data from forward masking experiments can be well explained by models which include a physiologically valid model of cochlear filtering and compression with increasing sound level (the compressive non-linearity), which is followed by a module that models either persistence (temporal window model, see Plack and Oxenham, 2002) or adaptation (e.g. the Seneff model described in Chapter 2; Oxenham, 2001). Plack and Oxenham (2002) have argued that the validity of the cochlear front end of such models is the major determining factor in their success. It would have been ideal to compare the ability of such models to account for the variety of effects on gap detection described here. While this was beyond the scope of the present thesis, it would be a potentially fruitful avenue for future study.

A number of models have been proposed to account for gap detection acuity with spectrally and intensively similar markers (what has been referred to as 'within-channel' gap detection). A common one is that of an energy detector (Plomp, 1964; Buus & Florentine, 1985; Florentine *et al.*, 1999). Within a given channel with a physiologically valid bandwidth, the gap is detected if the short term energy at the output of a short-term integrator over the period of the gap is statistically lower than that during an uninterrupted signal (Florentine *et al.*, 1999). Another form of such a model is the temporal window model (Moore *et al.*, 1988; Plack & Moore, 1991; Oxenham, 1997; Florentine *et al.*, 1999) in which the gap is detected by a change in 'loudness' rather than energy. The gap is detected when the decrement in level imposed by the gap exceeds the difference limen for intensity discrimination. Such a model assumes that there is a relationship between the gap threshold and the intensity difference limen, i.e. the gap is detected when the effective change in stimus intensity after temporal integration exceeds the difference limen for intensity. The temporal window model argues that the width of the temporal integrating window defines temporal resolution, not the absolute value of the gap threshold itself [Note that this is the strategy used to define the binaural temporal window in Chapter 4, using binaural 'gap' detection thresholds and the difference limen for interaural correlation]. In this latest version of the temporal window model (Plack *et al.*, 2002), the stimulus is passed through a physiologically realistic bank of auditory filters, after which it is subjected to a compressive non-linearity. The output of each filter is passed through a linear temporal integrating window with an equivalent rectangular duration of about 8-10ms (defined by Moore *et al.*, 1988), and the output is analyzed by a decision variable. Inclusion of the features which simulate cochlear processing has been demonstrated to be important in the accurate modelling of psychophysical observations from forward masking (Plack *et al.*, 2002), and gap detection (Florentine *et al.*, 1999). This model accounts well for gap detection acuity in within-channel conditions across marker centre-frequency and overall level. The temporal window model would probably also account for thresholds obtained with marker level differences, as it has been successful in accounting for decrement detection data where the markers differ in level (Plack & Moore, 1991). However, in order to account for gap detection sensitivity to marker duration, Schneider and Hamstra (1999) have suggested that a temporal window type of model must also include the effects of adaptation.

A temporal window model (with or without adaptation) has not been used to account for gap detection with dissimilar markers, although a multi-channel version (Heinz *et al.*, 1996) of the energy detector model (Forrest & Formby, 1996) has been fairly successful in accounting for the diminishing gap acuity as the markers are made dissimilar in frequency. In this model, the output from each channel driven by the gap stimulus is separately analyzed and provides a vote on the gap's presence, based on the change in energy at the time of the gap relative to random fluctuations present in an uninterrupted stimulus (Heinz, 1996). In order for such a model to account for the effects of spatial separation of gap markers on gap detection, it would need to be expanded to include both ears. A schematic for such a model is presented in Fig. 5.1. Multi-channel outputs from each cochlea are passed through a temporal integrating window. Some of these inputs converge, and code for ITD. The pattern of intensity change in all these channels can be evaluated by a decision variable. The decision process selectively weights channels depending on the goal of the task (integration or segregation). It is hypothesized that such a model might account for the data on spatial separation of gap markers without the need to postulate effects of low-level binaural interaction (i.e., those coding for ITDs and ILDs; see discussion Chapter 2). The effects of frequency separation, and presumably level differences, on gap detection would be attributed to the existence of fewer channels reporting the presence of the gap until larger gap durations. In such a model, gap detection for markers presented to different ears, or disparate frequency regions of the cochlea would have to be accomplished by a different mechanism.

There is no definitive explanation for the similarity of gap threshold obtained when markers are presented to different ears, or activate disparate frequency channels (Phillips *et al.*, 1997; Formby *et al.*, 1998a). In such conditions, there is always a discontinuity present in both the signal and the standard in the forced choice task. One explanation is that the task is accomplished by an inherently central, relative timing operation between the offset of the neural population responding to the leading marker, and the onset of the population responding to the trailing marker (Fitzgibbons et al., 1974; Phillips et al, 1997). Such an interval timing type of mechanism might also be used in a task like gap duration discrimination, or for gap detection between a visual flash and an auditory burst (Collyer, 1974). However, it is unlikely that there is a single relative timing operator with a set acuity, as acuity appears to be dependent on perceptual similarity. Thresholds for markers presented to different ears or disparate frequencies were lower than those obtained for markers differing on two dimensions, e.g. left-ear, low-frequency vs right-ear high-frequency (Taylor *et al.*, 1999). Thresholds for audiovisual markers (light & tone) are also worse than those for markers of disparate frequencies (5.2 kHz & 2.0 kHz) (Collyer, 1974).

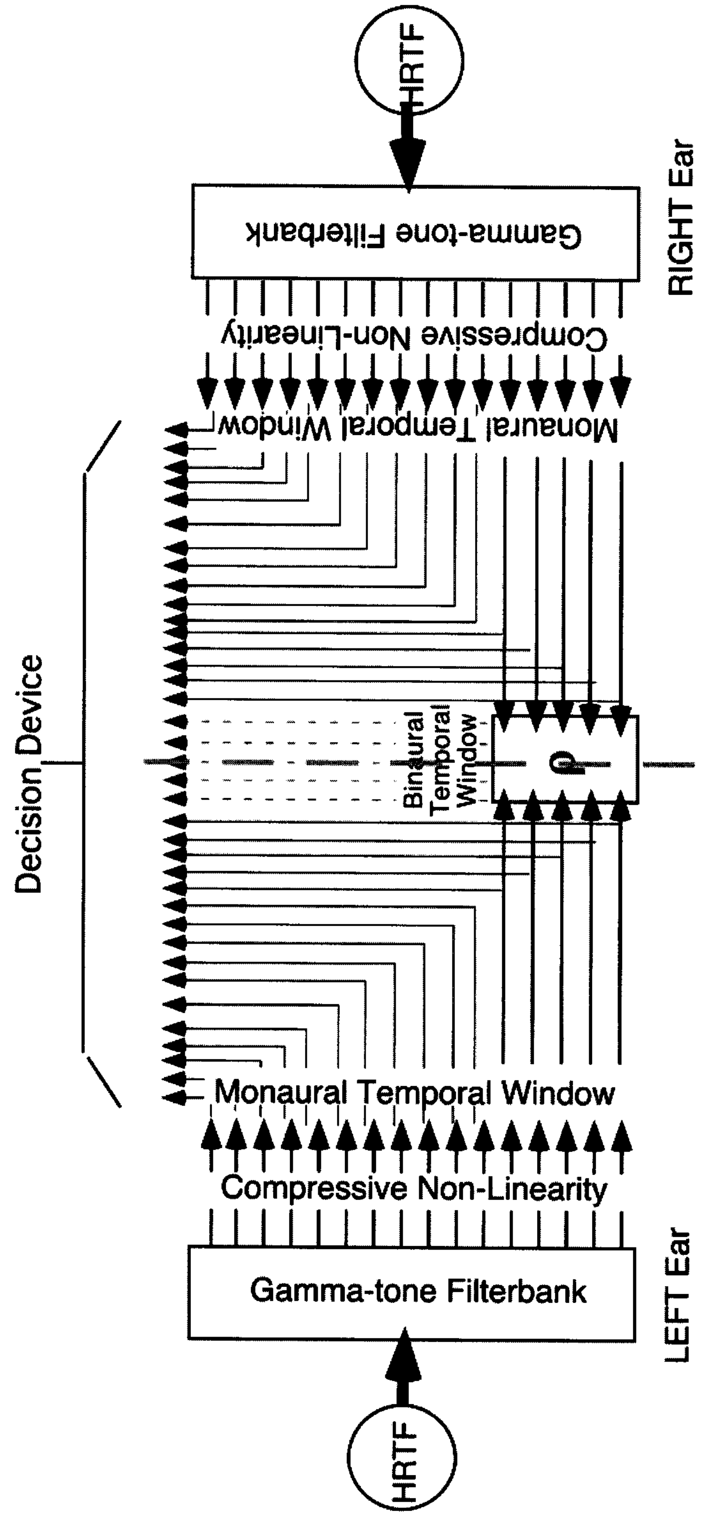


Figure 5.1
A possible model architecture for evaluating various temporal tasks (masking, gap detection, stream segregation). The decision device would differ depending on whether the task biased the listener towards integration or segregation.

As has been described, stimulus differences lead to release from masking and increases in gap detection (and gap duration discrimination) thresholds. As we will show here, the pattern of gap thresholds obtained with stimulus differences mimics the speed by which those stimulus differences lead to segregated percepts. Those resulting in immediate (primitive) stream segregation will result in the most elevated gap thresholds; those which require time to build-up to segregation will not elevate gap thresholds. In this sense, gap detection thresholds appear to track the temporal coherence boundary – that point in stimulus space beyond which segregation in mandatory.

Given the similarity in pattern of effect across stimulus dimensions, it is possible that performance on the tasks described here might be predicted based on evaluation of the output of a physiologically-valid, multichannel model of cochlear filtering, transduction and temporal integration. The task differences might be attributed to task-dependent decision variables that evaluate that output. Tasks for which performance biases the listener to try to segregate sounds (signal detection in noise, sequential stream segregation) might be performed by giving more weight to channels with greater output differences to the two sounds, while tasks which bias the listener to integrate across the sounds (gap detection, temporal asymmetry detection, perception of a galloping rhythm in stream segregation) might be performed by giving more weight to channels responding to both stimuli.

Depending on the duration of the gap (and of course on the spatial separation of wideband markers) there is a host of cues available in the neural representation of the stimulus that may contribute to detection. One is the presence of the leading marker at all (backward masking effects), the fidelity of the onset of the trailing marker (forward masking effects), binaural interaction information, and so on. Because the representations subserving such cues are possibly separable, it is possible that the listener (decision process) could use any one, or conjunction, of these cues to execute the gap detection task. That is, some listeners may, with practice be better able to access ‘low-level’ within-channel cues for detection; others may have difficulty accessing such cues and rely on a sort of interval-timing mechanism (the ‘between-channel’ relative timing process). This selective access and use of different representations (cues) may be what differentiates listeners with different patterns of dependence of gap threshold on stimulus parameters (e.g. leading marker duration). Practice effects may represent listeners learning to use one or the other representation efficiently, or may reflect a shift in which strategy is used. Thus, in the hypothetical model architecture shown in Fig. 5.1, listeners may have a similar underlying representation of the cues to the gaps presence, but would differ in the decision variable – i.e. in how the information available in different channels is used.

While such models do not 'place' the limit for the task at a specific processing level of the auditory system, presumably the biasing performed by the 'decision variable' would likely have neural modulatory effects at cortical areas. One very recent example (an unpublished abstract) of how this might occur comes from single-unit recordings of auditory cortical neurons in awake primates to stream segregation stimuli (ABA_ tone sequences) where A and B differ in frequency (Micheyl *et al.*, 2002). Recall, that early in the sequence the listener may perceive the sequence as integrated, while later in the sequence the perception may become segregated (build-up to segregation). The neural correlate of this was observed in the systematic decrease in the response to the 'B' tone, and enhancement in the response to the 'A' tone, over the duration of the sequence, in a neuron with a spectral receptive field centred on the 'A' tone. It is as if the sequence is perceived as integrated as long as neurons are responding to both 'A' and 'B', and becomes segregated when populations of neurons begin to selectively respond to either tone. Whether this biasing towards segregation over the duration of the sequence simply reflects the build-up of adaptive processes from earlier stages of auditory processes (i.e. the effect would occur in anaesthetized animals) or is the result of active descending processes (perhaps reflecting attention) which sharpen frequency tuning in order to aid segregation remains to be determined. It is this type of research, akin to that which has been ongoing in studies of visual attention, that will help unravel the basis of cognitive effects on auditory processing.

THE STRUCTURE OF AUDITORY SPACE AND THE COCKTAIL PARTY SCENARIO: HOW DO WE USE THE RELATIVE LOCATIONS OF SOUNDS TO SEGREGATE THEM?

This thesis began with the question of which spatial stimulus attributes are relevant to segregating of spectrally similar sounds? The experiments revealed that the classic binaural cues used in sound localization are relatively weak segregation cues. In the real world (outside of headphones), spatial separation is effective in separating sounds sources. This effectiveness comes largely from the differential activation of peripheral auditory channels that ensues when sounds come from different directions. Having two ears which are mutually shadowed by the head is advantageous, because when sounds are separated in azimuth, so that one ear will always have a more favourable signal-noise ratio. When two sound sources are separated across the listener's midline, the two sources will have optimal SNR at different ears, in turn optimizing the opportunity for selective scrutiny of either one of them. In general, the results are consistent with the conjecture that speech intelligibility of one voice in a cocktail party improves with spatial separation from other sources because the voice of interest is better coded (less masked) by the ear which is shadowed from the masking stimulus (Zurek, 1993; Phillips *et al.*, submitted). Given this, it is probably not surprising that those suffering from high-frequency hearing loss, such as the elderly, report particular problems with speech intellibility in multi-source environments (Pichora-Fuller *et al.*, 1995; Schneider *et al.*, 2000; Schneider *et al.*, 2002). Lower frequency components, which they can hear well, are less subject to the headshadow and thus benefit less from the improved signal to noise ratio at a given ear that occurs with spatial separation of sounds sources.

In a recent study, Hawley *et al.* (1999) presented speech intelligibility data for a target sentence located at the midline or the lateral pole in the presence of up to 3 competing sentences dispersed through azimuth from -90 to +90°. Much of their data could be well accounted for by the improved S/N at the near-signal ear. When targets and competitors were in the same hemifield (beyond 30 degrees) intelligibility was impaired, but if targets and competitors were located in different hemifields, their intelligibility was unimpaired. An interesting case that illustrates that separation distance is not necessarily the critical factor was that of a target located at the midline. Intelligibility was impaired if there were competitors at 90° in each hemifield, but was less impaired if two competitors were located at 60 and 90 degrees in the same hemifield. Peissig and Kollmeier (1997) reported a similar result. A midline located target can be processed through the ear that was contralateral, and thus shadowed by the interfering effects of the two competitors. In contrast, if both hemifields are occupied by competitors, both ears are equally masked and unable to selectively process the midline target, leading to impaired intelligibility.

In the case of signal that is spatial separated from a masking noise, there is some benefit in signal detection or intelligibility (if it is speech) that can not be accounted for by monaural listening with the ear with the best S/N. This benefit can be attributed to the binaural system but only accounts for about 2-3 dB (Saberi *et al.*, 1991; Hawley *et al.*, 1999) What is the role of these binaural cues in grouping and segregation? As has long been known they are crucial for localization, however they appear not to be important in the formation of auditory objects. Darwin and Hukin (1999) have argued for a model in which auditory objects are generated based on monaural features such as common onsets of frequency components and harmonicity. The ITD of these grouped sound components may be processed in parallel, and the average ITD across these components might be used to direct attention over time to the location of the already formed auditory object. This secondary processing role for ITD would be consistent with the failure to observe a primary effect of ITD in the results presented here.

APPENDIX A

CHAPTER 2, EXPERIMENT 1: WEIGHTED AVERAGE MODEL

In Chapter 2, Experiment 1, blocking one ear was used as a means of assessing the gap detection performance supported by one (the open) ear for each combination of leading and trailing marker locations. As the gap markers are separated across the midline, they will progressively differ in their spectral and intensive qualities at the open ear. Since it is known independently that marker level and spectral differences can affect gap thresholds (Oxenham, 2000; Experiment 3 in this thesis), it is of interest to establish the extent to which the pattern of binaural gap detection performance is predicted by performance at one or the other ear. In this regard, the monaural contributions to free-field performance could be used by the perceptual processor in any of a number of ways. As described in Chapter 2, the binaural thresholds might be limited by the 'Best Ear', 'Worst Ear', 'Leading marker ear', or some'Average' of those obtained through each of the two ears. The mean results for this analysis were presented in passing in Chapter 2; a more detailed description of these analysis is presented here.

Note that in Experiment 1, the left ear was blocked and so thresholds were obtained through the right ear. The functions were always symmetrical, so we could assume that the same pattern of results would have been obtained through the left ear (only flipped, i.e., reflected across the midline). For the purpose of evaluating the models, both left and right ear functions were required. In order to achieve this, the functions obtained under monaural listening was for midline located leading markers were simply flipped, and reused as if it was obtained through the other ear (see Fig A1, top). For leading markers located at –30°, -60°, -90°, and –150°, the functions actually obtained through the right-ear were used. To generate left-ear functions for these locations, the functions obtained through the right ear for leading markers located at +30°, +60°, +90°, +150° were flipped and treated as if they were obtained through the left-ear to leading marker locations at –30°, -60°, -90°, and –150°. The thresholds obtained at the left and right ear for a given marker pair could now be used to predict the pattern of gap detection

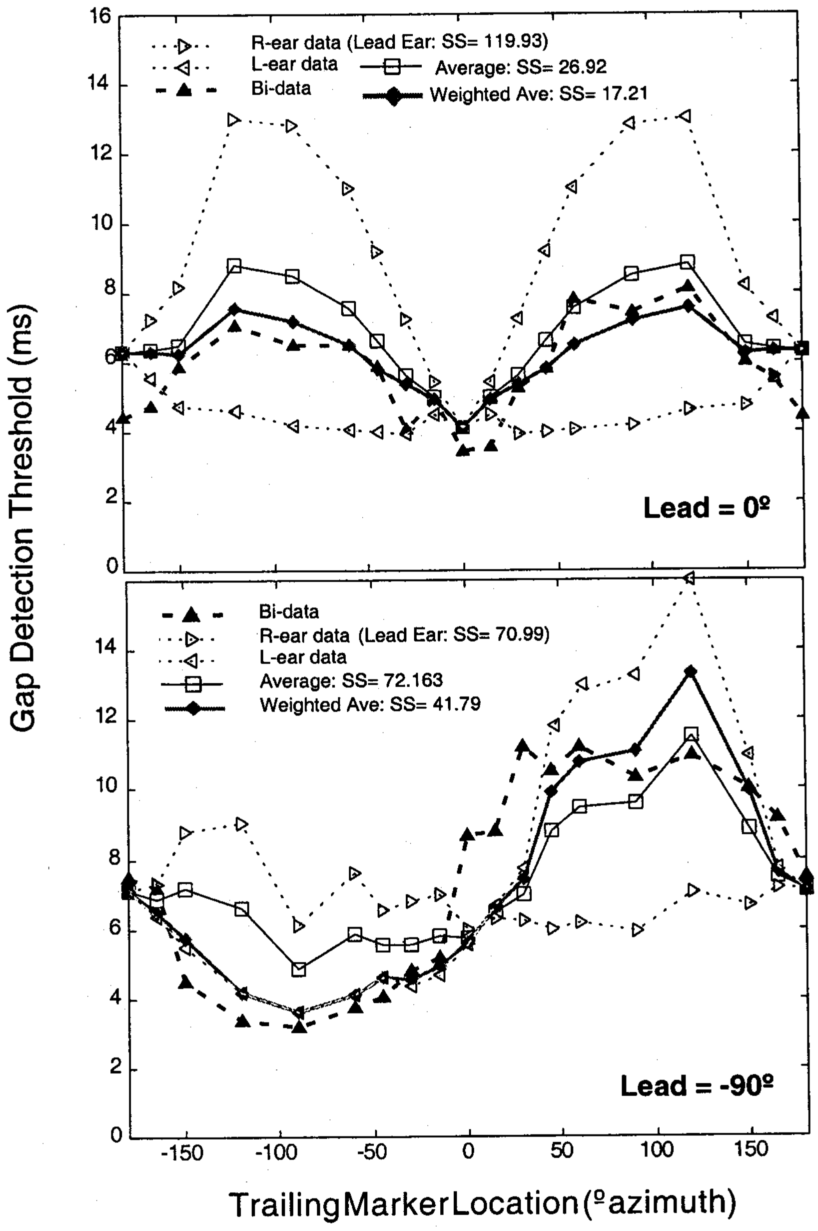


Figure A.1

Gap detection thresholds obtained under binaural and monaural (right and left ear) listening in Experiment 1 (Chapter 2) are plotted together. Note that the binaural thresholds typically fall in between that obtained from each of the two ears, except for conditions where both markers are in the same hemifield (bottom left side of curve), in which thresholds are determined by the near ear. A model in which thresholds are determined by the ear near the leading marker ('Lead-Ear') would predict binaural thresholds that would follow the curve for the left ear function. The 'average' model takes the linear average of the left and right ear functions. The 'weighted-average' model also takes an average; however, the left and right ear thresholds are weighted depending on the overal intensity of the markers at that ear for a given condition (see Text).

thresholds across trailing marker azimuth obtained for all leading markers under binaural conditions.

Models for combining the Monaural thresholds to predict the Binaural Thresholds

In the'Best Ear' model the lower of the right and left ear thresholds for a given condition was used in the prediction. Conversely, in the 'Worst-Ear' model, the highest threshold was used. In the 'Lead-Ear' model, the binaural thresholds were simply those obtained through the ear nearest the leading marker, with no influence of the other ear. Finally, two models were tested in which temporal limits of each ear were combined either linearly ("Average' model) or by some weighted average ("Weighted Average' model). The line of reasoning behind these two models is as follows. Let us assume that the level of representation at which the gap detection analysis is performed is that after information from the two ears is combined. When a good representation of the gap is available in the firing pattern of one auditory nerve and a poor representation is available in the other, convergence in the binaural system might cause the good representation of the gap to be contaminated by spikes from the other ear. In the average model, a linear average of the left and right ear thresholds was done for each condition. This was fairly successful in accounting for the binaural data, particularly for leading marker locations near the midline, but it failed for conditions when the leading marker was deep in one hemifield. Logically, since the near ear would be receiving far greater *sound intensity,* particularly for frequency components greater than 1000 Hz, that ear would probably deserve greater *weight* in the averaging process. The mean intensity difference between the markers at either ear is dependent on the headshadow. This idea lead to the creation of the Weighted-Average model.

In this model, the weight attributed to each ear for a given marker location was calculated based on a simple model of the head shadow, i.e. 50%L and 50%R for 0° azimuth, 0%L and 100%R for +90° azimuth, with a slope spanning 30° on either side of the midline ("Weighted Average Model"). The functions used are presented in column 2-3 of Table A1. This model assumes that depending on the relative locations of the markers, the weight of the gap representations at each ear will biased by the overall level of the sound reaching each ear. One problem was determining which location – that of the leading or the trailing marker – should influence the weighting of the two ears. Instead of arbitrarily choosing some value, the relative contribution of the leading and trailing marker locations was made a variable (X). Equation A1 was used to generate the weight for each ear that would be used in the weighted-average.

$$W_{ear} = L + (X * (T - L)) \qquad \text{(A1)}$$

In this equation, L refers to the value for the leading marker location for that ear's head-shadow function. The T refers to that value for the trailing marker location. The weighting factor, 'X', controlled how much the trailing marker location influenced the final weightings, and is the only flexible parameter in the model. When X is low, the leading marker location dicates all the weights (like a graded version of the 'Lead-Ear' model). As X increases, the location of the trailing marker is allowed to influence the final weightings of the two ears. The last part of the equation: factor* (trailweight-leadweight), takes into account how far away the trailing marker is from the leading marker while keeping the basic symmetry of auditory space (front-back/left-right) intact. To illustrate how this works, Table 1A presents a sample calculation of left and right ear weights across trailing marker location, for a leading marker located at -30°, given a low and high 'X' factor.

Evaluating the Models

For each leading marker condition, functions of gap threshold with trailing marker location were generated by combining the monaural thresholds as described above. For the weighted model, functions were generated for 11 values of X (0 – 1, in .1 steps). This process was completed based on the mean data from each subject and based on the group mean data (MEAN). In Figure A1, the binaural, left ear and right ear functions from the group mean data are plotted for a leading marker at either 0° or 90°. Superimposed are the predicted functions from the Average and Weighted-Average models. The Lead-ear model would simply follow the left-ear function. The 'Best-ear' and 'Worst-ear' models would simply follow the left-ear function for one hemifield of space and the right-ear

Table A1.

The Weighted-Average Model. Calculation of Model Weights.

Trailing Marker	**Head-Shadow Fn's**		**Lead= -30° Weight=.2**		**Lead= -30° Weight=.7**	
Location	L	R	Left Weights	Right Weights	Left Weights	Right Weights
180	50	50	70	30	57.5	42.5
-165	62.5	37.5	72.5	27.5	66.25	33.75
-150	75	25	75	25	75	25
-120	100	0	80	20	92.5	7.5
-90	100	0	80	20	92.5	7.5
-60	100	0	80	20	92.5	7.5
-45	100	0	80	20	92.5	7.5
-30	**75**	**25**	75	25	75	25
-15	62.5	37.5	72.5	27.5	66.25	33.75
0	50	50	70	30	57.5	42.5
15	37.5	62.5	67.5	32.5	48.75	51.25
30	25	75	65	35	40	60
45	0	100	60	40	22.5	77.5
60	0	100	60	40	22.5	77.5
90	**0**	**100**	**60**	**40**	**22.5**	**77.5**
120	0	100	60	40	22.5	77.5
150	25	75	65	35	40	60
165	37.5	62.5	67.5	32.5	48.75	51.25
180	50	50	70	30	57.5	42.5

Calculation of the weights for the condition Lead=-30°, Trail=90° for X=0.2

$W_{\text{left-ear}} = L + (X * (T - L)) = 75 + (0.2*(0-75)) = 60$

$W_{\text{Right-ear}} = L + (X * (T - L)) = 25 + (0.2*(100-25) = 40$

function for the other hemifield. Note that the 'Weighted-average' model accounts well for both conditions.

Three features of these model-generated functions were deemed important: 1) did a given model produce the same *pattern* of thresholds across trailing marker location for each of the leading markers, and 2) did it predict the *absolute value* of those thresholds 3) were the *errors* in prediction systematic (biased) or random (unbiased). The strength of the prediction was quantified by obtaining the sum of the squared deviations (SS) of the predicted thresholds from the actual thresholds obtained under binaural listening was determined. This measure took into account how well the prediction accounted for both the absolute value of gap thresholds obtained under binaural listening and the pattern of thresholds across conditions. The model which best minimized the SS across all leading marker locations was deemed to be the best model. This analysis was first done on the mean data collapsed across subjects. As shown in the bottom panel of Fig. A2, the Best-, Worst-, and Lead-Ear models did not predict the binaural data consistently across the various leading marker conditions. The overall SS was fairly well minimized by the Average model. However, this model predicted near-midline leading marker functions better than it predicted the data for hemifield-located markers. The weighted-average model predicted the functions best when the leading marker location received slightly more weight than the trailing marker locations (X=0.3-0.4). However, there were large individual differences in model success, as shown in the plots for each subject in Fig. A2. The binaural data obtained from Subject JS was best predicted by the lowest monaural threshold (Best-ear model). In contrast, the data of listener JH was best predicted by the Worst-ear model. The data of subject SB (the author) was closest to the mean, and thus was well predicted by the Weighted-average model similar to that which best predicted the mean (X=0.3-0.4). Finally, the data of listener SH was also best predicted by the Weighted-average model, but with a higher weighting (x=0.6).

The degree of prediction bias is also a good measure of model performance. A good model will be unbiased across the parameter of interest (in this case trailing marker location). That is, it should randomly over- or under-predict points rather than systematically overestimating some ranges of locations and underestimating others. Bias was examined by subtracting the predicted function from the actual binaural function and

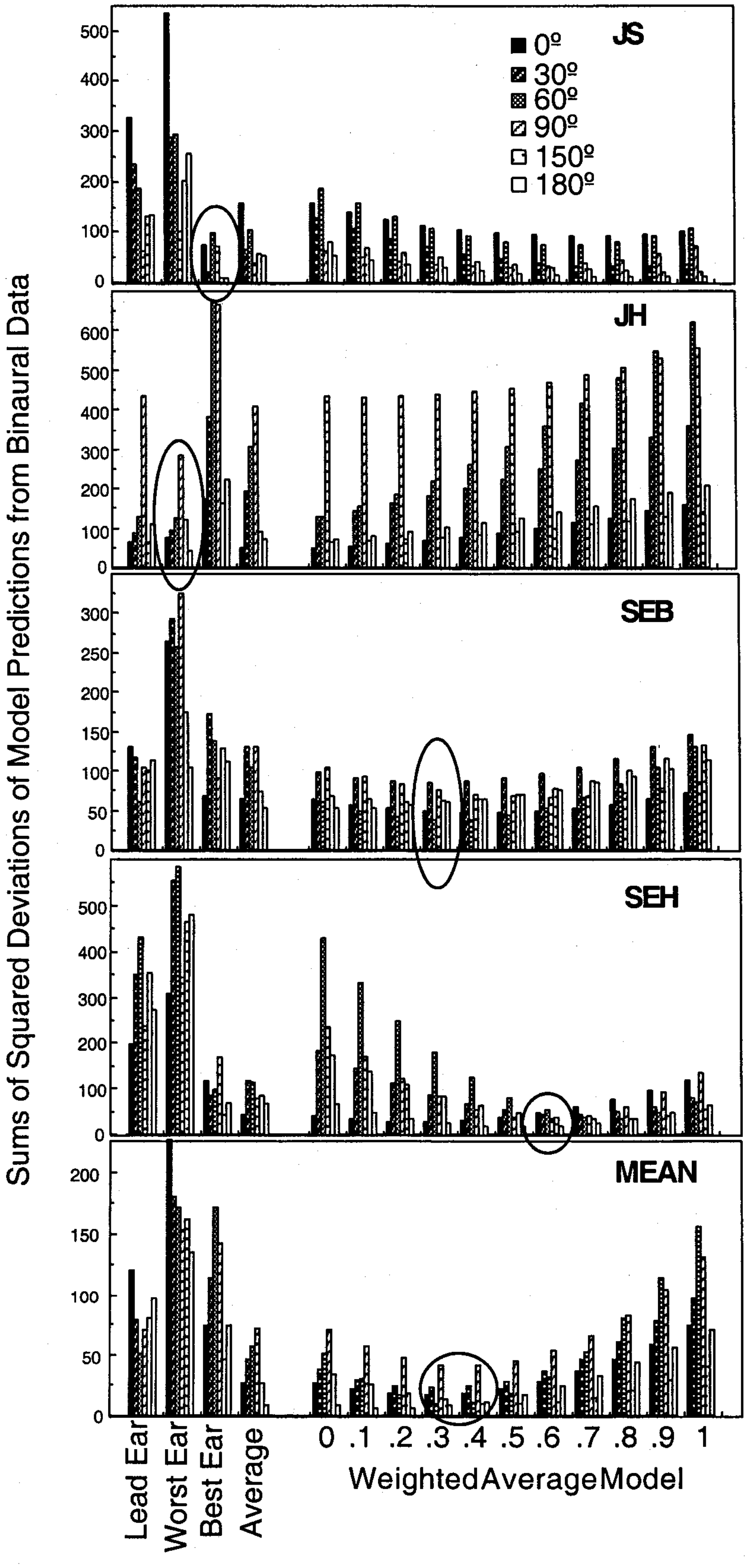


__*Figure A.2*__

This figure is the same as Fig. 2.5B, but additionally shows the data from the four individual listeners. Somewhat surprisingly, each listener's data were best predicted by a different model. The patterns of results obtained from SEB and SEH could both be best modelled by the weighted average model, although SEH required a higher weighting. However, JS and JH differed remarkably. JS acted as an ideal psychophysical observer who had indepdent access to each ear (thresholds were determined by the 'best-ear' under binaural listening). In contrast, listener JH seemed to be limited by the 'Worst-ear'.

examining the pattern of errors across trailing marker location. An example of this analysis using the mean data obtained for leading markers at 0°, 30°, 60°, and 90° is shown in Fig. A3. All of the models show some bias in their prediction of patterns of gap thresholds. Extreme cases include the Best-ear model which always overestimates gap thresholds for near-midline locations and underestimates gap thresholds for lateral locations; and the worst ear model, which always overestimates thresholds. The Weighted-ear model has the lowest absolute deviation from the actual binaural data across all conditions. As can be seen in Fig. A3, it also exhibits the most 0-line crossings of all the models in each leading marker condition, indicating that it is less systematic in its bias, since it is sometimes overestimating and sometimes underestimating.

From these analyses, it can be concluded that the Weighted-average model accounts best for the binaural data for all leading-marker locations and does so with the least bias. There are some problems with this analysis. The first is that it is based on data in which the binaural conditions were studied prior to the monaural conditions. It is possible that practice effects on the gap detection task influenced the absolute gap detection thresholds achieved under monaural conditions. Furthermore, it is limited to marker configurations in which the leading marker is short. In further experiments conducted after this analysis, it was determined that the pattern of thresholds obtained in Experiment 1 under monaural listening (low thresholds when the stimulus was quiet-loud, high thresholds when it was loud-quiet; intermediate or low thresholds under binaural conditions) was not found when longer leading markers were used. This does not dispute the results, only limits them.

THE WEIGHTED AVERAGE MODEL HAS THE LEAST BIAS IN PREDICTION

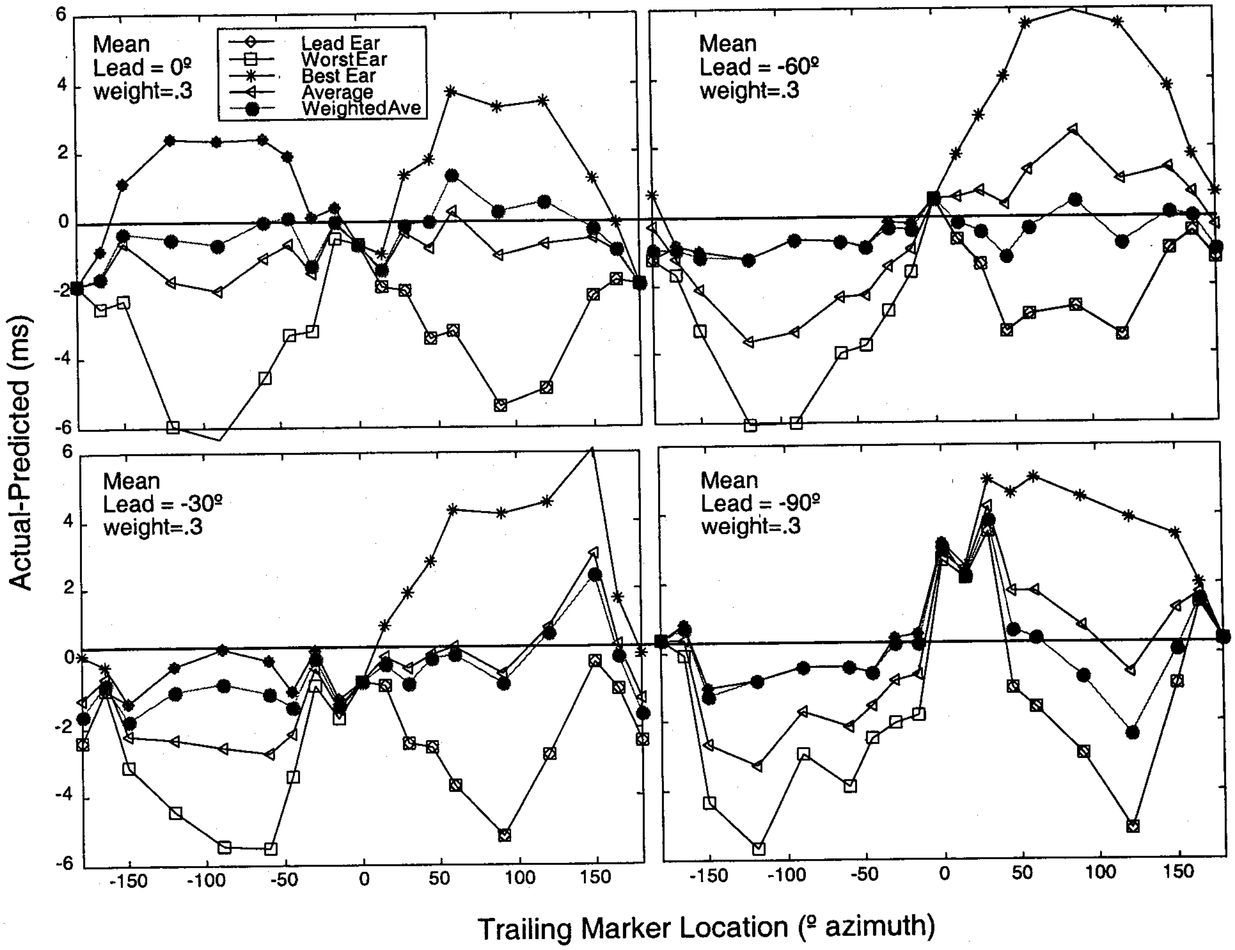


Figure A.3
This figure examines the prediction bias of the various models of combining gap thresholds from the two ears. The predicted threshold for a given condition is subtracted from the actual threshold obtained under binaural listening. A good model should have random bias. That is, it should randomly under and over predict each point. Note that the 'Best-', 'Worst-', and 'Lead-ear' models all have systematic bias to either under or over predict certain ranges of trailing marker location. Such biases are far less evident in the 'weighted-average' model.

REFERENCES

Ahissar, M., Ahissar, E., Bergman, H., & Vaadia, E. (1992). Encoding of sound-source location and movement: activity of single neurons and interactions between adjacent neurons in the monkey auditory cortex. J Neurophysiol, 67, 203-215.

Akeroyd, M. A., & Summerfield, A. Q. (1999). A binaural analog of gap detection. J Acoust Soc Am, 105, 2807-2820.

Allen, P. D., Virag, T. M., & Ison, J. R. (2002). Humans detect gaps in broadband noise according to effective gap duration without additional cues from abrupt envelope changes. J Acoust Soc Am, 112, 2967-2974.

Anstis, S., & Saida, S. (1985). Adaptation to auditory streaming of frequency-modulated tones. J Exp Psychol-Hum Percept Perform, 11, 257-271.

Beauvois, M. W. (1998). The effect of tone duration on auditory stream formation. Percept Psychophys, 60, 852-861.

Beauvois, M. W., & Meddis, R. (1991). A computer model of auditory stream segregation. Q J Exp Psychol A, 43, 517-541.

Beauvois, M. W., & Meddis, R. (1996). Computer simulation of auditory stream segregation in alternating-tone sequences. J Acoust Soc Am, 99, 2270-2280.

Beauvois, M. W., & Meddis, R. (1997). Time decay of auditory stream biasing. Percept Psychophys, 59, 81-86.

Belin, P., & Zatorre, R. J. (2000). 'What', 'where' and 'how' in auditory cortex. Nat Neurosci, 3, 965-966.

Benson, D. A., Hienz, R. D., & Goldstein, M. H., Jr. (1981). Single-unit activity in the auditory cortex of monkeys actively localizing sound sources: spatial tuning and behavioral dependency. Brain Res, 219, 249-267.

Bernstein, L. R., & Trahiotis, C. (1992). Discrimination of interaural envelope correlation and its relation to binaural unmasking at high frequencies. J Acoust Soc Am, 91, 306-316.

Bernstein, L. R., Trahiotis, C., Akeroyd, M. A., & Hartung, K. (2001). Sensitivity to brief changes of interaural time and interaural intensity. J Acoust Soc Am, 109, 1604-1615.

Blauert, J. (1997). Spatial Hearing. Cambridge: MIT Press.

Blauert, J., & Lindemann, W. (1986). Spatial mapping of intracranial auditory events for various degrees of interaural coherence. J Acoust Soc Am, 79, 806-813.

Blodgett, H. C., Jeffress, L. A., & Taylor, R. W. (1958). Relation of masked threshold to signal duration for various interaural phase combinations. Am J of Psychol, 71, 283-290.

Boehnke, S. E. (1998). The tuning of human auditory perceptual channels for sound location: evidence from gap detection. Unpublished Master's of Science, Dalhousie University, Halifax.

Boehnke, S. E., & McAlpine, D. (2001). Effect of stimulus level on ITD coding of low frequency cells in the guinea pig auditory midbrain. Assoc Res Otolaryngol Abs, 21723.

Boehnke, S. E., & Phillips, D. P. (1999). Azimuthal tuning of human perceptual channels for sound location. J Acoust Soc Am, 106, 1948-1955.

Boehnke, S. E., & Phillips, D. P. (2000). Human spatial perceptual channels arise from binaural mechanisms. Assoc Res Otolaryngol Abs, 4637.

Brand, A., Behrend, O., Marquardt, T., McAlpine, D., & Grothe, B. (2002). Precise inhibition is essential for microsecond interaural time difference coding. Nature, 417, 543-547.

Bregman, A. S. (1978). Auditory streaming is cumulative. J Exp Psychol Hum Percept Perform, 4, 380-387.

Bregman, A. S. (1990). Auditory Stream Segregation: The Perceptual Organization of Sound. Cambridge: MIT Press.

Bregman, A. S., Ahad, P. A., Crum, P. A., & O'Reilly, J. (2000). Effects of time intervals and tone durations on auditory stream segregation. Percept Psychophys, 62, 626-636.

Bregman, A. S., Ahad, P. A., & Van Loon, C. (2001). Stream segregation of narrow-band noise bursts. Percept Psychophys, 63, 790-797.

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. J Exp Psychol, 89, 244-249.

Broadbent, D. E. (1958). Perception and communication. London: Oxford University Press.

Brosch, M., & Schreiner, C. E. (1997). Time course of forward masking tuning curves in cat primary auditory cortex. J Neurophysiol, 77, 923-943.

Brugge, J. F., Anderson, D. J., & Aitkin, L. M. (1970). Responses of neurons in the dorsal nucleus of the lateral lemniscus of cat to binaural tonal stimulation. J Neurophysiol, 33, 441-458.

Brugge, J. F., Dubrovsky, N. A., Aitkin, L. M., & Anderson, D. J. (1969). Sensitivity of single neurons in auditory cortex of cat to binaural tonal stimulation; effects of varying interaural time and intensity. J Neurophysiol, 32, 1005-1024.

Brugge, J. F., Reale, R. A., & Hind, J. E. (1996). The structure of spatial receptive fields of neurons in primary auditory cortex of the cat. J Neurosci, 16, 4420-4437.

Brugge, J. F., Reale, R. A., & Hind, J. E. (1998). Spatial receptive fields of primary auditory cortical neurons in quiet and in the presence of continuous background noise. J Neurophysiol, 80, 2417-2432.

Buus, S., & Florentine, M. (1985). Gap detection in normal and impaired listeners: The effect of level and frequency. In A. Michelsen (Ed.), Time Resolution in Auditory Systems (pp. 159-179). London: Springer-Verlag.

Cant, N. B., & Hyson, R. L. (1992). Projections from the lateral nucleus of the trapezoid body to the medial superior olivary nucleus in the gerbil. Hear Res, 58, 26-34.

Carlile, S. (Ed.). (1996). Virtual auditory space: Generation and applications. New York: Landes Publishing Company.

Carlile, S., Leong, P., & Hyams, S. (1997). The nature and distribution of errors in sound localization by human listeners. Hear Res, 114, 179-196.

Carlile, S., & Pralong, D. (1994). The location-dependent nature of perceptually salient features of the human head-related transfer functions. J Acoust Soc Am, 95, 3445-3459.

Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. J Exp Psychol Hum Percept Perform, 27, 115-127.

Chatterjee, M., Fu, Q. J., & Shannon, R. V. (1998). Within-channel gap detection using dissimilar markers in cochlear implant listeners. J Acoust Soc Am, 103, 2515-2519.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. J Acoust Soc Am, 25, 975-979.

Cherry, E. C., & Bowles, J. A. (1960). Contribution to a study of the "cocktail party problem ". J Acoust Soc Am, 32, 884.

Cherry, E. C., & Taylor, W. K. (1954). Some further experiments upon the recognition of speech with one and with two ears. J Acoust Soc Am, 26, 554-559.

Clarey, J. C., Barone, P., & Imig, T. J. (1992). Physiology of the Thalamus and Cortex. In A. N. Popper & R. R. Fay (Eds.), The Mammalian Auditory Pathway: Neurophysiology (Vol. 2, pp. 232-334). New York: Springer-Verlag.

Clarey, J. C., Barone, P., Irons, W. A., Samson, F. K., & Imig, T. J. (1995). Comparison of noise and tone azimuth tuning of neurons in cat primary auditory cortex and medical geniculate body. J Neurophysiol, 74, 961-980.

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

Collyer, C. E. (1974). The detection of a temporal gap between two disparate stimuli. Percept Psychophys, 16, 96-100.

Crawley, E. J. (1999). Demonstrating "Overshoot" in the Gap Detection Task: Breaking down a Paradigmatic Barrier. Unpublished Doctoral Thesis, Binghampton University, State University of New York, Binghamton, NY, U.S.A.

Culling, J. F., & Colburn, H. S. (2000). Binaural sluggishness in the perception of tone sequences and speech in noise. J Acoust Soc Am, 107, 517-527.

Culling, J. F., Colburn, H. S., & Spurchise, M. (2001). Interaural correlation sensitivity. J Acoust Soc Am, 110, 1020-1029.

Culling, J. F., Marshall, D. H., & Summerfield, A. Q. (1998a). Dichotic pitches as illusions of binaural unmasking. II. The Fourcin pitch and the dichotic repetition pitch. J Acoust Soc Am, 103, 3527-3539.

Culling, J. F., & Summerfield, A. Q. (1998). Measurements of the binaural temporal window using a detection task. J Acoust Soc Am, 103, 3540-3553.

Culling, J. F., Summerfield, A. Q., & Marshall, D. H. (1998b). Dichotic pitches as illusions of binaural unmasking. I. Huggins' pitch and the 'binaural edge pitch'. J Acoust Soc Am, 103, 3509-3526.

Culling, J. F., & Summerfield, Q. (1995). Perceptual separation of concurrent speech sounds: absence of across-frequency grouping by common interaural delay. J Acoust Soc Am, 98, 785-797.

Dannenbring, G. L., & Bregman, A. S. (1976). Stream segregation and the illusion of overlap. J Exp Psychol Hum Percept Perform, 2, 544-555.

Darwin, C. J., & Hukin, R. W. (1997). Perceptual segregation of a harmonic from a vowel by interaural time difference and frequency proximity. J Acoust Soc Am, 102, 2316-2324.

Darwin, C. J., & Hukin, R. W. (1998). Perceptual segregation of a harmonic from a vowel by interaural time difference in conjunction with mistuning and onset asynchrony. J Acoust Soc Am, 103, 1080-1084.

Darwin, C. J., & Hukin, R. W. (1999). Auditory objects of attention: the role of interaural time differences. J Exp Psychol Hum Percept Perform, 25, 617-629.

Davis, K. A., Miller, R. L., & Young, E. D. (1996). Effects of somatosensory and parallel-fiber stimulation on neurons in dorsal cochlear nucleus. J Neurophysiol, 76, 3012-3024.

de Cheveigné, A. (2001). The auditory system as a "separation machine". In D. J. Breebaart & A. J. M. Houtsma & A. Kohlrausch & V. F. Prijs & R. Schoonhoven (Eds.), Physiological and Psychphysical Bases of Auditory Function: Proceedings of the 12th International Symposium on Hearing. Maastricht: Shaker Publishing.

Delgutte, B., Joris, P. X., Litovsky, R. Y., & Yin, T. C. (1999). Receptive fields and binaural interactions for virtual-space stimuli in the cat inferior colliculus. J Neurophysiol, 81, 2833-2851.

Deutsch, D., & Roll, P. L. (1976). Separate "what" and "where" decision mechanisms in processing a dichotic tonal sequence. J Exp Psychol Hum Percept Perform, 2, 23-29.

Divenyi, P. L., & Blauert, J. (1987). On creating precedent for binaural patterns: When is an echo an echo? In W. A. Yost & C. S. Watson (Eds.), Auditory Processing of Complex Sounds. Hillsdale, NJ: Lawrence Erlbaum.

Divenyi, P. L., & Danner, W. F. (1977). Discrimination of time intervals marked by brief acoustic pulses of various intensities and spectra. Percept Psychophys, 21, 125-142.

Divenyi, P. L., & Hirsh, I. J. (1974). Identification of temporal order in three-tone sequences. J Acoust Soc Am, 56, 144-151.

Durlach, N., & Colburn, H. S. (1978). Binaural phenomena. In E. C. Carterette & M. P. Friedman (Eds.), Handbook of Perception: Hearing (pp. Chap. 10). New York: Academic.

Durlach, N. I. (1978). Binaural signal detection: Equalization and Cancellation Theory. In J. V. Tobias (Ed.), Foundations of Modern Auditory Theory. New York: Academic.

Durlach, N. I., Gabriel, K. J., Colburn, H. S., & Trahiotis, C. (1986). Interaural correlation discrimination: II. Relation to binaural unmasking. J Acoust Soc Am, 79, 1548-1557.

Eddins, D. A., Hall, J. W., 3rd, & Grose, J. H. (1992). The detection of temporal gaps as a function of frequency region and absolute noise bandwidth. J Acoust Soc Am, 91, 1069-1077.

Egan, J. P. (1965). Masking-level differences as a function of interaural disparities in intensity of signal and of noise. J Acoust Soc Am, 38, 1043-1049.

Eggermont, J. J. (1995). Neural correlates of gap detection and auditory fusion in cat auditory cortex. Neuroreport, 6, 1645-1648.

Eggermont, J. J. (1999). Neural correlates of gap detection in three auditory cortical fields in the Cat. J Neurophysiol, 81, 2570-2581.

Eggermont, J. J. (2000). Neural responses in primary auditory cortex mimic psychophysical, across-frequency-channel, gap-detection thresholds. J Neurophysiol, 84, 1453-1463.

Eggermont, J. J., & Mossop, J. E. (1998). Azimuth coding in primary auditory cortex of the cat. I. Spike synchrony versus spike count representations. J Neurophysiol, 80, 2133-2150.

Elliot, L. L. (1962). Backward and forward masking of probe tones of different frequencies. J Acoust Soc Am, 34, 1116-1117.

Fedderson, W. E., Sandel, T. T., Teas, D. C., & Jeffress, L. A. (1957). Localization of high frequency tones. J Acoust Soc Am, 29, 988-991.

Fishman, Y. I., Reser, D. H., Arezzo, J. C., & Steinschneider, M. (2001). Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hear Res, 151, 167-187.

Fitzgibbons, P. J. (1983). Temporal gap detection in noise as a function of frequency, bandwidth, and level. J Acoust Soc Am, 74, 67-72.

Fitzgibbons, P. J., Pollatsek, A., & Thomas, I. B. (1974). Detection of temporal gaps within and between perceptual tonal groups. Percept Psychophys, 16, 522-528.

Fitzgibbons, P. J., & Wightman, F. L. (1982). Gap detection in normal and hearing-impaired listeners. J Acoust Soc Am, 72, 761-765.

Fitzpatrick, D. C., Ebert, C. A., Pollock, H. W., Pearson, J. M., & Crocker, W. D. (2002). Responses of auditory cortical neurons sensitive to interaural time differences (ITDs) to changes in interaural correlation. Paper presented at the 25th MidWinter Meeting of the Association for Research in Otolaryngology.

Fitzpatrick, D. C., & Kuwada, S. (2001). Tuning to interaural time differences across frequency. J Neurosci, 21, 4844-4851.

Fitzpatrick, D. C., Kuwada, S., & Batra, R. (2000). Neural sensitivity to interaural time differences: beyond the Jeffress model. J Neurosci, 20, 1605-1615.

Florentine, M., & Buus, S. (1984). Temporal gap detection in sensorineural and simulated hearing impairments. J Speech Hear Res, 27, 449-455.

Florentine, M., Buus, S., & Geng, W. (1999). Psychometric functions for gap detection in a yes-no procedure. J Acoust Soc Am, 106, 3512-3520.

Formby, C., Barker, C., Abbey, H., & Raney, J. J. (1993). Detection of silent temporal gaps between narrow-band noise makers having second-formantlike properties of voiceless stop/vowel combinations. J Acoust Soc Am, 93, 1023-1027.

Formby, C., & Forrest, T. G. (1991). Detection of silent temporal gaps in sinusoidal markers. J Acoust Soc Am, 89, 830-837.

Formby, C., Gerber, M. J., Sherlock, L. P., & Magder, L. S. (1998a). Evidence for an across-frequency, between-channel process in asymptotic monaural temporal gap detection. J Acoust Soc Am, 103, 3554-3560.

Formby, C., Sherlock, L. P., & Forrest, T. G. (1996a). An asymmetric roex filter model for describing detection of silent temporal gaps in sinusoidal markers. Audit Neurosci, 3, 1-20.

Formby, C., Sherlock, L. P., & Green, D. M. (1996b). Evaluation of a maximum likelihood procedure for measuring pure-tone thresholds under computer control. J Am Acad Audiol, 7, 125-129.

Formby, C., Sherlock, L. P., & Li, S. (1998b). Temporal gap detection measured with multiple sinusoidal markers: effects of marker number, frequency, and temporal position. J Acoust Soc Am, 104, 984-998.

Forrest, T. G., & Formby, C. (1996). Detection of silent temporal gaps in sinusoidal markers simulated with a single-channel envelope detector model. Audit Neurosci, 3, 21-33.

Forrest, T. G., & Green, D. M. (1987). Detection of partially filled gaps in noise and the temporal modulation transfer function. J Acoust Soc Am, 82, 1933-1943.

Furukawa, S., & Middlebrooks, J. C. (2001). Sensitivity of auditory cortical neurons to locations of signals and competing noise sources. J Neurophysiol, 86, 226-240.

Furukawa, S., Xu, L., & Middlebrooks, J. C. (2000). Coding of sound-source location by ensembles of cortical neurons. J Neurosci, 20, 1216-1228.

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

Gilkey, R. H., & Good, M. D. (1995). Effects of frequency on free-field masking. Hum Factors, 37, 835-843.

Gockel, H., & Carlyon, R. P. (1998). Effects of ear of entry and perceived location of synchronous and asynchronous components on mistuning detection. J Acoust Soc Am, 104, 3534-3545.

Gockel, H., Carlyon, R. P., & Micheyl, C. (1999). Context dependence of fundamental-frequency discrimination: lateralized temporal fringes. J Acoust Soc Am, 106, 3553-3563.

Goldberg, J. M., & Brown, P. B. (1969). Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. J Neurophysiol, 32, 613-636.

Good, M. D., & Gilkey, R. H. (1996). Sound localization in noise: the effect of signal-to-noise ratio. J Acoust Soc Am, 99, 1108-1117.

Grantham, D. W. (1995). Spatial hearing and related phenomena. In B. C. J. Moore (Ed.), Hearing (pp. 297-345). San Diego, CA: Academic Press, Inc.

Grantham, D. W., & Wightman, F. L. (1978). Detectability of varying interaural temporal differences. J Acoust Soc Am, 63, 511-523.

Grantham, D. W., & Wightman, F. L. (1979). Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation. J Acoust Soc Am, 65, 1509-1517.

Griffiths, T. D., Green, G. G., Rees, A., & Rees, G. (2000). Human brain areas involved in the analysis of auditory movement. Hum Brain Mapp, 9, 72-80.

Griffiths, T. D., Rees, A., Witton, C., Cross, P. M., Shakir, R. A., & Green, G. G. (1997). Spatial and temporal auditory processing deficits following right hemisphere infarction. A psychophysical study. Brain, 120 (Pt 5), 785-794.

Griffiths, T. D., Rees, A., Witton, C., Shakir, R. A., Henning, G. B., & Green, G. G. (1996). Evidence for a sound movement area in the human cerebral cortex. Nature, 383, 425-427.

Griffiths, T. D., Rees, G., Rees, A., Green, G. G., Witton, C., Rowe, D., Buchel, C., Turner, R., & Frackowiak, R. S. (1998). Right parietal cortex is involved in the perception of sound movement in humans. Nat Neurosci, 1, 74-79.

Grimault, N., Bacon, S. P., & Micheyl, C. (2002). Auditory stream segregation on the basis of amplitude-modulation rate. J Acoust Soc Am, 111, 1340-1348.

Grose, J. H., Hall, J. W., 3rd, & Buss, E. (1999). Modulation gap detection: effects of modulation rate, carrier separation, and mode of presentation. J Acoust Soc Am, 106, 946-953.

Grose, J. H., Hall, J. W., 3rd, & Buss, E. (2001a). Gap duration discrimination in listeners with cochlear hearing loss: effects of gap and marker duration, frequency separation, and mode of presentation. J Assoc Res Otolaryngol, 2, 388-398.

Grose, J. H., Hall, J. W., 3rd, Buss, E., & Hatch, D. (2001b). Gap detection for similar and dissimilar gap markers. J Acoust Soc Am, 109, 1587-1595.

Grothe, B. (1994). Interaction of excitation and inhibition in processing of pure tone and amplitude-modulated stimuli in the medial superior olive of the mustached bat. J Neurophysiol, 71, 706-721.

Grothe, B. (2000). The evolution of temporal processing in the medial superior olive, an auditory brainstem structure. Prog Neurobiol, 61, 581-610.

Grothe, B., & Sanes, D. H. (1993). Bilateral inhibition by glycinergic afferents in the medial superior olive. J Neurophysiol, 69, 1192-1196.

Haas, H. (1951). On the influence of a single echo on the intelligibility of speech. Acustica, 1.

Hall, J. W., 3rd, Grose, J. H., & Joy, S. (1996). Gap detection for pairs of noise bands: effects of stimulus level and frequency separation. J Acoust Soc Am, 99, 1091-1095.

Hanekom, J. J., & Shannon, R. V. (1998). Gap detection as a measure of electrode interaction in cochlear implants. J Acoust Soc Am, 104, 2372-2384.

Hanna, T. E., Robinson, D. E., Shiffrin, R. M., & Gilkey, R. H. (1982). Forward masking of diotic and dichotic clicks by noise. J Acoust Soc Am, 72, 1171-1177.

Hartmann, W. M., & Johnson, D. (1991). Stream segregation and peripheral channeling. Music Percept, 9, 155-184.

Hartung, K., & Trahiotis, C. (2001). Peripheral auditory processing and investigations of the "precedence effect" which utilize successive transient stimuli. J Acoust Soc Am, 110, 1505-1513.

Hawley, M. L., Litovsky, R. Y., & Colburn, H. S. (1999). Speech intelligibility and localization in a multi-source environment. J Acoust Soc Am, 105, 3436-3448.

Heffner, H., & Masterton, B. (1975). Contribution of auditory cortex to sound localization in the monkey (Macaca mulatta). J Neurophysiol, 38, 1340-1358.

Heffner, H. E. (1997). The role of macaque auditory cortex in sound localization. Acta Otolaryngol Suppl, 532, 22-27.

Heffner, H. E., & Heffner, R. S. (1990). Effect of bilateral auditory cortex lesions on sound localization in Japanese macaques. J Neurophysiol, 64, 915-931.

Heinz, M. G., Goldstein, M. H., Jr., & Formby, C. (1996). Temporal gap detection thresholds in sinusoidal markers simulated with a multi-channel, multi-resolution model of the auditory periphery. Auditory Neuroscience, 3, 35-56.

Hirsh, I. J. (1948). The influence of interaural phase on interaural summation and inhibition. J Acoust Soc Am, 20, 536-544.

Holube, I., Kinkel, M., & Kollmeier, B. (1998). Binaural and monaural auditory filter bandwidths and time constants in probe tone detection experiments. J Acoust Soc Am, 104, 2412-2425.

Hukin, R. W., & Darwin, C. J. (1995). Effects of contralateral presentation and of interaural time difference in segregating a harmonic from a vowel. J Acoust Soc Am, 98, 1380-1387.

Huynh, H., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. Journal of Educational Statistics, Vol 1(1), 69-82.

Imig, T. J., Bibikov, N. G., Poirier, P., & Samson, F. K. (2000). Directionality derived from pinna-cue spectral notches in cat dorsal cochlear nucleus. J Neurophysiol, 83, 907-925.

Imig, T. J., Irons, W. A., & Samson, F. R. (1990). Single-unit selectivity to azimuthal direction and sound pressure level of noise bursts in cat high-frequency primary auditory cortex. J Neurophysiol, 63, 1448-1466.

Irvine, D. R. (1992). Physiology of the Auditory Brainstem. In A. N. Popper & R. R. Fay (Eds.), The Mammalian Auditory Pathway: Neurophysiology (Vol. 2, pp. 153-231). New York: Springer-Verlag.

Ison, J. R., Castro, J., Allen, P., Virag, T. M., & Walton, J. P. (2002). The relative detectability for mice of gaps having different ramp durations at their onset and offset boundaries. J Acoust Soc Am, 112, 740-747.

Jain, M., Gallagher, D. T., Koehnke, J., & Colburn, H. S. (1991). Fringed correlation discrimination and binaural detection. J Acoust Soc Am, 90, 1918-1926.

Jeffress, L. A. (1948). A place theory of sound localization. Journal of Comparative Physiological Psychology, 41, 35-39.

Jeffress, L. A., Blodgett, H. C., Sandel, T. T., & Wood III, C. L. (1956). Masking of tonal signals. J Acoust Soc Am, 28, 416-126.

Jenkins, W. M., & Masterton, R. B. (1982). Sound localization: effects of unilateral lesions in central auditory system. J Neurophysiol, 47, 987-1016.

Jenkins, W. M., & Merzenich, M. M. (1984). Role of cat primary auditory cortex for sound-localization behavior. J Neurophysiol, 52, 819-847.

Kanold, P. O., & Young, E. D. (2001). Proprioceptive information from the pinna provides somatosensory input to cat dorsal cochlear nucleus. J Neurosci, 21, 7848-7858.

Kavanagh, G. L., & Kelly, J. B. (1986). Midline and lateral field sound localization in the albino rat (Rattus norvegicus). Behav Neurosci, 100, 200-205.

Kelly, J. B. (1980). Effects of auditory cortical lesions on sound localization by the rat. J Neurophysiol, 44, 1161-1174.

Kinney, J. A. S. (1961). Discrimination in auditory and visual patterns. Am J Psychol, 74, 529-541.

Kistler, D. J., & Wightman, F. L. (1992). A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. J Acoust Soc Am, 91, 1637-1647.

Kitzes, L. M., Wrege, K. S., & Cassady, J. M. (1980). Patterns of responses of cortical cells to binaural stimulation. J Comp Neurol, 192, 455-472.

Knudsen, E. I., & Konishi, M. (1978). A neural map of auditory space in the owl. Science, 200, 795-797.

Koehnke, J., Colburn, H. S., & Durlach, N. I. (1986). Performance in several binaural-interaction experiments. J Acoust Soc Am, 79, 1558-1562.

Kollmeier, B., & Gilkey, R. H. (1990). Binaural forward and backward masking: evidence for sluggishness in binaural detection. J Acoust Soc Am, 87, 1709-1719.

Kuhn, G. F. (1977). Model for the interaural time differences in the azimuthal plane. J Acoust Soc Am, 62, 157-167.

Kuwada, S., Batra, R., & Fitzpatrick, D. C. (1997). Neural Processing of Binaural Temporal Cues. In R. H. Gilkey & T. R. Anderson (Eds.), Binaural and spatial hearing in real and virtual environments (pp. 399-425). Mahwah: Lawrence Erlbaum Associates, Inc.

Langford, T. L., & Jeffress, L. A. (1964). Effect of noise cross correlation on binaural signal detection. J Acoust Soc Am, 36, 1455-1458.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. J Acoust Soc Am, 49, Suppl 2:467+.

Levitt, H., & Rabiner, L. R. (1967). Binaural release from masking for speech and gain in intelligibility. J Acoust Soc Am, 42, 601-608.

Lewald, J., Foltys, H., & Topper, R. (2002). Role of the posterior parietal cortex in spatial hearing. J Neurosci, 22, RC207.

Lister, J., Besing, J., & Koehnke, J. (2002). Effects of age and frequency disparity on gap discrimination. J Acoust Soc Am, 111, 2793-2800.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., & Guzman, S. J. (1999). The precedence effect. J Acoust Soc Am, 106, 1633-1654.

Litovsky, R. Y., & Delgutte, B. (2002). Neural correlates of the precedence effect in the inferior colliculus: effect of localization cues. J Neurophysiol, 87, 976-994.

Lorenzi, C., Gatehouse, S., & Lever, C. (1999). Sound localization in noise in normal-hearing listeners. J Acoust Soc Am, 105, 1810-1820.

Makous, J. C., & Middlebrooks, J. C. (1990). Two-dimensional sound localization by human listeners. J Acoust Soc Am, 87, 2188-2200.

McAlpine, D., Jiang, D., & Palmer, A. R. (2001). A neural code for low-frequency sound localization in mammals. Nat Neurosci, 4, 396-401.

McAlpine, D., Jiang, D., Shackleton, T. M., & Palmer, A. R. (1998). Convergent input from brainstem coincidence detectors onto delay-sensitive neurons in the inferior colliculus. J Neurosci, 18, 6026-6039.

McCabe, S. L., & Denham, M. J. (1997). A model of auditory streaming. J Acoust Soc Am, 101, 1611-1621.

McFadden, D., & Wright, B. A. (1987). Comodulation masking release in a forward-masking paradigm. J Acoust Soc Am, 82, 1615-1620.

Micheyl, C., Tian, B., Carlyon, R. P., & Rauschecker, J. P. (2002). The neural basis of stream segregation in the primary auditory cortex. Abstract of a seminar presented at the Hearing Research Centre, Boston University: Retrieved December 18, 2002, from the world wide web: http://www.bu.edu/hrc/cmAbstract.htm.

Middlebrooks, J. C. (1992). Narrow-band sound localization related to external ear acoustics. J Acoust Soc Am, 92, 2607-2624.

Middlebrooks, J. C. (1997). Spectral shape cues for sound localization. In R. H. Gilkey & T. R. Anderson (Eds.), Binaural and Spatial Hearing (pp. 77-97). Mawah, NJ: Lawrence Erlbaum Associates, Inc.

Middlebrooks, J. C., Clock, A. E., Xu, L., & Green, D. M. (1994). A panoramic code for sound location by cortical neurons. Science, 264, 842-844.

Middlebrooks, J. C., & Green, D. M. (1991). Sound localization by human listeners. Annu Rev Psychol, 42, 135-159.

Middlebrooks, J. C., & Knudsen, E. I. (1984). A neural code for auditory space in the cat's superior colliculus. J Neurosci, 4, 2621-2634.

Middlebrooks, J. C., Makous, J. C., & Green, D. M. (1989). Directional sensitivity of sound-pressure levels in the human ear canal. J Acoust Soc Am, 86, 89-108.

Middlebrooks, J. C., & Pettigrew, J. D. (1981). Functional classes of neurons in primary auditory cortex of the cat distinguished by sensitivity to sound location. J Neurosci, 1, 107-120.

Middlebrooks, J. C., Xu, L., Eddins, A. C., & Green, D. M. (1998). Codes for sound-source location in nontonotopic auditory cortex. J Neurophysiol, 80, 863-881.

Mills, A. W. (1958). On the minimal auditory angle. J Acoust Soc Am, 30.

Moore, B. C., Glasberg, B. R., Plack, C. J., & Biswas, A. K. (1988). The shape of the ear's temporal window. J Acoust Soc Am, 83, 1102-1116.

Moore, B. C. J. (1997). An introduction to the psychology of hearing (4th ed.). San Diego: Academic Press.

Munte, T. F., Kohlmetz, C., Nager, W., & Altenmuller, E. (2001). Neuroperception. Superior auditory spatial tuning in conductors. Nature, 409, 580.

Musicant, A. D., & Butler, R. A. (1984a). The influence of pinnae-based spectral cues on sound localization. J Acoust Soc Am, 75, 1195-1200.

Musicant, A. D., & Butler, R. A. (1984b). The psychophysical basis of monaural localization. Hear Res, 14, 185-190.

Musicant, A. D., & Butler, R. A. (1985). Influence of monaural spectral cues on binaural localization. J Acoust Soc Am, 77, 202-208.

Neff, D. L., Jesteadt, W., & Brown, E. L. (1982). The relation between gap discrimination and auditory stream segregation. Percept Psychophys, 31, 493-501.

Neti, C., Young, E. D., & Schneider, M. H. (1992). Neural network models of sound localization based on directional filtering by the pinna. J Acoust Soc Am, 92, 3140-3156.

Oldfield, S. R., & Parker, S. P. (1984a). Acuity of sound localisation: a topography of auditory space. I. Normal hearing conditions. Perception, 13, 581-600.

Oldfield, S. R., & Parker, S. P. (1984b). Acuity of sound localisation: a topography of auditory space. II. Pinna cues absent. Perception, 13, 601-617.

Oldfield, S. R., & Parker, S. P. (1986). Acuity of sound localisation: a topography of auditory space. III. Monaural hearing conditions. Perception, 15, 67-81.

Oxenham, A. J. (1997). Increment and decrement detection in sinusoids as a measure of temporal resolution. J Acoust Soc Am, 102, 1779-1790.

Oxenham, A. J. (2000). Influence of spatial and temporal coding on auditory gap detection. J Acoust Soc Am, 107, 2215-2223.

Oxenham, A. J. (2001). Forward masking: adaptation or integration? J Acoust Soc Am, 109, 732-741.

Palmer, A. R., & King, A. J. (1985). A monaural space map in the guinea-pig superior colliculus. Hear Res, 17, 267-280.

Parsons, J. E., Lim, E., & Voigt, H. F. (2001). Type III units in the gerbil dorsal cochlear nucleus may be spectral notch detectors. Ann Biomed Eng, 29, 887-896.

Pavani, F., Macaluso, E., Warren, J., Driver, J., & Griffiths, T. (2002). A common cortical substrate activated by horizontal and vertical sound movement in the human brain. Curr Biol, 12, 1584.

Peissig, J., & Kollmeier, B. (1997). Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners. J Acoust Soc Am, 101, 1660-1670.

Penner, M. J. (1975). Persistence and integration: Two consequences of a sliding integrator. Percept Psychophys, 18, 114-120.

Penner, M. J. (1977). Detection of temporal gaps in noise as a measure of the decay of auditory sensation. J Acoust Soc Am, 61, 552-557.

Perrott, D. R., & Williams, K. N. (1971). Auditory temporal resolution: Gap detection as a function of interpulse frequency disparity. Psychon Sci, 25, 73-74.

Phillips, D. P. (1985). Temporal response features of cat auditory cortex neurons contributing to sensitivity to tones delivered in the presence of continuous noise. Hear Res, 19, 253-268.

Phillips, D. P. (1999). Auditory gap detection, perceptual channels, and temporal resolution in speech perception. J Am Acad Audiol, 10, 343-354.

Phillips, D. P., & Brugge, J. F. (1985). Progress in neurophysiology of sound localization. Annu Rev Psychol, 36, 245-274.

Phillips, D. P., & Hall, S. E. (2000). Independence of frequency channels in auditory temporal gap detection. J Acoust Soc Am, 108, 2957-2963.

Phillips, D. P., & Hall, S. E. (2002). Auditory temporal gap detection for noise markers with partially overlapping and non-overlapping spectra. Hear Res, 174, 133-141.

Phillips, D. P., Hall, S. E., & Boehnke, S. E. (2002a). Central auditory onset responses, and temporal asymmetries in auditory perception. Hear Res, 167, 192-205.

Phillips, D. P., Hall, S. E., Boehnke, S. E., & Rutherford, L. E. (2002b). Spatial stimulus cue information supplying auditory saltation. Perception, 31, 875-885.

Phillips, D. P., Hall, S. E., Harrington, I. A., & Taylor, T. L. (1998). "Central" auditory gap detection: a spatial case. J Acoust Soc Am, 103, 2064-2068.

Phillips, D. P., & Irvine, D. R. (1981). Responses of single neurons in physiologically defined area AI of cat cerebral cortex: sensitivity to interaural intensity differences. Hear Res, 4, 299-307.

Phillips, D. P., Semple, M. N., Calford, M. B., & Kitzes, L. M. (1994). Level-dependent representation of stimulus frequency in cat primary auditory cortex. Exp Brain Res, 102, 210-226.

Phillips, D. P., Taylor, T. L., Hall, S. E., Carr, M. M., & Mossop, J. E. (1997). Detection of silent intervals between noises activating different perceptual channels: some properties of "central" auditory gap detection. J Acoust Soc Am, 101, 3694-3705.

Phillips, D. P., Vigneault-MacLean, B., Boehnke, S. E., & Hall, S. E. (submitted). Acoustic hemifields in the Hearing In Noise Test (HINT). Journal of the American Academy of Audiology.

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. J Acoust Soc Am, 97, 593-608.

Pinek, B., & Brouchon, M. (1992). Head turning versus manual pointing to auditory targets in normal subjects and in subjects with right parietal damage. Brain Cogn, 18, 1-11.

Plack, C. J., & Moore, B. C. (1990). Temporal window shape as a function of frequency and level. J Acoust Soc Am, 87, 2178-2187.

Plack, C. J., & Moore, B. C. (1991). Decrement detection in normal and impaired ears. J Acoust Soc Am, 90, 3069-3076.

Plack, C. J., Oxenham, A. J., & Drga, V. (2002). Linear and nonlinear processes in temporal masking. Acta Acust United Ac, 88, 348-358.

Plomp, R. (1964). Rate of decay of auditory sensation. J Acoust Soc Am, 36, 277-282.

Pollack, I., & Trittipoe, W. J. (1959a). Binaural listening and interaural noise correlation. J Acoust Soc Am, 31, 1250-1252.

Pollack, I., & Trittipoe, W. J. (1959b). Interaural noise correlation: Examination of the variables. J Acoust Soc Am, 31, 1616-1618.

Proctor, L., & Konishi, M. (1997). Representation of sound localization cues in the auditory thalamus of the barn owl. Proc Natl Acad Sci U S A, 94, 10421-10425.

Rajan, R., Aitkin, L. M., Irvine, D. R., & McKay, J. (1990). Azimuthal sensitivity of neurons in primary auditory cortex of cats. I. Types of sensitivity and the effects of variations in stimulus parameters. J Neurophysiol, 64, 872-887.

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc Natl Acad Sci U S A, 97, 11800-11806.

Rayleigh, L. J. W. S., 3rd Baron of Rayleigh). (1907). On our perception of sound direction. Philos Mag, 13, 214-232.

Recanzone, G. H., Guard, D. C., Phan, M. L., & Su, T. K. (2000). Correlation between the activity of single auditory cortical neurons and sound-localization behavior in the macaque monkey. J Neurophysiol, 83, 2723-2739.

Relkin, E. M., & Turner, C. W. (1988). A reexamination of forward masking in the auditory nerve. J Acoust Soc Am, 84, 584-591.

Rhode, W. S., & Greenberg, S. (1992). Physiology of the cochlear nuclei. In A. N. Popper & R. R. Fay (Eds.), The mammalian auditory pathway: neurophysiology (pp. 94-152). New York: Springer-Verlag.

Rhode, W. S., & Smith, P. H. (1986). Encoding timing and intensity in the ventral cochlear nucleus of the cat. J Neurophysiol, 56, 261-286.

Roberts, B., Glasberg, B. R., & Moore, B. C. J. (2002). Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. J Acoust Soc Am, 112, 2074-2085.

Robinson, D. E., & Jeffress, L. A. (1963). Effect of varying the interaural noise correlation on the detectability of tonal signals. J Acoust Soc Am, 35, 1947-1952.

Rogers, W. L., & Bregman, A. S. (1993). An experimental evaluation of three theories of auditory stream segregation. Percept Psychophys, 53, 179-189.

Rogers, W. L., & Bregman, A. S. (1998). Cumulation of the tendency to segregate auditory streams: resetting by changes in location and loudness. Percept Psychophys, 60, 1216-1227.

Roman, N., Wang, D. L., & Brown, G. J. (2002). "Speech Segregation Based on Sound Localization" (OSU-CISRC-6/02-TR16). Columbus, OH, USA: Department of Computer and Information Science, The Ohio State University.

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

Rose, M. M., & Moore, B. C. (2000). Effects of frequency and level on auditory stream segregation. J Acoust Soc Am, 108, 1209-1214.

Roth, G. L., Kochhar, R. K., & Hind, J. E. (1980). Interaural time differences: implications regarding the neurophysiology of sound localization. J Acoust Soc Am, 68, 1643-1651.

Ruggero, M. A. (1992). Physiology of the Auditory Nerve. In A. N. Popper & R. R. Fay (Eds.), The Mammalian Auditory Pathway: Neurophysiology (Vol. 2, pp. 34-93). New York: Springer-Verlag.

Saberi, K., Dostal, L., Sadralodabai, T., Bull, V., & Perrott, D. R. (1991). Free-field release from masking. J Acoust Soc Am, 90, 1355-1370.

Samson, F. K., Barone, P., Clarey, J. C., & Imig, T. J. (1994). Effects of ear plugging on single-unit azimuth sensitivity in cat primary auditory cortex. II. Azimuth tuning dependent upon binaural stimulation. J Neurophysiol, 71, 2194-2216.

Samson, F. K., Clarey, J. C., Barone, P., & Imig, T. J. (1993). Effects of ear plugging on single-unit azimuth sensitivity in cat primary auditory cortex. I. Evidence for monaural directional cues. J Neurophysiol, 70, 492-511.

Scharf, B. (1970). Critical bands. In T. J. V. (Ed.), Foundations of Modern Auditory Theory (pp. Chapter 5). New York: Academic Press.

Schneider, B. A., Daneman, M., Murphy, D. R., & See, S. K. (2000). Listening to discourse in distracting settings: the effects of aging. Psychol Aging, 15, 110-125.

Schneider, B. A., Daneman, M., & Pichora-Fuller, M. K. (2002). Listening in aging adults: from discourse comprehension to psychoacoustics. Can J Exp Psychol, 56, 139-152.

Schneider, B. A., & Hamstra, S. J. (1999). Gap detection thresholds as a function of tonal duration for younger and older listeners. J Acoust Soc Am, 106, 371-380.

Schnupp, J. W., Mrsic-Flogel, T. D., & King, A. J. (2001). Linear processing of spatial cues in primary auditory cortex. Nature, 414, 200-204.

Schwartz, I. R. (1992). The Superior Olivary Complex and Lateral Lemniscal Nuclei. In R. R. Fay & A. N. Popper (Eds.), The Mammalian Auditory Pathway: Neuroanatomy (Vol. 1, pp. 117-167). New York: Springer-Verlag.

Seneff, S. (1988). A joint synchrony/mean-rate model of auditory speech processing. Journal of Phonetics, 16, 55-76.

Shackleton, T. M., Arnott, R. H., & Palmer, A. R. (2002). Interaural correlation functions in the inferior colliculus of the guinea pig. Paper presented at the 25th MidWinter Meeting of the Association for Research in Otolaryngology.

Shackleton, T. M., McAlpine, D., & Palmer, A. R. (2000). Modelling convergent input onto interaural-delay-sensitive inferior colliculus neurones. Hear Res, 149, 199-215.

Shailer, M. J., & Moore, B. C. (1983). Gap detection as a function of frequency, bandwidth, and level. J Acoust Soc Am, 74, 467-473.

Shaw, E. A. (1974). Transformation of sound pressure level from the free field to the eardrum in the horizontal plane. J Acoust Soc Am, 56, 1848-1861.

Shaw, E. A., & Vaillancourt, M. M. (1985). Transformation of sound-pressure level from the free field to the eardrum presented in numerical form. J Acoust Soc Am, 78, 1120-1123.

Slaney, M. (1998). Auditory Toolbox: A MATLAB toolbox for auditory modelling work, Version 2. [Software]. Interval Research Corporation. Retrieved, 2002, from the World Wide Web: http://www.slaney.org/malcolm/pubs.htm.

Slattery, W. H., 3rd, & Middlebrooks, J. C. (1994). Monaural sound localization: acute versus chronic unilateral impairment. Hear Res, 75, 38-46.

Smiarowski, R. A., & Carhart, R. (1975). Relations among temporal resolution, forward masking, and simultaneous masking. J Acoust Soc Am, 57, 1169-1174.

Smith, R. L. (1977). Short-term adaptation in single auditory nerve fibers: some poststimulatory effects. J Neurophysiol, 40, 1098-1111.

Smith, R. L. (1979). Adaptation, saturation, and physiological masking in single auditory-nerve fibers. J Acoust Soc Am, 65, 166-178.

Smith, R. L. (1988). Encoding of sound intensity by auditory neurons. In G. M. Edelman & W. E. Gall & W. M. Cowan (Eds.), Auditory Function: Neurobiological Bases of Hearing (pp. 243-274). New York: Wiley.

Smith, R. L., Brachman, M. L., & Frisina, R. D. (1985). Sensitivity of auditory-nerve fibers to changes in intensity: a dichotomy between decrements and increments. J Acoust Soc Am, 78, 1310-1316.

Snell, K. B., & Hu, H. L. (1999). The effect of temporal placement on gap detectability. J Acoust Soc Am, 106, 3571-3577.

Stevens, S. S., & Newman, E. B. (1936). The localization of actual sources of sound. Am J Psychol, 48, 297-306.

Stillman, R. D. (1971). Characteristic delay neurons in the inferior colliculus of the kangaroo rat. Exp Neurol, 32, 404-412.

Taylor, T. L., Hall, S. E., Boehnke, S. E., & Phillips, D. P. (1999). Additivity of perceptual channel-crossing effects in auditory gap detection. J Acoust Soc Am, 105, 563-566.

Teller, D. Y. (1984). Linking propositions. Vision Res, 24, 1233-1246.

Tollin, D. J., & Henning, G. B. (1998). Some aspects of the lateralization of echoed sound in man. I. The classical interaural-delay based precedence effect. J Acoust Soc Am, 104, 3030-3038.

Tollin, D. J., & Henning, G. B. (1999). Some aspects of the lateralization of echoed sound in man. II. The role of the stimulus spectrum. J Acoust Soc Am, 105, 838-849.

Tollin, D. J., & Yin, T. C. (2002a). The coding of spatial location by single units in the lateral superior olive of the cat. II. The determinants of spatial receptive fields in azimuth. J Neurosci, 22, 1468-1479.

Tollin, D. J., & Yin, T. C. T. (2002b). The coding of spatial location by single units in the lateral superior olive of the cat. I. Spatial receptive fields in azimuth. J Neurosci, 22, 1454-1467.

Turgeon, M., Bregman, A. S., & Ahad, P. A. (2002). Rhythmic masking release: contribution of cues for perceptual organization to the cross-spectral fusion of concurrent narrow-band noises. J Acoust Soc Am, 111, 1819-1831.

van der Heijden, M., & Trahiotis, C. (1997). A new way to account for binaural detection as a function of interaural noise correlation. J Acoust Soc Am, 101, 1019-1022.

van Noorden, L. P. A. S. (1975). Temporal coherence in the perception of tone sequences. Unpublished Ph.D., Technische Hogeschool, Eindhoven, The Netherlands.

van Noorden, L. P. A. S. (1977). Minimum differences of level and frequency for perceptual fission of tone sequences ABAB. J Acoust Soc Am, 61, 1041-1045.

van Wieringen, A., & Wouters, J. (1999). Gap detection in single- and multiple-channel stimuli by LAURA cochlear implantees. J Acoust Soc Am, 106, 1925-1939.

Vliegen, J., Moore, B. C., & Oxenham, A. J. (1999). The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task. J Acoust Soc Am, 106, 938-945.

Vliegen, J., & Oxenham, A. J. (1999). Sequential stream segregation in the absence of spectral cues. J Acoust Soc Am, 105, 339-346.

Wallace, M. N., Rutkowski, R. G., Shackleton, T. M., & Palmer, A. R. (2000). Phase-locked responses to pure tones in guinea pig auditory cortex. Neuroreport, 11, 3989-3993.

Wallace, M. N., Shackleton, T. M., & Palmer, A. R. (2002). Phase-locked responses to pure tones in the primary auditory cortex. Hear Res, 172, 160-171.

Walton, J. P., Frisina, R. D., Ison, J. R., & O'Neill, W. E. (1997). Neural correlates of behavioral gap detection in the inferior colliculus of the young CBA mouse. J Comp Physiol [A], 181, 161-176.

Wightman, F. L., & Kistler, D. (1993). Sound Localization. In W. A. Yost & A. N. Popper & R. R. Fay (Eds.), Human Psychophysics (Vol. 3, pp. 155-192). New York: Springer-Verlag.

Wightman, F. L., & Kistler, D. J. (1989a). Headphone simulation of free-field listening. I: Stimulus synthesis. J Acoust Soc Am, 85, 858-867.

Wightman, F. L., & Kistler, D. J. (1989b). Headphone simulation of free-field listening. II: Psychophysical validation. J Acoust Soc Am, 85, 868-878.

Wightman, F. L., & Kistler, D. J. (1992). The dominant role of low-frequency interaural time differences in sound localization. J Acoust Soc Am, 91, 1648-1661.

Wightman, F. L., & Kistler, D. J. (1997a). Factors affecting the relative salience of sound localization cues. In R. H. Gilkey & T. R. Anderson (Eds.), Binaural and Spatial Hearing (pp. 1-23). Mawah, NJ: Lawrence Erlbaum Associates, Inc.

Wightman, F. L., & Kistler, D. J. (1997b). Monaural sound localization revisited. J Acoust Soc Am, 101, 1050-1063.

Williams, K. N., & Elfner, L. F. (1976). Gap detection with three auditory events -- a single-channel process. J Acoust Soc Am, 60, 423-428.

Williams, K. N., Elfner, L. F., & Howse, W. R. (1978). Auditory temporal resolution: effects of sensation level. J Aud Res, 18, 265-269.

Williams, K. N., & Perrott, D. R. (1972). Temporal resolution of tonal pulses. J Acoust Soc Am, Vol. 51, 644-647.

Woodworth, R. S. (1938). Experimental Psychology. New York: Holt.

Wright, B. A., Lombardino, L. J., King, W. M., Puranik, C. S., Leonard, C. M., & Merzenich, M. M. (1997). Deficits in auditory temporal and spectral resolution in language-impaired children. Nature, 387, 176-178.

Yates, G. K. (1995). Cochlear structure and function. In B. C. J. Moore (Ed.), Hearing (pp. 41-74). San Diego, CA: Academic Press, Inc.

Yin, T. C., & Kuwada, S. (1983). Binaural interaction in low-frequency neurons in inferior colliculus of the cat. III. Effects of changing frequency. J Neurophysiol, 50, 1020-1042.

Yin, T. C. T., & Chan, J. C. K. (1990). Interaural Time Sensitivity in Medial Superior Olive of Cat. J Neurophysiol, 64, 465-488.

Young, E. D., Nelken, I., & Conley, R. A. (1995). Somatosensory effects on neurons in dorsal cochlear nucleus. J Neurophysiol, 73, 743-765.

Zatorre, R. J., Bauffard, M., Ahad, P., & Belin, P. (2002). Where is 'where' in the human auditory cortex? Nature Neuroscience, 5, 905-909.

Zhang, W., Salvi, R. J., & Saunders, S. S. (1990). Neural correlates of gap detection in auditory nerve fibers of the chinchilla. Hear Res, 46, 181-200.

Zurek, P. M. (1993). Binaural advantages and directional effects in speech intelligibility. In G. A. Studebaker & I. Hochberg (Eds.), Acoustical Factors Affecting Hearing Aid Performance II. Boston: Allyn and Bacon.

Zwicker, E. (1965). Temporal effects in simultaneous masking by white-noise bursts. J Acoust Soc Am, 37, 653-663.