Lateralization perception

The role of interfering noise and temporally varying binaural cues

Darrin K. Reed

Lateralization perception

The role of interfering noise and temporally varying binaural cues

Von der Fakultät VI für Medizin und Gesundheitswissenschaften
der Carl von Ossietzky Universität Oldenburg
zur Erlangung des Grades und Titels eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
angenommene Dissertation

von Herr
Darrin K. Reed
geboren am 17. November 1982
in Nampa, Idaho USA

Gutachter: Prof. Dr. Steven van de Par
Weitere Gutachter: Prof. Dr. Volker Hohmann

Tag der Disputation: 23. March 2015

Contents

Summary 1

Zusammenfassung 3

1 General Introduction 5

1.1 Objectives . . . 9

1.2 Outline of this thesis . . . 9

2 Lateralization of noise bursts in interaurally correlated or uncorrelated background noise using interaural level differences 11

2.1 Introduction . . . 12

2.2 Methods . . . 13

2.2.1 Procedure . . . 14

2.2.2 Stimuli . . . 14

2.2.3 Listeners . . . 16

2.3 Psychophysical Experiments . . . 16

2.3.1 Experiment 1 Results . . . 16

2.3.2 Experiment 2 Results . . . 20

2.3.3 Discussion . . . 21

2.4 Binaural Localization Model . . . 24

2.4.1 Model Description . . . 24

2.4.2 Model Results & Discussion . . . 27

2.5 General Discussion . . . 29

3 Characterizing perceptual properties of a binaurally modulated stimulus 31

3.1 Introduction . . . 32

3.2 Experiment 1: Binaural modulation detection . . . 34

3.2.1 Methods . . . 34

3.2.2 Results . . . 36

3.3 Experiment 2: Binaural modulation frequency discrimination . . . 39

3.3.1 Methods . . . 40

3.3.2 Results . . . 41

3.4 Experiment 3: Modulation masking between monaural and binaural modulations 42
3.4.1 Methods 43
3.4.2 Results 45
3.5 Discussion 48
3.6 Conclusion 56

4 Lateralization of alternating interaural time difference stimuli: The role of the monaural envelope 57

4.1 Introduction 58
4.2 Experiment 1: Lateralization without monaural cues 59
4.2.1 Stimuli 60
4.2.2 Procedure 61
4.2.3 Results 62
4.3 Experiment 2: Influence of monaural envelope cues on lateralization 67
4.3.1 Stimuli 67
4.3.2 Procedure 67
4.3.3 Results 68
4.4 Experiment 3: Influence of monaural cues and bandwidth on ITD sensitivity . . 69
4.4.1 Stimuli 69
4.4.2 Procedure 70
4.4.3 Results 71
4.5 Experiment 4: Influence of spatial separation on AM source segregation 73
4.5.1 Stimuli 73
4.5.2 Procedure 73
4.5.3 Results 74
4.6 Discusion 75
4.7 Conclusion 80

5 General Conclusions 81

5.1 Suggestions for future research 85

Bibliography 88

Acknowledgments 97

Curriculum Vitae 98

Summary

It is truly remarkable that the human auditory system is able to extract a plethora of information from only two, time-varying pressure signals, especially considering that multiple sound sources are often occurring at the same time. From these two input signals, listeners can effortlessly attend to sound sources of interest while ignoring others in situations of communication, entertainment, and survival. Although humans today probably face a reduced evolutionary pressure as compared to that which motivated the need for two ears, we are still able to benefit from the result. The binaural system provides lateral estimates of sound source position resulting from timing and level differences at the ears, namely, the interaural time difference (ITD) and interaural level difference (ILD). In addition, humans are also sensitive to how similar the signals are between the ears, i.e. the interaural coherence, which influences the perceived compactness of the auditory scene. These elements of spatial processing have been demonstrated to contribute to the ability of listeners to make sense of complex auditory scenes.

The primary objective of this dissertation is the understanding of how the auditory system utilizes binaural information in listening scenarios that involve more than a single target source. In the first study, the ability of listeners to identify the absolute lateral position of an ILD-lateralized noise stimulus, which was simultaneously presented with an interfering background noise, was measured. As a result of the background noise, the physical ILD at the ears does not correspond to the target ILD, but rather corresponds to an ILD between that of the target ILD and the background noise ILD. Results of this study indicate that a temporal onset/offset asynchrony (TOA) can permit most listeners to use the target ILD (or at least an ILD much larger than the physical ILD at the ears) for lateralization. In the absence of a TOA, listeners tend to use the physical ILD for an interaurally correlated background noise, but, interestingly, many listeners are able to utilize the target ILD if the background noise is interaurally uncorrelated.

In the second study, the temporal acuity of the binaural system was measured using a stimulus that periodically alternated between two ITD values. Discrimination between this stimulus with temporally varying binaural cues and a stimulus with stationary binaural cues is possible for changes in binaural information that only last for a few milliseconds. This ITD-switching stimulus was reported to contain a modulation percept, which is truly binaural in nature, i.e. the temporal envelope in each ear corresponds to an unmodulated white noise. Experimental results show that this purely binaural modulation percept is discriminable in frequency and is also sufficiently compact for sidedness judgements. Since it has been shown in the monaural domain that modulation masking effects exist, it was of interest to investigate whether such masking interactions also exist between monaural and binaural modulations. Results indicate a relatively weak interaction, suggesting that monaural and binaural modulation sensitivities are largely independent.

In the third study, the ability of listeners to make absolute lateralization estimates of brief glimpses of ITD information was evaluated using the previously mentioned stimulus that rapidly switches between two ITD values. This lateralization experiment indicates that listeners are not only able to detect brief changes in ITD, but are also able to make a reasonably accurate lateralization estimate of these brief ITD-lateralized noise tokens. Results also show that the lateralization estimates with this broadband stimulus are not influenced by the presence of an amplitude modulation (AM) imposed onto the stimulus. For stimuli of smaller bandwidths, the monaural envelope cues, however, appear to influence the temporal weighting of the binaural cues within an AM cycle. Finally, the role of binaural information when discriminating between the envelope frequencies of two simultaneous AM noises was measured. Results indicate that binaural cues alone can help segregate the sound sources in order to identify the envelope frequency of at least one of the AM noises.

Zusammenfassung

Das menschliche auditorische System besitzt die Fähigkeit, eine Fülle an Informationen aus nur zwei, zeitvarianten Drucksignalen zu gewinnen. Bemerkenswert dabei ist, dass das auch gelingt, wenn mehrere unterschiedliche Quellen gleichzeitig Geräusche von sich geben. Aus den zwei Drucksignalen können Zuhörer ohne größere Anstrengung einzelnen Geräuschquellen folgen und andere Geräusche von etwa Gesprächen, Unterhaltungsmedien oder der Umgebung ignorieren. Der evolutionäre Druck, der im Laufe der Evolution dazu geführt hat, zwei Ohren statt nur einem zu entwickeln, lastet heutzutage wahrscheinlich nicht mehr auf dem Menschen, aber trotzdem profitieren wir immer noch von dieser Entwicklung. Das binaurale System liefert Abschätzungen über die laterale Position einer Geräuschquelle auf der Basis von Unterschieden in den Signalen zwischen beiden Ohren, nämlich interaurale Zeitunterschiede (interaural time difference, ITD) und interaurale Pegelunterschiede (interaural level difference, ILD). Außerdem ist der Mensch sensitiv auf die Ähnlichkeit der Signale beider Ohren. Die interaurale Koheränz beeinflusst dabei unter anderem die wahrgenommene Kompaktheit einer auditorischen Szene. Die eben beschriebenen Elemente der räumlichen Verarbeitung helfen dem auditorischen System nachweislich dabei, eine komplexe akustische Szene sinnvoll aufzulösen. Diese Dissertation verfolgt daher das Hauptziel, zu verstehen, wie das auditorische System binaurale Informationen in Hörsituationen nutzt, in denen mehr als eine Signalquelle vorhanden sind. Dazu wurden drei Hörexperimente mit normalhörenden Probanden durchgeführt.

In der ersten Studie wurde die Fähigkeit von Probanden gemessen, die absolute, laterale Position eines Zielrauschens mit aufgeprägter ILD in einem maskierenden Hintergrundrauschen zu bestimmen. Als Folge des maskierenden Rauschens, stimmte die resultierende, physikalische ILD an den Ohren nicht mit der ILD des Zielrauschens überein, sondern entsprach einer ILD zwischen der des Zielrauschens und des maskierenden Rauschens. Die Ergebnisse der Studie zeigen, dass eine Asynchronizität zwischen Anfang und Ende des Zielrauschens gegenüber dem des Hintergrundrauschens, den Probanden das Nutzen der ILD des Zielrauschens, oder zumindest einer deutlich größeren ILD als der physikalisch an den Ohren vorhandenen, zur Lateralisation erlaubt. Wurden Zielrauschen und Hintergrundrauschen synchron abgespielt, tendierten die meisten Probanden dazu, die physikalisch vorhandene ILD des Gesamtsignals zur Lateralisation zu nutzen, zumindest dann, wenn das Hintergrundrauschen interaural korreliert dargeboten wurde. Bei unkorrelliertem Hintergrundrauschen zwischen beiden Ohren, waren viele Probanden

wieder in der Lage die ILD des Zielrauschens zu nutzen.

In der zweiten Studie wurde die zeitliche Genauigkeit des binauralen Systems mit einem Stimulus gemessen, dem periodisch abwechselnd zwei unterschiedliche Werte als ITD aufgeprägt wurden. Dieser binaural zeitvariante Stimulus konnte selbst dann von einem binaural zeitinvarianten Stimulus unterschieden werden, wenn die zeitvarianten Anteile nur ein paar Millisekunden lang waren. Bei dem Stimulus mit wechselnder ITD wurde eine Modulation wahrgenommen, die rein binaural funktioniert, da die zeitliche Einhüllende des Signals, an jedem Ohr einzeln betrachtet, einem unmodulierten weißen Rauschen entspricht. Die Ergebnisse des Experiments zeigen, dass diese rein binaural wahrnehmbare Modulation in ihrer Frequenz unterscheidbar ist und auch ausreichend kompakt wahrgenommen wird, um zu entscheiden, ob das Signal links oder rechts lateralisiert. Frühere Studien haben gezeigt, dass es zwischen monauralen Modulationen zu Maskierungseffekten kommen kann. Deshalb wurde untersucht, ob diese maskierende Interaktion auch zwischen monauraler und binauraler Modulation auftritt. Die Ergebnisse der Studie zeigen nur eine schwache Interaktion, weshalb keine größere Abhängigkeit der Sensivitäten von monauralen und binauralen Modulationen anzunehmen ist.

In der dritten Studie wurde der im vorigen Abschnitt beschriebene Stimulus mit schnell abwechselnder ITD verwendet, um die Fähigkeit von Probanden zu testen, aus einem äußerst kurzen Zugriff auf die ITD, die absolute laterale Position eines Signals abzuschätzen. Die Ergebnisse der Studie deuten darauf hin, dass Probanden nicht nur in der Lage sind, kurze Veränderungen der ITD wahrzunehmen, sondern auch die laterale Position der kurzen Abschnitte einigermaßen genau zu bestimmen. Die Ergebnisse der Studie zeigen außerdem dass die Bestimmung der Lateralisation mit diesem breitbandigen Stimulus nicht von einer aufgeprägten Amplitudenmodulation beeinflusst wird. Bei schmalbandigen Stimuli hingegen, scheinen die monauralen Merkmale der Einhüllenden, die zeitliche Gewichtung der binauralen Merkmale innerhalb eines Durchlaufs der Amplitudenmodulation zu beinflussen. Zuletzt wurde die Unterscheidbarkeit der Einhüllendenfrequenzen von zwei gleichzeitig dargebotenen binauralen Rauschsignalen mit unterschiedlich modulierter Amplitude untersucht. Dabei zeigen die Ergebnisse, dass nur binaurale Merkmale dabei helfen können, die Geräuschquellen zu trennen, um zumindest eine der beiden Einhüllendenfrequenzen der amplitudenmodulierten Rauschen zu identifizieren.

1
General Introduction

The auditory system provides a large amount of information concerning the content and spatial impression of our surroundings. In normal daily situations, the acoustical signals reaching our ears are complex mixtures of multiple sound sources often combined with attributes of the given room acoustics. The exact manner in which the human auditory system can retrieve information from desired sound sources in these complex listening scenarios is largely still unknown. It is widely accepted, however, that listeners use a combination of cues, e.g. spectral, temporal, and spatial, to disentangle auditory scenes (Bregman, 1990). The investigations of this thesis will focus on the role of spatial cues, namely interaural level differences (ILD) and interaural time differences (ITD).

The ability of a listener to understand a particular talker in the presence of a distracting talker has been termed the 'cocktail party problem' (Cherry, 1953). As talkers in natural settings typically do not originate from the same spatial location, great effort has been afforded to the investigation of the binaural system as a contributor when processing complex auditory scenes. Traditionally, the investigations of the binaural system were largely directed at binaural masking level difference (BMLD) measurements, e.g. Licklider (1948). In various ways, BMLD studies have shown that disparate interaural phase differences between simultaneously presented target and masking sounds can improve the detection thresholds of the target. A small BMLD has also been shown for an interaurally uncorrelated masker provided that the target is interaurally in phase (Robinson and Jeffress, 1963). Note that although the improvement in BMLD thresholds as a result of using an interaurally uncorrelated masker is smaller than those obtained with an interaurally correlated masker and an interaurally anti-phasic target, the interaural coherence does appear to influence how the auditory system dissects complex auditory scenes. In a model by Faller and Merimaa (2004), selection of binaural cues in moments where the interaural coherence is high can predict performance from several psychophysical studies in literature. Changes in interaural coherence are directly related to the magnitude and location of the cross-correlation peak, which are assumed to be the basis for estimates of the ITD in cross-correlation models of spatial perception. However, there is no direct relationship between the normalized cross-correlation function and the ILD. This poses the question as to whether or not the coherence of the auditory scene influences ILD perception.

It is possible that the perceived position of a target sound can be altered by the presence of a background noise as compared to making a lateral estimate of the target in isolation. In these multi-source listening conditions, some studies have reported that listeners perceive the position of the target being "pushed" further away from the interfering noise (Thurlow and Jack, 1973, Canévet and Meunier, 1996, Carlile *et al.*, 2001, Braasch and Hartung, 2002). Other studies, however, have reported that the perceived position of the target is "pulled" toward the position of the interfering noise (Butler and Naunton, 1962, Heller and Trahiotis, 1996, Best *et al.*, 2007). As suggested by Lee *et al.* (2009), whether or not the target is pushed or pulled relative to the position of the interfering source could be related to how perceptually distinct the sound objects are, e.g. as a result of spectral and temporal differences. For ITD-lateralized noise sources that are both temporally and spectrally overlapping, selection of the appropriate peak in the cross-correlation function would permit a reasonably accurate estimate of the target's lateral position. For similar ILD-lateralized noise sources, the physical ILD at the ears is an an average of the individual sound source ILDs that are weighted depending upon the relative levels of the individual sound sources. Even if an Equalization-Cancellation (EC) calculation was adopted (Durlach, 1963, Breebaart *et al.*, 2001), which is sensitive to changes in ILD, only the overall ILD resulting from the multiple sound sources would be predicted. It is not clear whether this prediction of the overall ILD is appropriate for human listeners. If listeners are indeed able to use the target ILD (or at least not the physical ILD at the ears), the auditory system must be capable of compensating for the presence of the background noise. As suggested in a model by Braasch (2003), knowledge of the background noise in isolation can permit an accurate prediction of the target ILD. If the onsets and offsets of the target and background noise sources occur simultaneously, an assumption about the background noise would be necessary in order to maintain an accurate prediction of the target ILD.

The temporal acuity of the auditory system to detect changes in binaural information has been previously shown to be "sluggish" for a variety of listening tasks. For the tracking of movement in detail, the binaural system does not appear to possess the ability to exceed rates beyond 3-5 Hz (Perrott and Musicant, 1977, Grantham and Wightman, 1978). A similar low-pass characteristic was observed for sensitivity to sinusoidal changes in interaural correlation (Grantham, 1982). The required change in correlation at threshold steadily increased between 1 and 10 Hz until the maximum difference in correlation was reached at a rate of 50 Hz. Considering previously shown monaural sensitivity to amplitude modulation (AM) with nearly constant thresholds up to 10-20 Hz (Viemeister, 1979), the results of Grantham (1982) were interpreted to illustrate the binaural system as being sluggish compared to the monaural system. Sensitivity to time varying BMLD paradigms have also suggested that the auditory system is relatively insensitive to rapid changes in binaural information (Kollmeier and Gilkey, 1990, Culling and Summerfield, 1998). In contrast, Pollack (1978) has shown that rapid switching between a noise being interaurally homophasic or antiphasic can be detected for 2-4 ms durations of constant interaural phase relationship. Sensitivity to such brief changes in ITD and ILD has also been shown by Bernstein *et al.* (2001). Provided a sufficiently large interaural disparity, listeners were sensitive to interaural changes as short as 2 ms. Discrepancies in the length of the temporal window for the binaural system are likely related to the type of task presented (Kollmeier and Gilkey, 1990). Discrepancies in temporal acuity could also stem from differences in the construction of the stimuli. As demonstrated by Siveke *et al.* (2008), a stimulus that contains both time-varying ITD and interaural cross correlation can elicit a more acute temporal sensitivity as compared to stimuli with time-varying interaural cross correlation alone. It is interesting to note that by an appropriate choice for the peripheral model, the binaural model by Breebaart *et al.* (2001) can predict a range of psychophysical results for binaural tasks with a single time constant of 30 ms.

Although many studies have shed light on the sensitivity of listeners to the interaural relationships of isolated sound stimuli, it is also important to determine the role of binaural cues for processing cocktail party scenarios. It appears that for source *segregation*, monaural cues resulting from spectral separation provide a stronger benefit than spatial separation of concurrent speech sources (Buell and Hafter, 1991, Shackleton and Meddis, 1992, Culling and Summerfield, 1995). This does not mean that spatial cues cannot provide a benefit to source segregation but rather that the auditory system attributes a greater weight to spectral cues when forming auditory objects. Results from Darwin and Hukin (1999) suggest that listeners likely attend to auditory objects at a perceptual location instead of attending to the spectral components associated with common interaural cues. In other words, the auditory system seems to first form auditory objects from spectral and temporal cues (i.e. monaural information) and subsequently attribute a location to these objects instead of forming objects based on common binaural cues. Spatial information could, therefore, be more suited to source *streaming*, which occurs subsequent to the segregation of the spectro-temporal structure into auditory objects. Darwin and Hukin (2000) have shown that spatial separation can strongly influence how words are linked to an attended sentence. This suggests that binaural information can be important when connecting short-term auditory objects into perceptual streams. The ability to attend to a target stream in the presence of interfering sound information, however, is probably not determined by the specific value of an interaural parameter applied to the target. Sach and Bailey (2004) found evidence that the perceived lateral separation between the target and the interfering sound sources, rather than the specific differences in ITD or ILD, predicts the ability of listeners to hear out the target stream. This suggests that a perceived spatial separation would be required for binaural information to provide a benefit for source streaming. Because "*maintenance of the auditory stream is vulnerable to uncertainty of the target location*" (Kidd *et al.*, 2005), an ability to periodically reinforce the target position should provide a benefit when making judgements concerning the target stream. The resolution with which the auditory system can sample desired binaural glimpses would, therefore, limit the role of spatial information in the streaming of auditory objects.

To measure the temporal acuity of the auditory system for brief glimpses of binaural information, a stimulus inspired by Wagner (1991) and Bernstein *et al.* (2001) was developed for experiments of this thesis. The stimulus was a continuous, broadband Gaussian noise that periodically alternated between two different ITD values. Listeners from Bernstein *et al.* (2001) demonstrated high temporal acuity to changes in the binaural cues, but it is unknown whether or not these listeners could attribute an accurate lateral position to these brief changes in ITD. Absolute lateralization accuracy measurements were conducted using the ITD-switching stimulus developed for this thesis. Whether or not these brief changes in ITD can be accurately lateralized could potentially be influenced by monaural information. Schimmel *et al.* (2008) have shown that two temporally and spectrally overlapping ITD-lateralized sources could only be accurately lateralized when the target source had a high temporal peakedness. Although their result could be interpreted to suggest that monaural envelope cues are required for rapid processing of binaural information, whether or not envelope cues are truly essential for a temporally acute binaural system remains to be answered. A role of monaural envelope information for the processing of binaural information has been demonstrated by Dietz *et al.* (2013). In their study, listeners placed a greater perceptual weight to the instantaneous interaural phase difference occurring during the rising portion of a sinusoidal AM. It is uncertain, however, if this perceptual weighting will be observed for stimuli of greater bandwidth.

Temporal modulations in amplitude are known to be important for speech intelligibility. Only a few frequency bands of noise applied as an artificial carrier signal are sufficient to achieve high accuracy for consonant, vowel, and sentence recognition if temporal modulations of the speech below 50 Hz are maintained (Shannon *et al.*, 1995). Furthermore, degradations to the low-frequency modulation envelopes of speech reduce intelligibility (Drullman *et al.*, 1994). An impairment in the ability to detect these modulations could, therefore, also impair speech intelligibility. Detection of a target AM has been shown to degrade in the presence of an interfering AM (Houtgast, 1989, Ewert and Dau, 2000) where the ability to detect the target AM improves with greater separation in modulation frequency between the target and interfering AM. Upon listening to the ITD-switching stimulus developed for lateralization investigations of this thesis, it was observed that the stimulus can elicit a percept of modulation. Thompson and Dau (2008) have shown an impeded ability to detect AM in the presence of sinusoidally varying ILD cues. Since ILDs inherently cause changes in monaural level, it is impossible to determine if the elevation in AM detection thresholds was truly a result of the time-varying binaural information or simply a result of monaural modulations consequently produced by the binaural (ILD) modulation. Due to the synchronous and simultaneous switching between the two ITD-lateralized noises for the ITD-switching stimulus of this thesis, the modulation percept is purely binaural in nature. This stimulus, therefore, allows a more direct investigation as to whether or not an interference between modulation perception in the monaural and binaural domains exist.

1.1 Objectives

The overall aim of this thesis is to characterize the behavior of the binaural system in the presence of both target and interfering noise sources. The first objective is to determine if listeners are obligated to use the physical ILD cues at the ears for sound source lateralization or if the auditory system retains the ability to make proper lateral estimates corresponding to the target ILD in the presence of an interfering sound source. Furthermore, it is of interest to understand if the coherence of the interfering noise influences the manner in which listeners interpret the target position of an interaurally correlated, ILD-lateralized noise. A second objective is to measure the temporal acuity for which the auditory system can select glimpses of target binaural information. Since the binaural cues in a multi-talker listening scenario can rapidly change due to the spectro-temporally sparse nature of speech, knowledge of the timescale on which the binaural system operates can inform us about the utility of binaural information in these complex listening scenarios. The third objective is to investigate the role of monaural cues for the processing of binaural cues. Prior literature suggests that the monaural envelope could influence whether the binaural system can process rapid changes or possibly influence the perceptual weight of binaural cues depending on the temporal position within the envelope. The fourth objective is to measure if interference in the processing of binaural and monaural modulation information exists. The experiments presented here are directed at a better understanding of the interaction between the monaural and binaural pathways in lateralization, stream segregation, and modulation perception.

1.2 Outline of this thesis

The focus of **Chapter 2** is to determine if listeners are restricted to using the physical ILD at the ears when making lateral estimates of an ILD-lateralized noise burst in the presence of an interfering background noise that is spectrally and temporally overlapping. When the background noise can be observed in isolation, it is possible that the auditory system uses this information to estimate the target ILD when the target and background noise are presented together. However, when the target and background noise do not have a temporal onset/offset asynchrony (TOA), this would no longer be possible. Besides testing of a zero and non-zero TOA, whether or not the coherence of the background noise influences ILD perception was measured in these experiments. In contrast to the ITD, the ILD does not have a direct relationship to the normalized cross-correlation function. Subsequent to the psychophysical experiments, a basic extension to an EC model is implemented to account for the observed lateralization performance.

The percept of modulation resulting from a noise stimulus that rapidly alternates between two ITD values is explored in **Chapter 3**. Whether or not listeners can discriminate the alternating ITD cues from a stimulus containing stationary ITD cues with a similar long-term cross-correlation pattern is measured. Performance on the task is evaluated across a range of ITD-switching frequencies. Due to the percept of modulation that results from this ITD-switching stimulus as described in pilot studies, a secondary experiment is to measure if this modulation percept is discriminable in frequency and if the modulation percept occupies a lateral position corresponding to one of the ITDs. As the ability to detect a target AM has been shown to decrease in the presence of an interfering AM, it is conceived that the modulation percept resulting from the ITD switching, which is purely binaural in nature, could also interfere with detection of AM. Therefore, whether or not an interaction exists between the monaural and binaural modulation domains is measured.

Although prior studies have demonstrated that the temporal acuity of the binaural system appears to be able to detect changes in binaural information as brief as a few milliseconds, whether or not listeners can attribute an accurate lateral position to these brief tokens is unknown. In **Chapter 4**, lateralization performance is measured using the ITD-switching stimulus applied in Chapter 3. To determine if monaural envelope information influences the ability to lateralize these brief noise segments, the absolute measurements are also conducted when a diotic AM is imposed onto the ITD-switching stimulus. By setting the AM frequency and ITD-switching frequency equal, it is possible to temporally position the ITD target within the AM cycle to determine if a greater perceptual weight is placed on the binaural information during different parts of the AM cycle. Whether or not an emphasis to binaural information in different parts of the AM cycle occurs for stimuli of different bandwidths is also measured. The final experiment of this chapter determines if spatial separation between two simultaneously presented broadband AM noises provides a benefit for the task of discriminating the envelope frequencies.

Finally, **Chapter 5** summarizes the main findings of this thesis and suggests future research.

2

Lateralization of noise bursts in interaurally correlated or uncorrelated background noise using interaural level differences[1]

The interaural level difference (ILD) of a lateralized target source may be effectively reduced when the target is presented together with background noise containing zero ILD. It is not certain whether listeners perceive a position congruent with the reduced ILD or the actual target ILD in a lateralization task. Two sets of behavioral experiments revealed that many listeners perceived a position at or even larger than the presented target ILD when a temporal onset/offset asynchrony (TOA) between the broadband target and the broadband background noise was present. When no TOA was present, however, the perceived lateral position indicated a dependency on the coherence of the background noise for several listeners. With interaurally correlated background noise, listeners reported a reduced ILD resulting from the combined target and background noise stimulus. In contrast, several of the listeners made a reasonable estimate of the target ILD for interaurally uncorrelated, broadband background noise. No obvious difference in performance was seen for low- or high-frequency only stimuli. Extension of a weighting template to the output of a standard Equalization Cancellation (EC) model was shown to remove the lateral bias on the predicted target ILD resulting from the presence of background noise. Provided that an appropriate weighting template is applied based on knowledge of the background noise coherence, good prediction of the behavioral data is possible.

1 This chapter is based upon a revised manuscript submitted for publication to the *J. Acoust. Soc. Am.*

2.1 Introduction

Sound sources in isolation will generally provide reliable acoustic localization cues, but interfering noise and reverberation can adversely affect the localization cues arriving at the ears. The purpose of this study is to provide a better understanding of the perceptual effect on sound lateralization when the target sound is presented with background noise. In addition to psychophysical performance, an analytical model is provided to help facilitate the understanding of a possible computational mechanism employed by listeners when making lateral position judgements in the presence of background noise.

Localization acuity of the two primary cues used for sound localization on the horizontal plane, i.e. the interaural time difference (ITD) and the interaural level difference (ILD), has been well established for both free-field and headphone presentations (e.g. Grantham, 1995). These binaural cues are altered, however, when reverberation or interfering sound sources are presented in addition to the isolated target source (Rakerd and Hartmann, 1985, Shinn-Cunningham *et al.*, 2005, May *et al.*, 2011). The perceived position of target objects may change due to the alteration of these primary localization cues. In anechoic environments, the low-frequency ITD information appears to be the dominant localization cue (Wightman and Kistler, 1992), but for sound localization in reverberant or multi-source conditions, the high-frequency information may be more accurately used (Ihlefeld and Shinn-Cunningham, 2011, Rakerd and Hartmann, 2010, Brungart and Simpson, 2009). It is well known that the ILD is the primary binaural cue used for localization of high-frequency sounds; therefore, it could play a significant role for localization in complex acoustical situations.

Because non-anechoic conditions and situations with more than a single sound source can cause alterations of the interaural cues, it is interesting to compare the physical ILD entering the auditory system with the perceived laterality for a target ILD stimulus combined with a background noise. When a target noise stimulus with a given ILD is mixed with a background noise stimulus with zero ILD, the resulting combined stimulus will have an overall ILD that is smaller than the target ILD. This ILD of the target plus background noise will be denoted as the 'aggregate ILD' for the remainder of this paper. Although studies investigating sound localization accuracy in the presence of interfering noise have shown a dependence on the SNR (Good and Gilkey, 1996, Lorenzi *et al.*, 1999), it is not clear to what extent listeners utilize the actual target ILD, the aggregate ILD of the overall stimulus, or the low-frequency ITD for the localization task.

Previous headphone and free-field studies concerning sound localization tasks with background noise have reported conflicting conclusions. Some show a pushing effect, i.e. perceived position of the target being away from the interfering noise (Thurlow and Jack, 1973, Canévet and Meunier, 1996, Carlile *et al.*, 2001, Braasch and Hartung, 2002), while other studies have reported a pulling effect, i.e. perceived position of the target being toward the interfering noise (Butler and Naunton, 1962, Heller and Trahiotis, 1996, Best *et al.*, 2007). It is possible, however, that these discrepancies can be attributed to the different properties of the stimuli tested. Lee *et al.* (2009) showed that differences between pushing and pulling depend on the spectral and temporal relationships between the target and other sources. They suggest that spectrally and temporally similar sound elements will be bound together resulting in the target sound being pulled toward the other congruent sound elements. Pushing will occur to the extent that two distinct sound objects are perceived where the degree of pushing decreases with increasing temporal separation between the two sounds.

The coherence of the interfering noise has been shown to affect the just-noticeable difference (jnd) between ITDs and absolute localization accuracy of a target sound (Robinson and Egan, 1974, Cohen, 1981, Trahiotis and Bernstein, 1990). Since an ITD results in a shift of the interaural cross-correlation maximum, the reduction in spatial acuity can be attributed to the broadening of the cross-correlation peak due to a decrease in interaural coherence. The ILD, however, has no direct effect on the normalized cross-correlation function. This raises a question as to whether ILD-lateralized targets may also be susceptible to the perceptual changes resulting from interfering noise of varying degrees of coherence.

The goal of the current study was to investigate the processing of ILD cues for lateralization of a target sound that was combined with a spectrally overlapping interfering noise. More specifically we were interested in determining whether listeners use the physical ILD at the ears, i.e. resulting in a pulling effect, or if listeners are capable of using the ILD of the target alone. The investigation was composed of two separate sets of experiments: one focused on broadband stimuli and another that utilized low- and high-frequency noise stimuli in addition to the broadband stimuli in order to look at frequency-dependent effects on ILD perception. To supplement the behavioral investigation, we show how an equalization-cancellation (EC) model, (e.g. Durlach, 1963, Breebaart *et al.*, 2001), can be modified to predict lateralization performance trends seen in the psychophysical results.

2.2 Methods

Two separate sets of headphone lateralization experiments were conducted in order to better understand how ILD cues are processed when both a target and interfering source are present. Experiment 1 focused on broadband stimuli whereas Experiment 2 provided an extension to investigate the contribution of low- and high-frequency regions to ILD-based lateralization. The fundamental aim of this study was to determine if listeners simply use the ILD resulting from the combination of the two sources or if listeners are able to independently utilize the target ILD.

2.2.1 Procedure

To determine the perceived position of an ILD-lateralized target noise stimulus presented over headphones, a three-interval acoustic pointer task was used. An ILD acoustic pointer has been successfully used for reliable measurements of target source laterality in other studies (Bernstein and Trahiotis, 1985, Best *et al.*, 2007, Lee *et al.*, 2009). The acoustic pointer task was designed so that the first and third intervals contained an interaurally correlated target stimulus temporally centered within an interfering background noise. The second interval contained the acoustic pointer which was an interaurally correlated noise stimulus with the same spectral properties as the target in the first and third intervals but did not contain background noise. Within each trial, the target ILDs of the first and third intervals were always the same. The ILD of the second interval (containing the target alone) was adjusted left or right by the listener to match the perceived lateral position of the target in the first and third intervals by pressing one of two buttons on a standard computer keyboard. The starting ILD for the second interval was randomized between ±15 dB. The initial step size (4 dB) in the adaptive procedure was halved after two direction reversals until the minimum step size was 1 dB where two more reversals were completed. This adaptive adjustment of the acoustic pointer was similar to methods employed by Dietz *et al.* (2009). The reported ILD for each track was the average of the pointer ILD on each of the direction reversals while the step size was 1 dB. For Experiment 2 (described below), the number of reversals at the minimum step size was increased to six. [2]

Prior to beginning the experiments, participants were given specific instructions for adjusting the acoustic pointer. If two sound sources were perceived, listeners were instructed to adjust the pointer to the furthest lateral sound source. If only one sound source was perceived, listeners were instructed to adjust the pointer to the center of gravity of the sound object. During the experiments and in prior pilot studies with stimuli similar to those used in the experiments, listeners never reported confusion with the task due to hearing more than two sound sources or split images.

2.2.2 Stimuli

In all experiments, the target and background noise were spectrally overlapping Gaussian noise bursts; thus, no spectral cues were available to the listeners to help distinguish the target and background noise. The target stimulus was always 300 ms in duration and temporally centered within the background noise. A pause of 400 ms was inserted between each of the intervals. For each interval, new noise tokens were generated for both the target and background noise. Since this lead to slight variations in the lateral perception of the first and third intervals due

2 It was considered that increasing the number of reversals could possibly reduce the variability in the data. Subsequent analysis showed no change to the trends in the data nor any statistically significant reduction in the variance of the data.

to potentially different ILDs across critical bands, this method reduced the possibility of lateral bias from a single noise generation.

Experiment 1: Broadband stimuli

Both the target stimulus and background noise used in Exp. 1 were broadband (20 Hz - 20 kHz) Gaussian noise bursts. Parameters for the experiments were the coherence of the background noise and the TOA between the target and background noise. The background noise was either interaurally correlated or uncorrelated and always contained an ILD of 0 dB. The TOA between the onsets and offsets of the target and background noise was either 0 ms (synchronous) or 250 ms (asynchronous), i.e. the total stimulus duration was either 300 ms or 800 ms, respectively. The 300 ms target was temporally centered within the background noise.

The pedestal ILDs used for this experiment were ± 6 dB, ± 4 dB, ± 2 dB and 0 dB. For each adaptive measurement, the presented target ILD was randomly and uniformly distributed (± 1 dB) around one of the specified pedestal ILDs. This allowed for a uniform distribution of presented target ILDs from ± 7 dB. Four repetitions for each of the pedestal ILDs were completed for each experimental condition, i.e. 28 data points were collected for each test condition and each listener. The target stimuli were first generated at 70 dB SPL, and the ILD was applied by changing the levels at both ears in opposing directions by an equal amount in dB. The overall level of the acoustic pointer, i.e. the second interval, was roved over a 10 dB range around 70 dB SPL to ensure that listeners were unable to use monaural cues for the lateralization task. The SNR was fixed at 0 dB for all presentations.

Listeners were allowed to practice with the lateralization task on similar stimuli used in the actual experiment until they were comfortable with the task. This practice session typically lasted 20-30 minutes. Listeners were not provided feedback for their responses in either the practice session or the actual test runs.

Experiment 2: Low-frequency, high-frequency, and broadband stimuli

The application of the ILDs was identical between both Exp. 1 and Exp. 2. The stimuli used in the second experiment differed only in the frequency range of the target and background noise. For the low-frequency condition, the target and background noise were both limited to 20 Hz - 800 Hz. For the high-frequency condition, the target and background noise were limited to 2 kHz - 20 kHz. For comparison, the broadband conditions specified in Exp. 1 were also presented. Similar to Exp. 1, listeners were allowed to practice with the lateralization task prior to beginning the main experiments.

2.2.3 Listeners

Nine listeners (seven male, two female) participated in the broadband study for Exp. 1. Listeners ranged in age from 25 - 35 years (mean age 30). For Exp. 2, a separate set of nine listeners (five male, four female) was chosen with ages ranging from 19 - 32 (mean age 23). None of participants reported any evidence or history of hearing problems. As will be described in Sect. 2.3.2, six of the listening subjects from Exp. 1 also participated in a subset of the conditions in Exp. 2.

2.3 Psychophysical Experiments

2.3.1 Experiment 1 Results

In Fig. 2.1, the adjusted pointer ILD is plotted against the presented ILD for the population of listeners in the baseline condition, i.e. when no background noise was presented. The reported ILDs for each measurement track are plotted as individual symbols, and the thick grey line with a slope of one indicates a perceived lateral position equal to the target ILD in isolation. Using a similar plotting convention, Fig. 2.2 illustrates the results of the test conditions when background noise was present. For these plots, a dotted gray curve with an average slope of 0.5 for the plotted range is also provided, where this curve indicates the aggregate ILD. To estimate the lateralization tendencies for the baseline and test conditions, generalized linear regression models with one degree of freedom (slope) were fit to each set of data. The linear regression models are plotted for the respective listening conditions as solid black lines in Fig. 2.1 and Fig. 2.2. A slope of 1.11 dB/dB for the baseline condition indicates that listeners were accurate in their estimate of the target ILD with the acoustic pointer paradigm. To estimate statistically significant differences in lateralization tendencies from the baseline condition, a generalized linear hypothesis test that takes into account the underlying distribution of the data was employed.

The upper panels of Fig. 2.2 show the population performance for the asynchronous test conditions with a TOA of 250 ms. For both the interaurally uncorrelated background noise (Fig. 2.2(a), $t(1254) = -3.92, p < 0.05$) and the interaurally correlated background noise (Fig. 2.2(b), $t(1254) = -6.61, p < 0.05$), the population performed significantly differently than the baseline performance per a two-tailed, multiple-comparison with the Bonferroni adjustment. In these conditions, the slopes of 1.29 dB/dB for the interaurally uncorrelated background noise and 1.42 dB/dB for the interaurally correlated background noise indicate that participants judged the target at more lateral positions than one would expect from the target ILD alone. Both values are much larger than 0.5 dB/dB slope predicted for the aggregate ILD.

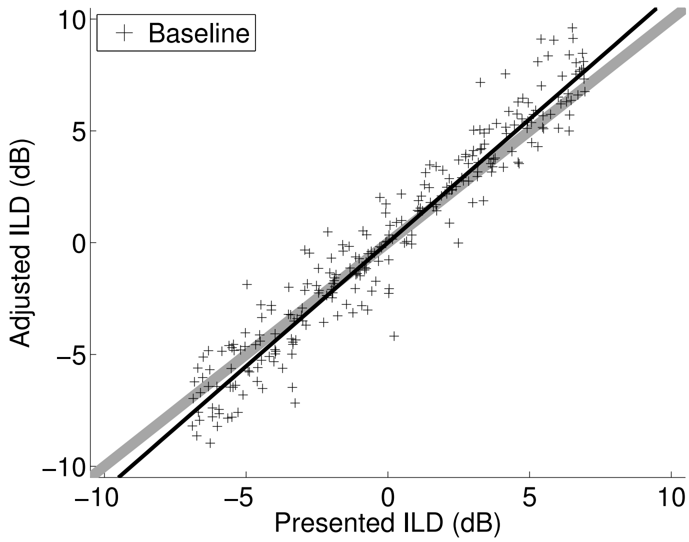

Figure 2.1: Scatter plots of the baseline lateralization performance when listeners from Exp. 1 estimated the lateral position of an isolated ILD-lateralized source without the presence of background noise. The solid gray line with a slope of one indicates the target ILD. The solid black line depicts the linear regression line fitted to the data.

When the target and the background noise were synchronous, the performance of the population depended on the coherence of the background noise as seen in the lower two panels of Fig. 2.2. For an interaurally correlated background noise, the slope of 0.68 dB/dB was significantly smaller (Fig. 2.2(d), $t(1254) = 8.86, p < 0.05$) than the baseline condition. However, for the interaurally uncorrelated background noise, the slope of 1.09 dB/dB was not significantly different from the baseline condition (Fig. 2.2(c), $t(1254) = 0.344, p = 0.996$). Furthermore, analysis reveals that for the two synchronous conditions, the population performed significantly differently between the interaurally correlated and uncorrelated background noise presentations (Fig. 2.2(c) vs. Fig. 2.2(d), $t(1254) = 8.45, p < 0.05$).

Figure 2.3 plots the slope values from the generalized linear regression models fit to individual data across the different listening conditions for both Exp. 1 and Exp. 2. Results for listeners from Exp. 1 are plotted as squares whereas results for individual listeners of Exp. 2 are plotted as circles. In a similar fashion, slopes of the generalized linear regression models fit to the population data are plotted in Fig. 2.4 for listeners of Exp. 1 and Exp. 2 (squares and circles, respectively). As will be described in Sect. 2.3.2, some of the listeners from Exp. 1 participated in a subset of listening conditions in Exp. 2. Slopes of these data are plotted as squares for the appropriate test conditions. The shaded symbols in Figs. 2.3 and 2.4 indicate a statistically significant difference in performance from the respective baseline performance. Statistical comparisons were conducted with the same generalized linear hypothesis test applied earlier.

The analysis showed that six of the nine listeners for the interaurally correlated background noise condition with a TOA of 0 ms performed significantly differently than their performance on the baseline condition. In contrast, three of the listeners performed significantly differently

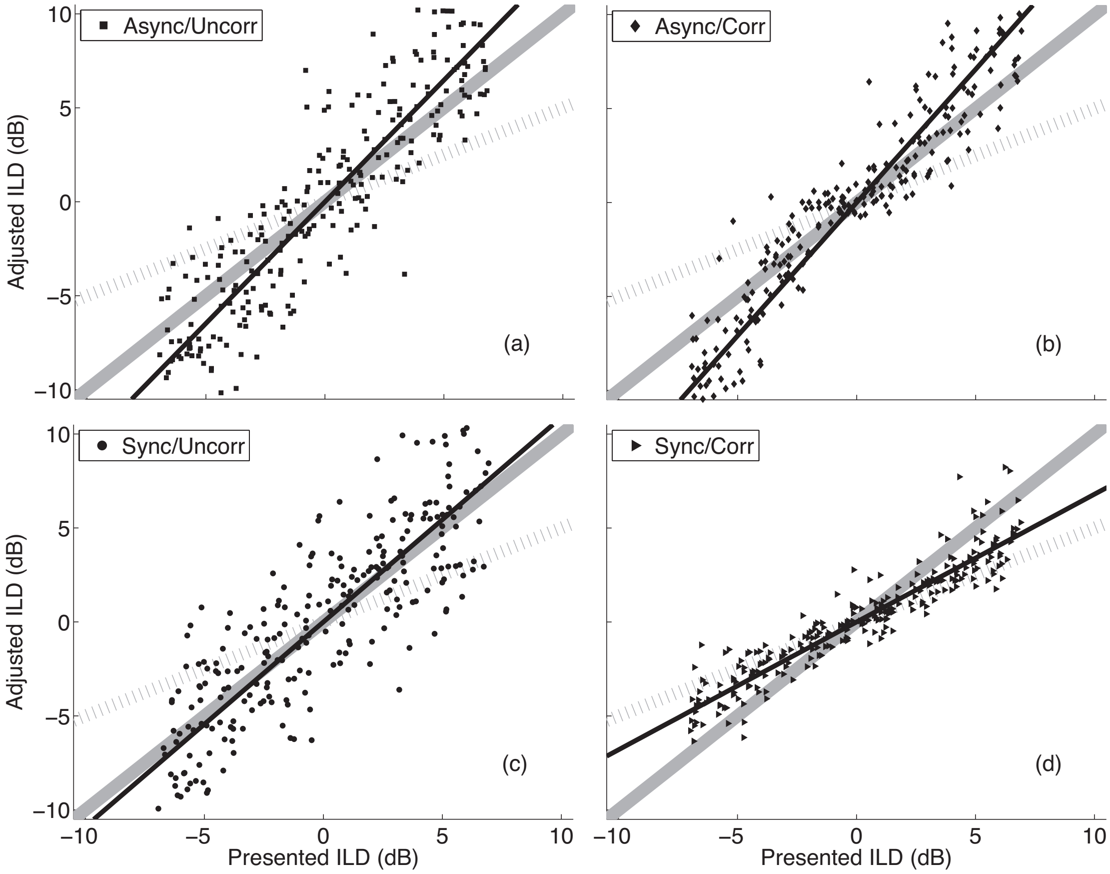

Figure 2.2: Scatter plots of the Exp. 1 population performance for each of the different background noise conditions: (a) interaurally uncorrelated and asynchronous with the target, (b) interaurally correlated and asynchronous with the target, (c) interaurally uncorrelated and synchronous with the target, (d) interaurally correlated and synchronous with the target. For all plots, the solid gray line with a slope of one indicates the target ILD whereas the dotted gray line with a slope of 0.5 indicates the aggregate ILD. The solid black lines depicts the linear regression lines fitted to the data for the respective background noise condition.

from their baseline performance when the background noise was interaurally uncorrelated and synchronous with the target. Only one of these three listeners, however, had a reduced slope near the aggregate ILD. When the background noise and the target were asynchronous, five listeners for an interaurally correlated background noise and two listeners for an interaurally uncorrelated background noise perceived a lateral position significantly beyond what would be expected for the target ILD in isolation.

To assess the within-subject variance, generalized linear models were computed for each of the four repetitions for each listener and compared using the same generalized linear hypothesis test applied in the population analysis. Fewer than 3% of the comparisons across the population of listeners resulted in significant differences ($p < 0.05$) based on the repetition number. We interpret this relative consistency between the repeated presentations as an indication that the variance in responses resulted from a blurred perceived position of the target and was not a result of listeners perceiving distracted targets at different locations from measurement to measurement. A between-subjects analysis was also conducted to investigate if the differences in individual slope values resulted from true differences in perceived lateral position or simply resulted from within-subject variance. For both of the asynchronous conditions, approximately 40% of the comparisons resulted in significant differences ($p < 0.05$). Furthermore, for the synchronous condition with an interaurally uncorrelated background noise, approximately 40% of the comparisons also resulted in significant between-subject differences. Such large numbers of statistically significant differences in performance indicate that differences between some of the listeners was a result of differences in perceived lateral positions. For the synchronous condition with an interaurally correlated background noise, none of the listeners performed significantly differently. The differences in individual slope values for this condition are not likely a result of perceived differences in lateral position but rather simply due to the variance in responses.

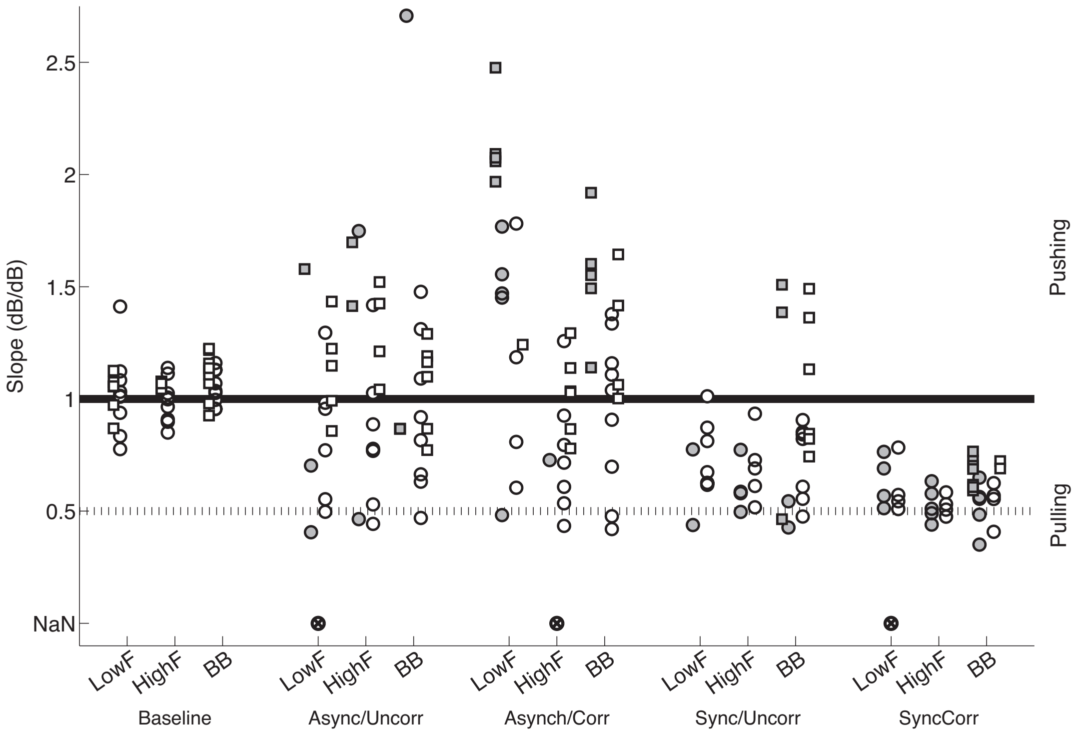

Figure 2.3: Slopes of the linear models fit to individual participant data for listeners of both Exp. 1 and Exp. 2. The different test conditions listed along the x-axis included differences in stimulus bandwidth (low-frequency, high-frequency, broadband), temporal relationship of the target to the background noise (asynchronous, synchronous), and background noise interaural coherence (uncorrelated, correlated). A slope of 1 dB/dB (indicated by the solid black line) represents a perceived position equal to the target ILD in isolation. For participants of the broadband experiments in Exp. 1, slopes are plotted as squares. Participants still available from Exp. 1 at a later date also completed the low- and high-frequency, asynchronous and baseline conditions as indicated by squares in the respective test conditions. Slopes for participants exclusive to Exp. 2 are plotted as circles. The shaded symbols indicate subjects whose performance for the test conditions was significantly different ($p < 0.05$) than their baseline performance. Conditions when an individual's data were unable to be explained by a linear model are indicated by crossed out symbols at the bottom of the plot with a value of NaN.

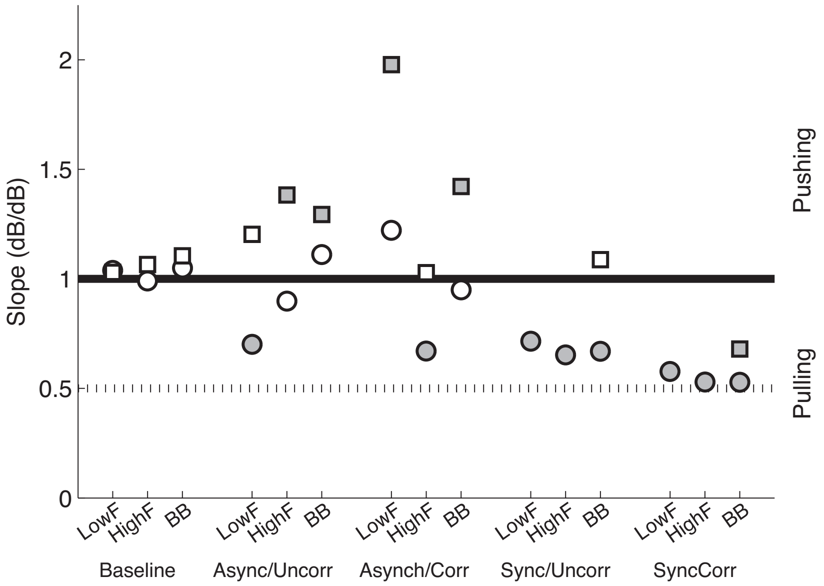

Figure 2.4: Slopes of the linear models fit to the population data. The details of the plot are illustrated in the same manner as applied in Fig. 2.3.

2.3.2 Experiment 2 Results

Similar to the analysis for Exp. 1, generalized linear models were fit to both individual and population data across the different listening conditions. The slopes of each linear regression model are plotted as circles in Fig. 2.3 for individual listeners and in Fig. 2.4 for the population of listeners. As will be described later in this section, six of the nine participants still accessible from Exp. 1 completed the low- and high-frequency conditions for the baseline and asynchronous configurations. Results for these data are plotted as squares in the respective listening conditions. For all symbols and listening conditions, shaded symbols indicate a statistically significant difference in performance from the baseline condition.

For the broadband stimuli, the population of listeners from Exp. 2 were accurate with their judgement of the target ILD in the baseline condition as seen by a slope of 1.05 dB/dB shown in Fig. 2.4 for the linear model fit. In contrast to the listeners of Exp. 1 for the broadband, asynchronous conditions, the population of listeners from Exp. 2 did not tend to perceive a position more lateral than what would be expected from the target ILD alone (i.e. slope > 1). Analysis with the same generalized linear hypothesis test applied in Exp. 1 revealed no statistically significant differences from the baseline performance for either the interaurally uncorrelated ($t(3758) = -0.944, p = 0.997$) or interaurally correlated ($t(3758) = 1.55, p = 0.869$) background noise conditions with slopes of 1.11 dB/dB and 0.95 dB/dB, respectively.

For the synchronous conditions with broadband noise, the listeners of Exp. 2 demonstrated similar tendencies as seen in Exp. 1 for the interaurally correlated background noise condition. A significant difference from the baseline performance was observed ($t(3758) = 5.91, p < 0.05$) with a linear model slope of 0.53 dB/dB indicating a perceived lateral position corresponding to the aggregate ILD. In contrast to the findings of Exp. 1, however, listeners of Exp. 2 also performed significantly differently for synchronous condition with interaurally uncorrelated background noise ($t(3758) = 8.12, p < 0.05$). For this condition, a slope of 0.67 dB/dB suggests that the population of listeners from Exp. 2 perceived a lateral position somewhere between what would be expected for the target ILD and the aggregate ILD. For the individual data plotted in Fig. 2.3, however, it can be seen that four of the nine listeners from Exp. 2 perceived a lateral position reasonably close to what would be expected for the target ILD in isolation when the background noise was interaurally uncorrelated and synchronous with the target noise burst.

The distribution of individual slopes for both the low- and high-frequency conditions are reasonably similar to their broadband conditions. The slopes for the baseline performance, 1.04 dB/dB for low-frequency and 0.99 dB/dB for high-frequency stimuli, showed negligible difference from the broadband condition. For all synchronous conditions of Exp. 2, the population performed significantly differently from the respective baseline condition. The linear models fit to these data indicate perceived lateral positions very close to the aggregate ILD. The slopes for the interaurally correlated background noise in the low-and high-frequency conditions were, respectively, 0.58 dB/dB and 0.53 dB/dB whereas the slopes for the interaurally uncorrelated background noise were slightly larger at 0.72 dB/dB and 0.65 dB/dB.

Since some discrepancy was observed in the performance for the broadband conditions between listeners of Exp. 1 and Exp. 2, six of the nine participants still accessible from Exp. 1 completed the low- and high-frequency conditions for the baseline and asynchronous configurations. Comparison between the two groups of listeners reveals a statistically significant difference in performance between the listeners of Exp. 1 and Exp. 2 for all of the low- and high-frequency, asynchronous conditions tested. Analysis shows that for all four of these listening conditions the population of listeners from Exp. 1 adjusted the pointer on average to a more lateral position than did the listeners of Exp. 2. This can be seen in Fig. 2.4 by the larger slope values for listeners in Exp. 1 (squares) as compared to listeners in Exp. 2 (circles) for the low- and high-frequency, asynchronous conditions.

2.3.3 Discussion

Based on the results of these experiments, it does not seem that listeners simply perceive a position corresponding to the aggregate ILD, although a range of inter-subject variability was observed. A use of something other than the aggregate ILD is rather interesting since the aggregate ILD is what could have been directly read out from the physical stimulus. This suggests that listeners possess some means to adjust or differently interpret the internal representation of the actual ILD at the ears.

Prior studies have shown an influence of an interfering noise source on the perceived spatial position of target sound objects. It appears that whether the perceived shift is pulled toward the interferer (Butler and Naunton, 1962, Heller and Trahiotis, 1996, Best *et al.*, 2007) or is pushed more laterally away from the interferer (Thurlow and Jack, 1973, Canévet and Meunier, 1996, Carlile *et al.*, 2001, Braasch and Hartung, 2002) depends on properties of the target and interferer combination. The current results indicate that the perceived lateral position of the target can be influenced depending on whether the target and background noise are synchronous or asynchronous. When the TOA was set to zero, the perceived position of the target tended to be pulled toward the position of the background noise (slope < 1). When the TOA was set to 250 ms, however, most subjects perceived a position at or even beyond the target ILD in isolation (slope ≥ 1). The current results also indicate that the coherence of the background noise can influence the ability of some listeners to use the target ILD when the target and background noise are synchronous. Experiment 2 emphasized that even within test conditions, how the target ILD was utilized depended on the listener. Additionally, Exp. 2 illustrated that the frequency range of the stimulus did not produce obvious differences in the ability of listeners to utilize the target ILD for these types of tasks.

Temporal synchrony and common onsets are considered to be important grouping cues for sound objects in mixed scenes (Darwin, 1984, Bregman, 1990, Shamma *et al.*, 2011). Furthermore, how listeners combine spatial information is influenced by the grouping of acoustic objects (Hill and Darwin, 1996, Best *et al.*, 2007). Asynchronous onsets and offsets between the target and background noise could, therefore, help to perceive the two noises in our experiments as distinct objects, which may facilitate the auditory system's ability to make an inference about the target ILD. As seen in Fig. 2.3, a vast majority of the participants in both studies were able to make a reasonable estimate of the target ILD for the asynchronous conditions. A perceived position more lateral than what would be expected from the target ILD in isolation, which was observed for several participants, is supported by the notion that perceptually distinct objects repel each other (Lee *et al.*, 2009).

Given the perceptual emphasis placed on signal onset/offset information for ILD sensitivity (Stecker and Brown, 2012) and the importance of common onsets for object identification in multi-source scenes (Bregman, 1990), it is reasonable that a target synchronous with a background noise would be grouped as one sound object. It is, therefore, not surprising to find the perceived position of the stimulus being pulled in the direction of the interaurally correlated background noise as seen in both Exp. 1 and Exp. 2. For two, independent broadband sounds spatialized over headphones using head-related transfer functions, Best *et al.* (2005) showed that pulling of the target towards the position of the interferer was strongest when the two sounds were temporally overlapping. In contrast, the degree of *repulsion* has been shown to depend on the onset disparity for both tones (Canévet and Meunier, 1996) and high-frequency narrowband noises (Meunier *et al.*, 1996) presented in the free-field. With overlapping spectral information and no temporal information, one would expect the two sound sources to be grouped into a single object. Thus, it was surprising that a target ILD estimate similar to the baseline performance occurred for many subjects when the background noise was interaurally uncorrelated and synchronous with the target. This result indicates that the coherence of the background noise can play a role in the auditory system's ability to identify the lateral position of the target.

Although several individuals from Exp. 2 illustrated the dependence on background noise coherence when the target and background noise were synchronous, the population of listeners in Exp. 2 did not. A clear reason for the discrepancy in performance between participants of Exp. 1 and Exp. 2 is not apparent; however, we can hypothesize that the two groups were possibly selected from different types of listeners. The participants of Exp. 1 were composed of experienced listeners in auditory perceptual investigations, whereas the participants of Exp. 2 were almost entirely naive listeners. It should be noted that all participants received training on the task prior to beginning the actual experiments. Since perceptually distinct sound sources have been reported to result in a lateral repulsion of the perceived target position (Lee *et al.*, 2009), it is possible that the difference in performance between Exp. 1 and Exp. 2 could be related to the inter-individual differences in salience of the target for the asynchronous conditions. Across all test conditions, the population of listeners from Exp. 1 tended to adjust the pointer to more lateral ILDs than the population of listeners of Exp. 2 as seen by larger slope values in Fig. 2.4. If indeed listeners from Exp. 1 heard the target and background noise as being more perceptually distinct, it is possible that these listeners could make a better estimate of the target ILD in the presence of a synchronous, interaurally uncorrelated background noise.

Comparison of the variance in responses for the synchronous conditions shows a statistically significantly larger variance for interaurally uncorrelated background noise as compared to interaurally correlated background noise ($F(251, 251) = 3.09, p < 0.001$), which is clearly visible in Fig. 2.2. Lateral sensitivity could be diminished when an interfering noise is interaurally uncorrelated versus when it is interaurally correlated as demonstrated by an increase in jnds for an ITD lateralization task (Cohen, 1981, Bernstein *et al.*, 2001). Additionally, this reduced certainty of the target's lateral position bears some similarity to observations in binaural masking level difference (BMLD) experiments. With interaurally uncorrelated noise maskers, a BMLD of only $\sim$3 dB has been found while much larger BMLDs are found for interaurally correlated noise maskers (Robinson and Jeffress, 1963). Higher detection thresholds and greater lateral uncertainty could be attributed to a less precise representation when the interaurally uncorrelated background noise was presented.

2.4 Binaural Localization Model

2.4.1 Model Description

Previous models designed to account for detection and lateralization performance of the binaural system are based on either a coincidence-counting mechanism (Jeffress, 1948, Colburn, 1973, Lindemann, 1986) or an equalization-cancellation (EC) mechanism (Durlach, 1963, Breebaart *et al.*, 2001). Physiological evidence shows that cells within the Lateral Superior Olive (LSO) are sensitive to ILDs and that these cells operate using an excitatory-inhibitory (EI) mechanism (Tollin and Yin, 2005), which has computational similarities to an EC calculation. An EC model used for signal detection as proposed by Breebaart *et al.* (2001) will be adapted here for the purpose of ILD-based lateralization when interfering noise is present.

The EC model in basic terms is composed of two processes: equalization and cancellation. The input signal is first transformed across a range of compensatory adjustments in order to 'equalize' the left and right channels, e.g. in level and in time. For each of these adjustments, the left and right channels are subtracted in the 'cancellation' process. Although this model was originally intended to address binaural masking level difference predictions, Durlach (1963) suggested that the model could be used for lateralization by selecting the compensatory adjustment that produce a minimum in the model output. While this method is sufficient for single sources, the presence of multiple sources requires an appropriate choice of the equalization in order to remove the lateral bias resulting from an interfering background noise. This consideration will be used here in our model extension to account for detecting a target ILD in the presence of an interfering ILD. We will assume that when a background noise is present, the EC model is adjusted to have uniform output as a function of the compensatory level adjustments in the equalization stage. This knowledge of the background noise is then applied as a template subsequent to the output of the cancellation stage so that the background noise creates no bias in the predicted lateralization estimate.

A general schematic of the proposed binaural interaction for the model is depicted in Fig. 2.5. Prior to the computation of the binaural processing, the left and right input signals are separated into different frequency bands using a fourth-order Gammatone filterbank implementation (Hohmann, 2002). A spacing of 0.5 equivalent rectangular bandwidth is used, which results in a total of 80 frequency bands ranging from 20 - 20,000 Hz. The respective left and right output signals from the Gammatone filterbank are subsequently passed through a series of compensatory level adjustments (ΔdB) that result in an array of α-dependent taps, where α represents the amplitude ratio in linear terms corresponding to the ILD. Using an EI-type element, the difference between the left and right signals is computed at each of the taps for every Gammatone filterbank channel. Subsequent to the EI-type element, an α-dependent weighting template $\beta(\alpha)$ is applied that results in a uniform model output across all values of α

and all frequency channels when the background noise is presented in isolation. The α which produces minimum activity for the target plus background noise stimulus is selected as the model's output ILD.

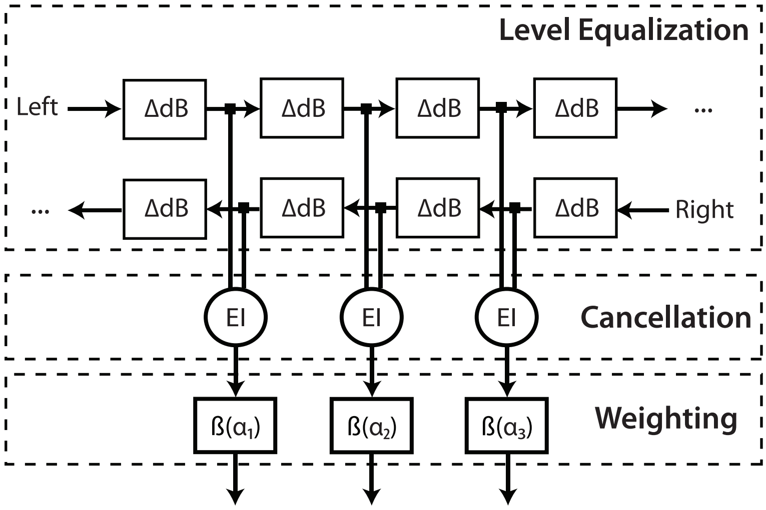

Figure 2.5: Block diagram illustrating EC binaural processing with the addition of an α-dependent weighting template. The equalization consists of a stepwise attenuation (ΔdB) for the left and right ear inputs, which result from the peripheral preprocessing model. The cancellation stage (EI) takes the outputs from these attenuation lines such that each cell has a characteristic ILD. The weighting stage takes the outputs from the EI units and applies an α-dependent (i.e. ILD-dependent) weighting template $\beta(\alpha)$. The template effectively removes the lateral bias resulting from the presence of a background noise such that an accurate prediction of the target ILD can be made.

The choice of the weighting template β is made such that the output at each of the EI elements (D^2) is equal to one across all values of α when the left and right input signals consist of only the background noise.

$$D^2 = \left[(\sqrt{\alpha}L - \frac{1}{\sqrt{\alpha}}R) * \beta\right]^2 = 1 \tag{2.1}$$

Solving for β with an interaurally uncorrelated noise input ($L \neq R$), we obtain an α-dependent weighting template denoted as $\beta_{uncorrelated}$:

$$\beta_{\text{uncorrelated}} = \sqrt{\frac{1}{\alpha + \frac{1}{\alpha}}} \tag{2.2}$$

Similarly, we can compute a weighting template ($\beta_{correlated}$) such that the model output is uniform across α for an interaurally correlated noise input ($L = R$):

$$\beta_{\text{correlated}} = \sqrt{\frac{1}{\alpha + \frac{1}{\alpha} - 2}} \tag{2.3}$$

The presence of only diotic background noise would cause $\beta_{correlated}$ to run to infinity at the compensatory α corresponding to the ILD of the background noise. Simulations indicated that limiting the weighting template to a large value does not disrupt the model performance. In general, it is possible that the background noise could possess a non-zero external ILD and could be similarly cancelled. For simplicity and conformity with the behavioral experiments, the background noise is computed here as having an ILD of 0 dB.

The effective result of applying β is that the lateral influence of a competing ILD-lateralized noise is mitigated such that the minimum in the model output now corresponds to the actual target ILD. Improving the accuracy of the target ILD estimate based upon knowledge of the background noise has already been demonstrated in a model by Braasch (2003). The model estimates the target ILD using the difference in power between the target plus interferer and the interferer alone. Without a temporal asynchrony, however, the model would presumably be only able to estimate the ILD of the target plus interferer interval, i.e. the aggregate ILD. For an interaurally correlated interferer, the prediction of the aggregate ILD would be a reasonable estimate given the psychophysical performance of our listeners. For an interaurally uncorrelated interferer, the results suggest that the aggregate ILD might not be the best estimate for many listeners. If the model has a priori knowledge of the background noise level in each ear, only then would a calculation of the target ILD be possible using the difference estimate suggested by Braasch (2003). The EC model implementation presented here only requires an assumption of the background noise coherence.

The general form of the decision variable for the stimuli of our experiments, i.e. an ILD-lateralized target combined with a background noise at zero ILD, could be written as:

$$D^2 = \left[((N_l + \sqrt{\alpha} S_o) - (N_r + \frac{1}{\sqrt{\alpha}} S_o)) * \beta \right]^2 \tag{2.4}$$

where S_o represents the ILD-lateralized target noise and (N_l, N_r) represent the left and right background noise signals, respectively. The choice of β is generalized and depends on the prior knowledge of the background noise coherence. When the coherence of the background noise is unknown to the listener due to a synchronous onset, we will adopt the assumption that the auditory system generally assumes the $\beta_{uncorrelated}$ weighting template. Only in the case when listeners can infer additional information about the coherence of the background noise, e.g. when an onset asynchrony is available, does the weighting template change.

The model was exposed to the same three-interval, adaptive acoustic pointer task as conducted in the psychophysical experiments previously described. For each interval, the α with minimum activity from Eq. 2.4 was determined for each frequency band. The overall ILD for each interval was calculated as the mean of the corresponding α values found in each individual frequency band. The ILDs calculated in the two test intervals were averaged and then compared with the calculated ILD of the pointer interval. The pointer was subsequently moved left or right for the next trial.

As suggested in the original implementation of the EC-model (Durlach, 1963), the addition of "jitter" to each of the left and right level compensations limits the perfect cancellation of the left and right channels. In our implementation of the jitter, a single value taken from a Gaussian distribution with a standard deviation of $10^{-30/20}$ was added independently to each value of α. This value was chosen such that the predicted lateral estimates were best matched to the behavioral results.

2.4.2 Model Results & Discussion

For the proposed implementation, we assume that the auditory system defaults to using the $\beta_{uncorrelated}$ weighting function when it has no evidence indicating the background noise coherence. The additional information about the background noise coherence from asynchronous onsets, however, could permit the auditory system to choose a different weighting strategy, e.g. $\beta_{correlated}$. Figure 2.6(a) shows the model performance for the broadband noise conditions when the analytical weighting functions derived above were applied. The open symbols indicate conditions when $\beta_{uncorrelated}$ was applied and the filled symbols indicate the stimulus condition when $\beta_{correlated}$ was applied. In Fig. 2.6(a) the open triangles illustrate that using $\beta_{uncorrelated}$ resulted in a prediction of the aggregate ILD when the background noise was interaurally correlated. However, $\beta_{uncorrelated}$ resulted in a prediction of the target ILD for the interaurally uncorrelated background noise (i.e. open circles). This dependence of the predicted target ILD on the coherence of the background noise is qualitatively similar to the observed psychophysical results for the broadband stimuli. Because the $\beta_{uncorrelated}$ weighting function was used when the background noise was interaurally uncorrelated, regardless if the target and background noise were synchronous or asynchronous, no distinction existed to predict the any differences in the performance based on the temporal asynchrony for interaurally uncorrelated background noise.

Use of the $\beta_{correlated}$ weighting function allowed for a prediction of the target ILD in the presence of a interaurally correlated background noise as illustrated by the filled diamonds in Fig. 2.6(a) that closely follow the target ILD. Neither weighting function $\beta_{uncorrelated}$ nor $\beta_{correlated}$ in this simple implementation were able to directly account for the pushing of the perceived target position beyond the expected target ILD. However, it is promising that the weighting function did account for the performance differences depending on the coherence of the background noise as seen in the psychophysical results.

For precise cancellation of the target noise, which will lead to a minimum in the cancellation pattern, the phase structure of the stimulus (specifically, that of the masker) needs to be represented. If only the envelopes of the left- and right-channel inputs are available, the random phase addition of the masker and target would lead to an envelope where the target is not a

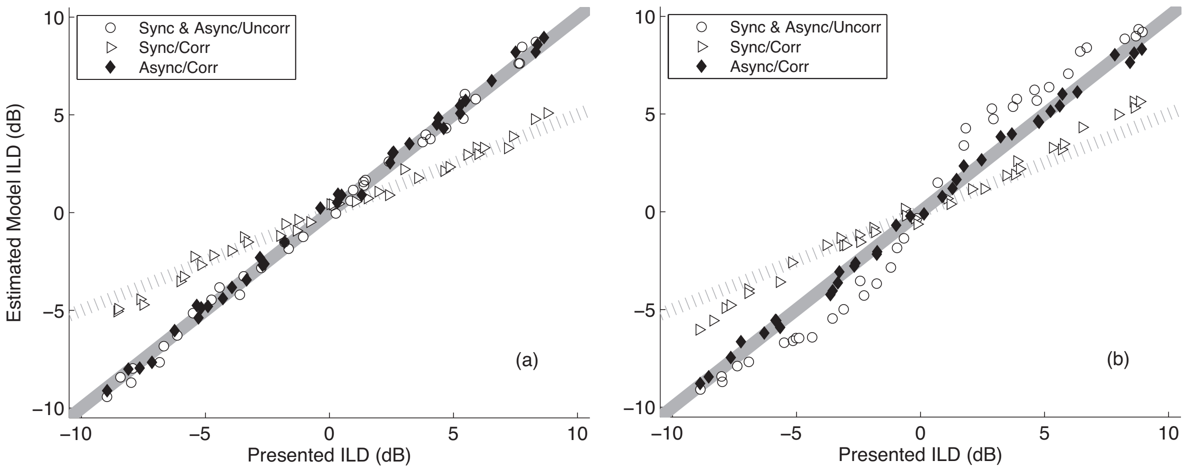

Figure 2.6: Model predictions for the (a) analytically calculated weighting functions and (b) modeled weighting functions accounting for an auditory periphery. The open symbols indicate conditions when the $\beta_{uncorrelated}$ weighting function was used, and the filled symbols indicate when the $\beta_{correlated}$ weighting function was used. The solid gray line with a slope of one indicates the target ILD whereas the dotted gray line with a slope of 0.5 indicates the aggregate ILD. The symbol shapes are the same as those used in the comparable conditions plotted in Fig. 2.2, however, only one symbol (∘) is shown for the interaurally uncorrelated noise conditions because the model used the same weighting function for the asynchronous and synchronous conditions.

linear addition to the masker anymore and, hence, perfect cancellation would not be possible. Tollin and Yin (2005) have shown low-frequency LSO neurons to be phase sensitive, thus, these predominantly ILD-sensitive neurons could provide the required cancellation. If, however, a more physiologically plausible peripheral stage that accounts for the loss of phase-locking ability at higher frequencies is applied prior to the EI-type elements, then the necessary cancellation for reporting the target ILD would no longer be possible at high frequencies.

Since the peripheral stage effectively modifies the ILD at the input to the binaural system, we evaluated an alternative model that included a model for inner-hair-cell processing. This implies that an alternative to the analytic weighting template could be numerically estimated as the inverse of the EC-model, which is preceded by a desired peripheral model, for a given background noise input. For each Gammatone frequency channel, the weighting template was computed as the inverse of the EC model response across α when a single background noise with zero ILD was provided as an input. This resulted in a weighting template that was a function of both internal level compensation and frequency. Similar to the analytic weighting template, jitter was added to each value of α such that the random values take from the Gaussian distribution had a standard deviation of $10^{-18/20}$. This amount of noise resulted in the best predictions with the psychophysical data for the broadband stimuli. For the implemented peripheral model, half-wave rectification was applied subsequent to the Gammatone filtering described in Sec. 2.4.1. Furthermore, a 5th-order, low-pass Butterworth filter with a cutoff frequency of 770 Hz was applied after the half-wave rectification as used in Breebaart *et al.* (2001).

The results of using this modeled weighting function to account for peripheral processing are illustrated in Fig. 2.6(b). It is apparent that some differences from the analytical weighting template are present although the general performance is maintained. Most importantly we see that this simple weighting assumption still allows for changes in the model's lateralization performance depending on the coherence of the background noise even when a more physiologically plausible auditory periphery is applied. Interestingly, a degree of lateral pushing (slope > 1) with interaurally uncorrelated background noise seen in the behavioral experiments can result from this modeled weighting template.

2.5 General Discussion

When the ILD of a target noise was altered by an interfering noise source, our psychophysical experiments illustrated that listeners' judgements did not coincide with the use of the actual ILD physically at the ears. It appears that the auditory system possesses a means to calculate the lateral position of a desired target although this ability seems to be listener dependent. Many listeners actually perceived a position beyond what would be expected from the target ILD alone when there was a temporal asynchrony between the target and background noise. This finding is in agreement with Lee *et al.* (2009) that perceptually distinct objects can result in spatial repulsion. For some of the listeners, it was even possible to make a reasonable estimate of the target ILD when the target and background noise were synchronous, so long as the background noise was interaurally uncorrelated. When the background noise was interaurally correlated, the perceived extent of laterality for all listeners was close to the physical ILD at the ears.

This observed difference in perceived laterality dependent upon the background noise coherence could be related to our natural acoustic surroundings. In everyday acoustical settings, interaurally uncorrelated noise always has an ILD close to zero. If this assumption is made about interaurally uncorrelated noise, it would be theoretically possible to calculate back what the target ILD must have been for an interaurally correlated target combined with the interaurally uncorrelated background noise. For an interaurally correlated interfering noise source, the assumed ILD of zero is no longer a realistic choice, thus, making it less reasonable to attempt an inference of the target ILD. Until more is known about the ILD of the interferer, e.g. due to an onset asynchrony, the best assumption is to simply use the ILD at the ears.

These assumptions about the background noise with respect to the interaural coherence motivated the described model structure. For an interaurally uncorrelated noise input with zero ILD, a weighting template was computed such that every α had an equal probability of being chosen. This resulted in the interaurally uncorrelated background noise having limited influence on the laterality percept of an interaurally correlated target. Only when an interaurally correlated background noise was perceptually distinct from the interaurally correlated target, e.g. temporally unique, did the background noise coherence assumption change. For interaurally correlated background noise, the α-dependent weighting template was computed such that the model output was maximally sensitive to changes from the background noise. The effective result of the template for either form of the weighting template was to reduce the lateral bias produced by the background noise so that the minimum in the model output was found at the target ILD.

Intersubject differences in lateralization performance could be comparable to different choices of the weighting template. The psychophysical study illustrated that even in some of the asynchronous conditions, some listeners were unable to accurately lateralize the target and were only able to use the physical ILD at the ears. It is interesting that the group of listeners in the second study were severely biased towards the aggregate ILD for the asynchronous condition with high-frequency noise, considering that the ILD is assumed to be a salient cue at high frequencies. This suboptimal use of the ILD cues at high-frequencies could possibly be related to most of these listeners not being able to use the target ILD in the condition with synchronous, interaurally uncorrelated background noise. Another consideration would be related to the relative experience with binaural listening experiments between the two groups of listeners. One could consider that additional listening experience provides an improved ability to select a better weighting function to minimize the interference of the background noise.

In summary, we have shown that the temporal disparity between the onsets/offsets of a target and background noise is an important parameter that helps listeners to use the target ILD instead of the physical ILD at the ears for a lateralization task. Although considerable inter-subject variability was observed for the asynchronous conditions, listeners were generally able to perceive a position at or more lateral than the actual target ILD in these asynchronous conditions. If instead the target and background noise were synchronous, the coherence of the background noise apparently influenced the ability of some listeners to use the target ILD. For interaurally correlated background noise, listeners consistently perceived a position near the aggregate ILD regardless of the frequency range of the noise stimuli tested. When the background noise was interaurally uncorrelated and broadband, however, several participants were able to make a reasonable estimate of the target ILD.

We also demonstrated that a simple assumption with respect to the weighting of the internal level adjustments of an EC model could account for the general trends in the psychophysical data. More specifically, the model provides a mechanism to overcome the lateral bias in the position of the minimum in EC activity when an interfering noise is present. The model defaults to using a template that results in uniform output across all compensatory level adjustments for an interaurally uncorrelated noise input that has a zero ILD. This default assumption results in the model's lateralization prediction to be dependent upon the background noise coherence as observed in the behavioral results. Even when an auditory peripheral model was applied prior to binaural interaction, this simple weighting assumption was still able to describe the background noise dependence.

3

Characterizing perceptual properties of a binaurally modulated stimulus[1]

The majority of everyday listening situations involve a complex mixture of multiple sound sources. Assuming a spectro-temporally sparse target signal, e.g. speech, the binaural cues in these complex listening environments are not simply due to a single sound source but are often rapidly switching between the most dominant source at any given moment. To investigate the perception of rapidly switching ITD cues, a noise stimulus that periodically alternates between two different values of interaural time difference (ITD) was created. This stimulus appears to evoke a purely binaural percept of modulation, which is the focus of the experiments presented here. Results indicate that listeners can reliably discriminate this ITD-switching stimulus from a stimulus composed of stationary ITD cues. Frequency discrimination of this ITD-switching stimulus was tested in a separate experiment, which showed that listeners are generally able to discriminate a 50%-change in modulation frequency for rates below 16 Hz. The final experiment investigated if modulation masking exists between the monaural and binaural auditory pathways. Although a statistically significant increase in modulation detection thresholds was observed when both types of modulations were presented together, the increase was relatively small, indicating only a weak interaction.

1 This chapter has been submitted to the *J. Acoust. Soc. Am.*

3.1 Introduction

The natural soundscapes that humans encounter are often complex auditory scenes composed of several different sound sources. We regularly face situations where speech from undesired talkers compete acoustically with other speech of interest in so-called cocktail party scenarios. Through the use of several cues, e.g. spectral, temporal, and spatial information, the auditory system can disentangle different sound sources and form acoustic streams (Bregman, 1990). While it appears that spectral cues might play a stronger role (Culling and Summerfield, 1995, Shackleton and Meddis, 1992), spatial separation of target and interfering sound objects provides an important benefit for speech intelligibility (for a review see Bronkhorst (2000), Litovsky (2012)).

Since speech is a spectro-temporally sparse signal, the speaker that dominates in a multi-talker environment will switch regularly within a given frequency band. Consequently, the primary binaural cues, namely the interaural time difference (ITD) and interaural level difference (ILD), will also switch regularly. Knowledge of the speed at which the binaural system can operate would be useful for predicting limits of the benefit provided by the binaural system on auditory tasks, e.g. streaming and speech segregation in complex environments. The effective speed of the binaural system has been addressed by previous investigations, however, time constants as long as 50 - 250 ms Grantham and Wightman (1979) and as short as 10 ms Akeroyd and Bernstein (2001) have been demonstrated. As suggested by Kollmeier and Gilkey (1990), the speed of the binaural system could be related to the task presented. Relatively subtle manipulations to stimuli could also influence the temporal acuity of the binaural system as demonstrated for time-varying interaural cross-correlation stimuli (Siveke *et al.*, 2008). It is worth noting that the model by Breebaart *et al.* (2001) can be used to predict several psychophysical results for binaural tasks using a single time constant of 30 ms.

We have adapted a stimulus from previous studies, which investigated the binaural temporal window for barn owls (Wagner, 1991) and humans (Bernstein *et al.*, 2001), to develop a stimulus that periodically alternates in ITD as shown by the schematic drawing in Fig. 3.1. The stimulus is composed of broadband, Gaussian noise that permits instantaneous switching between the two ITD values without producing monaural cues of the switching. The regular time interval with which the two ITDs alternate will be called the 'binaural modulation period'. The percentage of the binaural modulation period spent at any one of the two ITDs applied in the stimulus will be referred to in terms of the duty cycle.

Prior to the experiments presented here, listening subjects qualitatively described their percept of this ITD-switching stimulus via sketches and text. Although the stimulus parameters were not exhaustively explored in a formal study, it was noted that the listeners perceived two separate components when the ITDs were sufficiently far apart (approximately $\geq 200\mu s$). For the lateral position associated with the ITD containing the longer duty cycle, listeners reported

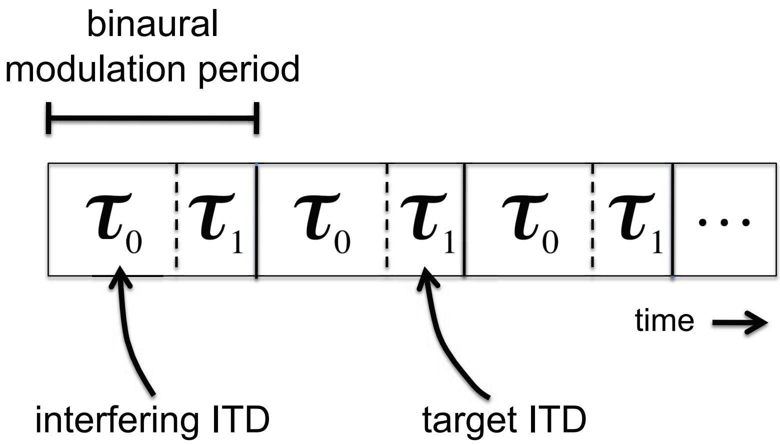

Figure 3.1: Schematic drawing of a purely binaural modulation stimulus. A continuous, broadband Gaussian noise periodically alternates between two different ITDs (τ_1 and τ_2). The duty cycle, i.e. the percentage of the binaural modulation period at a given ITD, is an experimental parameter that is manipulated in the experiments of this paper.

hearing a continuous noise. A separate sound object containing a percept of modulation was generally reported, however, only for the noise token with the shorter duty cycle. This reported modulation has characteristics described as something like a "helicopter noise" that the authors would describe as having a timbral quality similar to an amplitude modulation (AM) with a low-frequency noise carrier. It is worth reemphasizing that this stimulus does not contain any monaural cues due to the ITD switching. If either the left or right channel is presented in isolation, only a continuous, broadband noise is heard. This indicates that the binaural system mediates the buildup of the reported modulation percept. The purpose of this study is to investigate perceptual properties of this purely binaural modulation.

For the monaural system, temporal modulations in amplitude appear to be an important cue used in speech reception. Only a few frequency bands of noise applied as an artificial carrier signal are sufficient to achieve high accuracy for consonant, vowel, and sentence recognition if temporal modulations of the speech below 50 Hz are maintained (Shannon *et al.*, 1995). Furthermore, degradations to the low-frequency modulation envelopes of speech reduce intelligibility (Drullman *et al.*, 1994). Modulations of interfering noise have also been shown to be beneficial for speech reception since the valleys of noise provide glimpses of the target speech information (Bronkhorst and Plomp, 1992).

Due to the importance of monaural modulation cues, it was of interest to determine whether the purely binaural modulation interacts with the processing of monaural modulation cues. Sensitivity to AM depth has been well-studied where a low-pass behavior is observed for wideband carriers (Dau *et al.*, 1997, Viemeister, 1979). In a modulation masking paradigm when both a target and interfering amplitude modulations are present, the ability to detect the target modulation is a function of the separation in target and masker modulation frequencies (Houtgast, 1989, Ewert and Dau, 2000). Furthermore, the interaction of monaural and binaural modulation sensitivity has been previously investigated by Thompson and Dau (2008) where sinusoidally modulated ILD cues were used to impede AM detection. With this stimulus, however, the ILD modulation is not strictly binaural in nature as it also produces monaural cues due to changes in level, and it is possible that a purely monaural interaction may affect the AM detection in this configuration. The modulated ITD cues applied in the present paper do not produce monaural cues from the ITD switching, which permits an isolated binaural modulation percept. By combining this binaural modulation with a diotic AM, we investigated the potential interaction of monaural and binaural modulation sensitivities.

The first experiment of this study determined under what conditions a binaural modulation percept was created by switching ITD cues. The second experiment investigated whether the binaural modulation frequency could be discriminated and whether indeed a lateral position could be attributed to the binaural modulation percept as introspectively reported by listeners. Finally, the interference between this purely binaural modulation and amplitude modulation (AM) was measured in the third experiment.

3.2 Experiment 1: Binaural modulation detection

This experiment investigated the ability of listeners to detect the proposed binaural modulation percept resulting from rapidly switching ITDs. Listeners were required to distinguish between the ITD-switching stimulus and a stimulus with stationary ITD cues containing the same long-term correlation as the ITD-switching stimulus. Therefore, if the representation of binaural cues are subject to a limited temporal resolution, both intervals will sound the same. The duty cycle and binaural modulation rates were systematically varied to measure the sensitivity of listeners to the periodic changes in binaural cues.

3.2.1 Methods

An adaptive, three-alternative-forced-choice (AFC) paradigm was employed where the task of the listeners was to indicate the interval containing the ITD-switching stimulus. The stimulus was created by taking two separate Gaussian-noise signals each with a different ITD. Regular intervals were used to switch between these noise signals instantaneously and synchronously in both left and right channels. The ITD with the longer duty cycle of the binaural modulation period was always set to an ITD of 0 μs. The ITD for the noise with the shorter duty cycle was tested across several different target ITD values ($\pm150, \pm300$, and ±450 μs) to measure the detectability of the ITD switching as a function of spatial separation between the two ITDs.

The reference intervals in each presentation were designed to possess the same long-term cross-correlation as the binaurally modulated stimulus. This was done by summing two noises continuously presented at the same ITDs applied in the binaural modulation stimulus for the target interval. The two noises continuously presented at the specified ITDs were mixed at an effective signal-to-noise ratio (SNR) that was a function of the duty cycle used in the target interval. This SNR was computed using the following:

$$\mathrm{SNR} = 10 \cdot \log_{10}\left(\frac{duty\ cycle}{1 - duty\ cycle}\right) \tag{3.1}$$

where *duty cycle* is the fraction of the binaural modulation period when the target ITD (i.e. the ITD not positioned at 0 μs) is active. Since this mixing ratio is purely an estimate to achieve the spatial image due to the two different ITDs used in the binaural modulation interval, this effective SNR was roved ± 1 dB to limit listeners from using differences in the width of the spatial image to identify the reference intervals. Furthermore, the two ITDs within all intervals were roved ± 50 μs such that both ITDs shifted by the same magnitude and direction. The magnitude and direction of the rove was randomized between the intervals of a single presentation. This ITD roving was applied to both the target and reference intervals in order to avoid that listeners can detect the target based on a possible shift of the stimulus centroid.

The duty cycle of the non-zero ITD was adaptively adjusted in percent according to a two-down, one-up rule. The adaptive track began with a duty cycle of 20% and a starting step size of 5%. The step size was halved at every two reversals until a step size of 0.625% was reached upon which six additional reversals were completed before the track was terminated. The reported duty cycle was the average of the values at each of the final six reversals. If the listener reached the maximum value (35%) five times, the track would terminate without recording a threshold. The binaural modulation frequencies tested were 4 Hz to 128 Hz at octave spacing. Tracks were blocked on the binaural modulation frequency and ITD.

In an effort to look at the frequency range contributing to the binaural modulation, the detection study was conducted on the broadband, binaural modulation stimulus and, additionally, on low-pass (LPF) and high-pass (HPF) filtered versions of the broadband stimulus. For the LPF condition, the broadband stimulus was transformed into the frequency domain and all Fourier coefficients above 1 kHz were set to zero before transforming back into the time domain. Similarly for the HPF condition, all coefficients below 2 kHz were set to zero. Since the filtering was completed subsequent to the application of the ITD switching, artifacts that would provide monaural cues for the ITD switching were not introduced.

Participants completed two separate sessions for each frequency condition on different days where each session took approximately two hours. The presentation order for the different frequency conditions was randomized for each participant. Prior to the first session of each frequency condition, a practice session of approximately 15 minutes in duration was provided. All stimuli were 600 ms in duration, which included 50-ms raised cosine onset/offset ramps, and were calibrated to a level of 70 dB SPL. A pause of 400 ms was inserted between the different intervals for a given three-interval presentation. All stimuli in this paper were presented via headphones (Sennheiser HD-650). Six listening participants (three female, three male; average age 25 years), none of whom reported any evidence or history of hearing problems, completed this experiment.

3.2.2 Results

The duty cycle required for detection of the binaural modulation is illustrated in the Fig. 3.2 box plots for the broadband condition across the population of six listeners. Since no statistical difference was found in performance between the left and right hemisphere via a one-way analysis of variance (ANOVA), the data are grouped on magnitude of the ITD. The same folding of the data across the midline was done for the LPF (Fig. 3.3) and HPF (Fig. 3.4) conditions. The percentage of all tracks where a threshold was not determined (i.e. the percentage of tracks skipped) is specified for each individual test condition at the top of each plot. As a reference, the total number of tracks possible for a given test condition was 24, i.e. six participants, two repeats, two (positive and negative) ITD values.

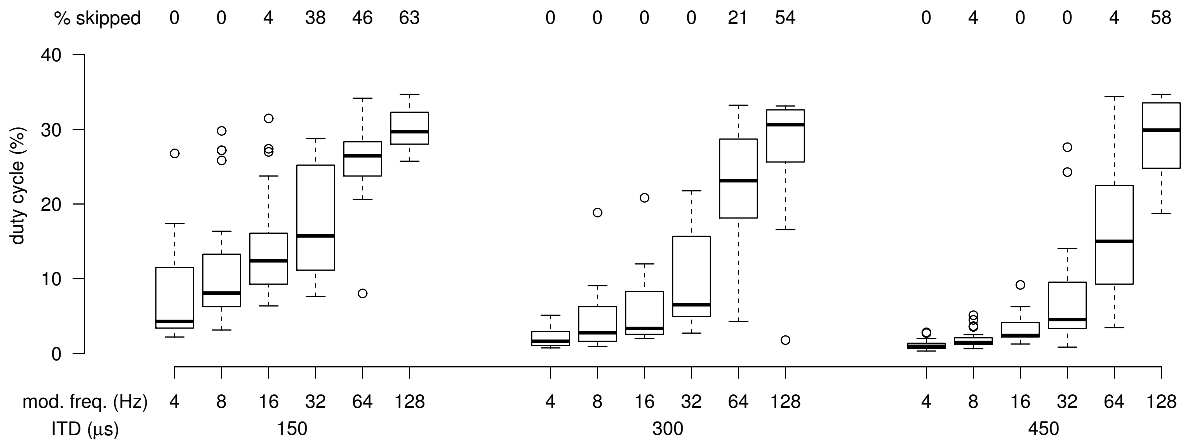

Figure 3.2: Performance for the *broadband* binaural modulation detection task presented as box plots. The different test conditions are plotted along the x-axis and are grouped based on magnitude of spatial separation between the two ITDs used in the binaural modulation. Six different modulation rates were tested for each magnitude of ITD. Above the data for each condition is a value representing the percentage of tracks skipped due to the adaptive tracking variable repeatedly exceeding a duty cycle of 35%.

Looking at Fig. 3.2, we see that the mean duty cycle required for detection increases with increasing modulation frequency. Furthermore, thresholds tend to decrease as the spatial separation between the two ITD-lateralized noises increases. Detection of the ITD-switching stimulus appears to become more difficult at the higher modulation frequencies as seen by the considerable percentage of skipped tracks. The unequal sample sizes due to the skipped tracks

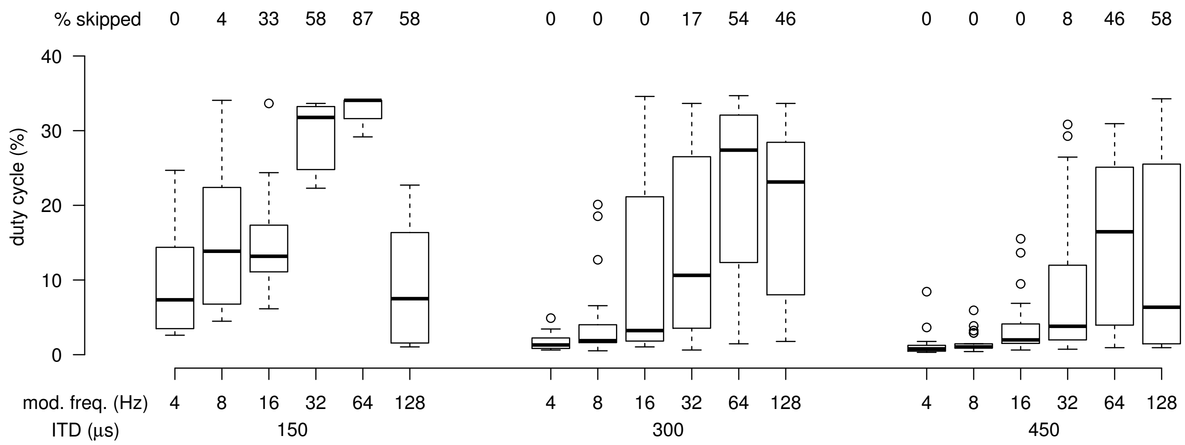

Figure 3.3: Performance for the *low-pass* filtered (LPF) binaural modulation detection task. Plots follow the same convention described in Fig. 3.2.

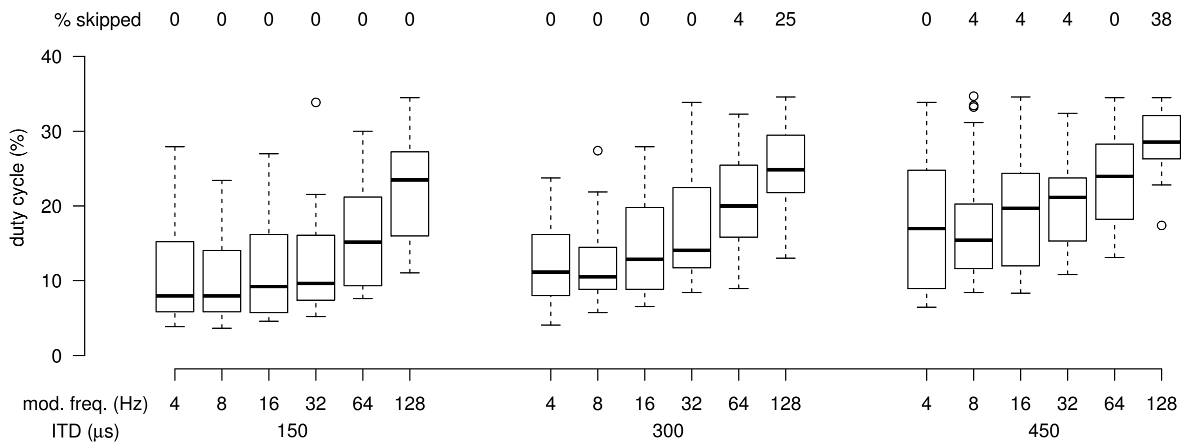

Figure 3.4: Performance for the *high-pass* filtered (HPF) binaural modulation detection task. Plots follow the same convention described in Fig. 3.2.

is likely the reason that Levene's test for equality of variances across all test conditions was found to be violated ($F = 5.58$, $p < 0.001$). Furthermore, the distribution of the data has a questionable shape of normality due to the natural limit of the tracking variable (duty cycle). Since manipulations of the data, such as log transformations, can be problematic (Erceg-Hurn and Mirosevich, 2008), statistical methods employing trimmed means and bootstrapping were used for the data analysis here. The trimmed mean is the mean of the data distribution after values exceeding a specified quantile value on both sides of the distribution are truncated. The combination of both trimmed means and a percentile bootstrap method is an effective method for controlling the Type-I error when the shape of the distribution cannot be assumed (Wilcox, 2012, pg. 387).

A two-way analysis for within-by-within design was carried out across the different ITDs and binaural modulation frequencies to compare the trimmed means of the marginal distributions of the broadband data. For the subsequent analyses, 20% trimming and 599 bootstrap samples were applied, which have been shown to be appropriate settings for these types of analyses (Wilcox, 2012, pg. 162). Hochberg's method was applied for the control of familywise error rate in post-hoc analysis. For the reporting of the statistical results, $\hat{\Psi}$ represents the estimated difference between the measures of location, e.g. the trimmed mean.

A multiple comparison analysis indicated a statistically significant decrease in the mean duty cycle required for detection as ITD increased. With respect to the 150 μs conditions, decreases in threshold by an average of 6.04% (95% CI:[2.84, 9.13]) for the 300 μs conditions and by an average of 7.72% (95% CI: [5.22, 10.7]) for the 450 μs conditions were reported. With regard to the influence of the binaural modulation frequency, a statistically significant difference was found for nearly all combinations of conditions where a general increase in thresholds was observed for increasing binaural modulation frequency. The statistically significant interactions reported in the post-hoc analysis indicate that the difference in thresholds between lower binaural modulation frequencies was larger for the 150 μs condition than it was for the 300 or 450 μs ITD conditions.

Considering the results in terms of duration of the noise at the non-zero ITD, i.e. the ITD attributed to the binaural modulation percept, it does not appear that there was necessarily a constant duration required for detection. For the conditions with a 150 μs ITD, the durations steadily decreased with increasing modulation frequency from 16 ms for the 4 Hz binaural modulation frequency to 2.4 ms for the 128 Hz binaural modulation frequency. At the larger ITD values, however, it appears that a relatively constant duration at the target ITD was required: ranging from 2.3 to 4.7 ms for the 300 μs ITD conditions and 1.8 to 2.3 ms for the 450 μs ITD conditions.

The same statistical analysis was conducted on the data of the LPF conditions plotted in Fig. 3.3. Similar to the broadband conditions, a statistically significant decrease in thresholds with increasing ITD was seen for all combinations of ITD conditions. With respect to the 150 μs ITD conditions, decreases in threshold by an average of 6.15% (95% CI:[2.95, 9.87]) for the 300 μs conditions and by an average of 9.37% (95% CI: [6.66, 12.7]) for the 450 μs conditions were reported. Statistically significant differences based on binaural modulation frequency were observed that indicated a general increase in thresholds with increasing modulation frequency. For the 128 Hz condition, however, a statistically significant *reduction* in thresholds from the 64 Hz to 128 Hz condition was found.

Although a statistically significant interaction was reported for the LPF data, these interactions were nearly all related to differences in thresholds for the 128 Hz binaural modulation frequency across the various ITD conditions. It is particularly interesting that a trend of decreasing thresholds was seen between the 64 Hz and 128 Hz conditions. Similar to the broadband condition, a majority of the tracks were skipped for the 128 Hz binaural modulation frequency, but the average thresholds were noticeably lower for the LPF condition. It is possible that listeners attempted to use a different cue to complete the task since the binaural modulation percept is likely weak or nonexistent at high modulation rates. Furthermore, the average duration of the non-zero ITD required for the detection of the 128 Hz binaural modulation was below 1 ms for both the 150 and 450 μs ITDs in the LPF condition. Although the possibility of a different cue will be discussed in more detail later in this chapter, it is quite unlikely that the binaural system is fast enough to detect periodic changes in binaural information for such brief durations.

For the HPF conditions in Fig. 3.4, a statistically significant *increase* in thresholds with increasing ITD was observed. A significant difference was not seen, however, between the 150 μs and 300 μs conditions. With respect to the 300 μs conditions, an average increase in thresholds by 4.10% (95% CI:[2.46, 6.22]) for the 450 μs conditions was seen, whereas with respect to the 150 μs ITD conditions, an average increase in thresholds by 7.07% (95% CI:[5.02, 9.31]) for the 450 μs conditions was seen. Post-hoc analysis indicated a statistically significant increase in thresholds with increasing binaural modulation frequency. A statistically significant interaction between ITD and binaural modulation frequency was not observed for the HPF data.

When the data were grouped based on the different spectral conditions, the homogeneity of variance assumption was not violated. For this reason a standard one-way ANOVA was possible, and a statistically significant difference in the spectral condition was reported ($F(2, 1091) = 62.0, MSE = 5828, p < 0.05$). Post-hoc Tukey's HSD analysis shows that the thresholds for the HPF data were significantly higher than the broadband ($\hat{\Psi} = 7.10$, 95% CI:[5.40, 8.80]) and the LPF ($\hat{\Psi} = 6.36$, 95% CI:[4.72, 8.00]) thresholds. Furthermore, the broadband and LPF data were not statistically different from each other. It is worth pointing out that although the thresholds for the HPF data were significantly higher than the other two spectral conditions, the mean thresholds were actually lower than the broadband and LPF data for the 150 μs ITD in most of the binaural modulation frequency conditions.

3.3 Experiment 2: Binaural modulation frequency discrimination

The results of the first experiment indicated that detection of the ITD switching was possible across a range of binaural modulation frequencies. At the very high modulation rates, however, it seems there was possibly another cue other than the modulation percept that could have been used for detecting the ITD-switching stimulus, e.g. a change in the source width related to changes in the long-term cross-correlation function. The second experiment was designed to require the proposed binaural modulation percept resulting from the alternating ITDs to complete the listening task. The focus was to determine if the binaural modulation component described by listeners in the pilot investigations contains a modulation which is discriminable in rate. Since it is conceivable that the described modulation in this stimulus is perceivable, yet, not necessarily localizable, a secondary goal of this experiment was to determine if the binaural modulation percept also possesses a discriminable lateral position.

3.3.1 Methods

For Exp. 2, listening participants were presented with two separate intervals of the ITD-switching stimulus that was described in Sect. 3.1. The ITD with the longer duty cycle of the binaural modulation period was set to an ITD of 0 μs. The magnitude of the ITD for the noise with the shorter duty cycle, i.e. the ITD that contained the binaural modulation percept in question, was laterally positioned at 300 μs. In a given two-interval presentation, the signs of the laterally positioned ITD values were opposing where the side of the first interval (i.e. left or right) was randomized. The task of the listeners was to determine which side contained the slower modulation. This paradigm was employed to test if the proposed binaural modulation is both localizable in lateral position and discriminable in modulation frequency.

The experiment was conducted as a method of constant stimulus where each test condition was repeated 20 times in a given session. In a single session, the pedestal binaural modulation periods tested were 240, 120, and 60 ms (approximately 4, 8, and 16 Hz modulation frequencies, respectively). Although higher binaural modulation rates were shown to be detectable in Exp. 1, pilot studies with the frequency discrimination paradigm indicated 16 Hz was the uppermost frequency worth testing. The difference between the modulation rates of the intervals in a two-interval presentation were specified as a number of semitones different from a given pedestal binaural modulation frequency. The number of semitones different from a given pedestal binaural modulation frequency employed in this experiment was $\pm 3, \pm 7, \pm 11, \pm 15$ or ± 19 semitones. The duty cycle for the non-zero ITD, which corresponds to the position of the binaural modulation percept being investigated, was fixed at 25%.

Due to the limited temporal resolution of the auditory system, the percept of modulation is likely to be less salient or not heard at all with increasing rates as shown in Exp. 1. Since feedback was provided to the listeners in this experiment, it was possible that an inference could be made concerning the rate of modulation when one interval contained a modulation too high to be detected. For this reason, 20 catch-trials (9% of total trials) were added for each pedestal binaural modulation frequency such that the ITD-switching stimulus was present in only one of the two intervals. Listeners were instructed to respond using a third response alternative ('catch') when one interval contained no perceivable modulation. Furthermore, the listeners were specifically instructed that correctness of responses for the catch-trials was important and should not be ignored.

The catch-trials were created with the same techniques used for the reference intervals of Exp. 1

using two noises with non-interrupted ITDs. The ratio of summing these two noises together was determined by Eq. 3.1. The same ±1 dB precautionary roving of the SNR and the ±50 μs roving of the ITDs were applied. Prior to the formal experiment, a practice session of approximately 20 minutes in duration was provided. Participants completed three separate sessions on different days where each session took approximately two hours in total. The stimuli were 600 ms in duration, which included 50-ms raised cosine onset/offset ramps, and were calibrated to a level of 70 dB SPL. A pause of 400 ms was inserted between the first and second intervals of a given presentation. Seven listening participants (four female, three male; average age 26 years) participated in this study; all six participants from Exp. 1 were included here.

3.3.2 Results

The ability to discriminate the side and frequency of the binaural modulation percept produced by the ITD-switching stimulus is plotted in Fig. 3.5 for the population of listeners. The data are plotted as values of d' where the open symbols represent the test conditions in which the binaural modulation frequency was correctly discriminated. Each symbol represents the various number of semitones different from the respective binaural modulation frequencies tested: 4 Hz (circles), 8 Hz (triangles), 16 Hz (squares). The closed symbols indicate the catch-trials for each binaural modulation frequency where the task of the listener was to identify that one of the two intervals did not contain a modulation. A horizontal line at d' = 1, which represents a percent-correct score of 63%, is also provided as a reference. A one-way ANOVA revealed that no statistically significant difference between the three different sessions was observed. For this reason, data across the three repetitions are grouped together for the following analysis.

Overall, listeners scored above chance (33%), ranging from 65% to 85% correct across a large-majority of the conditions. Assuming a value of d' = 1 as the discrimination threshold, it appears that a difference of only seven semitones (i.e. the second open symbol to the left or right from the pedestal frequency) was required for listeners to discriminate from the 4 and 8 Hz pedestal frequency conditions. This difference equates to approximately a 50%-change in binaural modulation frequency required for discrimination of the side and frequency. Accuracy in the catch-trial conditions was also reasonably accurate with approximately 78% of responses being correct for each of the pedestal binaural modulation frequency conditions tested.

For the 16 Hz pedestal frequency, a difference of eleven semitones was required to achieve a value of d' > 1 for discrimination between the 16 Hz binaural modulation frequency. Accurate discrimination in the 16 Hz pedestal condition, however, was only possible for the binaural modulation frequencies tested below the pedestal frequency as seen in the dramatic drop in performance for the conditions above 16 Hz. This inability to discriminate at even very large

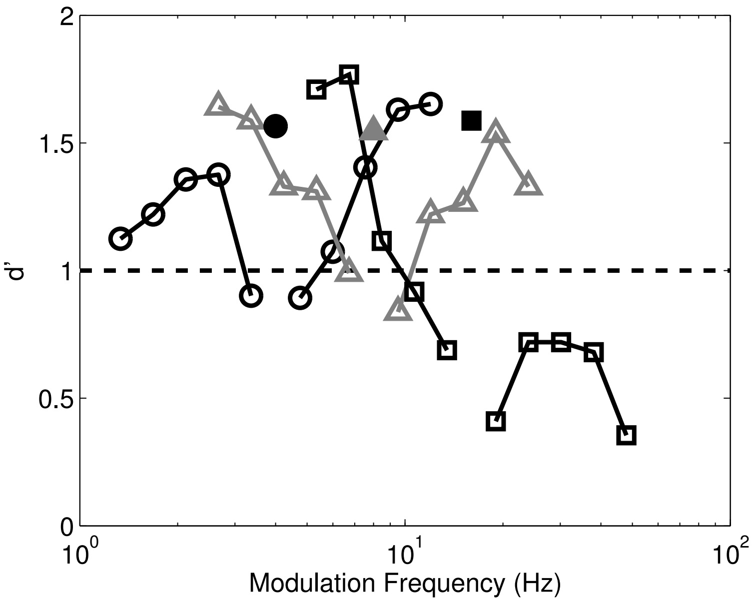

Figure 3.5: Results indicated as d' across the population of listeners in the binaural modulation frequency discrimination task for different pedestal modulation frequencies: 4 Hz (circles), 8 Hz (triangles), and 16 Hz (squares). Open symbols represent performance on test conditions where the listeners discriminated between two binaural modulations. Filled symbols represent catch-trials where the listeners detected an interval without binaural modulation. As a reference, a horizontal line at d' = 1 is indicated that corresponds to a percent-correct score of 63%.

differences in frequency is likely due to an inability to perceive a distinct modulation, thus, resulting in a confusion with the catch-trials. A slight drop in performance is also seen for the upper-most frequency (i.e. the +19 semitones condition) in the 8 Hz condition, although discrimination was still possible for the difference of 15 semitones indicating that a binaural modulation frequency of 19 Hz still maintained a perceivable modulation.

Analysis of individual performance shows that one of the seven participants was only slightly better than chance for many of the conditions tested. Two additional experimental sessions were conducted for this participant to see if more training was required, but no improvement in performance was observed. Although the results from this participant reduce the d' values across the population data, the overall trends are not distorted; thus, the data are included for completeness.

3.4 Experiment 3: Modulation masking between monaural and binaural modulations

Detection of AM can be attributed to purely monaural processing; however, the modulation percept from the ITD-switching stimulus can only result subsequent to binaural processing.

This leaves the question as to whether the sensitivity to monaural modulation is affected by the presence of binaural modulation, and conversely, is sensitivity to this binaural modulation affected by monaural modulation. Because sensitivity to temporal changes in amplitude does not require binaural processing, the sinusoidal AM used in this set of experiments will be referred to as a "monaural modulation" whereas the perceived modulation resulting from the ITD-switching stimulus will be explicitly referred to as a "binaural modulation".

3.4.1 Methods

The third experiment is broken into two separate investigations: (a) the influence of binaural modulation on the ability to detect monaural modulation and (b) the influence of monaural modulation on the ability to detect binaural modulation.

Experiment 3a - Masking of monaural modulation with binaural modulation

For this experiment, a three-AFC task was used where listeners needed to detect the interval containing a diotic monaural modulation (i.e. AM) applied to the binaural modulation stimulus. Both the target and reference intervals contained the binaural modulation stimulus. Construction of the binaural modulation stimulus was such that the ITD with the longer duty cycle (and reportedly a continuous noise percept) was positioned at an ITD of 0 μs. The ITD with the duty cycle fixed at 25% (and reportedly containing a modulating percept) was laterally positioned with a magnitude of 300 μs, but the sign of the ITD was randomized in each interval.

The target interval was the one interval where an AM was imposed onto the binaural modulation stimulus. The modulation depth, m, of the AM was adaptively adjusted in dB (20 log m) according to a two-down, one-up rule. The modulation depth at the beginning of the adaptive track was -8 dB. The starting step-size was 8 dB and was halved at every two reversals until a step size of 0.5 dB was reached upon which six additional reversals were completed before the track was terminated. The reported value for the modulation depth at threshold was the average value on each of the final six reversals.

The monaural modulation frequency was either 4, 8 or 16 Hz and began with a random starting phase. In a similar fashion as used in Exp. 2, differences in modulation frequencies were specified in semitones. The binaural modulation frequency applied in a given measurement track was specified as a fixed number of semitones different ($\pm 24, \pm 16, \pm 8$ or 0 semitones) from the monaural modulation frequency. Tracks were blocked on the monaural modulation frequency such that both the monaural and binaural modulation frequencies were fixed throughout an entire track. To determine a baseline performance for comparison, thresholds for AM detection with diotic, broadband Gaussian-noise were collected for each monaural modulation frequency.

All conditions were repeated three times in separate sessions for each listener where each session took approximately 1.5 hours. The presentation order for the different monaural frequency conditions was randomized for each participant and each session. Prior to the formal experiment, a practice session of approximately 20 minutes in duration was conducted. The stimuli were 600 ms in duration, which included 50-ms raised cosine onset/offset ramps, and were calibrated to a level of 70 dB SPL subsequent to the application of AM. A pause of 300 ms was inserted between the three intervals of a given presentation. The same six listeners from Exp. 1 participated in this experiment.

Experiment 3b - Masking of binaural modulation with monaural modulation

For this experiment, a three-AFC task was used where listeners needed to detect the interval containing a binaurally modulating stimulus which was used as the carrier signal to a fully modulated AM. A fully modulated AM masker was present in all intervals. The ITDs in the target binaural modulation stimulus switched between 0 and 300 μs, but the sign of the 300 μs ITD was randomly chosen in each interval, i.e. the target modulation could be randomly lateralized to the left or right of the midline. The carrier signal for the fully modulated AM in the reference intervals was created with the same technique described in Exp. 1 using two noises with non-interrupted ITDs. The ratio of summing these two noises together was determined by Eq. 3.1. The same ± 1 dB precautionary roving of the SNR and the ± 50 μs roving of the ITDs were applied.

The duty cycle of the binaural modulation period was adaptively adjusted in percent according to a two-down, one-up rule. The adaptive track began with a duty cycle of 20% and a starting step size of 10%. The step size was halved at every two reversals until a step size of 0.625% was reached upon which six additional reversals were completed before the track was terminated. The reported duty cycle at threshold was the average value on each of the final six reversals. To ensure that the binaural modulation remained at the more lateral ITD, the maximum duty cycle allowed was 40%. If the listener reached the maximum value five times, the track would terminate without recording a threshold.

The binaural modulation frequency was either 4, 8 or 16 Hz and began with a random "starting phase" of the binaural modulation period, i.e. the stimulus could start at any point within the binaural modulation period. The monaural modulation frequency was specified as a number of semitones different ($\pm 24, \pm 16, \pm 8$ or 0 semitones) from the binaural modulation frequency. Tracks were blocked on the binaural modulation frequency such that both the monaural and binaural modulation frequencies were fixed throughout an entire track. To determine a baseline performance for comparison, thresholds for binaural modulation detection without an AM envelope were collected for each binaural modulation frequency. The stimuli were 600 ms in duration, which included 50-ms raised cosine onset/offset ramps, and were calibrated to a level of 70 dB SPL subsequent to the application of AM. A pause of 300 ms was inserted between the three intervals of a given presentation. The same listening participants and training protocol related to Exp. 3a were used here.

3.4.2 Results

Experiment 3a - Masking of monaural modulation with binaural modulation

The results of Exp. 3a are shown in Fig. 3.6. Filled symbols represent the baseline AM detection performance with a diotic, broadband, Gaussian-noise carrier signal. Open symbols represent the test conditions when a binaural modulation was present. The three different monaural modulation frequencies tested are plotted using different shapes: 4 Hz (circles), 8 Hz (triangles), and 16 Hz (squares). Data for the 8 Hz condition are plotted in grey strictly for the motivation of readability.

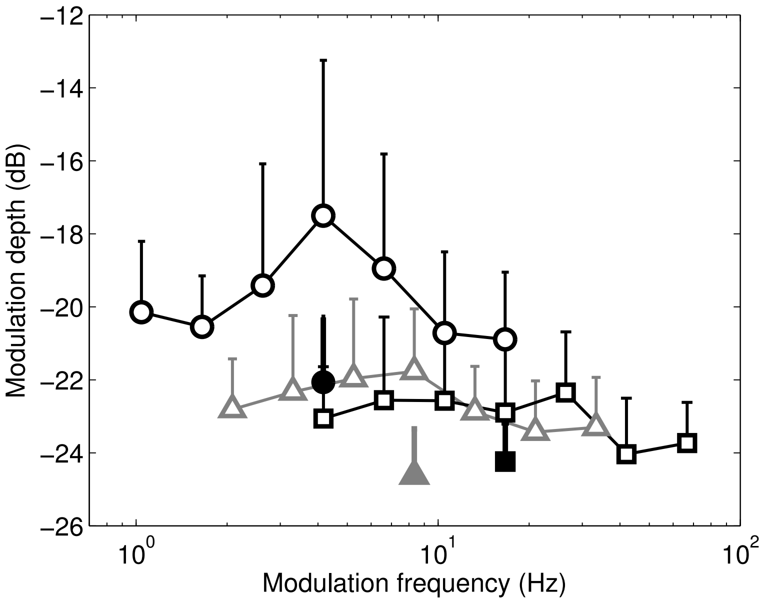

Figure 3.6: Results of Exp. 3a plotted as average modulation depth thresholds (in dB) for AM detection as a function of modulation frequency for the population of listeners. The baseline AM detection thresholds are plotted as filled symbols for the different pedestal modulation frequencies plotted in different shapes: 4 Hz (circles), 8 Hz (triangles), 16 Hz (squares). Test conditions when the binaural modulation was present are plotted as open symbols at the binaural modulation frequencies for each of the three different AM frequencies. Error bars extending above the mean value indicate one standard deviation.

Levene's test for equality of variances was found to be violated for the present analysis, ($F = 1.637$, $p = 0.033$). For this reason, the same trimmed mean plus percentile bootstrap methods as applied in Sect. 3.2.2 were used for the statistical analysis here. Since this violation of homogeneity could be considered borderline, an analysis using traditional ANOVA methods was also conducted. The results from the ANOVA concur with the results reported in this section.

To test the effect of the binaural modulation on the detectability of a monaural modulation, a one-way analysis with the data grouped as either a baseline or test condition revealed that monaural modulation detection performance was statistically significantly poorer by 2.4 dB (95% CI:[1.80, 2.99]) when a binaural modulation masker was present. Comparison of the different test conditions was conducted via a two-way, within-by-within design against the two parameters: pedestal modulation frequency and number of semitones different. A multiple comparison analysis using Hochberg's method to correct for the familywise error rate indicate that the 4 Hz pedestal modulation frequency required a significantly larger modulation depth as compared to the 8 Hz ($\hat{\Psi} = 2.60$, 95% CI:[2.19, 3.02]) and 16 Hz ($\hat{\Psi} = 2.96$, 95% CI:[2.52, 3.34]) modulation frequencies. Post-hoc analysis reveals that the ±16, $+24$ semitone different conditions all had significantly lower thresholds than the zero semitone different condition. This potential modulation frequency tuning is most apparent in the 4 Hz condition as seen by the peaked threshold for the test conditions at the pedestal AM frequency. In general, the amount of masking was quite small when compared to the amount of masking resulting from an AM masker, e.g. Ewert and Dau (2000). This implies that the interaction between the monaural and binaural modulation sensitivities is relatively weak.

Experiment 3b - Masking of binaural modulation with monaural modulation

The results of Exp. 3b are shown in Fig. 3.7. Filled symbols represent the baseline binaural modulation detection performance when no added monaural modulation was imposed. Open symbols represent the test conditions when a monaural modulation was present. The three different binaural modulation frequencies tested are plotted using different shapes: 4 Hz (circles), 8 Hz (triangles), and 16 Hz (squares). Data for the 8 Hz condition are plotted in grey strictly for the motivation of readability.

Although Levene's test for equality of variance was not found to be violated for this data set, the data have a questionable normal distribution due to the natural limit of the adjustment variable (duty cycle). For this reason, the same trimmed mean plus percentile bootstrap methods as applied in Sect. 3.2.2, which do not require normally distributed data, are used for the statistical analysis here.

To test the effect of the monaural modulation on the detectability of the binaural modulation, a one-way analysis with the data grouped as either a baseline or test condition revealed that the increase from a duty cycle of 2.51% for the baseline conditions to 4.56% for the test conditions was statistically significant ($\hat{\Psi} = 2.05$, 95% CI:[1.21, 2.76]). This near doubling in the required

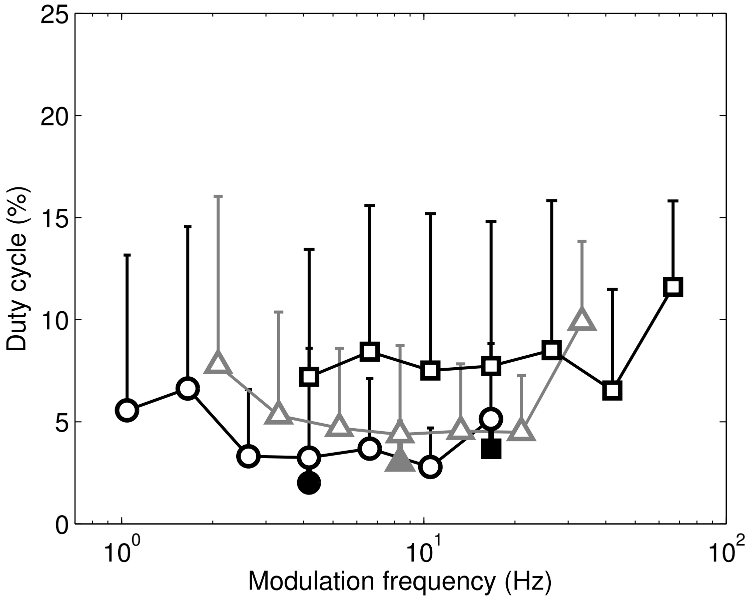

Figure 3.7: Results of Exp. 3b plotted as duty cycle thresholds (in percent) for binaural modulation detection as a function of modulation frequency for the population of listeners. Data are plotted the same as described in Fig. 3.6

duty cycle needed for detection of the binaural modulation, i.e. doubling in the duration of the noise at the target ITD, poses a potentially greater implication than the relatively small increase in AM thresholds seen in the previous experiment.

Comparison of the test conditions based on the pedestal modulation frequency and the number of semitones different was conducted using a two-way, within-by-within design. A multiple comparison analysis using Hochberg's method to correct for the familywise error rate indicated that the 4 Hz pedestal modulation frequency required a significantly smaller duty cycle as compared to the 8 Hz ($\hat{\Psi} = 1.27$, 95% CI:[0.71, 2.15]) and 16 Hz ($\hat{\Psi} = 3.25$, 95% CI:[2.41, 4.41]) modulation frequencies. The duty cycle required for the 8 Hz condition was also significantly smaller ($\hat{\Psi} = 1.81$, 95% CI:[1.22, 2.59]) than the 16 Hz condition. Assuming a constant duty cycle, the non-zero ITD would be present for the same total duration throughout the entire stimulus, which is independent of the modulation frequency. The observed increase in thresholds with increasing modulation frequency, therefore, suggests that detection of alternating ITDs improves when the target ITD is presented for longer periodic durations, even at the consequence of fewer ITD transitions due to the slower modulation rate. With regard to the statistically significant effect of the number of semitones, post-hoc analysis reveals that this difference resulted from the elevated thresholds for the +24 semitone conditions. This indicates that the possible tuning of the modulation detection interference seen in Exp. 3a was not present when the binaural modulation was used as the target modulation.

3.5 Discussion

The experiments presented in this study investigated the perception of a binaural modulation resulting from the periodic switching between noise tokens with different ITDs. The first experiment explored the detectability of this ITD-switching stimulus and illustrated that the binaural system is fast enough to discriminate between the proposed binaural modulation stimulus and a stimulus composed of two continuously presented ITDs. This was seen particularly for the larger interaural disparities between the two ITDs where changes in the binaural cues of only 2 ms in duration were sufficient for listeners to identify the ITD switching. The pilot experiment described in Sect. 3.1 indicated that this stimulus can produce two simultaneous sound streams: one being a modulation percept and the other being a continuous noise percept. Since the detection of ITD switching in the first experiment may not have been based solely on perceived modulation cues, the second experiment required listeners to discriminate both the frequency and sidedness of the modulation percept. Listeners were quite accurate in their ability to discriminate between the binaural modulation frequency for this stimulus, but the rolloff in performance above 16 Hz indicates that the salience of the modulation percept was best for slow ITD-switching rates. In the third experiment, the modulation detection interference produced by the binaural modulation on the ability to detect AM was shown to be very small although some indication of frequency tuning characteristics was found. When AM was used to mask sensitivity to the binaural modulation, a doubling in the required duration of the noise at the ITD corresponding to the binaural modulation percept was necessary for detection. No indication of tuning was seen, however, in this modulation masking paradigm.

The detection of the binaural modulation percept, as explored in Exp. 1, illustrated a general increase in the required duty cycle for detection when the modulation frequency increased. This trend was seen almost exclusively with the exception of the low-frequency spectral condition where thresholds surprisingly decreased for the 128 Hz binaural modulation frequency. The average duty cycle required for detection in this 128 Hz condition resulted in the duration of the target ITD noise to be less than 1 ms. Since it is unlikely that the binaural system possesses a means to detect a modulating ITD pattern for such short durations, a more reasonable theory is that listeners were unable to hear the modulation and resorted to using a different cue. It is suspected that this alternative cue was related to the perceived source widening due to changes in the long-term cross correlation.

Because the ITD switching occurred synchronously and simultaneously, there were small portions (only a few samples) at the beginning and end of the target ITD noise tokens which were uncorrelated due to the imposed ITD. At low binaural modulation rates these fringes imposed little change to the coherence within a given ITD noise token since the duration of the noise token was much longer than the short fringes. For the high binaural modulation rates, however, the durations of the noise tokens were very brief, and the fringes consumed a non-negligible portion of the noise token. The implication that this had on the stimulus was an overall decrease in coherence. Furthermore, this could have also caused a difference in coherence between the target and reference intervals of the three-AFC paradigm even though the long-term cross-correlation patterns remained essentially the same.

Considering that a majority of trials for the 128 Hz binaural modulation frequency in the broadband and LPF conditions were skipped, the salience of the coherence cue and/or the implemented roving limited the utility of using discrepancies in coherence for accurately completing the listening task. It is still not understood why this drop in threshold was seen only for the LPF condition and not the broadband condition. Considering an optimal selection of audio information, it seems that the auditory system would utilize only the low-frequency information when presented this task in the broadband condition and show a similar drop in thresholds for the 128 Hz condition. One would be left to assume that this optimal selection was not possible here, and influence from other frequency bands or interaction of information across frequency inhibited the auditory system from making the best choice.

For the HPF condition of Exp. 1, it is unlikely that the temporal structure of the binaural modulation stimulus could provide the necessary cue to detect the target modulation since high-frequency ITD sensitivity is restricted to envelope cues. If we consider that the cross-correlation pattern between the envelopes of the respective left/right high-frequency peripheral filter pairs are very broad, it would be questionable if the binaural system is sensitive enough to detect small shifts in the cross-correlation pattern resulting from the binaural modulation. Detecting such small shifts at very high modulation rates would be even more difficult, but the very few skipped tracks across nearly all binaural modulation frequencies in the HPF spectral condition (a stark contrast to the number of tracks skipped in the broadband and LPF conditions) indicates that the cue used by listeners was quite reliable. It is possible that the cue based on changes in coherence mentioned earlier served as the more useful cue in the HPF condition.

A numerical analysis of the binaural modulation target stimulus and the respective reference interval stimulus was conducted to investigate if the change in coherence, i.e. peak long-term envelope cross-correlation value, between the target and reference intervals could explain the relatively good detection performance in the HPF condition. For this analysis, a basic auditory periphery was simulated using a Gammatone filterbank from Hohmann (2002) where half-wave rectification and a fifth-order low-pass Butterworth filter at 770 Hz were subsequently applied to each Gammatone channel. The normalized cross-correlation pattern was then computed using the respective left- and right-channel outputs from the simulated auditory periphery for each of the different frequency bands. The cross-correlation coefficient with the largest magnitude was calculated for both the test interval (ρ_{targ}) and reference interval (ρ_{ref}). An average of these peak cross-correlation values was computed across 100 instances of each stimulus for each combination of binaural modulation frequency and ITD magnitude. For simplicity, the duty cycle was held constant at 25% and the stimuli were computed without the roving applied in

Exp. 1.

The results of the numerical analysis are plotted in Fig. 3.8 for each ITD magnitude (columns) and each binaural modulation frequency (rows) where the x-axis in each plot represents the Gammatone channel center frequency and the y-axis is drawn for measures of coherence (ρ). The solid black curves are calculated as $\Delta\rho_{targ} = 1 - \rho_{targ}$ whereas the grey curves are calculated as ρ_{targ} - ρ_{ref}. Listeners should be most sensitive to changes in coherence when $\Delta\rho_{targ}$ (black curve) is close to zero since sensitivity to changes in coherence is best from a reference coherence of one where a change in coherence of only 0.02 is can be detected (Pollack and Trittipoe, 1959). If one assumes that detection of the binaurally modulating interval could be possible if the difference in coherence between the reference and target intervals (ρ_{targ} - ρ_{ref}) is larger than the reference coherence ($\Delta\rho_{targ}$), we can estimate in which test conditions the coherence cue might be useful for the auditory task. For the range of reference coherence values in these data, we interpret from the interaural correlation discrimination data of Culling *et al.* (2001) that this is a reasonable estimate.

Results indicate that for most of the high-frequency Gammatone channels, the envelope coherence is generally very high ($\Delta\rho_{targ} < 0.05$), which is a range of values where humans would be quite sensitive to changes in coherence (Pollack and Trittipoe, 1959). With increasing binaural modulation frequency, however, $\Delta\rho_{targ}$ tends to increase while ρ_{targ} - ρ_{ref} tends to decrease. Combination of both trends would likely lead to an increasing difficulty for discrimination based on coherence in the high-frequency spectral channels as the binaural modulation frequency increased. This resembles the general increase in detection thresholds with increasing binaural modulation frequency seen in the data of Exp. 1 (see Figs. 3.2 - 3.4).

Looking at the high-frequency Gammatone channels in Fig. 3.8, we see that the number of frequency channels where ρ_{targ} - ρ_{ref} is larger than $\Delta\rho_{targ}$ decreases with increasing ITD. This trend would support a reduced ability to discriminate the reference and target intervals based on coherence as the ITD increased, which matches the increase in detection thresholds with increasing ITD as observed in the HPF data of Exp. 1 (see Fig. 3.4). Although we cannot completely rule out detection of the ITD switching based on something other than the coherence, trends in the numerical analysis indicate that the coherence could potentially predict the tendencies observed in the psychophysical data. Assuming an optimal detector, it would be expected that this high-frequency coherence cue would be available for the broadband stimulus. Previous work, however, has shown that low-frequency spectral regions dominate ITD processing and interfere with processing of high-frequency ITD cues (Bernstein and Trahiotis, 2004). With respect to the psychophysical data presented here, this low-frequency dominance for ITD processing is supported by the statistically similar results of broadband and LPF performance, which were both shown to be significantly different than the HPF condition.

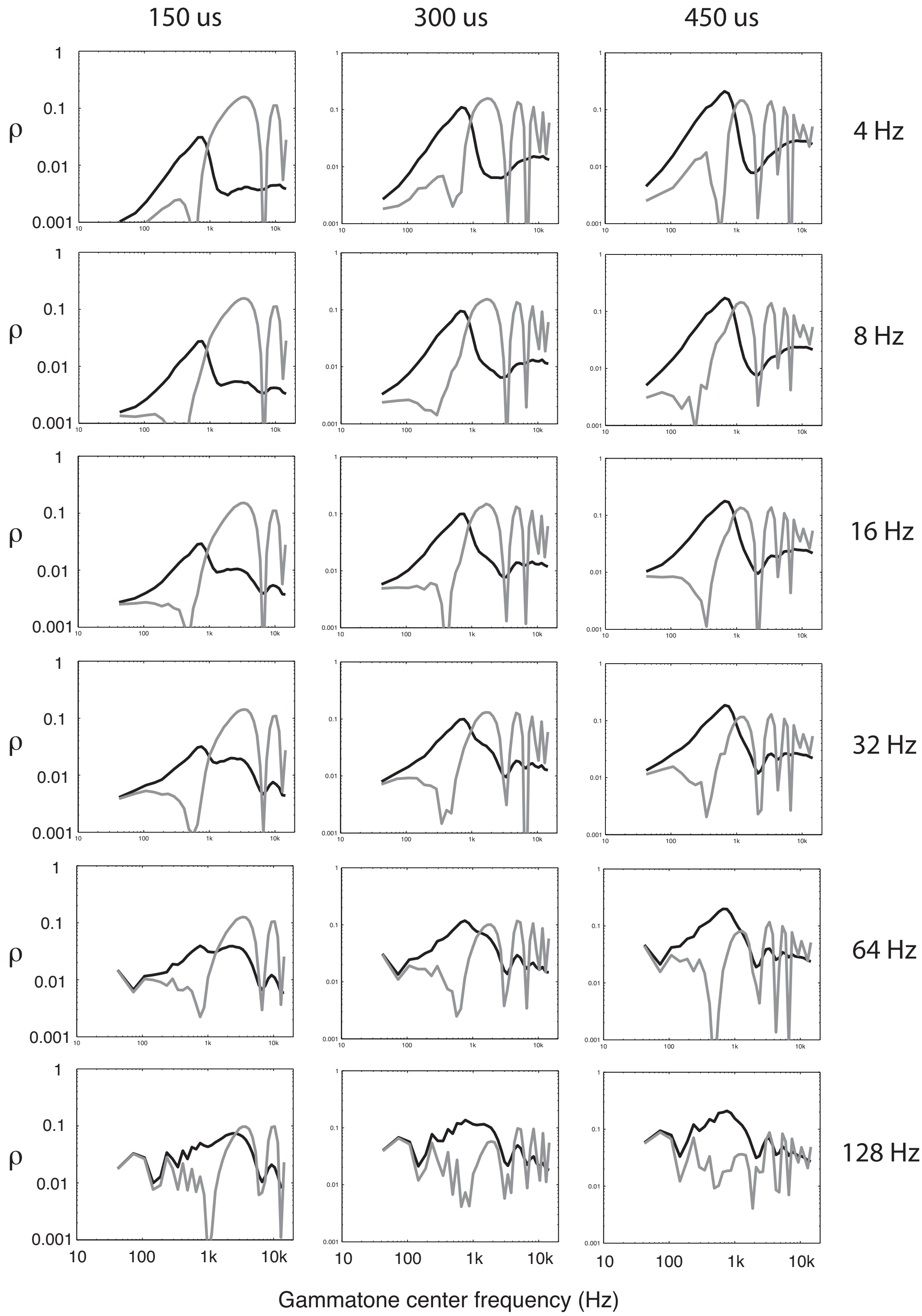

Figure 3.8: Numerical analysis of coherence as a potential cue for the binaural modulation detection task in Exp. 1 across different modulation frequencies (rows) and ITD magnitudes (columns). The black line represents the difference from a perfectly coherent noise for the target ITD-switching stimulus, i.e. $1 - \rho_{targ}$. The grey line indicates the difference in coherence between the target and reference interval, i.e. ρ_{targ} - ρ_{ref}. The Gammatone filter center frequencies are indicated along the x-axis, and values of coherence are indicated for the y-axis.

Turning our attention towards the low-frequency Gammatone channels (< 1 kHz), we see that ρ_{targ} - ρ_{ref} was never greater than $\Delta\rho_{targ}$. This would suggest that detection based on the coherence would be quite difficult and detection of the modulation percept would likely be the more useful cue for completing the task. It is possible at high binaural modulation rates, however, that the modulation percept disappears and the coherence remains as the only available cue. The skipped tracks could suggest the poor salience of this coherence cue in the low-frequency Gammatone channels. The drop in thresholds between the 64 Hz and 128 Hz binaural modulation frequency conditions for the LPF data could possibly be rationalized by the increase in ρ_{targ} - ρ_{ref} to values potentially more salient for discrimination. Again, due to the considerable number of skipped tracks in the 128 Hz condition, the utility of this coherence cue was still questionable.

In general, this analysis suggests that coherence could provide a salient cue for detecting the periodically alternating ITD cues in some conditions. Since some of the lowest thresholds were obtained in conditions where the modulation percept was likely the best cue, listeners probably transitioned to using the coherence cue when the modulation cue became less salient. The rate at which this transition would have occurred is not precisely known. However, the diminished ability of listeners in Exp. 2 to discriminate the binaural modulation frequency for frequencies greater than ∼20 Hz could provide a rough estimate. This frequency range is not too far from findings of Grantham and Bacon (1991) where the effect of sinusoidally modulated ILD cues in an AM masking paradigm were measured. They suggested that the sinusoidally modulated ILD cues below 32 Hz were useful for detection, but at higher modulated ILD rates, listeners likely resorted to using long-term interaural coherence differences.

Because it is possible that a cue other than the binaural modulation percept was actually used in the first experiment, Exp. 2 investigated whether the modulation percept was indeed present by measuring if it was salient enough to be discriminated both in frequency and in lateral position. In order for the binaural system to discriminate between different rates of this binaural modulation, it is necessary that the temporal pattern in the binaural display is sufficiently represented. One could, therefore, interpret the sharp rolloff in accuracy for the upper binaural modulation frequencies of Exp. 2 as the binaural system possessing an insufficient resolution to perceive the temporal pattern of these discrete changes in ITD. The reduced ability to detect the binaural modulation pattern would then likely result in a confusion with the catch-trials where listeners were instructed to report not hearing a modulation. This confusion with the catch-trials should not be directly attributed to an inability to detect the ITD switching per se since the results from Exp. 1 indicate that the minimum duty cycle required for detecting the 16 Hz binaural modulation in the broadband condition (5.8%) was much less than the 25% duty cycle used in Exp. 2. Conversion of the 5.8% duty cycle for the 16 Hz condition results in a duration of 5.2 ms for the target ITD, which is comparable to the very brief (2-8 ms), yet detectable, changes in ITD that was shown in Bernstein *et al.* (2001). Results of Exp. 2, however, suggest that the temporal pattern in the binaural display is only sufficiently represented for frequencies near or below 16 Hz for a task that requires frequency discrimination of the binaural temporal pattern.

A noticeable rolloff in frequency discrimination performance was also seen for the lower binaural modulation frequencies in Exp. 2. As one decreases the modulation frequency of this binaural modulation stimulus, it could be argued that the percept turns to something resembling movement of a single source between two locations rather than the separate continuous and modulating sound streams. In other words, at slower modulation frequencies the binaural system is fast enough to track the movement to discrete points in space. This would be in agreement with Grantham and Wightman (1978) where the ability of listeners to follow dynamically changing ITD cues in detail rapidly declines with increasing frequency between 2-5 Hz depending on the listener. If it is necessary to apply a switching rate that exceeds the speed at which the binaural system can track source movement in detail in order to produce the modulation percept, the decrease in d' observed for the lower frequencies used in the 4 Hz binaural modulation condition of Exp. 2 could possibly be attributed to a weaker modulation percept at these slower switching rates. For the very low modulation rates, the short duration of the stimulus as compared to the binaural modulation period could have resulted in the stimulus being perceived at only one ITD, and consequently, listeners confused these stimuli with the stationary ITD cues of the catch-trials.

Experiment 3 investigated whether a reciprocal interaction between binaural modulation sensitivity and monaural modulation sensitivity exists. Although the statistical results of both studies in Exp. 3 indicated elevated thresholds from the baseline condition, the strength of masking in Exp. 3a was very weak. The presence of the binaural modulation produced by our ITD-switching stimulus only elevated the AM detection thresholds by 2.4 dB. This is much less than the 10-15 dB of masking observed when the target AM was masked by a separate AM (Ewert and Dau, 2000). A larger amount of masking ($\sim$ 10 dB) was also shown in Thompson and Dau (2008) for an AM detection task when a sinusoidally modulated ILD (i.e. binaural modulation) was present. Both Ewert and Dau (2000) and Thompson and Dau (2008) demonstrate frequency selectivity for modulations where increasing the frequency separation between the target and masking modulations improves AM detection. The frequency selectivity for modulation sensitivity shown in Thompson and Dau (2008), however, was much broader in shape than the tuning curves shown in Ewert and Dau (2000). This broader tuning for AM detection when a binaural modulation was present bears resemblance to what was seen in Exp. 3a of the present study. For both Exp. 3a and the Thompson and Dau (2008) study, the amount of masking decreased by only a few dB when the modulation frequency difference increased from zero to two octaves. Although the overall amount of masking was much smaller for our Exp. 3a as compared to Thompson and Dau (2008), the broad *shape* of binaural masking is comparable to our findings using a modulating ITD binaural masker. For complex listening situations encountered in daily life, it is probably beneficial that the interaction between binaural and monaural modulation sensitivities is weak because they both represent independent properties of the auditory scene. This implies that random fluctuations of ITD cues in normal listening environments should not seriously degrade a listener's ability to utilize amplitude modulation cues, which have been shown to be beneficial for speech intelligibility.

The fact that this ITD-switching stimulus produces a percept of modulation is interesting in itself. Particularly when considering that the modulation percept comes from the ITD with the *shorter* duty cycle whereas a continuous noise is perceived at the ITD with the longer duty cycle. The continuous noise percept could suggest that the binaural system may not be sensitive to short gaps in spatial information coming from a given position. This could be similar to the so-called "continuity illusion" (Warren *et al.*, 1988) where a temporally disjoint tone can sound continuous if a sufficiently loud noise masks the onsets and offsets of the individual tone segments. With regard to the continuity of the noise in the ITD-switching stimulus, one could hypothesize that top-down mechanisms help to fill in missing binaural information (due to the rapid fluctuations of bottom-up binaural cues in realistic listening scenarios) in order to minimize any disruption to the target audio streams. For this to occur, however, the perceptual description of our ITD-switching stimulus suggests that the target stream must occupy a majority of the temporal switching pattern. Although not specifically explored in this paper, the percept for a binaural modulation stimulus with a 50% duty cycle is somewhat ambiguous and listener dependent. Informal listening experiments revealed that some listeners perceived two continuous noises at the two ITDs whereas other listeners commented of hearing two modulating sounds.

A clear rationale for the percept of modulation at the ITD with the shorter duty cycle is not well-established. Peaks in the long-term cross-correlation pattern of the stimulus can describe the perceived lateralization of the two ITD positions (Jeffress, 1948, Josupeit, 2011). One could consider that modulation of cues in the binaural display somehow results in a perceived modulation or roughness quality to the sound. The perceived modulation pattern would need to fall within the temporal resolution of the binaural system's sensitivity to changes in binaural cues. Although it has not been explicitly tested here, informal listening suggests that the formation of this binaural modulation precept does not necessarily require switching between only two fixed ITDs. It appears that switching between two distributions centered around two mean ITD values possibly results in a comparable modulation percept.

The existence of such a binaural modulation raises the question as to how one should organize the monaural and binaural modulation sensitivities in an auditory model. As illustrated by Thompson and Dau (2008), sensitivity to their ILD-based binaural modulation could be accounted for in the monaural pathway. For our ITD-based binaural modulation, however, it is necessary that the modulation sensitivity originates subsequent to the binaural interaction. It is more easily conceived that a stimulus modulating in level (e.g. ILD) produces a larger amount of masking on a task requiring a sensitivity to changes in signal amplitude (e.g. AM) because the modulated ILDs produce AM modulations in each ear. Since it is possible that the majority of the masking resulting from the ILD-based modulation comes from the monaural level cues, the broad tuning of masking or the shape of the masking curve from the ILD modulation could be interference from the periodic side-to-side movement, i.e. the binaurally modulating component. This broad tuning was also seen in Exp. 3a of the present study. Since both studies show interference due to laterally modulated stimuli, it is possible that any stimulus that results in periodic side-to-side movement would lead to similar interference.

The near doubling in thresholds for Exp. 3b indicate a potentially stronger interaction between the monaural and binaural modulation sensitive pathways as compared to the study of Exp. 3a. Notably, however, no indication of frequency selective modulation interference was observed. The larger overall interference of AM on the ability to detect the ITD-switching stimulus could suggest that the presence of AM degrades the temporal representation of ITD cues in the binaural display or, potentially, that the presence of AM requires longer glimpses of continuously presented ITD information in order to detect changes in ITD.

In light of the inability of one participant from Exp. 2 being able to properly discriminate the rate of binaural modulations tested, further implications for this result are worth considering. Results indicate that this listener also produced the highest mean threshold duty cycle (16.3%) required for detection across all conditions of Exp. 1. This duty cycle value was statistically significantly higher than four of the other five participants by 4-8% meaning that this participant required a considerably longer duration at the target ITD in order to detect the ITD switching. Furthermore, this participant had the second-highest threshold duty cycle (10.1%) required for detection when a monaural masker was present in Exp. 3b. This threshold was statistically significantly higher than four of the other five participants by 6 to 8%. In contrast, this listener performed similarly to the population of listeners for Exp. 3a where the task did not require a sensitivity to the ITD switching. Taking these results as a whole could suggest that this listener was less sensitive to the purely binaural modulation as compared to the other participants. While individual differences in binaural tasks such as localization (Andéol *et al.*, 2013) and binaural masking level differences (Buss *et al.*, 2007) are not uncommon, the ability to discriminate purely binaural temporal changes could be utilizing different properties of the binaural system. The ability (or inability) to use binaural temporal changes could have an influence on listening tasks in complex, cocktail-party scenarios. With attempts being made to identify correlates to individual differences in psychophysical tasks and objective measures such as electroencephalography (Ruggles *et al.*, 2011, Choi *et al.*, 2014), it is worth considering the utility of this ITD-switching stimulus for investigating individual differences in binaural processing.

3.6 Conclusion

A purely binaural modulation was described and utilized in experimentation aimed at characterizing perceptual properties of the alternating ITD stimulus used in this study. The ability to discriminate this binaural modulation from a stimulus without the periodic ITD switching is a function of modulation rate and ITD as demonstrated in Exp. 1. Although the modulation percept appears to be a salient cue for slower binaural modulation rates, numerical analysis demonstrated that the coherence of the signal could be a potential cue for detecting the ITD-switching stimulus at the higher modulation rates and also when the stimulus is bandlimited to high frequencies. Results of Exp. 2 confirm that the ITD-switching stimulus does contain a perceivable modulation at binaural modulation rates below 20 Hz as indicated by the ability of subjects to discriminate binaural modulation rates. Additionally, the results of Exp. 2 show that this purely binaural modulation percept contains a sufficiently compact lateral position to make accurate sidedness judgements. This suggests that these very brief tokens of ITD information by themselves allow for the build up of a separate auditory stream. The final set of experiments in Exp. 3 investigated if interference exists between monaural AM and purely binaural modulation sensitivities. It appears that the modulation due to ITD switching had a weak, adverse effect on sensitivity to AM. Furthermore, the tuning of modulation masking was very broad as compared to AM modulation-masking curves. The thresholds for sensitivity to our binaural modulation nearly doubled in the presence of AM suggesting a potentially stronger interference of monaural modulation on binaural modulation sensitivity. There is clearly less interference between these two modulation domains as compared to purely monaural modulation sensitivity. This suggests that the modulations in the monaural and the binaural domain are represented mostly independently. Better understanding of the source of this binaural modulation percept can possibly provide insight into the role of the binaural system for glimpsing of spatially disparate streams of information.

4

Lateralization of alternating interaural time difference stimuli: The role of the monaural envelope[1]

A high temporal acuity of the binaural system would help to resolve inherent fluctuations in binaural information that occur in complex auditory scenes. Using a broadband noise stimulus that rapidly alternates between two different values of interaural time difference (ITD), the ability of the binaural system to hear the lateral position resulting from one of the ITD values was investigated. Results show that listeners are able to accurately lateralize brief noise tokens of only 3-7 ms in duration. In two subsequent experiments, the role of an amplitude modulation (AM) imposed on the ITD-switching stimulus used in the first experiment was tested. For wideband stimuli, the temporal position of the ITD target relative to the phase of the AM did not influence absolute lateralization or detection performance. When the stimuli were narrowband, however, detection of the ITD target was best when temporally positioned in the rising portion of the AM. The fourth experiment determined if two simultaneously presented, ITD-lateralized AM noises could be discriminated in modulation frequency. Results indicate that the spatial separation enhances modulation discriminability, particularly in the range of modulation frequencies normally found in speech.

1 An extended version of this chapter will be submitted to the *J. Acoust. Soc. Am.*

4.1 Introduction

To accurately extract target sound information, the auditory system must operate on a time scale that can accommodate the inherent interaural cue fluctuations occurring in daily life settings, e.g. target speech within an environment of interfering noise sources. Since a range of different time constants for binaural processing have been reported, it is likely that the duration of the temporal window depends on the task presented to the auditory system (Kollmeier and Gilkey, 1990, Bernstein *et al.*, 2001). For some auditory tasks, it is possible that temporal integration of information is necessary to complete the listening objective whereas for other tasks only brief glimpses of binaural information might be required. Regardless of the mechanistic requirements, it has been shown that the binaural system is sluggish in the tracking of detailed movements (Perrott and Musicant, 1977, Grantham and Wightman, 1978) and for masking level difference measurements (Grantham and Wightman, 1979, Kollmeier and Gilkey, 1990). In contrast, the binaural system has been shown to be fast enough to detect very brief changes in interaural time difference (ITD) and interaural level difference (ILD) (Bernstein *et al.*, 2001) and to detect brief periods of interaural phase agreement Pollack (1978). It is notable that the proper choice of peripheral parameters has permitted a single time constant of 30 ms to predict results for a range of binaural tasks (Breebaart *et al.*, 2001).

For speech intelligibility in a multi-talker auditory scene, a fast auditory system that can select pieces of the target sound would be important. Due to the spectro-temporally sparse nature of speech, the dominating sound source at any moment in time may alternate quickly and, consequently, so will the associated binaural cues. This concept of "glimpsing" for speech perception is illustrated by the improvement in speech intelligibility for an interrupted or modulated masker as compared to a steady noise masker (Miller and Licklider, 1950, Peters *et al.*, 1998, Buss *et al.*, 2009). A model for speech perception was proposed by Cooke (2006) where brief spectro-temporal segments of the input signal are further processed given a certain criterium, e.g. when a favorable signal-to-noise ratio (SNR) is present.

It is readily assumed that a variety of cues - temporal, spectral, spatial, etc - are used to disentangle complex auditory scenes (Bregman, 1990). For speech intelligibility, spectral cues potentially contribute more than spatial cues (Buell and Hafter, 1991, Culling and Summerfield, 1995, Shackleton and Meddis, 1992) although it is possible that the salience of spatial information for speech-source segregation largely depends on the listener and the amount of practice (Drennan *et al.*, 2003). Nevertheless, spatial separation of speech sources is known to provide an advantage in speech intelligibility (for a review see Bronkhorst (2000), Litovsky (2012)). How the auditory system selects and groups the spectro-temporal glimpses is not precisely known. Creation of a spectro-temporal binary mask based on spatial information of the target source has been applied to enhance automatic speech recognition in the presence of reverberation and other interfering sources (Roman *et al.*, 2006). Knowledge about the temporal acuity for processing brief tokens of binaural information may also contribute to developing new source separation techniques for automatic speech recognition. To determine if the auditory system is capable of attributing an absolute lateral position to brief glimpses of binaural information, a broadband stimulus inspired by Wagner (1991) and Bernstein *et al.* (2001) was constructed. This stimulus was composed of noise tokens with periodically alternating ITDs as will be described in Sect. 4.2.1.

Several studies have illustrated that a greater perceptual weight is placed on the binaural information during the onset of a sound as compared to the binaural cues during the ongoing part of the sound (Houtgast and Aoki, 1994, Freyman *et al.*, 1997, Stecker *et al.*, 2013). Presumably, such an emphasis to binaural information at the onsets of stimuli could be attributed to the auditory system selecting moments of a target stimulus prior to any adverse influence from room reflections (Devore *et al.*, 2009). These onsets could also possibly contain brief moments when the interaural coherence is high, providing optimal times to sample ITD and ILD cues (Faller and Merimaa, 2004). Since complex auditory scenes often contain amplitude fluctuations, the rising portions of the ongoing envelope may provide a similar benefit in reverberant environments as proposed for the onsets of sounds. Dietz *et al.* (2013) have shown that for a binaural beat stimulus, which is modulated in amplitude at the same rate as the binaural beat, listeners predominantly perceive the stimulus as being lateralized to the side associated with the interaural phase difference during the rising portion of the amplitude modulation (AM) cycle. Furthermore, Schimmel *et al.* (2008) has shown that the temporal envelope of sound stimuli plays a role in the ability of listeners to discriminate the lateral position between two different sounds that are simultaneously presented over headphones. Both of these studies could be interpreted as a role for amplitude modulations during the readout of binaural information.

The experiments presented here explored the temporal acuity of the auditory system for extracting brief glimpses of binaural information. Whether or not the binaural system can both detect and accurately lateralize noise tokens with rapidly changing interaural differences was investigated in Exp. 1. The second and third experiments were directed at determining the influence of monaural envelope information on the ability of listeners to report absolute lateral position (Exp. 2) and to discriminate sidedness (Exp. 3). Experiment 4 tested if the binaural system could provide a benefit when discriminating between the envelope frequencies of two, AM stimuli that were simultaneously presented.

4.2 Experiment 1: Lateralization without monaural cues

The first experiment investigated the temporal acuity of the auditory system for selecting brief, yet repeated, segments of binaural information. Using a stimulus that rapidly and periodically alternates between two ITDs, it is possible to estimate how long the auditory system needs to accurately report the lateral position for one of the two ITDs used in the stimulus. As part of this experiment, the temporal acuity was measured with respect to the spatial disparity between the two ITDs presented.

4.2.1 Stimuli

As shown in the schematic drawing in Fig. 4.1, the stimulus of this experiment periodically alternated between two different ITD values at a regular interval designated as the 'ITD-switching period'[2]. The stimulus was adapted from stimuli used in previous studies to determine a binaural temporal window for barn owls (Wagner, 1991) and humans (Bernstein *et al.*, 2001). Broadband Gaussian noise was used for each of the sounds at τ_o and τ_1, thus, switching between the two ITD values can occur simultaneously and synchronously without causing a detectable cue for the ITD switching. If either the left or right channel was presented in isolation, only a continuous, broadband noise was heard.

Prior to this experiment, listening participants qualitatively described their percept of this ITD-switching stimulus via sketches and text. Although the stimulus parameters were not exhaustively explored in a formal study, it was noted that the listeners perceived two separate components when the ITDs were sufficiently far apart (approximately $\geq 200\mu s$). For the position associated with the ITD containing the longer duty cycle of the ITD-switching period, listeners reported hearing a continuous noise. A separate sound object containing a percept of modulation was generally reported, however, only for the noise token with the shorter duty cycle.

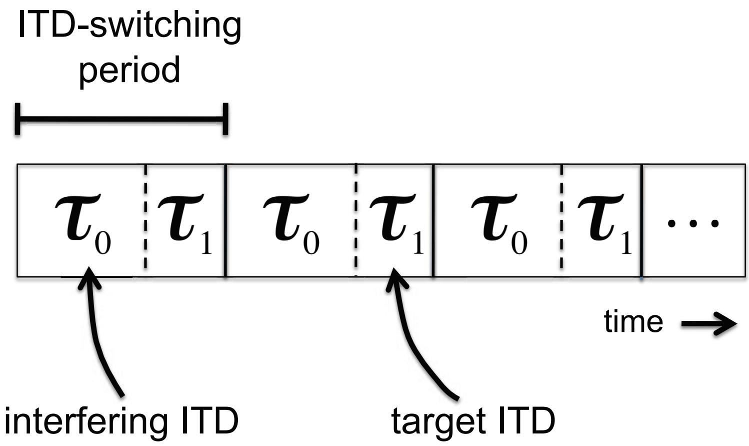

Figure 4.1: Schematic drawing of an ITD-switching stimulus where continuous, broadband Gaussian noises periodically alternate between two different ITDs (τ_1 and τ_2). For the experiments here, the nomenclature of the 'target ITD' refers to a non-zero ITD whereas the 'interfering ITD' was set to zero. The duty cycle, i.e. the percentage of the ITD-switching period at the target ITD, was an experimental parameter that was manipulated in the experiments of this study.

2 Note that the 'ITD-switching period' has previously been referred to as the 'binaural modulation period' in Chapter 3. The latter term was used in the previous chapter since the focus was on the modulation aspects of the alternating ITD stimulus.

4.2.2 Procedure

A three-interval acoustic pointer task was employed to measure if brief noise tokens at a given ITD can be accurately lateralized. The first and third intervals contained 600 ms of the ITD-switching stimulus described above. The second sound interval contained an ITD-lateralized acoustic pointer, which was a 300 ms broadband noise token continuously presented at a single ITD. Durations of all three intervals included 50 ms (Hann) onset/offset ramps, and a pause of 300 ms was inserted between each interval. In this experiment and all subsequent experiments of this study, stimuli were developed in MATLAB and digitally created at a sampling rate of 48 kHz with the digital-to-analog conversion via an RME Fireface UC. Furthermore, all stimuli were presented using headphone playback (Sennheiser HD-650) and were calibrated to 70 dB SPL. Seven listening participants (four female, three male; average age 26 years), none of whom reported any evidence or history of hearing problems, completed this experiment.

For the test stimuli of this experiment, one of the two ITDs was always set to zero while the other ITD was randomly and uniformly distributed between $\pm 625\ \mu s$. Lateralization performance for 20 different non-zero ITDs was measured for each test condition and for each listening participant. The noise with a non-zero ITD was always presented as the first noise token of the periodically alternating ITD pattern. The ITD-switching periods employed were 240, 120, 60, 30, and 15 ms, which lead to ITD-switching frequencies of 4.2, 8.3, 17, 33, and 67 Hz, respectively. The duty cycle of the noise with a non-zero ITD ranged from 2.5% to 75%. Although listeners were explicitly instructed that the lateral position of the modulation percept was the target position, the duration of the noise with the non-zero ITD will be denoted as the *'target duration'* for simplicity. It should be pointed out that this non-zero ITD does not necessarily coincide with the position attributed to the modulation percept. Values for the target duration ranged from 3 to 60 ms. In total, 14 different test conditions were presented to all listeners noting that not all combinations of duty cycle and target duration were tested. To assess listeners' lateralization abilities with this acoustic pointer paradigm, a baseline condition was also presented. For this baseline condition, the first and third intervals did not contain any ITD switching, i.e. the ITD was held constant throughout the entire interval.

In all test conditions of this experiment, listeners were specifically directed to adjust the pointer (i.e. the second interval) to match the perceived lateral position of the modulating component of the stimulus. The ITD of the second interval was increased or decreased by the listener by pressing one of two buttons on a standard computer keyboard. The starting ITD for the second interval was randomized between $\pm 700\ \mu s$. The initial step size ($200\ \mu s$) in the adaptive procedure was halved after two directional reversals of the acoustic pointer until the minimum step size was $25\ \mu s$ upon which six more reversals were completed before the track was terminated. The reported ITD for each measurement was the average of the pointer ITD value on each of the reversals while the step size was $25\ \mu s$. Both ILD and interaural phase difference acoustic pointers have been successfully used for reliable measurements of perceived target source laterality with adaptive paradigms similar to the one applied in our investigation (Dietz *et al.*, 2009). In order to maintain a direct relationship with the interaural parameters of the acoustic pointer and the parameters of the test stimului, an ITD acoustic pointer (laterally restricted to $\pm 1000\ \mu s$) was employed. Listeners did not report any problems with maintaining a focused image of the pointer even at large ITDs where the periodic nature of the cross-correlation function within interaural sub-band pairs could potentially result in an ambiguous lateral position.

In light of the perceived modulation percept resulting from the ITD-switching stimulus as described in Sect. 4.2.1, specific instructions were provided in order to disambiguate which ITD position the listeners should use for the lateralization task. The following instructions were provided to the listeners (via printout) throughout the duration of all test conditions:

1. Try to hear the modulating sound source and adjust the pointer to this source.
2. If you hear two modulating sources, then adjust the pointer to the furthest lateral modulating source.
3. If you hear two sources but neither are modulating, then adjust the pointer to the furthest lateral source.
4. If you hear a single sound source, then adjust the pointer to the center of gravity.

Prior to beginning the actual experiment, all participants completed four separate training conditions to familiarize themselves with the stimuli and the specified instructions. The training session lasted approximately one hour in total. Additionally, prior to every test condition, participants completed a brief, 5-minute training session in which the approximate range of target ITD values and the exact same parameters (i.e. ITD-switching period and duty cycle) of the subsequent experimental conditions were applied. Rationale for this additional training was to provide exemplary test stimuli so that listeners could establish a general lateralization strategy given the specified instructions for the upcoming test stimuli.

4.2.3 Results

Scatter plots of the population data are presented in Fig. 4.2 for all test conditions and the baseline condition. The solid black line with a slope of one indicates an adjusted ITD commensurate with the target ITD (i.e. the non-zero ITD). Generalized linear regression models with one degree of freedom (slope) were fit to the population data for each listening condition and are plotted as dashed black lines in each plot. The ITD-switching frequency, duty cycle, and target duration are written in each plot. Recall that the target duration specifies the duration of the noise at the non-zero ITD. The plots are organized such that experimental parameters

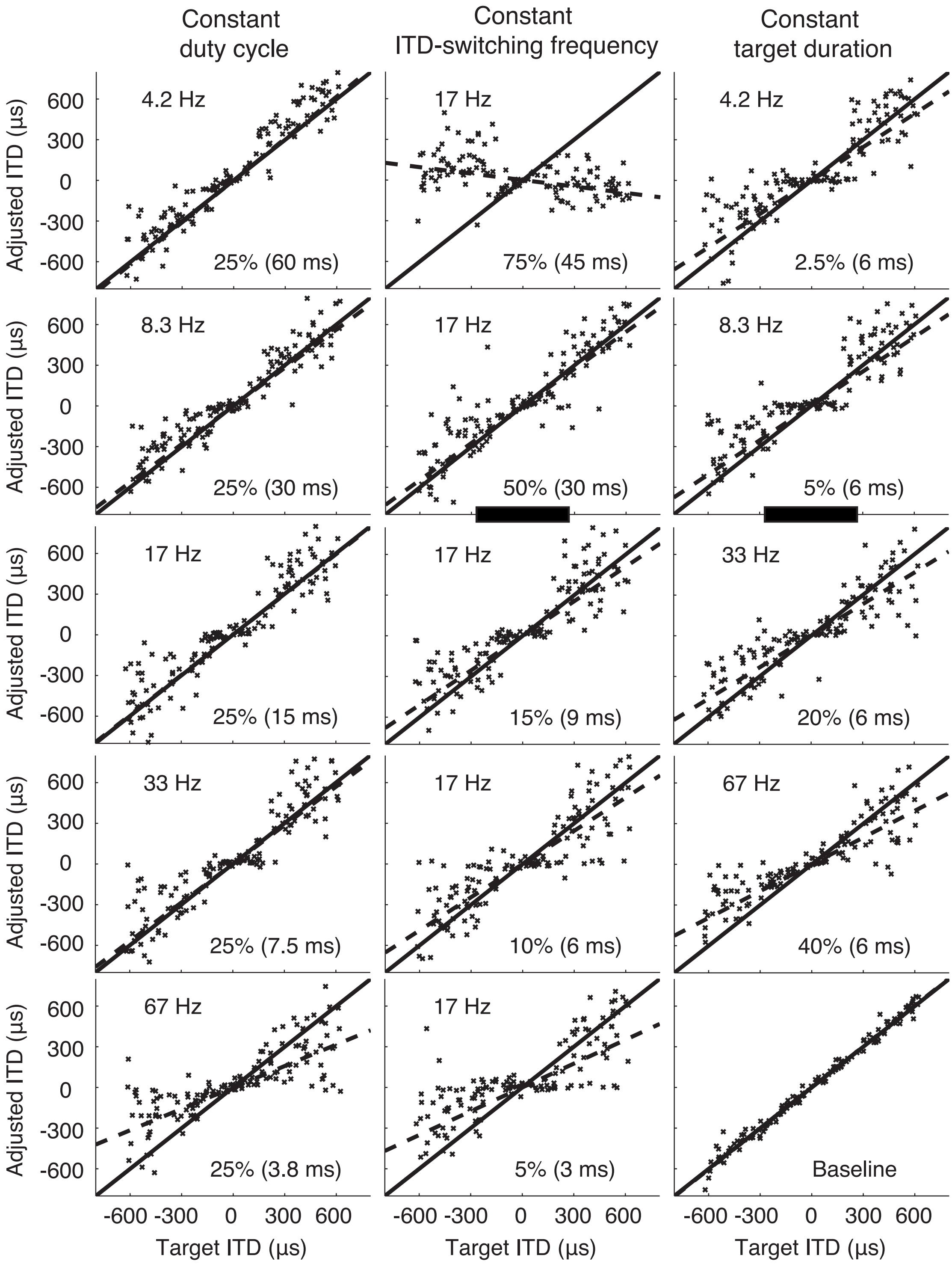

Figure 4.2: Scatter plots of data collected across all subjects and all listening conditions from Exp. 1. The solid line with a slope of one indicates adjustment of the ITD pointer to the target ITD. The dashed line depicts the generalized linear model fit to each set of data. For each panel, the experimental parameters are written: ITD-switching frequency (top-left), the duty cycle (bottom-middle), and the target duration (bottom-right). The panels are organized in columns with constant duty cycle (left column), constant ITD-switching frequency (middle column), and constant target duration (right column). The black rectangles in the middle and right columns are visual aids to specify where a panel from an adjacent column can be inserted to follow the organization of the experimental parameters for the given column.

are held constant for each column of plots: duty cycle (left), ITD-switching frequency (middle), and target duration (right). The two black rectangles drawn in the second and third columns are provided as visual aids to indicate positions where a plot from an adjacent column can be inserted which contains experimental parameters corresponding to the vertical arrangement of the given column. In general, many of the data points fall reasonably close to the target ITD for most of the test conditions although the variance in the data is greater than for the baseline condition. The slopes of the linear models fit to the data appear to be reduced for conditions with shorter target durations. Inspecting the target ITD values near the midline (e.g., less than ± 200 μs) shows that the data in several test conditions follow a reduced slope as compared to the slope for the more lateral ITDs. When the generalized linear regression models for each of the test conditions were fit to the data between ± 200 μs, the slopes for the midline data resulted in an average reduction in slope of 34% when compared to the slope of the linear model fit to the full range of ITDs. In contrast, computation of the generalized linear models for the data greater than ± 200 μs resulted in only a small increase in the slope by an average of 0.9%. Due to the noticeable change in slope for midline data, third-order generalized models were also fit to the different data conditions. The simple first-order linear fit accounted for essentially the same amount of the variance in the data ($\bar{R}^2 = 0.710$ compared to $\bar{R}^2 = 0.714$ for the third-order fit), thus, were used for subsequent analysis.

The slope values of each linear regression fit across all data conditions are plotted in Figs. 4.3 - 4.5. Data for the test conditions are plotted based on a constant duty cycle of 25% (Fig. 4.3), a constant ITD-switching frequency of 17 Hz (Fig. 4.4), and a constant target duration of 6 ms (Fig. 4.5). For ease of interpretation within each figure, the data for two conditions are presented in multiple figures, e.g. 15 ms target duration with a 17 Hz ITD-switching frequency and 10% duty cycle with a 6 ms target duration. In all plots, a slope equal to one would indicate a perceived position corresponding to the noise containing a non-zero ITD, and a slope of zero would indicate that the pointer was adjusted to match the ITD positioned at 0 μs. For all test conditions, the slopes of the linear regressions fit to the individual data are indicated by triangles. Slopes of the linear regressions fit to the population data are indicated by crosses. The baseline lateralization abilities where no ITD switching was present are indicated by open circles in Figs. 4.3 - 4.5. Comparison of the performance between the test conditions and the baseline condition was conducted using a generalized linear hypothesis test that takes into account the underlying distribution of the data. Shaded triangles indicate when an individual performed significantly differently ($p < 0.05$) than their respective baseline performance per a two-tailed, multi-comparison with Bonferroni adjustment. Asterisks above a given test condition indicate statistically significant differences for the population performance.

Although the range of slope values for the different test conditions indicate some differences in lateralization performance across listeners, a consistent trend of improved lateralization accuracy with increasing target duration was readily apparent. As seen in Fig. 4.3, a target duration of 7.5 ms was sufficient for accurate lateralization across the population. Fig. 4.4 illustrates, however,

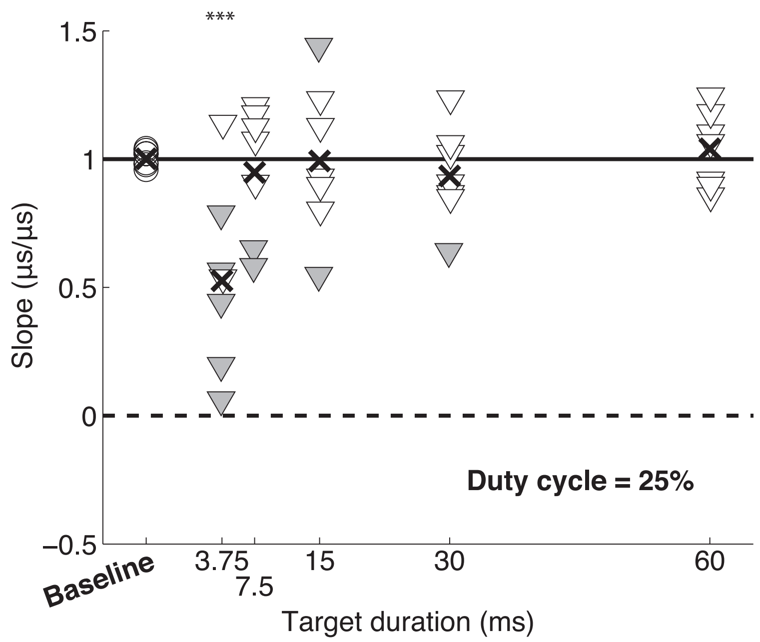

Figure 4.3: Slopes of the linear models fit to the test conditions of Exp. 1 when the duty cycle was held constant at 25%. The slope values are plotted versus the target duration along the x-axis. A slope of unity represents an adjustment of the acoustic pointer to a position corresponding to the noise with a non-zero ITD whereas a slope of zero indicates a position commensurate with the noise at an ITD of 0 μs. Triangles indicate slopes for individual listeners where the filled symbols indicate a significant difference ($p < 0.05$) from the individual's baseline performance. The crosses represent the slope of the linear model fit to the data for the population of participants where the asterisks signify a statistical difference from the baseline performance (*** for $p < 0.001$, ** for $p < 0.01$).

that a target duration of only 3 ms (i.e. 5% duty cycle for 17 Hz ITD-switching frequency) was sufficient for three of the listeners. For a majority of the listeners, a target duration of 6 ms was sufficient for reasonably accurate lateralization, especially at the lower modulation frequencies as seen in Fig. 4.5. In contrast, one listener in particular required a target duration greater than 30 ms. A look at the individual performance showed that the data for this listener consistently resulted in the smallest slopes out of the population of listeners. Furthermore, the slopes near zero in Fig. 4.4 for duty cycle values of 5% and 10% (respectively, target durations of 3 and 6 ms) indicate that this listener had difficulty even detecting the non-zero ITD for such short target durations.

Due to the specific instructions provided to the listeners, an interesting result from this experiment is an affirmation that the percept of modulation was attributed to the ITD with the smaller duty cycle as seen in the 75% duty cycle condition of Fig. 4.4. If both ITD locations contained a modulation percept or if neither ITD location contained a modulation percept, listeners should not have unanimously adjusted the acoustic pointer to the more centrally located noise token. It is interesting that all listeners adjusted the acoustic pointer in the 75% duty cycle condition such that the linear regression model had a negative slope.

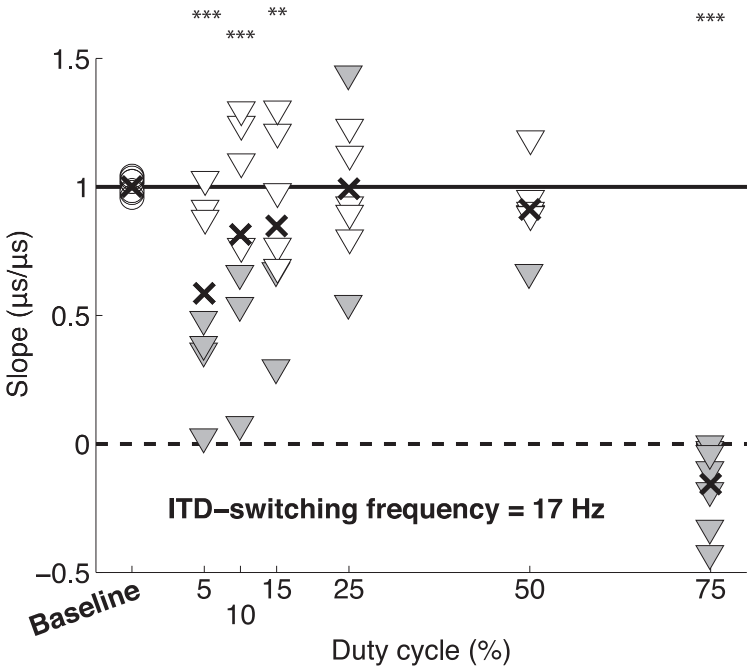

Figure 4.4: Slopes of the linear models fit to the test conditions of Exp. 1 when the ITD-switching frequency was held constant at 17 Hz. The slope values are plotted versus the duty cycle along the x-axis. The plot follows the same conventions as used in Fig. 4.3.

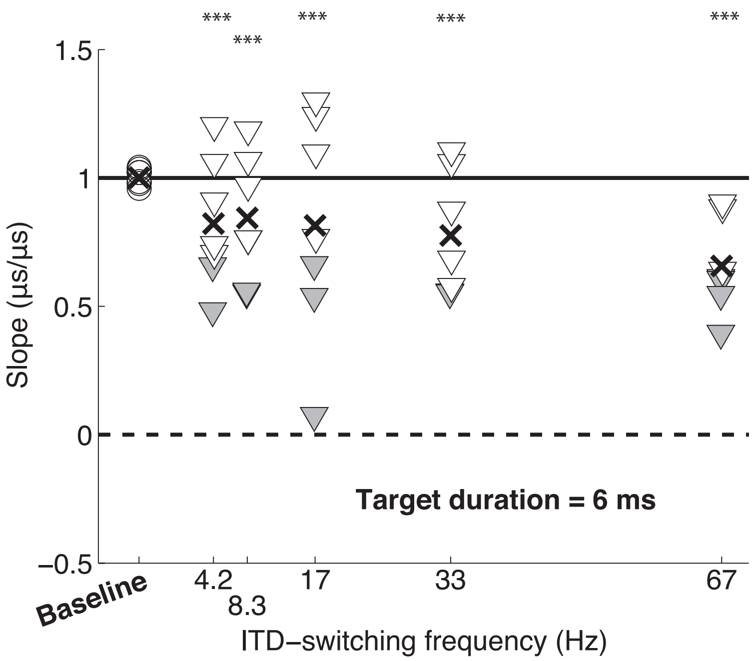

Figure 4.5: Slopes of the linear models fit to the test conditions of Exp. 1 when the target length was held constant at 6 ms. The slope values are plotted versus the ITD-switching frequency along the x-axis. The plot follows the same conventions as used in Fig. 4.3.

4.3 Experiment 2: Influence of monaural envelope cues on lateralization

The second experiment was designed to determine if AM (a monaural cue) influences the ability of listeners when making an absolute lateralization judgement of brief noise segments in a stimulus similar to the one used in Exp. 1. The hypothesis is that the auditory system is possibly more sensitive to binaural information during the rising portions of an AM (Dietz *et al.*, 2013) and/or that monaural envelope cues possibly 'trigger' a fast readout of binaural information (Schimmel *et al.*, 2008).

4.3.1 Stimuli

The stimuli for the second study were constructed using the same broadband, ITD-switching stimulus used in Exp. 1. A sinusoidal AM, however, was additionally imposed onto this ITD-switching stimulus. To determine if the temporal position of an ITD target within the phase of the AM influences absolute lateralization accuracy, the AM rate and the ITD-switching rate were identical. This ensured that the ITD target was located at a constant temporal position of the AM phase throughout the duration of the stimulus. Three fixed positions within the AM phase were tested: the rising ($\phi_{mod} = \pi/2$), peak ($\phi_{mod} = \pi$), and falling ($\phi_{mod} = 3\pi/2$) portions of the AM envelope. The noise containing the target ITD was temporally centered at each of the specified fixed positions in the phase of the AM. A fourth condition where the monaural and binaural modulation rates were not equal was also tested. The AM frequency for this condition was chosen such that, throughout the duration of the stimulus, the target ITD was presented twice at the same positions within AM phase during each interval. For all four conditions, the starting phase of the AM was randomized in every trial. Furthermore, the starting phase of the binaural modulation, i.e. the temporal position within the ITD-switching period at the beginning of the stimulus, was randomized in the fourth condition.

4.3.2 Procedure

The same seven participants from Exp. 1 completed a three-interval acoustic pointer task that was identical to the procedure in the first study. Absolute lateralization performance was measured across the same range of ITD values (± 625 μs) where 20 trials per test condition were collected for each listening participant as conducted in Exp. 1. The ITD of the other noise was always set to 0 μs. For this experiment, the binaural modulation period was set to 60 ms for all test conditions. When the binaural and monaural modulation rates were different, the monaural modulation period was set to 75 ms. Durations of the noise at the non-zero ITD were tested at 3, 6, and 9 ms for the three different fixed positions of the sinusoidal envelope (i.e. rising, peak, falling). The same target durations were used for the randomized position with the

AM when the monaural and binaural modulation frequencies were not equal. A target duration of 15 ms was tested only for the fixed position during the rising portion of the AM. Similar to Exp. 1, the duration of all stimuli was set to 600 ms which included 50 ms (Hann) onset/offset ramps. The test stimuli were calibrated to a level of 70 dB SPL subsequent to the application of the AM and onset/offset ramps.

Because the imposed AM limited the ability to disambiguate the two ITD positions based on the modulation percept as specified in the Exp. 1 instructions, the listeners were simply instructed to adjust the pointer to the furthest lateral sound object. If only one sound was perceived, listeners were instructed to use the center of gravity. Prior to beginning the actual experiment, all participants completed four separate training sessions, which lasted approximately one hour in total. A brief, 5-minute training session prior to each experimental condition was conducted as described in Exp. 1.

4.3.3 Results

The results of Exp. 2 are plotted in Fig. 4.6 using the same statistical analyses and plotting conventions applied in Figs. 4.3 - 4.5. Data are grouped on different durations for the target ITD (i.e. 3, 6, 9, and 15 ms). Within each data group, the different positions of the target ITD with respect to the phase of the AM are indicated by different columns in each group (i.e. rise, peak, fall, rand). The data indicated by 'rand' are test conditions when the binaural and monaural modulation frequencies were not identical. For comparison, the baseline performance recorded in Exp. 1 was used to determine significant differences in lateralization performance for the individual test conditions.

Results of this experiment reveal the same trend of improved accuracy with increasing target duration as found in Exp. 1. In contrast, however, a target duration of only 6 ms was sufficient for accurate lateralization across the population. For the shortest target duration (3 ms), listeners generally illustrated a perceived position closer to the midline than would be expected for the target ITD although multiple listeners were still quite accurate in this test condition. Grouping the data based on the different target ITD positions and comparing the generalized linear models fit to the data groups revealed no statistically significant differences in lateralization performance between the four different ITD positions. Analysis of the generalized linear models fit to individual data also revealed no trend with respect to the target ITD position within the phase of the AM.

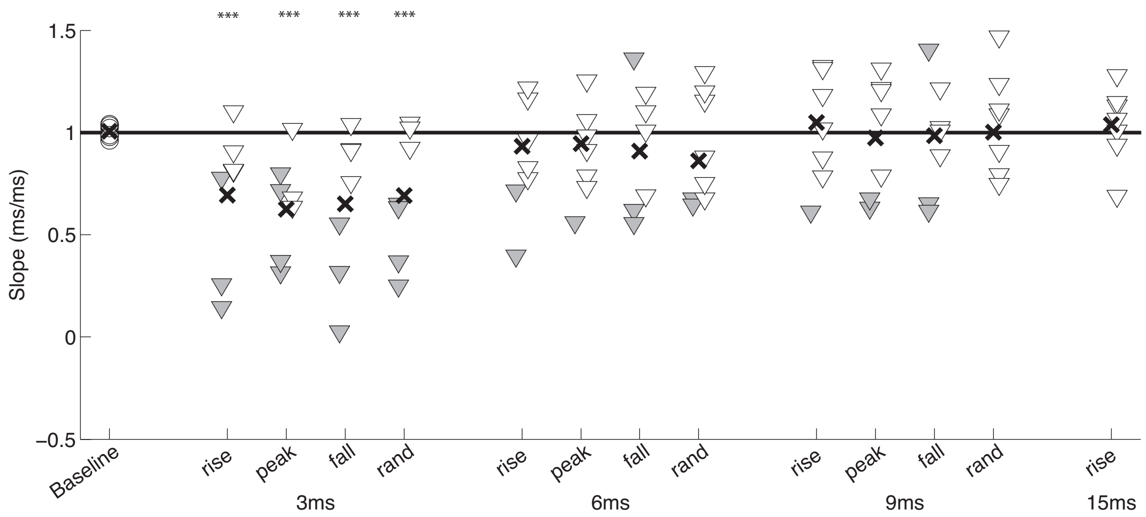

Figure 4.6: Slopes of the linear models fit to the different test conditions of Exp. 2. The plot was designed in the same manner as in Fig. 4.3. The different test conditions are grouped on the different durations of the target ITD noise. For the 3, 6 and 9 ms target ITD duration conditions, the different temporal positions within the phase of the AM are indicated for the rising, peak, and falling portions (*rise*, *peak*, and *fall*). The test condition when the AM and the ITD switching occurred at different frequencies is indicated by the label *rand*. For the 15 ms duration, the ITD target was only tested for the rising portion of the AM cycle.

4.4 Experiment 3: Influence of monaural cues and bandwidth on ITD sensitivity

No difference was observed in Exp. 2 for the absolute lateralization performance depending on the various temporal positions of the target ITD within the AM cycle. Since a dependency on the phase of AM has been previously demonstrated using tonal stimuli (Dietz *et al.*, 2013), it was conceived that the bandwidth of the stimulus could play a role. In order to quantify the bandwidth dependence, ITD just noticeable difference (jnd) values were measured in Exp. 3. This paradigm permitted several experimental conditions to be measured with a much shorter testing time as compared to that required for the previous absolute lateralization measurements.

4.4.1 Stimuli

The stimulus was similar to that used in Exp. 2, but the broadband, ITD-switching stimulus serving as the carrier signal for the AM was filtered into four different values of equivalent rectangular bandwidth (ERB) as specified by Glasberg and Moore (1990). Values of 1, 2, 4, and 8 ERB filters were used where the upper and lower bounds for all filters were logarithmically centered around 500 Hz. The stimuli were created by taking a one-second segment of the broadband modulation stimulus from Exp. 1 and transforming it into the frequency domain. A brickwall filter was implemented by setting all coefficients outside the upper and lower bounds defined by a given ERB filter to zero. The stimulus was then transformed back into the time domain at which point the AM was applied. The time domain signal was trimmed to the desired stimulus length of 600 ms, and subsequently 50-ms onset/offset (Hann) ramps were applied. The stimuli were calibrated to a presentation level of 70 dB SPL after the AM and onset/offset ramps were applied. It is worth emphasizing that artifacts due to switching between the two ITD values were not present since the filtering was applied subsequent to creating the broadband ITD-switching stimulus. Therefore, monaural cues of the ITD switching were still not present in these bandlimited conditions. The duration of the noise at the target ITD was set to 9 ms which resulted from a duty cycle of 15% for the 17 Hz ITD-switching frequency applied in all presentations of this experiment.

4.4.2 Procedure

Instead of measuring absolute lateralization accuracy as done in Exp. 2, lateral acuity from the midline was used to measure the influence of the temporal position of the ITD target within the AM cycle. An adaptive, two-interval, two-alternative forced choice paradigm was employed where the first interval was a diotic noise on which an AM was imposed and the second interval contained the test stimulus. Both intervals were restricted to the same spectral region. The task of the listener was to determine if the target ITD was to the left or right by (presumably) comparing a lateral shift from the first interval. The side of the target ITD was randomized from trial to trial.

The initial magnitude of the target ITD was set to 100 μs and was adaptively adjusted using a two-down, one-up procedure. The magnitude of the target ITD was adjusted with a starting step-size of 50 μs and was halved after every two reversals until a minimum step-size of 3.125 μs was reached upon which six more reversals were completed. The reported ITD-jnd for each measurement was the average of the target ITD values on each of the direction reversals at the minimum step size. In order to avoid that listeners could learn the reference position from successive presentations, lateral roving was applied. This ensured that listeners were required to use the reference interval on each presentation. For each trial, a roving value taken from a uniform distribution between ± 25 μs was applied equally to both intervals. This ensured that only the absolute ITDs were roved and that the roving did not alter the magnitude of the difference in ITD values between the first interval and the target ITD of the second interval.

Thresholds were measured for the four different temporal positions within the AM phase used in Exp. 2 (e.g. rising, peak, falling, random) for each of the four different ERB values. Six thresholds were measured for each of the 16 experimental conditions for each participant. Of the seven participants in Exp. 2, only six were available for Exp. 3. Prior to beginning the actual experiment, listening participants completed a training session lasting approximately 30 minutes in order to become familiarized with the experimental paradigm.

4.4.3 Results

The ITD-jnd thresholds across all conditions of Exp. 3 are plotted as box plots in Fig 4.7 where the ITD-jnd value indicates the spatial separation between the two ITDs applied in the second interval. The data are grouped on the different bandwidth conditions with increasing bandwidth moving from left to right. Within each group, the different data sets represent different temporal positions of the target ITD within the phase of the AM indicated by *rise*, *peak*, *fall*, *rand*.

Levene's test (Fox and Weisberg, 2011) for equality of variances across all test conditions was found to be violated, ($F(3, 572) = 6.69$, $p < 0.001$). For this reason, a two-way analysis for within-by-within design was carried out across the different ITDs and modulation frequencies to compare the trimmed means of the marginal distributions of the broadband data. A benefit of this method is that the assumptions of normality and homogeneity of variance are not required. The trimmed mean is the mean of the data distribution after values exceeding a specified quantile value on both sides of the distribution are truncated. For the subsequent analyses, 20% trimming and 599 bootstrap samples were applied, which have been shown to be appropriate settings for this type of analysis (Wilcox, 2012, pg. 162). Hochberg's method was applied for the control of familywise error rate in post-hoc analysis. For the reporting of the statistical results, $\hat{\Psi}$ represents the estimated difference between the measures of location, e.g. the trimmed mean.

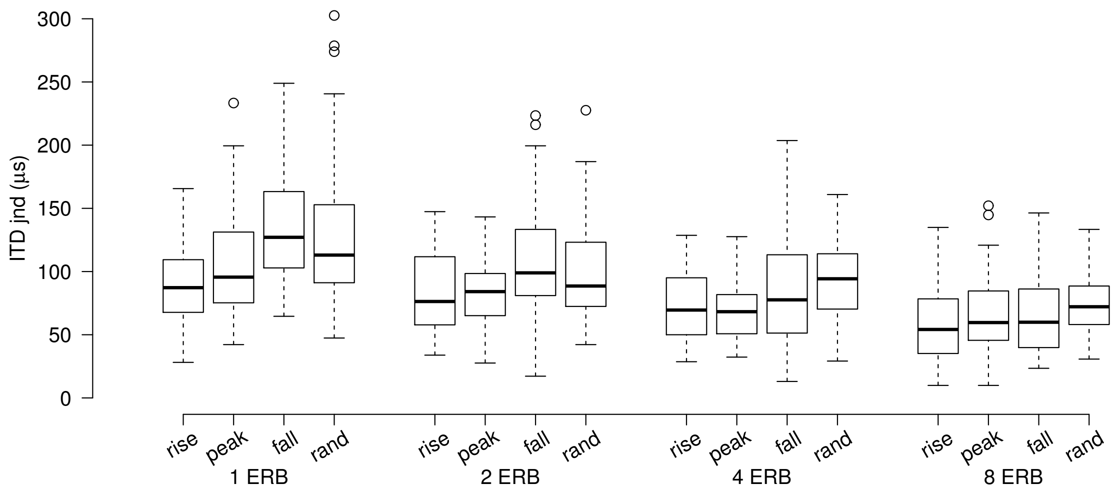

Figure 4.7: Data from Exp. 3 plotted as boxplots for the ITD-jnd thresholds across the different bandwidths (e.g. 1, 2, 4, 8 ERB). For each bandwidth, the position of the ITD target within the phase of the AM for the rising, peak, and falling portions are specified (*rise*, *peak*, and *fall*). The test condition when the AM and the ITD switching occurred at different frequencies is indicated by the label *rand*.

The analysis showed statistically significant differences for both main factors: number of ERB and phase position. With regard to the number of ERB, the spatial separation between the two ITDs of the stimulus required for detection decreased as the bandwidth increased. All conditions were statistically significantly different from each other ($p< 0.05$) where the trimmed-mean ITD-jnd steadily decreased from 108 μs for the 1-ERB condition to 63 μs for the 8-ERB condition. Concerning the main factor of different phase positions, the only two conditions that were not significantly different from each other were the *fall* vs. *rand* and the *rise* vs. *peak*. This illustrates that thresholds were generally lower for the rising or peak positions of the envelope as compared to the falling position or in the condition when the ITD switching was not synchronous with the AM.

More interestingly, however, was the difference in thresholds between the falling and rising phase positions depending on the bandwidth as indicated by the statistically significant interaction between the two main factors. The difference between the *rise* and *fall* thresholds for the 1-ERB condition was significantly larger than the 8-ERB condition ($\hat{\Psi} = 44.3$ μs, 95% CI:[19.2 68.6]). Only a very small difference was found between the two phase conditions within the 8-ERB condition (i.e. the trimmed-mean thresholds for the *rise* and *fall* conditions were respectively 56.3 μs and 60.7 μs). Therefore, this statistically significant interaction was likely a result of the increasing differences between the *rise* and *fall* thresholds for decreasing bandwidths (i.e. the mean threshold for the *rise* and *fall* conditions were respectively 86.3 μs and 130 μs for the 1-ERB condition). Pairwise comparison across all test conditions provide further validation by showing a statistically significant difference between the *rise* and *fall* thresholds for the 1-ERB condition ($\hat{\Psi} = 46.8$ μs, 95% CI:[26.6 70.7]) but not for the 8-ERB condition ($\hat{\Psi} = 7.84$ μs, 95% CI:[-7.46 28.3]). This 50% increase in thresholds for the smallest bandwidth suggests that the increased weight to binaural cues in the rising portion of an AM could be dependent upon the signal bandwidth. A similar difference in ITD-jnd thresholds based on bandwidth was found between the *rise* and *fall* thresholds for the 1-ERB and 4-ERB conditions ($\hat{\Psi} = 39.7$ μs, 95% CI:[12.2 73.8]).

Grouping the data based on whether the AM rate and the ITD-switching rate were the same or different illustrates statistically significantly lower thresholds when the the AM and ITD-switching occur synchronously ($\hat{\Psi} = -20.3$ μs, 95% CI:[-11.2 -29.7]). This suggests that ITD switching not synchronous with the monaural envelope rate can have an adverse effect on ITD thresholds. Additionally, the statistically significant interaction between the *rise* and *rand* thresholds for the 1-ERB and the 8-ERB conditions ($\hat{\Psi} = 34.3$ μs, 95% CI:[3.10 66.2]) could imply that the elevation in thresholds due to non-sycnhronous AM and ITD-switching rates could be more pronounced at narrower stimulus bandwidths.

4.5 Experiment 4: Influence of spatial separation on AM source segregation

The previous three experiments evaluated sensitivity to brief segments of ITD information using stimuli with specific properties unlikely to be encountered in natural auditory scenes. The final experiment applied a stimulus with properties more directly applicable to multi-talker listening situations, which normally consist of multiple sources with independent temporal envelope patterns. In order to discriminate between the modulation frequencies of two simultaneously presented AM signals, it is necessary that the temporal patterns are sufficiently represented. This experiment was directed at determining if the binaural system plays a role in the auditory system's ability to discriminate between the sinusoidal envelope frequencies of two simultaneously presented AM signals.

4.5.1 Stimuli

The stimuli used for this study were two independent, broadband Gaussian noises that were ITD-lateralized in equal magnitude but with opposing sign. Subsequently, a diotic AM was imposed independently onto both of these spatially lateralized noises. The difference in AM frequency between the two noises was parametrically varied.

4.5.2 Procedure

A method of constant stimulus paradigm was employed to determine the ability of listening participants to discriminate between the rates of two simultaneously presented and ITD-lateralized AM signals. The task of the listeners was to specify if the slower modulation was on the right or left. Three different ITD magnitudes were used (25, 50, and 150 μs). Since the AM signals maintained the same ITD magnitude but with opposing sign, these ITD magnitudes resulted in a spatial separation of 50, 100 and 300 μs between the two AM sounds. The pedestal modulation frequencies tested were 2, 4, 8, and 16 Hz. The difference between the test modulation frequency was specified as a number of semitones different from the pedestal modulation frequency being tested. Values of 5, 7, 11, 13, 15, 17 and 21 semitones were tested for each pedestal modulation frequency.

In total, 80 repetitions were collected across several dates for each experimental condition. Measurements were blocked on the pedestal modulation frequency and ITD magnitude so that the target modulation frequency and ITD were held constant throughout a single measurement. The starting phase of each AM signal was randomized from trial to trial. Each measurement was a single 1.5-second interval of the two AM signals that included 50 ms onset/offset (Hann) ramps applied after the summation of the two AM signals. The sound output of the combined

AM signals was calibrated to 70 dB SPL subsequent to the addition of AM and the onset/offset ramps. Due to the extensive duration of this experiment (12 hours per participant), only five participants completed this study – four of whom completed the first three experiments. Prior to beginning the experiment, all participants completed a short training session (20 minutes) to become familiar with the experimental task.

4.5.3 Results

The percent-correct scores across the population of listeners were converted to d' values for the two-alternative, forced choice paradigm, and the results are plotted in Fig 4.8. The different symbols indicate different values of spatial separation: triangles (300 μs), circles (100 μs), and squares (50 μs). For clarity of interpretation, the different pedestal modulation frequencies (grouped on spatial separation) alternate in grayscale color. The different number of semitones for a given pedestal modulation frequency are connected with a solid curve for the respective condition. A horizontal line at d' = 1, which represents a percent-correct score of 76%, is also provided as a reference. Negative values of d' were set to zero.

In general, this task appears to be quite difficult, particularly at higher modulation frequencies. The population of listeners were unable to exceed a d' > 1 for most of the conditions. The best acuity for discriminating between the modulation rates of two simultaneous AM stimuli was for the greatest spatial separation (300 μs) and at the slowest modulation rate (2 Hz) where a difference of 13 semitones from the pedestal modulation frequency was required.

Statistical analysis with a 3-way analysis of variance (ANOVA) was computed with the raw percent-correct scores across the population of listeners. Analysis either based on d' scores calculated per Fig. 4.8 or based on the mean d' scores calculated from each of the test conditions for each participant resulted in a violation of homogeneity of variance. Similar bootstrap methods used in Sect. 4.4.3 were applied to check the severity of the violated assumption. The same qualitative trends were found for both statistical methods. For this reason, the results from the simple 3-way ANOVA based on d' scores across the population of listeners are reported here.

As seen in Fig. 4.8, d' scores appear to decrease with increasing modulation frequency and increase with increasing number of semitones. Statistical analysis confirms these tendencies where statistically significant effects were observed for both the pedestal modulation frequency ($F(3,384)=58.6$, $p < 0.001$) and the number of semitones different ($F(7,384)=43.4$, $p < 0.001$). An increase in spatial separation between the two AM stimuli provided improved frequency discrimination as seen by a statistically significant increase in d' scores with increasing ITD magnitude ($F(2,384)=50.8$, $p < 0.001$). Statistically significant interactions were also reported for the main effects of semitone difference and pedestal modulation frequency ($F(21, 384) = 4.74$, $p < 0.001$) and for the main effects of semitone difference and spatial separation ($F(14,384)=5.10$, p < 0.001). These interactions were a result of a flattening of the d' scores with increasing modulation frequency and decreasing spatial separation. Although this task was particularly difficult for the higher modulation rates, sufficient spatial separation (e.g. 300 μs) between the two AM stimuli permitted listeners to discriminate a difference of 21 semitones across all pedestal modulation rates tested. Differences in discrimination thresholds were observed between the listeners, but the same trends across experimental parameters were consistent across listeners.

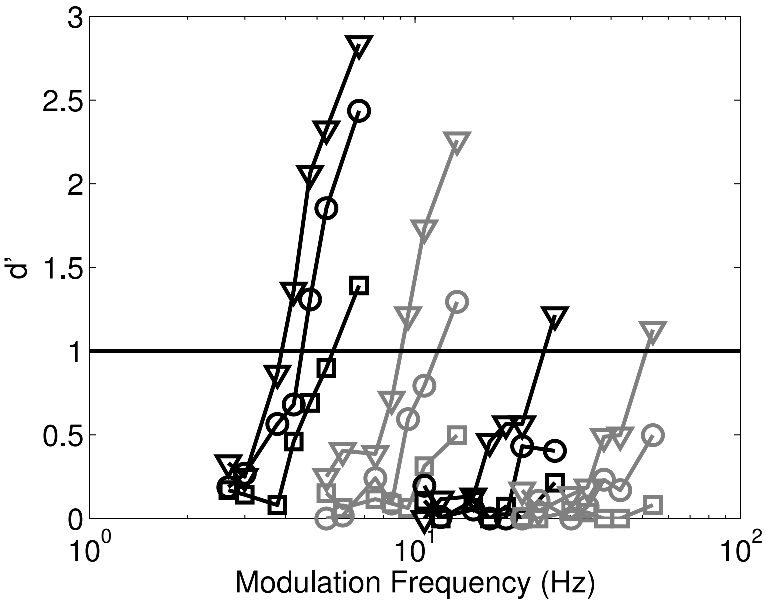

Figure 4.8: Frequency discrimination between two simultaneous AM noises with percent correct scores converted to values of d'. The reference frequency was held constant at 2, 4, 8, or 16 Hz, for which results are shown, from left to right, by the groups of curves with alternating black, grey, black, grey shading, respectively. Each of the open symbols along a continuous curve represent a different separation in number of semitones from the specified reference frequency. The different symbols represent different amounts of spatial separation between the two AM noises: triangles (300 μs), circles (200 μs) and squares (100 μs). As a reference, a horizontal line at d' = 1 is indicated that corresponds to a percent-correct score of 76%.

4.6 Discusion

The experiments presented here investigated the temporal acuity of the auditory system for selecting brief segments of binaural information. Using a stimulus that rapidly alternates between two values of ITD, the first experiment illustrated that the auditory system can accurately lateralize noise tokens as short as 7.5 ms in duration. The second and third experiments focused on the role of AM with respect to binaural sensitivity. Although the temporal position of the ITD target within the AM cycle did not influence absolute lateralization performance for broadband stimuli, the auditory system appears to be more sensitive to binaural cues during the rising portion of the AM for narrow bandwidths, e.g. 1 ERB. The final experiment explored the binaural system's potential to provide a benefit for the task of discriminating between two simultaneously presented envelope frequencies. Results show that greater spatial separation does indeed aid in the frequency discrimination task and that performance is best for lower modulation rates.

It has been suggested that ringing due to the responses of the filters in the auditory periphery could serve as a limitation to temporal processing, particularly for lower frequencies (Duifhuis, 1973, Plack and Moore, 1990). Since ITD sensitivity is accepted to be largely a low-frequency cue, one could expect the auditory periphery to impose a temporal limit for processing ITD cues. Indeed, the apparent sensitivity to such brief changes in binaural cues as seen in our experiments (3 - 7 ms) suggests that the binaural system is not adding a considerable limitation in temporal resolution beyond that expected from ringing in peripheral processing. For frequencies below 1 kHz, this ringing has been suggested to account for gap detection in the monaural domain which has thresholds of 2-3 ms (Shailer and Moore, 1983). The ability to process rapidly changing cues is in close agreement with thresholds obtained in the binaural domain for detecting brief (2 ms) changes in ITD information (Bernstein *et al.*, 2001) and for detecting 2-4 ms periods of interaural phase agreement (Pollack, 1978).

This rapid processing of binaural cues is in contrast to the idea of a sluggish binaural system. As proposed by Kollmeier and Gilkey (1990), however, differences in the temporal processing of binaural cues could be a result of testing different binaural abilities. For example, the binaural system might be less sensitive to time-varying changes in the width of the spatial image, which is arguably the cue used for masking level difference measurements (Grantham and Wightman, 1979, Kollmeier and Gilkey, 1990, Culling and Summerfield, 1998) and interaural correlation measurements (Grantham, 1982, Akeroyd and Summerfield, 1999) where sluggishness has been observed. It is worth noting that even the manner of implementation for a time-varying interaural correlation stimulus results in different binaural temporal resolutions (Siveke *et al.*, 2008). With regard to lateral estimates, the limited ability to track movement in detail (Perrott and Musicant, 1977, Grantham and Wightman, 1978) is possibly a different auditory task than attributing a lateral position to a brief noise segment as was required in Exp. 1. The relative insensitivity to detailed movement agrees with the qualitative description given by listeners since the stimulus was reported to produce a percept of two different sound objects as compared to a single source switching back-and-forth between the two lateral positions. It is possibly elements of binaural sluggishness which cause the binaural system to mediate the buildup of two separate sound streams from the ITD-switching stimulus described in Sect. 4.2.1 as compared to producing a percept of a single sound source rapidly switching in location.

Since listeners were explicitly instructed to adjust the pointer to the position corresponding to the furthest lateral modulation percept, we can infer from the 75% duty cycle condition that listeners attributed the modulation percept of the ITD-switching stimulus to the ITD position occupying the smaller percentage of the ITD-switching period. If modulating sources were perceived at both ITD positions, the instructions should have caused the listeners to adjust the pointer to the more lateral ITD. Instead, all listeners adjusted the pointer to a position closely corresponding to the noise at an ITD of zero. A close look at the slope values fit to the data for the 75% duty cycle condition indicates slightly negative values for all listeners. A negative slope effectively results in the perceived position of the target modulation being pushed further from the expected position of the other ITD-lateralized noise. Lee *et al.* (2009) has suggested that this type of lateral repulsion can occur when two perceptually distinct sound sources are presented together, thus, supporting the notion that listeners perceived two distinct streams of sound information as a result of this ITD-switching stimulus. Slopes greater than one for the other listening conditions in Figs. 4.3 - 4.5 could also suggest the presence of lateral repulsion. For the 50% duty cycle condition, the modulation percept of the stimulus was listener dependent per verbal reports. Some listeners suggested hearing two modulating sounds whereas some heard two stationary sound objects. Regardless, listeners all adjusted the pointer to the more lateral ITD position.

It is interesting that this stimulus simultaneously produces a modulating noise percept and a separate sound stream that is perceived as being continuous. One could consider that the sluggishness of the binaural system results in an inability to detect brief breaks in spatial information, thus, a percept of continuous noise at the ITD with the longer duty cycle. Alternatively, the continuous percept could be likened to the continuity illusion where a temporally disjoint tone can sound continuous if a sufficiently loud noise masks the onsets and offsets of the individual tone segments (Warren *et al.*, 1988). The more puzzling question is to the origin of the sound stream with a modulating component. The authors can only hypothesize that the percept of modulation is a result of the peak and trough patterns one would observe in a (short-term) cross-correlation pattern of the stimulus. Why the modulation percept would be attributed to the ITD occupying the shorter percentage of the duty cycle could potentially be related to an enhanced sensitivity to onsets of new glimpses of binaural information.

When the interaural disparity between the two ITDs of the ITD-switching stimulus was small, the ability of listeners to attribute an accurate lateral position to the target ITD was largely diminished. This is illustrated by the reduced slope for target ITD values between ± 200 μs observed in Fig. 4.2 for many of the test conditions. As indicated by the findings of the pilot study mentioned in Sect. 4.2.1, listeners likely did not perceive two separate sound objects for these smaller target ITD values. Given the instructions of Exp. 1 to adjust the pointer to the middle of a single perceived sound source, this reduced slope for smaller ITD magnitudes was a reasonable outcome.

Although no monaural information via temporal envelopes was provided in Exp. 1, results from Schimmel *et al.* (2008) suggest that envelope information might influence sensitivity to binaural cues. In their study it was shown that the peakedness of the stimulus influenced the ability of listeners to discriminate between the sidedness of a simultaneously presented harmonic tone complex and a bandpass noise. The authors argued that the monaural cues were necessary to facilitate a fast selection of the binaural cues. The ability of listeners in Exp. 1 of the current study to accurately attribute a lateral position to very brief segments of ITD information in the absence of a particular monaural temporal structure indicates that the envelope information is not a requirement for a fast binaural readout. Envelope information has, however, been shown to influence the sensitivity to interaural phase difference cues (Dietz *et al.*, 2013). The enhanced sensitivity to binaural information during the rising portion of an AM as demonstrated in Dietz *et al.* (2013) was not observed in the absolute lateralization task of Exp. 2 using broadband stimuli. Results of Exp. 3, however, suggest that the bandwidth of the stimulus can influence the role of AM for sensitivity to binaural cues. For narrow bandwidths, listeners were more sensitive to the ITD target in the rising portion of the AM as compared to when the ITD target was temporally located in the falling portion of the AM. An emphasis of the rising portion was not seen, however, when the bandwidth of the ITD-switching stimulus was increased to 4 or 8 ERB. The relationship of bandwidth on the emphasis of binaural cues during the rising portion of an AM is consistent with the results of Dietz *et al.* (2013) where tonal stimuli were used.

As suggested by recordings from both the medial superior olive and the inferior colliculus of adult Mongolian gerbils in Dietz *et al.* (2014), the increased weighting of spatial cues during the rising portion of the AM could be explained by neural adaptation. Quantifying the contribution of the instantaneous interaural phase difference during the different portions of the modulation cycle illustrated a clear weighting of the binaural cues during the rising portion of the AM. This onset emphasis could also possibly be explained by ringing along the basilar membrane which has been shown to influence the temporal resolution of the auditory system at low frequencies (Duifhuis, 1973, Plack and Moore, 1990). The implication this ringing has on the ITD-switching stimulus is that the ITD prior to the ITD transition will persist for a brief time, where the duration is dependent upon the bandwidth of the given peripheral filter. The ITD subsequent to the ITD transition will dominate only once the ringing has decayed such that the energy ratio between the noises for the new ITD and the persisting ITD is sufficiently large. For the ITD target positioned in the rising portion of the AM, it is possible that the target ITD can be more quickly represented due to the reduced energy from the interfering ITD prior to the transition. In contrast, the energy of the interfering ITD is relatively greater when the ITD target is positioned in the falling portion of the AM.

This general reduction in ITD thresholds with increasing bandwidth found in Exp. 3 agrees with Klumpp and Eady (1956) where smaller ITD jnds were observed for noises with larger bandwidths as compared to those with narrower bandwidths. It is possible that the influence of bandwidth could be related to the straightness measure introduced by Stern *et al.* (1988). Although the weighting of straightness was used to predict the bandwidth dependence on the perceived absolute position of ITD-lateralized stimuli, it is conceivable that the increased weighting for stimuli with a larger band of frequencies containing a consistent ITD would result in an improvement in ITD thresholds. The reduction in thresholds with increasing bandwidth, however, does not in itself explain the dominance of the rising AM portion at smaller bandwidths. For the narrow bandwidths in Exp. 3, ITD thresholds were elevated independent of the temporal position in the AM cycle, which is presumably due to the straightness weighting (Stern *et al.*, 1988). At these elevated thresholds, a potential improvement in thresholds could be achieved by using the more optimal rising portion where the influence of ringing is reduced and/or where neural adaptation is pronounced. For wider bandwidths, the auditory system is provided with more opportunities across frequency to detect the changes in ITD such that detection in at least one frequency channel could be possible even during the falling portion of the AM. This would result in detection of the changing ITD cues reaching a lower limit, e.g. due to peripheral and central processes, regardless of the temporal position of the ITD target. Improvements in performance as a result of wider bandwidths have been modeled by Buus *et al.* (1986) and suggest that the auditory system has the ability to make looks across frequency to improve detection of target information. Similar improvements in performance with increasing bandwidth have also been observed with a binaural detection paradigm (Langhans and Kohlrausch, 1992). Without further examination, it is not clear if the added bandwidth provides filters with a faster response and less deleterious effects of ringing and/or if the additional bandwidth simply provides more frequency channels to detect glimpses of target ITD information.

As noted by Bernstein *et al.* (2001), several of the studies showing a sluggishness of the binaural system, (e.g. Grantham and Wightman (1979), Kollmeier and Gilkey (1990), Culling and Summerfield (1998), Akeroyd and Summerfield (1999)) used stimuli where the time varying binaural cues were restricted to relatively narrow bandwidths. The elevated thresholds observed for narrowband stimuli as compared to broadband stimuli in Exp. 3 also suggest that the bandwidth of a stimulus could play some role in the measured temporal acuity of the binaural system. This bandwidth dependency would also be in agreement with the result of Schimmel *et al.* (2008) where lateral discrimination between spatial configuration of two simultaneously presented, ITD-lateralized stimuli was not possible when the spectrum was limited to one auditory filter.

In realistic listening scenarios, there are often multiple sound sources that contain different temporal patterns of changing amplitude. Due to the importance of low modulation frequencies for speech intelligibility (Drullman *et al.*, 1994), the ability to identify a target modulation from other modulations based on the temporal pattern could play a role in the ability of listeners to disentangle complex auditory scenes. Although a strong role of spatial cues for AM detection is questionable (Kopčo and Shinn-Cunningham, 2008), the results of Exp. 4 show that greater spatial separation improved the ability of listeners to discriminate between the modulation frequencies of two simultaneously presented AM noises. This demonstrates that the binaural system in its own right is capable of segregating multiple, modulated sources into separate streams. Results for the slower modulation frequencies tested in Exp. 4 indicate that a frequency difference of only 3-5 Hz is necessary to discriminate between the two AM signals assuming a spatial separation of 100 or 300 μs. Since the temporal modulations of speech typically fall into these slower modulation frequencies, it is possible that differences in AM frequency of only a few Hz combined with reasonable spatial separation of the modulated sources could help listeners to segregate sounds into separate streams.

One final point to be made concerns the fixed modulation pattern in Exp. 1, i.e. the non-zero ITD noise token was always the first to be presented in the alternating ITD sequence. Considering the onset dominance for lateralization (Houtgast and Aoki, 1994, Freyman *et al.*, 1997, Devore *et al.*, 2009, Stecker *et al.*, 2013), one could suggest that lateralization performance was predominantly due to the first ITD noise token presented. We would argue, however, that the fixed temporal order was not the absolute factor in listeners' adjustments of the acoustic pointer in Exp. 1 for two reasons. First, the instructions given to the listeners specified that they should adjust the pointer to the modulating sound object. Without sufficient repetitions of the alternating ITD, the modulation percept would not be sufficiently built up and the lateral position for the target sound object could not yet be made. Furthermore, the choice of listeners to adjust the pointer in the 75% duty cycle condition to the non-zero ITD, i.e. not the first ITD noise token to be presented in the sequence, suggests that the ability to adjust the pointer to the modulating sound object did not directly depend on the temporal order of which ITD was presented first. The second reason why we don't believe the temporal order largely influenced lateralization performance is because little difference in performance was observed in Exp. 2 where the starting ITD noise token was randomized. Had the starting order influenced lateralization, less accurate judgements of lateral position should have been observed in Exp. 2, which was not the case.

4.7 Conclusion

The experiments presented here illustrate that the binaural system does not impose great constraints beyond monaural temporal limitations on the ability of the auditory system to detect brief glimpses of ITD information. Experiment 1 showed that listeners are able to attribute a reasonably accurate lateral position to these brief ITD tokens in the absence of any particular monaural envelope structure. Results of Exp. 2 and Exp. 3 indicate that the temporal position of an ITD target within the phase of a diotic AM has little influence on detection and absolute lateralization given a sufficiently wide stimulus bandwidth. As the bandwidth of the stimulus reduces to only 1 or 2 ERB, it appears that the ability to detect an ITD target in the rising portion of an AM is better than if the ITD target is temporally positioned in the falling portion of an AM. For envelope frequency discrimination, an increased spatial separation between the simultaneous AM noises provided a benefit as shown in Exp. 4. This result illustrates that the binaural system alone can aid in the task of disentangling a mixture of stimuli in complex acoustic scenes.

5

General Conclusions

The aim of the research presented in this thesis was directed at the performance of the binaural system in listening scenarios composed of multiple sound sources. The following were primary investigations conducted for this thesis: (1) Absolute lateralization performance was evaluated for ILD-lateralized noise bursts that were temporally and spectrally overlapping. (2) Temporal acuity of the auditory system to brief changes in ITD was measured. (3) An interaction between the binaural and monaural pathways of modulation sensitivity was investigated. (4) An ability to attribute an accurate lateral estimate of brief ITD noise tokens was determined. (5) Whether or not monaural envelope cues influence sensitivity to binaural tokens was demonstrated.

In **Chapter 2** absolute lateralization performance was measured for an ILD-lateralized target presented simultaneously with an interfering noise containing the same long-term spectrum and zero ILD. The presence of a temporally overlapping background noise stimulus in addition to the target noise burst results in the physical ILD at the ears being a mixture of both the target ILD and background noise ILD. Since common onsets are known to be useful for the grouping of sound information into auditory objects, a temporal onset/offset asynchrony (TOA) could presumably aid the auditory system in making a better estimate of the lateral position corresponding to the target ILD. Indeed, the TOA applied in the experiments of this study allowed listeners to estimate a position corresponding to the target ILD or a position much greater than the reduced ILD expected from the combined target and background noise ILDs. This suggests that the auditory system possesses some means of compensating for the presence of the background noise. Furthermore, in order to make this compensation, it is likely that the auditory system assumes (in the absence of a salient cue suggesting otherwise) that the background noise continues during the target noise burst and that the level of the background noise in each ear remains constant.

In a lateralization model by Braasch (2003), knowledge about the background noise in isolation was used to make an estimate of the target ILD for a stimulus when the target was combined with background noise. If, however, the background noise in isolation cannot be observed, this model would presumably only be able to estimate the ILD resulting from sum of the target and background noise. In the experiments of Chapter 2 for conditions when no TOA was present, we indeed found lateralization to correspond with this sum-based ILD provided that the background noise was interaurally correlated. Results when the background noise was interaurally uncorrelated, however, illustrated that many listeners were able to use the target ILD or at least an ILD that was larger than the physical ILD at the ears. It was conceived that this dependence on the background noise coherence could be related to our natural acoustic surroundings. In everyday acoustical settings, interaurally uncorrelated noise always has an ILD close to zero. If the auditory system makes this assumption about interaurally uncorrelated noise, it would be possible to infer the ILD of an interaurally correlated target. For an interaurally correlated background noise, the assumed ILD of zero is no longer a realistic choice, and it would not be reasonable to attempt an inference of the target ILD. Until more is known about the ILD of the interferer, e.g. due to an onset asynchrony, the best assumption is to simply use the ILD at the ears.

Because the model from Braasch (2003) is only capable of estimating the target ILD in a target plus background noise stimulus if knowledge of the background noise in isolation is available, an alternative solution would be necessary to account for lateralization performance with a synchronous, interaurally uncorrelated background noise. An extension to the EC model (Durlach, 1963, Breebaart *et al.*, 2001) was, therefore, proposed in this thesis. The EC model can be used to make lateral estimates by taking the compensatory adjustment from the equalization stage that results in a minimum at the output of the cancellation stage. By making the output of the model uniform for a given background noise, the lateral bias as a result of the presence of the background noise can be removed, and the target ILD can be accurately estimated. For the model simulations, one weighting template was generated for an interaurally uncorrelated background noise while another was generated for the interaurally correlated background noise. Each template resulted in uniform model output when the background noise with corresponding interaural coherence was presented in isolation. By assuming that the auditory system defaults to using the template generated for an interaurally uncorrelated background noise unless information about the coherence of the background noise can be determined, e.g. from a TOA, the general trends observed in the psychophysical data can be predicted. It was also shown that even when a more realistic peripheral model was incorporated, which included the loss of phase locking at the auditory nerve, numerical computation of a weighting template that resulted in uniform model output could still achieve similar predictions of the trends in the psychophysical data. It was conjectured that the individual differences observed in the psychophysical data could possibly be a result of different choices for an internal weighting template.

In **Chapter 3**, the temporal acuity of the binaural system for detecting brief periodic changes in ITD was measured. The stimulus was a broadband noise that periodically alternated between two different ITD values. Listeners demonstrated an ability to detect brief changes in ITD that were only 3-5 ms in duration. Such brief durations suggest that for this auditory task the binaural system does not impose a considerably longer processing time beyond the temporal acuity of the monaural system. It was observed that this ITD-switching stimulus elicited two perceptual sound streams: a continuous noise percept and a modulating source percept. The ability to discriminate both the frequency and the sidedness of the perceived modulation was evaluated. Results showed that the purely binaural modulation percept resulting from this stimulus contained a periodic structure that was discriminable for frequencies below approximately 20 Hz. At the slower modulation frequencies, a 50% change in modulation frequency was necessary for accurate discrimination of both frequency and sidedness of this modulation. This experiment illustrates that a purely binaural modulation percept can arise from brief, instantaneous changes in interaural cues.

As a result of this purely binaural modulation percept, it was conceived that a potential interaction between monaural and binaural modulation sensitivities could exist. Modulation interference due to sinusoidally varying ILD cues have been shown (Thompson and Dau, 2008), however, ILD cues inherently generate monaural level cues. In contrast, the ITD-switching stimulus produces a perceived modulation that is purely binaural in nature. Results of the modulation interference experiment indicate a very small interference of the binaural modulation on AM detection. Additionally, a very broad tuning of the interference was found. These results suggest that the rapid fluctuation of binaural cues in cocktail party scenarios should have a limited effect on the ability of listeners to detect the envelope modulations of speech. The interference of AM on the ability of listeners to detect the purely binaural modulation appears to have a relatively larger influence. A near doubling in the duration of the noise at the ITD corresponding to the percept of modulation was required for detection when AM was present. This has a potentially greater implication on the ability to detect periodic glimpses at a particular ITD since the result suggests the presence of AM could interfere with the temporal acuity of the binaural system. In this condition where AM was applied as a modulation masker, no indication of tuning in the modulation masking was apparent. Even with the doubling of these thresholds, the temporal acuity of the binaural system in the presence of AM permitted detection of changes in binaural information of less than 10 ms in duration. These two experiments, taken as a whole, would suggest that the modulations in the monaural and the binaural domain are, for the most part, independently represented.

Although results of Chapter 3 and work by Bernstein *et al.* (2001) have demonstrated that listeners are able to detect very brief changes in ITD, whether or not an accurate lateral position can be attributed to these brief changes is not known. In **Chapter 4**, the temporal acuity of the auditory system for making lateralization estimates of brief glimpses of ITD information was determined by using the ITD-switching stimulus applied in Chapter 3. For these absolute lateralization measurements, the population of listeners required glimpses of 7.5 ms in duration for the change in ITD in order to make accurate lateralization estimates of the target. Some of the individual listeners could accurately lateralize noise targets of only 3-6 ms in duration. For sensitivity to such brief durations, this suggests that the binaural system does not necessarily require longer processing times than that which is required for the monaural system. Given the explicit instructions provided to listeners such that they should indicate the lateral position of the modulating sound source, the results of this lateralization study further validated that listeners perceive a modulation percept only at the ITD occupying a smaller percentage of the ITD-switching period. This stimulus illustrates that the binaural system, by itself, can mediate the buildup of two separate auditory streams.

The role of monaural envelope cues has been suggested to influence the auditory system's readout of binaural cues. It is possible that binaural cues during the rising slopes of AM are given a greater perceptual weight (Dietz *et al.*, 2013) and/or that envelope cues act as an indicator of more optimal times when to sample the binaural cues (Schimmel *et al.*, 2008). By imposing an AM onto the ITD-switching stimulus, the influence of the monaural envelope on the ability of listeners to lateralize brief changes in ITD was measured. Although no influence of the monaural envelope was observed for absolute lateralization performance with a broadband stimulus, a second experiment with bandlimited versions of the stimulus did show an influence of the monaural envelope on ITD-jnd sensitivity. For a stimulus bandwidth of 1 ERB, listeners demonstrated a greater sensitivity to the ITD changes during the rising edge of the monaural envelope. In contrast, no influence of the monaural envelope was observed when the bandwidth was increased to 4 ERB or more. Although the emphasis to binaural cues during the rising portion of the AM could potentially be related to neural adaptation (Dietz *et al.*, 2014), there is no clear explanation for the absence of this effect at larger bandwidths. It can only be hypothesized that with larger bandwidths the auditory system has a greater number of frequency channels to observe the changes in ITD, and with a sufficient number of channels to detect the changes, a lower limit resulting from peripheral processing is reached.

The final experiment of Chapter 4 determined whether spatial separation of two simultaneously presented broadband AM noises provides a benefit for an envelope frequency discrimination task. Although the task was generally quite difficult, increasing the spatial separation between the two AM noises improved the ability of listeners to discriminate the two AM frequencies. This illustrates that the binaural system alone can provide a benefit when segregating multiple sources that are temporally entangled. Since performance on the task was better for the slower modulation rates (generally in the range of speech modulations), it is interesting to suggest that purely binaural information provides a benefit in the segregation of AM sources, e.g. speech.

Taken as a whole, the results of this thesis further validate the remarkable capabilities of the auditory system. Because the human brain likely attempts to utilize a full range of cues (spectral, temporal, spatial, visual, memorial, contextual, etc.) when resolving complex auditory tasks, it is only possible to reflect on potential interpretations of the largely binaurally-focused results presented in this thesis with regard to their relationship in the processing cocktail party scenarios. For auditory object formation, stronger contribution of monaural cues over binaural cues is likely beneficial. Since a variety of issues – energetic masking, hearing damage, occlusion, etc. – can impede a single ear, the requirement of two ears for the fundamental task of source segregation would place a severe limitation on listeners in the presence of the aforementioned issues. Binaural information, however, could provide additional cues that strengthen (and potentially weaken) the various auditory streams that arise from the monaural cues. By attending to glimpses of information corresponding to the location of the target auditory steam, the binaural system provides an additional perceptual layer in which to organize the auditory image. Therefore, an auditory system that is capable of compensating for the deleterious effects on binaural cues resulting from background noise and one that is able to rapidly select brief tokens of binaural information to maintain perceptual streams would likely be important when making sense of the soundscapes in daily life.

5.1 Suggestions for future research

Individual differences in performance were observed in multiple studies presented here, most notably in the absolute lateralization experiments. Differences in listener performance are not that uncommon and have been observed for a variety of binaural tasks (Andéol *et al.*, 2013, Buss *et al.*, 2007). While degradations to the hearing sensory organ, e.g. resulting in pure-tone threshold deficiencies, can lead to a more obvious drop in performance due to the reduced capacity to obtain auditory input signals, differences observed for supra-threshold measurements possibly require a more complex explanation. It could be argued that these supra-threshold capabilities have a greater practical relevance since the levels of sound targets within acoustic scenes are predominantly much greater than normal thresholds of hearing. If it is true that fundamental differences in a given listener's auditory pathway, e.g. as a result of age or neuropathy, can result in differences in auditory object formation and streaming, a better understanding of mechanisms behind these differences could provide more insight into the crucial functionality behind cocktail party processing. Although these listener-dependent differences will likely emerge from more than just binaural tasks, the results from the studies of this thesis could suggest directions in which these differences might be assessed.

From the ILD investigations in Chapter 2, a range in lateralization performance was observed across the population of listeners. Some listeners were only able to use the physical ILD at the ears whereas other listeners in the same test condition were able to utilize the target ILD or at least an ILD much larger than that physically found at the ears. This could suggest that some listeners have a deficit in their ability to compensate for the presence of background noise when making such lateral judgements. If perceptually distinct objects indeed result in more laterally disparate perceptual locations (Lee *et al.*, 2009), it would be interesting to determine if differences in the ability to process the ILD cues in target plus background noise scenarios can reflect differences in the ability to perform tasks requiring source segregation and streaming. Additionally, if the compensation for the presence of background noise when making lateral estimates is influenced by the coherence of the background noise, as proposed by the model extension in Chapter 2, then individual sensitivities to interaural coherence could influence the ability to compensate for the background noise. Differences in sensitivity to interaural coherence have already been demonstrated between young and old listeners (Whitmer *et al.*, 2012), and it would be interesting to identify if these differences can be associated with individual abilities to utilize the target ILD in target plus background noise listening scenarios.

The ITD-switching stimulus applied in Chapters 3 and 4 could also help identify individual differences in binaural processing. While some listeners were able to detect and accurately lateralize brief ITD tokens of only a few milliseconds, another listener required a duration that was an order of magnitude larger. It would be interesting to determine if the differences in performance are directly related to binaural processing or are possibly related to subtle deficiencies in monaural processing. Since the modulation percept from this ITD-switching stimulus is purely binaural in nature, individual differences in sensitivity to this modulation could be associated with differences in binaural processing capabilities. It is yet to be determined if such sensitivity could be attributed to some listeners possessing a more temporally acute binaural system or to some listeners possessing a stronger and more faithful representation of binaural temporal patterns. Electroencephalography recordings have been employed to identify neural signals associated with different auditory tasks in the hope of finding correlates between psychophysical performance and the recorded signals (Ruggles *et al.*, 2011, Choi *et al.*, 2014). Whether or not the ITD-switching stimulus could be utilized for similar interests is worth further investigation. As demonstrated by Siveke *et al.* (2008), the manner in which time-varying interaural coherence is generated can influence the temporal acuity of listeners in psychophysical measurements and can affect single-cell responses in the dorsal nucleus of the lateral lemniscus of gerbils. Therefore, it would be worth testing if the absence of weighting to the ITD cues during the rising slope of the AM for broadband stimuli (Experiments 2 and 3 in Chapter 4) is influenced by the manner in which the time-varying ITD cues are generated. Continuously varying binaural cues from Dietz *et al.* (2013, 2014) (where weighting of the binaural cues at the rising slope of the AM was found) could possibly utilize a different type of binaural processing as compared to that required for the instantaneous ITD transitions of the experiments presented here. An attempt to identify correlates between differences found in scalp recordings and differences found in psychophysical measures when different types of time-varying binaural cues are employed could be worth investigating. These correlates could potentially provide insight into which types of time-varying binaural features yield more salient information in particular circumstances of complex listening scenes.

Bibliography

Akeroyd, M. A. and Bernstein, L. R. (**2001**), “The variation across time of sensitivity to interaural disparities: Behavioral measurements and quantitative analyses,” *J. Acoust. Soc. Am.* **110**(5), pp. 2516–2526. (Cited on page 32)

Akeroyd, M. A. and Summerfield, A. Q. (**1999**), “A binaural analog of gap detection,” *J. Acoust. Soc. Am.* **105**(5), pp. 2807–2820. (Cited on pages 76 and 79)

Andéol, G., Macpherson, E. A., and Sabin, A. T. (**2013**), “Sound localization in noise and sensitivity to spectral shape,” *Hearing Research* **304**(0), pp. 20–27. (Cited on pages 55 and 85)

Bernstein, L. R. and Trahiotis, C. (**1985**), “Lateralization of low-frequency, complex waveforms: The use of envelope-based temporal disparities,” *J. Acoust. Soc. Am.* **77**(5), pp. 1868–1880. (Cited on page 14)

Bernstein, L. R. and Trahiotis, C. (**2004**), “The apparent immunity of high-frequency “transposed” stimuli to low-frequency binaural interference,” *J. Acoust. Soc. Am.* **116**(5), pp. 3062–3069. (Cited on page 50)

Bernstein, L. R., Trahiotis, C., Akeroyd, M. A., and Hartung, K. (**2001**), “Sensitivity to brief changes of interaural time and interaural intensity,” *J. Acoust. Soc. Am.* **109**(4), pp. 1604–1615. (Cited on pages 7, 8, 23, 32, 52, 58, 59, 60, 76, 79, and 84)

Best, V., Gallun, F. J., Carlile, S., and Shinn-Cunningham, B. G. (**2007**), “Binaural interference and auditory grouping,” *J. Acoust. Soc. Am.* **121**(2), pp. 1070–1076. (Cited on pages 6, 13, 14, and 22)

Best, V., van Schaik, A., Jin, C., and Carlile, S. (**2005**), “Auditory spatial perception with sources overlapping in frequency and time,” *Acta Acustica united with Acustica* **91**(3), pp. 421–428. (Cited on page 22)

Braasch, J. (**2003**), “Localization in the Presence of a Distracter and Reverberation in the Frontal Horizontal Plane: III. The Role of Interaural Level Differences,” *Acta Acustica united with Acustica* **89**(4), pp. 674–692. (Cited on pages 6, 26, and 82)

Braasch, J. and Hartung, K. (**2002**), “Localization in the Presence of a Distracter and Reverberation in the Frontal Horizontal Plane: I. Psychoacoustical Data," *Acta Acustica united with Acustica* **88**, pp. 942–955. (Cited on pages 6, 12, and 22)

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001**), "Binaural processing model based on contralateral inhibition. I. Model structure," *J. Acoust. Soc. Am.* **110**(2), pp. 1074–88. (Cited on pages 6, 7, 13, 24, 28, 32, 58, and 82)

Bregman, A. S. (**1990**), *Auditory Scene Analysis*, MIT Press, Cambridge, MA. (Cited on pages 5, 22, 32, and 58)

Bronkhorst, A. W. (**2000**), "The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple-Talker Conditions," *Acta Acustica united with Acustica* **86**(1), pp. 117–128. (Cited on pages 32 and 58)

Bronkhorst, A. W. and Plomp, R. (**1992**), "Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing," *J. Acoust. Soc. Am.* **92**(6), pp. 3132–3139. (Cited on page 33)

Brungart, D. S. and Simpson, B. D. (**2009**), "Effects of bandwidth on auditory localization with a noise masker," *J. Acoust. Soc. Am.* **126**(6), pp. 3199–3208. (Cited on page 12)

Buell, T. and Hafter, E. (**1991**), "Combination of binaural information across frequency bands," *J. Acoust. Soc. Am.* **90**(4), pp. 1894. (Cited on pages 7 and 58)

Buss, E., Hall, J. W., and Grose, J. H. (**2007**), "Individual differences in the masking level difference with a narrowband masker at 500 or 2000Hz," *J. Acoust. Soc. Am.* **121**(1), pp. 411–419. (Cited on pages 55 and 85)

Buss, E., Whittle, L. N., Grose, J. H., and Hall, J. W. (**2009**), "Masking release for words in amplitude-modulated noise as a function of modulation rate and task," *J. Acoust. Soc. Am.* **126**(1), pp. 269–280. (Cited on page 58)

Butler, R. A. and Naunton, R. F. (**1962**), "Some Effects of Unilateral Auditory Masking upon the Localization of Sound in Space," *J. Acoust. Soc. Am.* **34**(8), pp. 1100–1107. (Cited on pages 6, 12, and 22)

Buus, S., Schorer, E., Florentine, M., and Zwicker, E. (**1986**), "Decision rules in detection of simple and complex tones," *J. Acoust. Soc. Am.* **80**(6), pp. 1646–1657. (Cited on page 79)

Canévet, G. and Meunier, S. (**1996**), "Effect of adaptation on auditory localization and lateralization," *Acta Acustica united with Acustica* **82**(1), pp. 149–157. (Cited on pages 6, 12, and 22)

Carlile, S., Hyams, S., and Delaney, S. (**2001**), "Systematic distortions of auditory space perception following prolonged exposure to broadband noise," *J. Acoust. Soc. Am.* **110**(1), pp. 416–424. (Cited on pages 6, 12, and 22)

Cherry, E. C. (**1953**), "Some Experiments on the Recognition of Speech, with One and with Two Ears," *J. Acoust. Soc. Am.* **25**(5), pp. 975–979. (Cited on page 5)

Choi, I., Wang, L., Bharadwaj, H., and Shinn-Cunningham, B. (**2014**), "Individual differences in attentional modulation of cortical responses correlate with selective attention performance," *Hear Res* **314**, pp. 10–9. (Cited on pages 55 and 86)

Cohen, M. F. (**1981**), "Interaural time discrimination in noise," *J. Acoust. Soc. Am.* **70**(5), pp. 1289–1293. (Cited on pages 13 and 23)

Colburn, H. S. (**1973**), "Theory of binaural interaction based on auditory-nerve data. I. General strategy and preliminary results on interaural discrimination," *J. Acoust. Soc. Am.* **54**(6), pp. 1458–1470. (Cited on page 24)

Cooke, M. (**2006**), "A glimpsing model of speech perception in noise," *J. Acoust. Soc. Am.* **119**(3), pp. 1562–1573. (Cited on page 58)

Culling, J. F., Colburn, H. S., and Spurchise, M. (**2001**), "Interaural correlation sensitivity," *J. Acoust. Soc. Am.* **110**(2), pp. 1020–1029. (Cited on page 50)

Culling, J. F. and Summerfield, Q. (**1995**), "Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay," *J. Acoust. Soc. Am.* **98**(2), pp. 785–797. (Cited on pages 7, 32, and 58)

Culling, J. F. and Summerfield, Q. (**1998**), "Measurements of the binaural temporal window using a detection task," *J. Acoust. Soc. Am.* **103**(6), pp. 3540–3553. (Cited on pages 7, 76, and 79)

Darwin, C. J. (**1984**), "Perceiving vowels in the presence of another sound: Constraints on formant perception," *J. Acoust. Soc. Am.* **76**(6), pp. 1636–1647. (Cited on page 22)

Darwin, C. J. and Hukin, R. W. (**1999**), "Auditory objects of attention: the role of interaural time differences," *J Exp Psychol Hum Percept Perform* **25**(3), pp. 617–29. (Cited on page 7)

Darwin, C. J. and Hukin, R. W. (**2000**), "Effectiveness of spatial cues, prosody, and talker characteristics in selective attention," *J. Acoust. Soc. Am.* **107**(2), pp. 970–977. (Cited on page 7)

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**), "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," *J. Acoust. Soc. Am.* **102**(5), pp. 2892–2905. (Cited on page 33)

Devore, S., Ihlefeld, A., Hancock, K., Shinn-Cunningham, B., and Delgutte, B. (**2009**), "Accurate Sound Localization in Reverberant Environments Is Mediated by Robust Encoding of Spatial Cues in the Auditory Midbrain," *Neuron* **62**(1), pp. 123–134. (Cited on pages 59 and 80)

Dietz, M., Ewert, S. D., and Hohmann, V. (**2009**), "Lateralization of stimuli with independent fine-structure and envelope-based temporal disparities," *J. Acoust. Soc. Am.* **125**(3), pp.

1622–1635. (Cited on pages 14 and 62)

Dietz, M., Marquardt, T., Salminen, N. H., and McAlpine, D. (**2013**), "Emphasis of spatial cues in the temporal fine structure during the rising segments of amplitude-modulated sounds," *Proc Natl Acad Sci U S A* **110**(37), pp. 15151–6. (Cited on pages 8, 59, 67, 69, 78, 84, and 86)

Dietz, M., Marquardt, T., Stange, A., Pecka, M., Grothe, B., and McAlpine, D. (**2014**), "Emphasis of spatial cues in the temporal fine-structure during the rising segments of amplitude modulated sounds II: Single neuron recordings," *Journal of Neurophysiology* . (Cited on pages 78, 84, and 86)

Drennan, W. R., Gatehouse, S., and Lever, C. (**2003**), "Perceptual segregation of competing speech sounds: The role of spatial location," *J. Acoust. Soc. Am.* **114**(4), pp. 2178–2189. (Cited on page 58)

Drullman, R., Festen, J. M., and Plomp, R. (**1994**), "Effect of temporal envelope smearing on speech reception," *J. Acoust. Soc. Am.* **95**(2), pp. 1053–64. (Cited on pages 8, 33, and 79)

Duifhuis, H. (**1973**), "Consequences of peripheral frequency selectivity for nonsimultaneous masking," *J. Acoust. Soc. Am.* **54**(6), pp. 1471–1488. (Cited on pages 76 and 78)

Durlach, N. I. (**1963**), "Equalization and Cancellation Theory of Binaural Masking-Level Differences," *J. Acoust. Soc. Am.* **35**(8), pp. 1206–1218. (Cited on pages 6, 13, 24, 27, and 82)

Erceg-Hurn, D. M. and Mirosevich, V. M. (**2008**), "Modern robust statistical methods: an easy way to maximize the accuracy and power of your research," *Am. Psychol.* **63**(7), pp. 591–601. (Cited on page 37)

Ewert, S. D. and Dau, T. (**2000**), "Characterizing frequency selectivity for envelope fluctuations," *J. Acoust. Soc. Am.* **108**(3), pp. 1181–1196. (Cited on pages 8, 33, 46, and 53)

Faller, C. and Merimaa, J. (**2004**), "Source localization in complex listening situations: selection of binaural cues based on interaural coherence," *J. Acoust. Soc. Am.* **116**(5), pp. 3075–89. (Cited on pages 6 and 59)

Fox, J. and Weisberg, S. (**2011**), *An R Companion to Applied Regression*, Sage, second edition ed. (Cited on page 71)

Freyman, R. L., Zurek, P. M., Balakrishnan, U., and Chiang, Y.-C. (**1997**), "Onset dominance in lateralization," *J. Acoust. Soc. Am.* **101**(3), pp. 1649–1659. (Cited on pages 59 and 80)

Glasberg, B. R. and Moore, B. C. (**1990**), "Derivation of auditory filter shapes from notched-noise data," *Hear Res* **47**(1-2), pp. 103–38. (Cited on page 69)

Good, M. D. and Gilkey, R. H. (**1996**), "Sound localization in noise: The effect of signal-to-noise ratio," *J. Acoust. Soc. Am.* **99**(2), pp. 1108–1117. (Cited on page 12)

Grantham, D. W. (**1982**), "Detectability of time-varying interaural correlation in narrow-band noise stimuli," *J. Acoust. Soc. Am.* **72**(4), pp. 1178–1184. (Cited on pages 6, 7, and 76)

Grantham, D. W. (**1995**), "Spatial Hearing and Related Phenomena," in *Handbook of Perception and Cognition*, edited by B. C. J. Moore, London, UK: Academic, pp. 297–339. (Cited on page 12)

Grantham, D. W. and Bacon, S. P. (**1991**), "Binaural modulation masking," *J. Acoust. Soc. Am.* **89**(3), pp. 1340–1349. (Cited on page 52)

Grantham, D. W. and Wightman, F. L. (**1978**), "Detectability of varying interaural temporal differences," *J. Acoust. Soc. Am.* **63**(2), pp. 511–23. (Cited on pages 6, 53, 58, and 76)

Grantham, D. W. and 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**(6), pp. 1509–1517. (Cited on pages 32, 58, 76, and 79)

Heller, L. M. and Trahiotis, C. (**1996**), "Extents of laterality and binaural interference effects," *J. Acoust. Soc. Am.* **99**(6), pp. 3632–3637. (Cited on pages 6, 13, and 22)

Hill, N. I. and Darwin, C. J. (**1996**), "Lateralization of a perturbed harmonic: Effects of onset asynchrony and mistuning," *J. Acoust. Soc. Am.* **100**(4), pp. 2352–2364. (Cited on page 22)

Hohmann, V. (**2002**), "Frequency analysis and synthesis using a Gammatone filterbank," *Acta Acustica united with Acustica* **88**, pp. 433–442. (Cited on pages 24 and 49)

Houtgast, T. (**1989**), "Frequency selectivity in amplitude-modulation detection," *J. Acoust. Soc. Am.* **85**(4), pp. 1676–1680. (Cited on pages 8 and 33)

Houtgast, T. and Aoki, S. (**1994**), "Stimulus-onset dominance in the perception of binaural information," *Hearing Research* **72**(1), pp. 29–36. (Cited on pages 59 and 80)

Ihlefeld, A. and Shinn-Cunningham, B. G. (**2011**), "Effect of source spectrum on sound localization in an everyday reverberant room," *J. Acoust. Soc. Am.* **130**(1), pp. 324–33. (Cited on page 12)

Jeffress, L. A. (**1948**), "A place theory of sound localization," *J Comp Physiol Psychol* **41**(1), pp. 35–9. (Cited on pages 24 and 54)

Josupeit, A. (**2011**), "Auditory processing of complex signals - Contribution of monaural and binaural cues to source discrimination and segregation," Master's thesis, University of Oldenburg. (Cited on page 54)

Kidd, G., Arbogast, T. L., Mason, C. R., and Gallun, F. J. (**2005**), "The advantage of knowing where to listen," *J. Acoust. Soc. Am.* **118**(6), pp. 3804–3815. (Cited on page 8)

Klumpp, R. G. and Eady, H. R. (**1956**), "Some Measurements of Interaural Time Difference Thresholds," *J. Acoust. Soc. Am.* **28**(5), pp. 859–860. (Cited on page 78)

Kollmeier, B. and Gilkey, R. H. (**1990**), "Binaural forward and backward masking: Evidence for sluggishness in binaural detection," *J. Acoust. Soc. Am.* **87**(4), pp. 1709–1719. (Cited on pages 7, 32, 58, 76, and 79)

Kopčo, N. and Shinn-Cunningham, B. G. (**2008**), "Influences of modulation and spatial separation on detection of a masked broadband target," *J. Acoust. Soc. Am.* **124**(4), pp. 2236–2250. (Cited on page 79)

Langhans, A. and Kohlrausch, A. (**1992**), "Spectral integration of broadband signals in diotic and dichotic masking experiments," *J. Acoust. Soc. Am.* **91**(1), pp. 317–326. (Cited on page 79)

Lee, A. K. C., Deane-Pratt, A., and Shinn-Cunningham, B. G. (**2009**), "Localization interference between components in an auditory scene," *J. Acoust. Soc. Am.* **126**(5), pp. 2543–2555. (Cited on pages 6, 13, 14, 22, 23, 29, 77, and 86)

Licklider, J. C. R. (**1948**), "The Influence of Interaural Phase Relations upon the Masking of Speech by White Noise," *J. Acoust. Soc. Am.* **20**(2), pp. 150–159. (Cited on page 5)

Lindemann, W. (**1986**), "Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals," *J. Acoust. Soc. Am.* **80**(6), pp. 1608–1622. (Cited on page 24)

Litovsky, R. Y. (**2012**), "Spatial Release from Masking," *Acoustics Today* **8**(2), pp. 18–25. (Cited on pages 32 and 58)

Lorenzi, C., Gatehouse, S., and Lever, C. (**1999**), "Sound localization in noise in normal-hearing listeners," *J. Acoust. Soc. Am.* **105**(3), pp. 1810–1820. (Cited on page 12)

May, T., van de Par, S., and Kohlrausch, A. (**2011**), "A Probabilistic Model for Robust Localization Based on a Binaural Auditory Front-End," *Audio, Speech, and Language Processing, IEEE Transactions on* **19**(1), pp. 1–13. (Cited on page 12)

Meunier, S., Bodden, M., Canévet, G., and Grabke, J. (**1996**), "Auditory adaptation and localization effect of frequency and bandwith," *Acta Acustica united with Acustica* **82**(Suppl. 1), pp. S215. (Cited on page 22)

Miller, G. A. and Licklider, J. C. R. (**1950**), "The Intelligibility of Interrupted Speech," *J. Acoust. Soc. Am.* **22**(2), pp. 167–173. (Cited on page 58)

Perrott, D. R. and Musicant, A. D. (**1977**), "Minimum auditory movement angle: Binaural localization of moving sound sources," *J. Acoust. Soc. Am.* **62**(6), pp. 1463–1466. (Cited on pages 6, 58, and 76)

Peters, R. W., Moore, B. C. J., and Baer, T. (**1998**), "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," *J. Acoust. Soc. Am.* **103**(1), pp. 577–587. (Cited on page 58)

Plack, C. J. and Moore, B. C. J. (**1990**), "Temporal window shape as a function of frequency and level," *J. Acoust. Soc. Am.* **87**(5), pp. 2178–2187. (Cited on pages 76 and 78)

Pollack, I. (**1978**), "Temporal switching between binaural information sources," *J. Acoust. Soc. Am.* **63**(2), pp. 550–558. (Cited on pages 7, 58, and 76)

Pollack, I. and Trittipoe, W. J. (**1959**), "Binaural Listening and Interaural Noise Cross Correlation," *J. Acoust. Soc. Am.* **31**(9), pp. 1250–1252. (Cited on page 50)

Rakerd, B. and Hartmann, W. M. (**1985**), "Localization of sound in rooms, II: The effects of a single reflecting surface," *J. Acoust. Soc. Am.* **78**(2), pp. 524–33. (Cited on page 12)

Rakerd, B. and Hartmann, W. M. (**2010**), "Localization of sound in rooms. V. Binaural coherence and human sensitivity to interaural time differences in noise," *J. Acoust. Soc. Am.* **128**(5), pp. 3052–3063. (Cited on page 12)

Robinson, D. E. and Egan, J. P. (**1974**), "Lateralization of an auditory signal in correlated noise and in uncorrelated noise as a function of signal frequency," *Perception & Psychophysics* **15**(2), pp. 281–284. (Cited on page 13)

Robinson, D. E. and Jeffress, L. A. (**1963**), "Effect of Varying the Interaural Noise Correlation on the Detectability of Tonal Signals," *J. Acoust. Soc. Am.* **35**(12), pp. 1947–1952. (Cited on pages 5 and 23)

Roman, N., Srinivasan, S., and Wang, D. (**2006**), "Binaural segregation in multisource reverberant environments," *J. Acoust. Soc. Am.* **120**(6), pp. 4040–4051. (Cited on page 58)

Ruggles, D., Bharadwaj, H., and Shinn-Cunningham, B. G. (**2011**), "Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication," *Proceedings of the National Academy of Sciences* **108**(37), pp. 15516–15521. (Cited on pages 55 and 86)

Sach, A. and Bailey, P. (**2004**), "Some characteristics of auditory spatial attention revealed using rhythmic masking release," *Perception & Psychophysics* **66**(8), pp. 1379–1387. (Cited on page 7)

Schimmel, O., van de Par, S., Breebaart, J., and Kohlrausch, A. (**2008**), "Sound segregation based on temporal envelope structure and binaural cues," *J. Acoust. Soc. Am.* **124**(2), pp. 1130–45. (Cited on pages 8, 59, 67, 77, 79, and 84)

Shackleton, T. M. and Meddis, R. (**1992**), "The role of interaural time difference and fundamental frequency difference in the identification of concurrent vowel pairs," *J. Acoust. Soc. Am.* **91**(6), pp. 3579–3581. (Cited on pages 7, 32, and 58)

Shailer, M. J. and Moore, B. C. J. (**1983**), "Gap detection as a function of frequency, bandwidth, and level," *J. Acoust. Soc. Am.* **74**(2), pp. 467–473. (Cited on page 76)

Shamma, S. A., Elhilali, M., and Micheyl, C. (**2011**), "Temporal coherence and attention in auditory scene analysis," *Trends Neurosci* **34**(3), pp. 114–23. (Cited on page 22)

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**), "Speech recognition with primarily temporal cues," *Science* **270**(5234), pp. 303–4. (Cited on pages 8 and 33)

Shinn-Cunningham, B. G., Kopco, N., and Martin, T. J. (**2005**), "Localizing nearby sound sources in a classroom: Binaural room impulse responses," *J. Acoust. Soc. Am.* **117**(5), pp. 3100–15. (Cited on page 12)

Siveke, I., Ewert, S. D., Grothe, B., and Wiegrebe, L. (**2008**), "Psychophysical and physiological evidence for fast binaural processing," *J Neurosci* **28**(9), pp. 2043–52. (Cited on pages 7, 32, 76, and 86)

Stecker, G. C. and Brown, A. D. (**2012**), "Onset- and offset-specific effects in interaural level difference discrimination," *J. Acoust. Soc. Am.* **132**(3), pp. 1573–1580. (Cited on page 22)

Stecker, G. C., Ostreicher, J. D., and Brown, A. D. (**2013**), "Temporal weighting functions for interaural time and level differences. III. Temporal weighting for lateral position judgments," *J. Acoust. Soc. Am.* **134**(2), pp. 1242–1252. (Cited on pages 59 and 80)

Stern, R. M., Zeiberg, A. S., and Trahiotis, C. (**1988**), "Lateralization of complex binaural stimuli: A weighted-image model," *J. Acoust. Soc. Am.* **84**(1), pp. 156–165. (Cited on pages 78 and 79)

Thompson, E. R. and Dau, T. (**2008**), "Binaural processing of modulated interaural level differences," *J. Acoust. Soc. Am.* **123**(2), pp. 1017–1029. (Cited on pages 8, 34, 53, 54, and 83)

Thurlow, W. R. and Jack, C. E. (**1973**), "Some determinants of localization-adaptation effects for successive auditory stimuli," *J. Acoust. Soc. Am.* **53**(6), pp. 1573–1577. (Cited on pages 6, 12, and 22)

Tollin, D. J. and Yin, T. C. (**2005**), "Interaural phase and level difference sensitivity in low-frequency neurons in the lateral superior olive," *J Neurosci* **25**(46), pp. 10648–57. (Cited on pages 24 and 28)

Trahiotis, C. and Bernstein, L. R. (**1990**), "Detectability of interaural delays over select spectral regions: Effects of flanking noise," *J. Acoust. Soc. Am.* **87**(2), pp. 810–813. (Cited on page 13)

Viemeister, N. F. (**1979**), "Temporal modulation transfer functions based upon modulation thresholds," *J. Acoust. Soc. Am.* **66**(5), pp. 1364–1380. (Cited on pages 7 and 33)

Wagner, H. (**1991**), "A temporal window for lateralization of interaural time difference by barn owls," *J Comp Physiol A* **169**(3), pp. 281–9. (Cited on pages 8, 32, 59, and 60)

Warren, R. M., Wrightson, J. M., and Puretz, J. (**1988**), "Illusory continuity of tonal and infratonal periodic sounds," *J. Acoust. Soc. Am.* **84**(4), pp. 1338–1342. (Cited on pages 54 and 77)

Whitmer, W. M., Seeber, B. U., and Akeroyd, M. A. (**2012**), "Apparent auditory source width insensitivity in older hearing-impaired individuals," *J. Acoust. Soc. Am.* **132**(1), pp. 369–379. (Cited on page 86)

Wightman, F. L. and Kistler, D. J. (**1992**), "The dominant role of low-frequency interaural time differences in sound localization," *J Acoust Soc Am* **91**(3), pp. 1648–61. (Cited on page 12)

Wilcox, R. R. (**2012**), *Introduction to robust estimation and hypothesis testing*, Statistical modeling and decision science, Academic Press, Amsterdam ; Boston, 3rd ed. (Cited on pages 37, 38, and 71)

Acknowledgments

I must first thank Steven van de Par for his tremendous guidance during this doctoral endeavor. As both a mentor and a great friend, your support was essential to the success of my work. I am also indebted to Tobias May and Björn Menke for helping to provide a smooth stay in Oldenburg, even though many of our fun evenings together resulted in suboptimal mornings. A big thank you is owed to all of the Akus who provided a friendly atmosphere at work and on group excursions. Thank you to Stefan Klockgether for translating the abstract of this thesis into German since my German abilities are sadly still terrible. I also owe great thanks to the people that gave me different opportunities to pursue my passion in audio: Jason Ringelstetter and Chris Parks at the Tonic Room Recording and Mastering Studios, Dave Jensen at the Morrison Center, Juha Merimaa and David Romblom at Sennheiser, and the whole family of friends at HeadRoom. A particular thank you is necessary for Juha since his words encouraged me to pursue my PhD and to strongly consider this opportunity under Steven in Oldenburg. Thank you to Rob Maher for giving this meager student from an AES conference the first opportunity to scientifically pursue the world of audio with which I find myself obsessed. Finally, I give utmost thanks to the loving people that have carried me throughout this journey. Without my parents (Kenton and Sandra Reed), none of what I have achieved in life would be possible. Without my beautiful girlfriend (Erika Iwamura), I would have gone completely mad while pursuing my degree... this time would have been terribly less fun without the spontaneous life we have lived together.

Curriculum Vitae

Personal:	
name	Darrin Kiyoshi Reed
date of birth	November 17th, 1982
place of birth	Nampa, Idaho USA

Education:	
2011 - 2014	**Ph.D. candidate** at the University of Oldenburg, Germany
2008 - 2009	**Master of Science (M.Sc.)** in Electrical Engineering, Montana State University, USA - Thesis: *Virtual audio localization with simulated early reflections and generalized head-related transfer functions* (Advisor: Prof. Dr. Rob C. Maher)
2001 - 2007	**Bachelor of Science (B.Sc)** in Electrical Engineering, Boise State University, USA

Declaration

I have completed the work independently and used only the indicated facilities. This dissertation is my own work. All sources of information have been acknowledged by means of references.

This dissertation has neither as a whole nor in part been published or submitted to assessment in a doctoral procedure at another university.

Oldenburg, den 30. December 2014

....
(Darrin K. Reed)