Characterizing binaural processing of amplitude modulated sounds

Ph.D. thesis by

Eric R. Thompson

Technical University of Denmark
2009

Printed in Denmark.

Contents

List of Figures

Preface

What an adventure it has been! After five years in Denmark, I have been through some amazing experiences and some amazingly frustrating experiences. From welcoming our son into the world, to having my wife get deported, to making friends from all over the world, to fighting with the kids' schools over admissions, it is surprising that there was any time to get any real work done. Thanks largely to the support that I have received from family and friends, some work did get done, and this thesis is the result. As with any research, more questions were opened than were answered, so I hope that this will be the starting point for many future projects.

I would like to begin by acknowledging and thanking my advisor, Torsten Dau. The passion that he exudes for his research and his engagement with his students played a big role in my deciding to pursue this degree. You have built an incredible group here in a short time, and I am proud to have been a part of it. I hope that we will continue to collaborate, at least informally, in the future.

A big "tak" to Torben Poulsen for his help with translating the abstract. Thanks go to the rest of the staff and students at CAHR for making a great environment for working and playing. I've enjoyed the scientific discussions with all of you as well as the social events, sometimes even combined as one. I look forward to crossing paths with you again.

I had a long and productive external research stay at Boston University, and would like to thank Barbara Shinn-Cunningham and the rest of the Hearing Research Center, particularly the binaural gang, for hosting me. We will be working together as I continue my career in Boston. Along with this trip, I must thank the Denmark-America Fund and the Idella Foundation for providing grants to help with some of the costs associated with this trip.

Thank you to Gerald Kidd and Chris Mason for their patience and support while I finished up the writing of this thesis. I look forward to our future work together.

I also want to thank our parents, Robert and Karen Thompson and Richard and Deborah Conklin, for their loving support and financial support. Without your help, none of this would have been possible. Thank you for your faith in me.

Finally, I would like to thank my wife, Amy, and my children, Emma, Mia and Aidan, who have given me their love and support, and provided me with wonderful distractions to make sure that I never lost sight of what is truly important. You have all sacrificed a lot for me during these past years, and it is to you that I dedicate this thesis.

Eric R. Thompson October 30, 2009

Abstract

Two important aspects of human hearing are the ability to process fluctuations in level, or amplitude modulation processing, and the ability to use differences in the acoustic signals arriving at the two ears, or binaural processing. In this thesis, psychoacoustic experiments related to interactions of binaural hearing and amplitude modulation processing are presented.

If there is a modulation phase difference between the signals in each ear, then there is a fluctuating interaural level difference, which can be used as a cue for signal detection. The minimum modulation depth required to discriminate between interaurally homophasic and antiphasic amplitude modulation was measured as a function of modulation frequency using tonal, interaurally correlated noise, and interaurally uncorrelated noise carriers. The listeners were sensitive to the interaural modulation phase at all measured modulation frequencies from 2 to 128 Hz. In the presence of a modulation masker, the listeners' modulation detection thresholds showed bandpass shapes as a function of the masker modulation frequency.

It is often easier to detect a signal when a masker is perceived to come from a different spatial location than when masker and signal are perceived at the same location. In order to test whether this also applies to the detection of amplitude modulation, masked modulation detection thresholds were measured using separate carriers for the masker and target modulation, where the interaural time difference of the two carriers could be controlled independently. The data showed that the listeners could hear the masker and target coming from different lateral positions, but this had no effect on the modulation detection thresholds.

Reverberation in a room filters the amplitude modulations of an input signal, reducing the modulation depth and changing the modulation phase. Differences in the reverberation paths from a source to a listener's two ears can create different

filters with an interaural modulation phase difference at some modulation frequencies. Monaural and binaural modulation detection thresholds were measured in different simulated reverberant environments. The data showed that there can be a binaural advantage in modulation detection at modulation frequencies at which there is an interaural modulation phase difference.

The effect of an interaural modulation phase difference on consonant identification in reverberant environments was investigated by comparing consonant confusions made in monaural and binaural listening conditions. The data showed a consistent binaural advantage with three impulse responses. The details of which consonants showed the largest improvement from monaural to binaural listening may help show what modulation frequency ranges are important for which consonants or groups of consonants.

The results will be used in the development and testing of models of binaural processing and perception. Such models may be useful for predicting speech intelligibility in rooms, for preprocessing a signal for an automatic speech recognizer, and as a test bed for new hearing appliance algorithms.

Resumé

Hørelsen besidder to væsentlige egenskaber. Den ene er evnen til at kunne opfatte ændringer i lydens niveau dvs. at kunne opfatte amplitude modulation. Den anden er evnen til at udnytte forskelle i de lyde der ankommer til de to ører, dvs. binaural hørelse. Denne rapport drejer sig om psykoakustiske undersøgelser af sammenhængen mellem binaural hørelse og opfattelsen af amplitudemodulation.

Hvis der er en faseforskel i modulationen ved de to ører (dvs. interaural forskel), vil der være en fluktuerende niveauforskel mellem ørerne som personen kan udnytte i forbindelse med detektion. I denne undersøgelse er den mindste modulationsdybde som er nødvendig for at kunne skelne mellem amplitudemodulation i medfase og modfase blevet bestemt som funktion af modulationsfrekvensen. Undersøgelsen er gennemført med toner, med interaural korreleret støj og med interaural ukorreleret støj. Testpersonerne kunne høre interaural modulationsfase for alle målte modulationsfrekvenser fra 2 Hz til 128 Hz. Hvis der blev anvendt et maskerende modulationssignal, fik lytternes modulations detektionstærskel båndpaskarakter hvor båndpasområdet fulgte det maskerende signals modulationsfrekvens.

Det er ofte nemmere at detektere et signal når den maskerende lyd (maskeren) opleves at komme fra et andet sted end hvis signal og masker kommer fra samme position. For at undersøge om dette også gælder for detektion af amplitudemodulation, blev modulations detektionstærsklen bestemt under anvendelse af separate lyde for masker og signal. Den interaurale tidsforskel mellem masker og signal kunne varieres uafhængigt af andre parametre. Resultaterne viste at lytterne kunne høre at masker og signal kom fra forskellige retninger i horisontalplanet, men det havde ingen indflydelse på modulations detektionstærsklen.

Efterklang i et rum ændrer amplitudemodulationen af et input signal idet modulationsdybden reduceres og modulationsfasen ændres. Forskelle i lydudbredelsen

fra en kilde til en lytters to ører kan bevirke at lyden optræder med en interaural modulationsfaseforskel ved visse modulationsfrekvenser. På denne baggrund blev der målt monaurale og binaurale modulationsdetektionstærskler i simulerede efterklangsomgivelser. Resultaterne viste at der kan være en binaural forbedring i modulations detektionstærsklen ved modulationsfrekvenser hvor der er en interaural modulationsfaseforskel.

Indflydelsen af en interaural modulationsfaseforskel på identifikation af konsonanter i efterklangsomgivelser blev undersøgt ved at sammenligne konsonantforvekslinger fra monaural og binaural lytning. Resultaterne viste en konsistent binaural fordel for tre forskellige impulsresponser. En analyse af hvilke konsonanter der viste den største forbedring fra monaural til binaural lytning kan belyse hvilke modulationsfrekvensområder der er vigtige for de enkelte konsonanter eller for grupper af konsonanter.

Resultaterne vil blive anvendt til udvikling og test af modeller for binaural lydopfattelse. Sådanne modeller er nyttige for forudsigelse af taleforståelighed i rum, for forbehandling af signaler til talegenkendelses procedurer og ved afprøvning af nye algoritmer til anvendelse i høreapparater.

List of acronyms and symbols

n-AFC	n-alternative forced choice
AM	Amplitude modulation
AM_m	Monaural amplitude modulation
AM_0	Diotic amplitude modulation
AM_π	Interaurally antiphasic amplitude modulation
ANOVA	Analysis of variance
B	Binaural presentation mode
BMLD	Binaural masking level difference
BRIR	Binaural room impulse response
CV	Consonant-Vowel
CVC	Consonant-Vowel-Consonant
D_{50}	Deutlichkeit (definition)
DC	Direct current
EC	Equalization-Cancellation
EI	Excitation-Inhibition
ERD	Equivalent rectangular duration
f_m	Modulation frequency
IACC	Interaural cross-correlation
ID	Identification
IID	Interaural intensity difference
ILD	Interaural level difference
IMPD	Interaural modulation phase difference

IM	Intensity modulation
IPD	Interaural phase difference
IR	Impulse response
ITD	Interaural time difference
L	Left ear, monaural
m	Amplitude modulation depth
MDI	Modulation detection interference
MTF	Modulation transfer function
N_0	Diotic noise
N_u	Interaurally uncorrelated noise
R	Right ear, monaural
RASTI	Rapid speech transmission index
RM-ANOVA	Repeated measures analysis of variance
S_0	Diotic signal
S_π	Interaurally antiphasic signal
SII	Speech intelligibility index
SNR	Signal-to-noise ratio
SPL	Sound pressure level
STI	Speech transmission index
T_{30}, T_{60}	Reverberation time
TMTF	Temporal modulation transfer function
U_{50}	Useful to detrimental ratio
VC	Vowel-Consonant
VCV	Vowel-Consonant-Vowel

Chapter 1

General Introduction

All of the sounds that humans experience in their daily routine fluctuate in level at various rates over time. Those level fluctuations often communicate information about the sound source to the listener. For example, as the mouth of a speaker opens and closes, shaping the speech sounds, the level of the signal will rise and fall, correspondingly. Those level fluctuations, along with other acoustic cues, e. g., pitch fluctuations, carry information about the speech signal, like the timing of syllable breaks and what consonant sounds were spoken. The human ability to hear those temporal level fluctuations is limited, and there have been many studies in the past that have investigated these limitations on human performance.

Some of the previous studies investigating this temporal resolution of the auditory system have looked at the ability to detect a silent gap introduced into a broadband noise or tone (e. g., Plomp, 1964; Formby and Muir, 1988; Moore et al., 1993). Others have investigated the ability to detect sinusoidal amplitude modulation (AM) imposed on broadband-noise or tonal carriers (e. g., Viemeister, 1979; Formby and Muir, 1988; Kohlrausch et al., 2000), or on narrowband noise carriers (e. g., Fleischer, 1982; Dau et al., 1997a) as a function of the modulation frequency. These studies, which were all based on monaural (or diotic) listening in anechoic conditions, generally showed a limited temporal resolution in that temporal gaps of less than about 3 ms cannot be detected, and sensitivity to AM decreases with increasing modulation frequency above about 150 Hz.

Binaural listening offers many advantages to the human listener. For example, the spatial location of a sound source, relative to the listener, can often be determined when listening with two ears, and there can be an improvement in signal detection

thresholds when a signal and a masking sound have different interaural parameters. These interaural parameters can be described by interaural timing or phase differences, interaural level or intensity differences, and in terms of the interaural cross-correlation. The temporal acuity of binaural listening has also been investigated with analogs of gap detection (Akeroyd and Summerfield, 1999), in which the interaural correlation of a sound was briefly changed, and with modulation detection using modulated ITDs (Grantham and Wightman, 1978), modulated IIDs (Grantham, 1984) and modulated interaural correlation (Grantham, 1982). In general, these studies found that the minimum binaural 'gap' duration required for detection was longer than the minimum monaural gap, and that the sensitivity to modulated interaural cues was reduced at lower modulation rates than in the monaural AM studies.

Previous researchers have also noted that a room sounds less reverberant when listening binaurally than when listening monaurally (e. g., Koenig, 1950), and that there is a significant binaural advantage in speech intelligibility in reverberant environments (e. g., Moncur and Dirks, 1967; Nábělek and Robinson, 1982). However, the two most commonly used models for predicting speech intelligibility in rooms, the speech intelligibility index (SII) and the speech transmission index (STI) are based on monaural listening. These models may underestimate speech intelligibility in situations where there is a significant binaural advantage in understanding speech. Knowledge about how a human listener uses binaural listening to improve on speech intelligibility in complex listening environments should be integrated into the models in order to improve their predictions. Some of the humans' binaural advantage comes from the ability to selectively listen to the sound in the ear with the best signal-to-noise ratio, or 'better-ear listening', but there can also be new cues created through a combination of the two ears' signals in the brain that may provide information about the input sound that would not be available when listening monaurally. In this thesis, binaural cues that may be used to improve on speech intelligibility are investigated to determine several basic psychophysical thresholds, as well as their potential use in a speech intelligibility task and applicability to the STI model.

The STI assumes that level fluctuations must be transmitted successfully to the listener in order for speech to be intelligible. It is calculated based on the degree of attenuation of sinusoidal intensity modulations when a modulated signal is transmitted

through a room, as a function of the modulation frequency in several audio-frequency bands (IEC 60268-16, 2003). Reverberation and noise in a room will cause a reduction in the modulation depth of a transmitted signal, effectively filling in the dips in the signal's intensity envelope and reducing the intelligibility of a transmitted speech signal.

The reduction in the modulation depth of a signal transmitted through a room as a function of the modulation frequency is described by the modulation transfer function. This function is specific to a particular source and receiver position in the room, and can be measured *in situ*, or calculated from the impulse response of the room. For a given modulation frequency, the modulation transfer function has an attenuation, usually expressed in dB, and a phase shift. Both of these parameters can be different in the two ears for a given source because of the different locations of the two ears and the presence of the head between the ears. When there is an interaural modulation phase and/or attenuation difference, a fluctuating interaural intensity difference results. This interaural intensity fluctuation may provide a binaural cue that results in improved modulation detection thresholds as compared to monaural listening in the same environment. This binaural cue may also be a part of the improvement in speech intelligibility when listening binaurally in reverberant environments.

The purpose of the experiments presented in this thesis was to investigate interactions between amplitude (or intensity) modulation processing and binaural listening in the human auditory system. More specifically, the purpose was to test the hypothesis that interaural level fluctuations can be used to enhance modulation detection thresholds and speech intelligibility in complex listening environments. The results of these investigations will help to further develop binaural models, and may help to improve on speech intelligibility predictions in reverberant rooms using binaural listening. The experiments are presented in the ensuing chapters as described in the following paragraphs.

First, the baseline abilities of the auditory system to detect fluctuating interaural level differences were investigated. **Chapter 2** presents psychoacoustic experiments measuring the listeners' sensitivity to interaural level fluctuations created with interaural modulation phase differences. An interaural level difference can give a perception of a lateralization of the sound towards the ear with the higher level. By

modulating the interaural level difference of a sound, it can appear to move back and forth between the ears with slow modulations, or to have a broad, diffuse sound image with faster modulations. The experiments presented in this chapter were designed to measure the temporal acuity of the auditory system in processing interaural level fluctuations, and to investigate modulation-frequency tuning in the processing of interaural level fluctuations. The results of these experiments should demonstrate situations in which the thresholds for modulation detection are significantly better when there is an interaural modulation phase difference than would be expected from 'better-ear listening'. The modulation-frequency tuning experiment shows whether two sounds can be perceptually segregated based on their respective rates of interaural fluctuations. This may be relevant for predicting speech intelligibility in fluctuating background noise in reverberant environments.

When two sound sources are spatially separated, it can be easier to hear the details of each sound signal than when they are colocated. **Chapter 3** discusses the effects of a perceived spatial separation of a masker carrier and a target carrier on the ability to detect amplitude modulation imposed on the target when a different amplitude modulation is imposed on the masker. The first experiment presents data on the perceived lateralization of two temporally-interleaved carriers as a function of the interaural timing difference of each carrier. This verifies whether the two carriers can be robustly perceived as coming from separate spatial locations. The second experiment measured amplitude modulation detection thresholds, using the same carriers used in the lateralization experiment, as a function of the masker modulation frequency content and the interaural timing difference of the masker carrier. This shows whether a perceived spatial separation of a target and masker is enough to reduce the effect of the masker on the detectability of envelope fluctuations on the target.

Chapter 4 presents an investigation into the effect of interaural modulation phase differences that result from dichotic (i. e., different in the two ears) impulse responses on sinusoidal intensity modulation detection thresholds. Modulation detection thresholds were measured both monaurally and binaurally with three impulse responses. This experiment was a first extension of the experiments presented in Chap. 2 to realistic listening environments, and to test the hypothesis

that interaural modulation phase differences that result from dichotic reverberation can create interaural level fluctuations that enhance modulation detection thresholds over monaural modulation detection thresholds.

In **Chap. 5**, the results of an experiment measuring consonant identifications in reverberant environments are presented. Vowel-consonant-vowel speech tokens were convolved with dichotic impulse responses and presented both monaurally and binaurally. Consonant confusion matrices were generated from the responses and compared for the monaural-left, monaural-right and binaural listening conditions. This experiment was designed to test the hypothesis that reverberant environments that create interaural modulation phase differences will also show a binaural advantage in a speech-related task.

Finally, **Chap. 6** presents a general discussion of the results presented in this thesis, including some implications of the experimental results for models of binaural processing and perception.

Chapters 2-5 are written as journal articles, and have either been already published, submitted, or are planned to be submitted shortly after submission of this thesis. Each chapter is written to be read separately. Therefore, there can be repetitions of background information between the chapters.

Chapter 2

Binaural processing of modulated interaural level differences

This chapter was originally published as Thompson and Dau (2008)

Abstract

Two experiments are presented that measure the acuity of binaural processing of modulated interaural level differences (ILDs) using psychoacoustic methods. In both experiments, dynamic ILDs were created by imposing an interaurally antiphasic sinusoidal amplitude modulation (AM) signal on high-frequency carriers, which were presented over headphones. In the first experiment, the sensitivity to dynamic ILDs was measured as a function of the modulation frequency using pure-tone, and interaurally correlated and uncorrelated narrowband noise carriers. The intrinsic interaural level fluctuations of the uncorrelated noise carriers raised the ILD modulation detection thresholds with respect to the pure-tone carriers. The diotic fluctuations of the correlated noise carriers also caused a small increase in the thresholds over the pure-tone carriers, particularly with low ILD modulation frequencies. The second experiment investigated the modulation frequency selectivity in dynamic ILD processing by imposing an interaurally uncorrelated band-pass noise AM masker in series with the interaurally antiphasic AM signal on a pure-tone carrier. By varying the masker center frequencies relative to the signal modulation frequency, broadly tuned, band-pass-shaped patterns were obtained. Simulations with an existing

binaural model show that a low-pass filter to limit the binaural temporal resolution is not sufficient to predict the results of the experiments.

2.1 Introduction

Information in sound signals is carried not only by the fine structure of the sound, but also by the intensity fluctuations of its envelope. In a reverberant environment, reflections can reduce the depth of those envelope fluctuations and can change their phase. The effective amount of envelope modulation and modulation phase transmitted to a receiver can be derived from the source-receiver impulse response as a function of the modulation frequency (Schroeder, 1981). This complex modulation transfer function (MTF) shows the modulation attenuation and phase shift as a function of modulation frequency for the particular source-receiver transmission path. A normal human auditory system has two working ears, thereby receiving information from a given source via two transmission paths and through two MTFs. Interaural differences in the modulation phase and/or depth can create fluctuating interaural level differences (ILDs) and interaural time differences (ITDs). In order to understand how ILD fluctuations are perceived and to begin to understand the binaural processing of envelopes in reverberation, artificial stimuli were generated in the present study with sinusoidal amplitude modulation and a controlled interaural modulation phase difference. The stimuli were presented over headphones to listeners in psychoacoustic tests.

An ILD is usually perceived as a lateralization of the sound source toward the ear with the higher intensity sound. When an ILD changes slowly, the sound is perceived to move, while more rapid ILD fluctuations are usually perceived as a stationary sound source with a broad or diffuse sound image (e. g., Blauert, 1972; Grantham, 1984; Griesinger, 1997). This is analogous to the ability of the auditory system to follow slow intensity fluctuations in monaural or diotic stimuli, and the perception of roughness with more rapid fluctuations (e. g., Terhardt, 1968).

Dynamic ILDs can be created by imposing amplitude modulation with an interaural modulation phase difference. However, a static interaural modulation phase difference can also be interpreted as a static envelope ITD, corresponding to the phase

difference divided by the angular modulation frequency. High-frequency sounds, which cannot be lateralized based on the ITD of their fine structure (e. g., Klumpp and Eady, 1956; Mills, 1960), can be lateralized based on the ITD of their envelopes (e. g., Klumpp and Eady, 1956; Henning, 1974; Nuetzel and Hafter, 1981; Bernstein and Trahiotis, 1994). A static modulation phase difference could then create a percept of a sound lateralized toward the leading ear instead of creating a moving or diffuse sound image, depending on the envelope ITD. However, with a phase difference of π, as was used in the present study, it is unclear which ear should be leading because of the temporal symmetry of the sinusoid. For complex sounds with random interaural level fluctuations, those fluctuations may actually be encoded internally as a combination of time-varying ITDs and ILDs. In situations with ambiguous localization cues, such as with a π interaural phase difference, onset cues may dominate localization of the ongoing signal (Buell et al., 1991).

The temporal acuity of the auditory system is often measured by determining the threshold of detection of the sinusoidal modulation of a physical parameter as a function of the modulation frequency, referred to as a 'temporal modulation transfer function' (TMTF). For example, the TMTF with diotic amplitude modulation (AM) was measured by Viemeister (1979) with broadband noise carriers, by Fleischer (1982) and Dau et al. (1997a) with narrowband noise carriers, and by Kohlrausch et al. (2000) with pure-tone carriers. Other studies have investigated the temporal acuity to dynamic interaural parameters, such as interaural time or phase differences (e. g., Grantham and Wightman, 1978; Witton et al., 2000) and interaural correlation (e. g., Grantham, 1982). Grantham (1984), Grantham and Bacon (1991) and Stellmack et al. (2005) measured the acuity of the binaural system in the detection of modulated ILDs, generated with interaurally antiphasic sinusoidal AM signals.

The TMTFs measured with pure-tone and diotic broadband noise carriers show a high sensitivity to slow AM, with a minimum modulation depth, m, required for detection of around 0.04 (often discussed on a dB scale as $20 \log_{10} m$, here -28 dB) for low-frequency modulations. As the modulation rate increases, larger modulation depths are required for detection, thereby exhibiting an overall low-pass characteristic (e. g., Viemeister, 1979; Kohlrausch et al., 2000). However, the thresholds measured with narrowband noise carriers can exhibit high-pass as well as

low-pass characteristics, depending on the bandwidth of the noise (Fleischer, 1982; Dau et al., 1997a). This led Dau and colleagues to propose a modulation filterbank model with band-pass filters acting on the envelope of a stimulus (Dau et al., 1997a,b), which can simulate AM detection performance with narrowband as well as broadband noise. Bacon and Grantham (1989), Houtgast (1989) and Ewert et al. (2002) made more direct measurements of modulation-frequency selectivity by measuring AM detection thresholds in the presence of a noise AM masker. These measurements also showed a band-pass characteristic with approximately constant filter bandwidth relative to the filter center frequency (constant Q-value).

Grantham (1984) also reported a low-pass shape in his ILD modulation detection thresholds. Those data were obtained through measurements of the threshold of *discriminability* of interaurally antiphasic AM from homophasic AM imposed on interaurally uncorrelated, band-pass noise carriers. In contrast to the diotic TMTFs described above, the modulation depths required to discriminate the ILD modulation with low-frequency AM were quite high, around $m = 0.15$ (-16 dB). In another study, Grantham and Bacon (1991) measured monaural and ILD modulation *detection* thresholds with broadband noise carriers and unmodulated reference intervals. Their monaural AM detection thresholds were very similar to those from Viemeister (1979) with thresholds of around -28 dB for low modulation frequencies. The ILD modulation detection thresholds were almost identical to the monaural thresholds, thereby showing 12 dB greater sensitivity to the modulation than reported by Grantham (1984). This increase in performance can be attributed to the difference in paradigm (AM detection vs. discrimination). Since relatively small AM depths can be detected monaurally, characterization of binaural processing of modulated ILDs should only be done with the elimination of the monaural AM cues through an AM discrimination paradigm (as done by Grantham, 1984).

Grantham and Bacon (1991) also measured monaural and binaural frequency tuning in the envelope domain by measuring the TMTF in the presence of an AM masker. One diotic broadband-noise carrier, with a diotic tonal or narrowband-noise amplitude modulator (the masker), was added to a second diotic broadband-noise carrier with an interaurally antiphasic sinusoidal amplitude modulator (the signal). They reported a bandpass tuning in the masked detection thresholds, but could not

conclude whether that tuning was due to monaural or binaural processing. Grantham (1984) described diotic AM as creating an 'up-and-down flutter' with perceived changes in level or roughness, and antiphasic AM as creating a 'side-to-side flutter' with a perception of motion or broadening between the ears. Assuming that the detection of the signal interval in their 1991 study was based on a comparison of the perceived motion or width of the two presentation intervals, a *diotic* masker, with no interaural fluctuations itself, would be perceived as motionless and narrow, and should have had little effect on the detectability of the modulated ILD signal. A *dichotic* masker, which does generate interaural fluctuations, would be better to measure the masked sensitivity to ILD modulations, along with a task of discriminating between interaurally antiphasic and homophasic AM (AM_π-AM_0), where the monaural cues have been made ambiguous. In this way, the results and any modulation frequency tuning could be attributed to purely binaural processing.

Stellmack et al. (2005) measured the sensitivity to ILD modulations using high-frequency (5 kHz) pure-tone and narrowband noise carriers (30- and 300-Hz-wide, interaurally correlated and uncorrelated), and an AM_π-AM_0 discrimination task. The thresholds measured with the pure-tone carrier were approximately constant at about -20 dB up to about $f_m = 100$ Hz, where the sensitivity worsened with increasing f_m until the threshold could no longer be determined above $f_m = 500$ Hz. There was a small increase in thresholds (up to 7 dB with the 30-Hz-wide carrier) when using the correlated noise carriers, relative to the pure-tone carrier data, particularly with low modulation rates ($f_m < 20$ Hz). This increase was much smaller than the increase in thresholds seen with uncorrelated noise carriers, and was described as independent of the intrinsic carrier fluctuations. Therefore, the main focus of their paper was on the thresholds measured with uncorrelated narrowband noise carriers.

Narrowband Gaussian noises fluctuate randomly with envelope frequencies up to the bandwidth of the noise (see, e. g., Lawson and Uhlenbeck, 1950; Price, 1955). Monaurally, those inherent fluctuations can make it more difficult to detect an imposed AM, as compared to AM imposed on a pure-tone (i. e., flat envelope) carrier, especially when the AM frequency is less than the bandwidth of the noise carrier. This increase in threshold can be viewed as the result of masking of the signal AM by the intrinsic envelope fluctuations of the carrier. Binaurally, presenting interaurally uncorrelated

narrowband noises to each ear creates a dynamic ILD and the perception of a randomly moving or broad sound, depending on the bandwidth. The modulation spectrum of the ILD fluctuations is governed by the frequency content of the envelopes of the stimuli. The difference between the ILD modulation thresholds measured with uncorrelated and correlated noise carriers by Stellmack et al. (2005) also showed that the intrinsic ILD fluctuations from the uncorrelated carriers had the largest effect on thresholds for AM frequencies within the bandwidth of the noise. This suggests that there might be a frequency-selective mechanism in the processing of ILD fluctuations, similar to the monaural modulation AM processing from Dau et al. (1997a), but with broader frequency tuning.

The goal of the present study was to further investigate the modulation frequency tuning of the processing of ILD fluctuations. In the first experiment, detailed in section 2.3, the measurements of sensitivity to modulated ILDs from Stellmack et al. (2005) with narrowband noise carriers were repeated with an additional carrier bandwidth (3-Hz, added to the 30- and 300-Hz-wide carriers) and a lower modulation frequency range (2 to 128 Hz instead of 4 to 600 Hz). A 3-Hz-wide Gaussian noise carrier has intrinsic modulations that can easily be followed (as loudness fluctuations monaurally, or as motion interaurally), where the 30- and 300-Hz-wide carriers are perceived with more roughness or width from the higher intrinsic modulation frequencies. The addition of the 3-Hz-wide carrier also enabled a comparison with all three carrier bandwidths (3-, 31-, and 314-Hz-wide) used by Dau et al. (1997a). With the additional data from the present study, a different interpretation of the results than that of Stellmack *et al.* is proposed, which includes more emphasis on the threshold differences with diotic carriers.

Section 2.4 details a second experiment for directly measuring the modulation frequency tuning for ILD fluctuations, using experimental design elements from Bacon and Grantham (1989), Houtgast (1989), Ewert and Dau (2000) and Ewert et al. (2002). In addition, simulations were made with an existing binaural computational model (from Breebaart et al., 2001a), which was designed mainly for static interaural conditions, but includes a sliding integrator window (low-pass filter) to limit the temporal resolution. This enables it to predict some signal detection thresholds with dynamic interaural conditions (see Breebaart et al., 2001c). The simulations should

show whether an existing binaural model can predict similar thresholds to those of a human listener when used as an artificial observer.

2.2 General Methods

Two psychoacoustic experiments were performed in order to investigate the sensitivity of the binaural system to modulated ILDs. In both experiments, the listener's task was to discriminate between a stimulus with interaurally antiphasic AM (AM_π; subscript indicating the interaural modulation phase) and a stimulus with homophasic (diotic) AM (AM_0). The AM frequency and depth was the same in all stimulus intervals of a 3-interval, 3-alternative forced-choice trial, with only an interaural difference in modulation phase in the signal interval. The stimuli were defined as in Eq. 2.1 with carriers x_L and x_R (subscripts L and R for left and right ears, respectively):

$$\begin{array}{rl} \text{Left:} & [1 + m\sin(2\pi f_m t + \phi_L)]\, x_L(t) \\ \text{Right:} & [1 + m\sin(2\pi f_m t + \phi_R)]\, x_R(t) \end{array}, \tag{2.1}$$

where m is the modulation depth, f_m is the modulation frequency, and $\phi_{L/R}$ is the initial modulation phase for the respective ear's stimulus. The reference intervals were defined by Eq. 2.1 with $\phi_R = \phi_L$ (AM_0) and the signal interval was defined with $\phi_R = \phi_L + \pi$ (AM_π). The instantaneous ILD of a stimulus is defined as the ratio of the envelopes (as in Stellmack et al., 2005):

$$\mathrm{ILD}(t) = 20\log_{10}\left(\frac{E_L(t)}{E_R(t)}\right). \tag{2.2}$$

A diotic sound does not create any ILDs itself. Therefore, the instantaneous ILD with an AM_π signal imposed on a diotic carrier is simply the ratio of the modulators:

$$\mathrm{ILD}(t) = 20\log_{10}\left(\frac{1 + m\sin(2\pi f_m t + \phi_L)}{1 - m\sin(2\pi f_m t + \phi_L)}\right) \tag{2.3}$$

and the maximum ILD is a function of the modulation depth only:

$$\mathrm{ILD}_{max} = 20 \log_{10} \left(\frac{1+m}{1-m} \right). \tag{2.4}$$

Interaurally uncorrelated noises produce stochastic ILD fluctuations, which add linearly (on a dB-scale) to the deterministic signal ILD modulation. These random ILD fluctuations will change the distribution of ILDs and the maximum ILD of the stimulus.

The two experiments differed in specifics of the stimuli (e. g., modulation phase and carrier), which will be presented in sections 2.3 and 2.4, but the general methods were the same.

2.2.1 Test subjects

Four test subjects were used for all experiments. They were not paid directly for their participation, but were all involved at the research center, and included the author of the present work. All had pure-tone audiometric thresholds of $15\,\mathrm{dB\,HL}$ or better for octave frequencies between $250\,\mathrm{Hz}$ and $8\,\mathrm{kHz}$. They were all experienced in psychoacoustic measurements and particularly in AM detection experiments. That experience was mostly with monaural or diotic stimuli, so their experience with the detection of interaural fluctuations was limited. All listeners were encouraged to listen to example stimuli and performed a limited set of training runs of approximately one hour duration.

2.2.2 Equipment

All signals were generated and presented using the AFC-Toolbox for MATLAB® (The MathWorks), developed at the University of Oldenburg, Germany and the Technical University of Denmark, at a sample rate of $44.1\,\mathrm{kHz}$ through a sound card (RME DIGI 96/8 PAD) and headphones (Sennheiser HD-580). The test subjects sat in a sound insulated booth with a computer monitor, which displayed instructions and visual feedback, and a keyboard for response input.

2.2.3 Procedure

A 3-interval, 3-alternative forced-choice (3-AFC) paradigm was used with an adaptive 1-up, 2-down tracking rule, which should converge at the 70.7% correct point on the psychometric function (Levitt, 1971). In a given track, the modulation frequency of the AM signal was fixed and the modulation depth was varied to find the AM depth required for identification of the signal. During each trial, a computer monitor displayed a window with three buttons, representing the three stimuli. Each button was highlighted when the corresponding interval was played. The signal interval was randomly selected with equal probability of occurrence from the three presentation intervals. The test subject responded via the computer keyboard and received immediate feedback on whether the response was correct or incorrect. All tracks were assembled in one long experiment that the test subject could start and stop at will after the completion of any track. The typical duration of a session was about 30 minutes, but could be longer or shorter depending on the circumstances.

Each track started with a modulation depth of -2 dB ($20 \log_{10} m$). The step size started at 4 dB, and was halved after every second reversal until the final step size of 1 dB was reached after the fourth reversal. The track continued for six further reversals with this step size, and the threshold was determined as the mean of those last six reversals. Each test subject completed four repetitions for each modulation frequency and set of experimental parameters, and the results shown are the mean and standard deviation of all test subjects and repetitions. If the test subject could not identify the correct interval with the maximum modulation depth ($m = 0$ dB) twice in a row, the track was skipped and the experiment continued to the next track.

2.2.4 Common stimulus parameters

All stimuli were centered at 5 kHz, so that all frequency components would lie well above the range of frequencies in which interaural timing differences in the fine structure of the carriers would affect the lateralization of the stimuli (e. g., Klumpp and Eady, 1956; Mills, 1960). The stimuli were gated simultaneously in the two ears, and presented at a level of 65 dB SPL, with 300 ms of silence between intervals. In

tracks where noise carriers were used, a new noise sample was generated for each presentation.

2.3 Experiment I: Modulation discrimination with narrowband noise carriers

The first experiment was designed to measure the sensitivity of binaural processing to modulated ILDs, using an experimental design based on the diotic AM detection measurements from Dau et al. (1997a). Parts of this experiment are a repetition of similar experiments performed by Stellmack et al. (2005).

2.3.1 Specific stimulus details

ILD modulations were created by applying an interaurally antiphasic sinusoidal AM to pure-tone and narrowband noise (3-, 30-, and 300-Hz-wide) carriers, centered at 5 kHz. In order to eliminate monaural modulation cues, an AM_π-AM_0 discrimination paradigm was used. The noise carriers were generated by creating a 1-s-long independent Gaussian noise sample for each interval in the time domain and setting the frequency components outside of the passband to zero in the spectral domain. Measurements were made with interaurally correlated (symbol N_0) and uncorrelated (N_u) noise carriers. Sinusoidal AM was applied to the carrier as given in Eq. 2.1 with $\phi_L = \phi_R = 0$ (AM_0) in the reference intervals and $\phi_L = 0$ and $\phi_R = \pi$ (AM_π) in the signal interval. With this choice of AM phase parameters, the change in modulation phase in the right ear could have been used as a monaural cue for signal detection. Therefore, a control experiment was performed to measure the modulation depth required for discrimination of monaural AM phase change (referred to as AM_m disc) using a pure-tone carrier and the right ear only. For these tracks, the signal interval had an initial modulation phase of π (negative-going zero-crossing) and the reference intervals started with a modulation phase of zero (positive-going zero-crossing).

Thresholds were measured with AM frequencies (f_m) in octave steps from 2 to 32 Hz and 128 Hz. At the highest modulation frequency used (128 Hz), an interaural modulation phase difference of π is equivalent to an ITD of ± 3.9 ms. Since this is well

above the ecologically relevant range of ITDs (approx. $650\,\mu$s for humans, Feddersen et al., 1957), and because the test subjects informally reported hearing a diffuse sound image, not a static lateralization of the sound, it is assumed that listeners did not use the static envelope ITD to localize the sound source. Therefore, the experiments will be discussed in terms of dynamic ILDs. Tracks with $f_m \geq 8\,\text{Hz}$ had a stimulus duration of $500\,\text{ms}$, including $50\,\text{ms}$ $\cos^2$ onset and offset ramps, while tracks with $f_m = 2$ or $4\,\text{Hz}$ had a duration of $1000\,\text{ms}$ in order to reduce the interference of the windowing function on the desired envelope frequency components (the stimulus windows of Stellmack et al., 2005 were $1\,\text{s}$ long with $150\,\text{ms}$ ramps). The intervals were separated by $300\,\text{ms}$ of silence, where the intervals of Stellmack et al. (2005) were only demarcated by the ramps with no additional separating silence. Previous measurements from Dau (1996) showed that listeners could not reliably discriminate (defined there as $P_c > 33\%$) between monaural AM phase with full modulation (i. e., $m = 1$) for $f_m > 12\,\text{Hz}$, using a 5-kHz pure-tone carrier. Therefore, the AM_m phase discrimination threshold was only measured in the present study with $f_m \leq 8\,\text{Hz}$. More recent results from Sheft and Yost (2007) showed that some listeners could only discriminate modulation starting phase with broadband noise carriers up to about $f_m = 12.5\,\text{Hz}$, while others were still able to perform the task up to around $50\,\text{Hz}$.

Two additional control measurements were made with an AM *detection* paradigm, where only the signal interval in a 3-interval trial had an applied AM and the two reference intervals were unmodulated. Monaural AM_m and interaural AM_π detection thresholds were measured with a 3-Hz-wide carrier in order to demonstrate the importance of eliminating monaural cues in the binaural experiment. Diotic noise carriers were used for this AM_π detection measurement.

2.3.2 Results

The results from the four test subjects were similar in shape and value, so the plots shown in Figs. 2.1-2.3 display the mean and standard deviation over all test subjects and runs. The modulation depths required for detection or discrimination of monaural AM and modulated ILDs are plotted in dB ($20\log_{10} m$) as a function of the signal modulation frequency for the pure-tone (Fig. 2.1b) and narrowband noise (Fig. 2.2) carriers. Note that the ordinates are shown with larger modulation depths and therefore

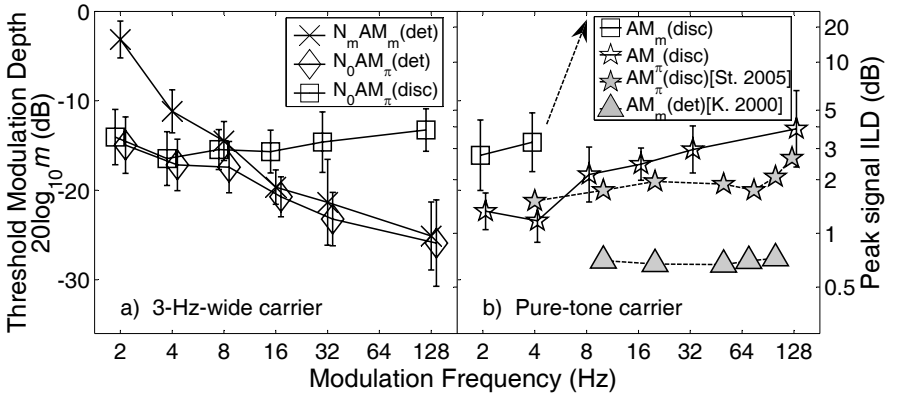

Figure 2.1: Amplitude modulation detection and discrimination thresholds in dB for various monaural and binaural conditions. The left ordinate labels (modulation depth) apply to all curves in both panels, and the right ordinate labels (peak ILD) applies to the binaural (AM_π) thresholds in both panels. Panel a: Thresholds measured with a 3-Hz-wide noise carrier. Monaural AM detection (N_mAM_m, X's), ILD modulation detection (N_0AM_π, diamonds), and ILD modulation discrimination (N_0AM_π-N_0AM_0, squares). Panel b: Thresholds measured with a pure-tone carrier for monaural AM phase discrimination (AM_m, squares) and ILD modulation discrimination (AM_π-AM_0, open stars). The data shown with shaded symbols are ILD modulation discrimination (shaded stars, adapted from Stellmack et al., 2005, Fig. 3), and monaural AM detection thresholds (shaded triangles, adapted from Kohlrausch et al., 2000, Fig. 2, 5 kHz carrier). Note that the data in panel a are offset around the AM frequency for visual clarity of the error bars.

poorer sensitivity toward the top of the axis. In Figs. 2.1 and 2.2, there is also a second ordinate on the right of each plot showing the peak *signal* ILD in dB, which applies to all of the AM_π data. This peak ILD was calculated from the modulation depth at threshold according to Eq. 2.4. As discussed above, the actual peak ILD seen with the uncorrelated noise carriers varied around this value. In the following, the results are presented in terms of the modulation depth at threshold in dB, unless otherwise noted. The data points are offset slightly from the AM frequency in Figs. 2.1a and b, and 2.2b and d for visual clarity of the error bars. A 2-way ANOVA with repeated measures was used to compare the data curves with a threshold of $p < 0.05$ required for significance.

Figure 2.1a shows the thresholds obtained for the monaural N_mAM_m detection (X's), N_0AM_π detection (diamonds) and N_0AM_π discrimination (squares) with a 3-Hz-wide carrier. The monaural curve shows the AM frequency selectivity described

by Dau et al. (1997a) in that the highest thresholds (-3 dB at $f_m = 2$ Hz) are at a frequency within the bandwidth of the noise (i. e., $f_m < 3$ Hz), and the thresholds for AM detection decrease with increasing f_m (down to -25 dB at $f_m = 128$ Hz). Discrimination of AM_π from AM_0 with the 3-Hz-wide carrier is approximately constant with thresholds between -16 and -13 dB for all measured f_m. This shows that while the low-frequency modulations ($f_m < 8$ Hz) are obscured by the fluctuations of the carrier in each ear (up-and-down), the ILD modulation (side-to-side) is still easily detectable with harmonic ILD oscillations with amplitudes of about 2.8 to 4 dB (see the right ordinate in Fig. 2.1a). The $\mathrm{N}_0\mathrm{AM}_\pi$ *detection* threshold demonstrates how the auditory system switches from monaural to binaural cues, depending on cue salience. For low AM frequencies ($f_m < 8$ Hz), where the monaural cues are obscured by the envelope fluctuations of the 3-Hz-wide carrier, there is no significant difference between the $\mathrm{N}_0\mathrm{AM}_\pi$ detection and discrimination thresholds. For $f_m > 8$ Hz, where the carrier fluctuations have a relatively small influence on the monaural detectability, the $\mathrm{N}_0\mathrm{AM}_\pi$ detection and $\mathrm{N}_m\mathrm{AM}_m$ thresholds have no significant differences.

In the control experiment to ensure that the test subjects could not perform the binaural discrimination tasks based solely on a monaural AM phase cue, thresholds for monaural AM phase discrimination could only be measured for $f_m \leq 4$ Hz (squares in Fig. 2.1b). At $f_m = 8$ Hz, the discrimination task could not reliably be performed by any of the test subjects, even at full modulation depth ($m = 0$ dB) and the arrow indicates that no threshold was measurable. These data correspond well to those from Dau (1996) and with some of the listeners from Sheft and Yost (2007), but can not help to explain how some listeners in the latter study were still able to discriminate modulation starting phase at rates up to about 50 Hz. The thresholds for monaural modulation phase discrimination in the present study for the 2-Hz and 4-Hz AM signals showed which of the results from the other measurements presented in this experiment (Figs. 2.1 and 2.2) could have been influenced by a monaural modulation phase cue. Those measurements were repeated informally with only the right ear's signal to verify that the tasks could not be performed monaurally at the measured threshold levels, and none of the listeners tested were able to do so.

The ILD modulation discrimination threshold curve measured with a pure-tone carrier (open stars in Fig. 2.1b) shows an overall low-pass tendency with thresholds around -23 dB (1.2 dB peak ILD) for $f_m = 2$ and 4 Hz and increasing to about -13 dB (4 dB peak ILD) for $f_m = 128$ Hz. Stellmack et al. (2005) reported a flatter threshold shape with thresholds around -20 dB (1.7 dB peak ILD) from $f_m = 4$ to almost 100 Hz (shaded stars in Fig. 2.1b), above which the threshold increased. The monaural TMTF with a pure-tone carrier (shaded triangles in Fig. 2.1b), reported by Kohlrausch et al. (2000), shows that the auditory system is much more sensitive to envelope fluctuations (thresholds around -28 dB up to $f_m > 100$ Hz) than to the ILD fluctuations caused by an interaural modulation phase inversion.

The data measured with narrowband noise carriers (3-, 30-, and 300-Hz-wide) are plotted twice in Fig. 2.2. The left panels show the data grouped by carrier bandwidth, and the right panels by carrier interaural correlation. In all panels, the squares represent data for the 3-Hz-wide carrier, the triangles for the 30-Hz-wide, and the circles for the 300-Hz-wide carrier data. Open symbols indicate interaurally correlated carriers (N_0), and shaded symbols are for interaurally uncorrelated carriers (N_u). In addition, the corresponding data from Stellmack et al. (2005) for the 30- and 300-Hz-wide carriers are shown in panels c and e, respectively, with diamond symbols and dashed lines, and the pure-tone (AM_π) thresholds are replotted from Fig. 2.1b in Figs. 2.2b and d with stars. The 30- and 300-Hz-wide carrier data show a good agreement with the data from Stellmack et al. (2005), with only small differences that could be the result of the procedural differences discussed above (e. g., stimulus length, windowing).

By grouping the thresholds by carrier bandwidth (Fig. 2.2a, c and e), a strong effect of the interaural carrier correlation can be seen. The N_uAM_π thresholds are much higher than the N_0AM_π thresholds. With the 3-Hz-wide carriers (Fig. 2.2a), the N_uAM_π thresholds show an almost constant offset of about 8 dB from the N_0AM_π curve. The differences between the thresholds measured with wider bandwidth carriers (Fig. 2.2c and e) increase with increasing modulation frequency from 7 to 12 dB with the 30-Hz-wide carriers and from 4 to 11 dB with the 300-Hz-wide carriers.

A 2-way ANOVA with repeated measures was calculated on the N_0AM_π data (Fig. 2.2b) with carrier bandwidth and modulation frequency as factors. The analysis

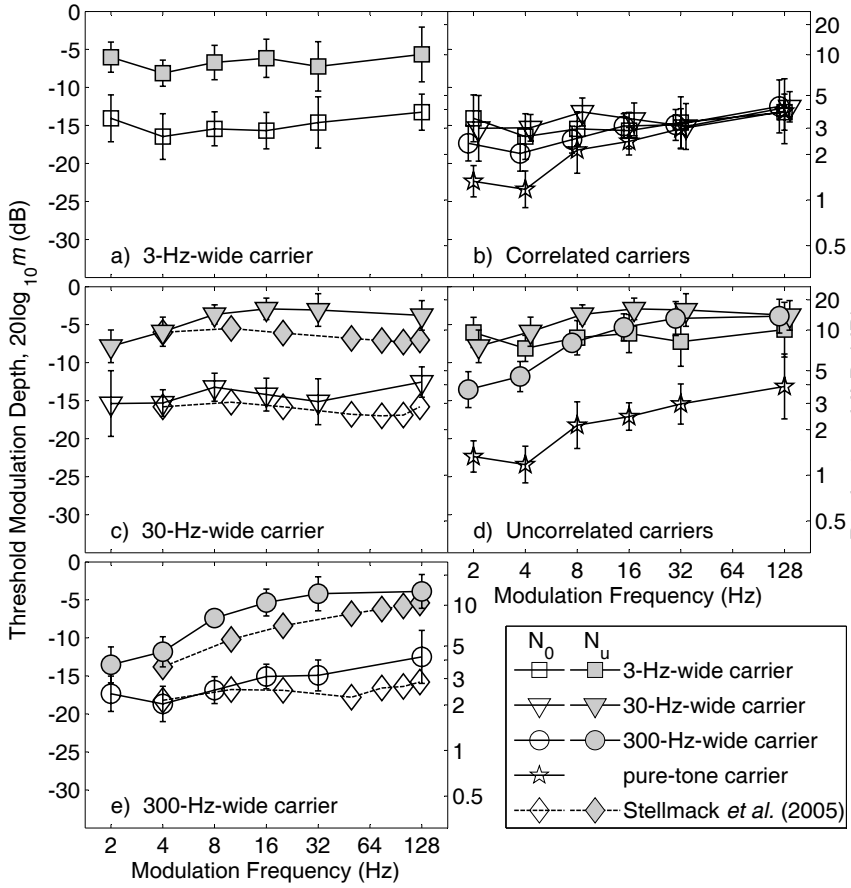

Figure 2.2: ILD modulation discrimination thresholds measured with narrowband noise carriers. The left and right ordinate labels apply to all curves in all panels, with the AM depth (left ordinate) converted to peak ILD in the right ordinate according to Eq. 2.4. In all panels, the interaural correlation of the carriers is indicated by the shading. Open symbols indicate correlated carriers and shaded symbols indicate uncorrelated carriers. The symbols indicate the carrier bandwidth: squares for 3-Hz-wide, triangles for 30-Hz-wide, and circles for 300-Hz-wide. The pure-tone thresholds from Fig. 2.1b are replotted with open star symbols in panels b and d. External data from Stellmack et al. (2005) are indicated by diamonds for both 30- and 300-Hz-wide data. The thresholds for the 3-, 30-, and 300-Hz-wide carriers are grouped by bandwidth in panels a, c and e, respectively. The data measured with correlated carriers are replotted in panel b, and with uncorrelated carriers in panel d. Note that the data in panels b and d are offset around the AM frequency for visual clarity of the error bars.

showed no significant effect of bandwidth ($p = 0.13$), but a significant effect of modulation frequency ($p < 0.05$) and a significant interaction between bandwidth and modulation frequency ($p < 0.05$). Adding the pure-tone AM_π data to the analysis as an additional bandwidth yielded a significant effect of bandwidth ($p < 0.01$). An ANOVA on the data measured with interaurally uncorrelated noise carriers (N_uAM_π; Fig. 2.2d) showed a significant effect of bandwidth, even without the pure-tone carrier data, and of modulation frequency, as well as a significant interaction between the factors ($p < 0.01$ for both factors and the interaction).

2.3.3 Discussion

The modulation depths required to discriminate AM_π from AM_0 imposed on diotic noise carriers were significantly larger than those required with pure-tone carriers, particularly with low modulation rates ($f_m < 16$ Hz). The discrimination thresholds measured with interaurally uncorrelated noise carriers were even higher than those measured with the correlated noise carriers. The increase in thresholds when using noise carriers instead of pure-tone carriers can be considered as masking of the modulation signal by the intrinsic envelope fluctuations of the noise carriers themselves, since a pure-tone carrier does not have these random fluctuations. Figure 2.3a shows the *difference* in thresholds measured with correlated noise carriers and the pure-tone carrier. Note that this reflects an increase in the threshold of ILD modulation discrimination as the result of *diotic* fluctuations of the noise carriers, which do not create ILD fluctuations themselves. Looking at the differences between the N_uAM_π thresholds and the pure-tone carrier thresholds (plotted in Fig. 2.3c), the 3- and 30-Hz-wide carriers show the greatest difference for f_m below the bandwidth of the carrier, beyond which the difference decreases monotonically toward an asymptotic value of about 7 to 9 dB. The threshold difference with the 300-Hz-wide carrier increases from about 8 dB with $f_m = 2$ Hz to 12 dB with $f_m = 16$ Hz, and then decreases again to 9 dB at $f_m = 128$ Hz.

Stellmack et al. (2005) also observed an increase in AM_π-AM_0 discrimination thresholds with the diotic 30- and 300-Hz-wide carriers, but did not report an effect of the carrier-envelope frequency content on the shape of the increase, and therefore focused on the difference between the thresholds with uncorrelated and correlated

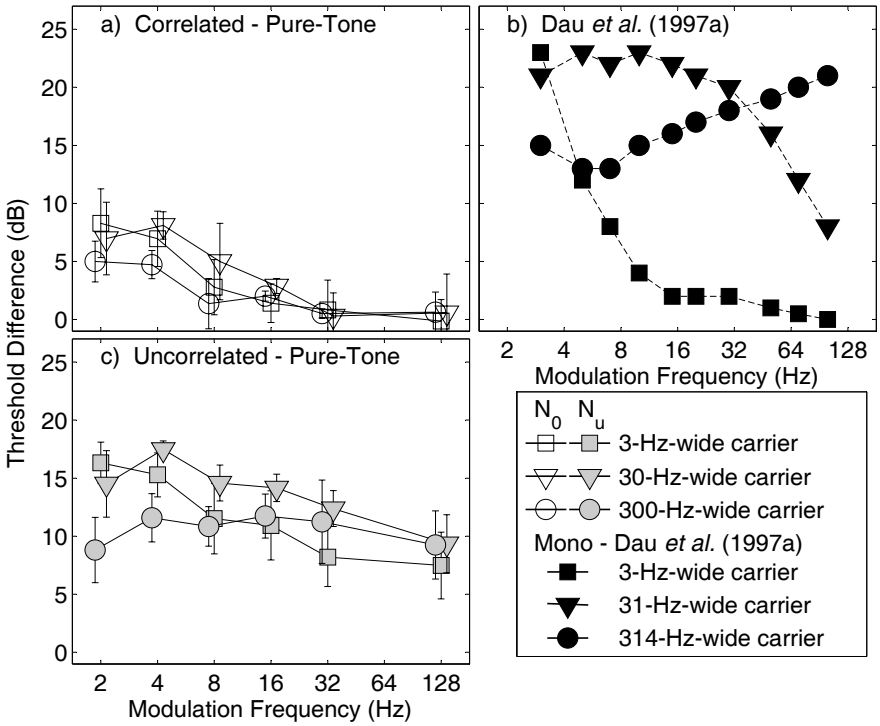

Figure 2.3: The difference in discrimination thresholds measured with narrowband noise carriers and pure-tone carriers (AM_π, stars from Fig. 2.1b). Panel a shows the difference for thresholds measured with correlated noise carriers (N_0AM_π, from Fig. 2.2b). Panel b shows the difference between monaural AM detection with narrowband noise and pure-tone carriers (data adapted from Dau et al., 1997a). Panel c shows the difference for thresholds measured with uncorrelated noise carriers (N_uAM_π, from Fig. 2.2d). The error bars in panels a and c show the standard deviation of the mean threshold difference across listeners.

noise carriers. However, there is a significant interaction between the diotic noise bandwidths and the modulation frequency in the present study, which can be seen in the threshold differences for $f_m < 16\,\text{Hz}$ (see Fig. 2.2b). These differences could suggest a dependence on the carrier envelope spectrum, but this needs to be investigated in further studies.

A control experiment was performed to investigate the difference between the thresholds measured in the AM_π discrimination experiments with the pure-tone and the diotic narrowband noise carriers. The ILD modulation discrimination threshold was measured with a 30-Hz-wide 'low-noise noise'[1] (Pumplin, 1985) carrier. Measuring with a low-noise noise carrier tests the hypothesis that the difference in thresholds between AM_π discrimination with a pure-tone carrier and with a narrowband noise carrier is caused by the envelope fluctuations of the noise and not by the noise's broader bandwidth per se. A pairwise t-test of the results showed no significant difference between thresholds measured with the low-noise noise carrier and the pure-tone carrier (thresholds at $-19.0\,\text{dB}$ and $-17.8\,\text{dB}$, respectively, $p = 0.43$), but there was a significant difference between the low-noise noise and the Gaussian noise thresholds ($-14.7\,\text{dB}$, $p < 0.05$). This suggests that it is the fluctuations in level of the Gaussian noise carriers that impede the discrimination of AM_π from AM_0.

The diotic Gaussian noise carrier and the pure-tone carrier do not create any ILDs themselves. Therefore, the binaural sensitivity to the AM_π signal is only limited by the internal variability of the auditory system, or by 'internal noise'. Since there is a significant increase in thresholds when using diotic Gaussian noise carriers, this suggests that the internal noise increases when the envelopes of the carriers fluctuate, or that the encoding of fluctuating envelopes is noisier than the encoding of steady envelopes. The range of modulation frequencies over which there is an increased

[1] Low-noise noise is a band-pass noise for which the phase angles of the individual frequency components have been optimally selected to minimize the fourth moment of the signal, thereby minimizing the envelope fluctuations. Kohlrausch et al. (1997) produced low-noise noise using an iterative process of band-pass filtering the noise and normalizing the noise with its envelope. Each filtering step produces envelope fluctuations, and each normalization step produces a flat envelope, but a broader frequency spectrum. After many iterations, a noise is produced which has both a flat envelope and the desired bandwidth. The control experiment was performed with one listener, a 30-Hz-wide carrier, and a 4 Hz AM_π signal.

threshold with the correlated noise carriers, and the differences between the thresholds from the three carrier bandwidths should provide some insight into how the interaural processing differences can be modeled.

Interaurally uncorrelated Gaussian noise carriers cause large stochastic fluctuations in ILD. This 'external' ILD variability is in addition to the internal noise described above. Therefore, it is not surprising that the ILD modulation discrimination thresholds are higher with $\mathrm{N}_u\mathrm{AM}_\pi$ than with $\mathrm{N}_0\mathrm{AM}_\pi$. It is unknown how the effects of the external and internal variances with Gaussian noise carriers combine in the auditory system. Therefore, the data obtained with *pure-tone* carriers were used as the reference threshold in the present study. This is in contrast to Stellmack et al. (2005), who compared the thresholds with uncorrelated and correlated noise carriers, leaving out the extra effect of the diotic level fluctuations on the measurements. The thresholds measured with the uncorrelated carriers are up to 18 dB higher than those measured with the pure-tone carrier, particularly at AM frequencies below the bandwidth of the carrier. By comparing with the pure-tone carrier thresholds instead of with the diotic noise carriers, the shapes of the threshold difference curves with uncorrelated noise carriers from Fig. 2.3c have similar aspects to those from monaural experiments (Fig. 2.3b; adapted from Dau et al., 1997a), but also large differences. The monaural curves (Fig. 2.3b) with 3-Hz and 31-Hz-wide carriers drop off quickly toward zero with f_m greater than the carrier bandwidth, indicating relatively sharp modulation frequency tuning. The binaural curves (Fig. 2.3c) also roll off with f_m greater than the carrier bandwidth, but do not roll off as quickly as the monaural curves, and seem to reach a plateau at about 8 dB, even with f_m much greater than the carrier bandwidth. This indicates broader much tuning in the binaural domain than in the monaural domain, as also suggested by Stellmack et al. (2005). Grantham and Bacon (1991) argued against a band-pass ILD modulation tuning after measuring detection thresholds with a 16-Hz AM_π signal in the presence of a diotic noise AM masker. At that AM frequency, in the present study, there were no significant differences between the AM_π discrimination thresholds with correlated noise or pure-tone carriers, even though the 30- and 300-Hz-wide carriers have envelope frequency components around 16 Hz. This suggests that the diotic masker in their study probably did not have a

significant effect on the AM_π detection threshold. Therefore, their data did not provide conclusive evidence for or against band-pass ILD modulation tuning.

The qualitative similarities between the binaural and monaural masking curves suggest that an element could be introduced in a binaural model that is similar to the monaural modulation filterbank from Dau et al. (1997a). However, it appears that the tuning of the ILD modulation filters in such a model must be broader than those of the monaural filterbank.

2.4 Experiment II: Masked modulation detection

The results of the first experiment suggested that there may be modulation frequency selectivity in the processing of ILD fluctuations. Therefore, further experiments were performed to directly measure the shape of this tuning. These experiments were based on similar experiments performed with diotic signals by Ewert et al. (2002), where a sinusoidal signal AM was masked by a narrowband noise modulator applied to a common pure-tone carrier.

2.4.1 Specific stimulus details

An interaurally uncorrelated, band-pass Gaussian-noise masker modulation was applied to the envelope of pure-tone carriers in a discrimination task, according to Eq. 2.5.

$$\begin{aligned} x_L(t) &= a\sin(2\pi f_c t)\,[1+N_L(t)]\,[1+m\sin(2\pi f_m t+\phi_L)] \\ x_R(t) &= a\sin(2\pi f_c t)\,[1+N_R(t)]\,[1+m\sin(2\pi f_m t+\phi_R)] \end{aligned} \tag{2.5}$$

where a controls the presentation level, f_c is the carrier frequency (in this case, 5 kHz), the subscripts L and R indicate left or right ear, and $N_{L/R}$ is the masking noise modulator (power set in this study to -10 dB re 1), spectrally centered at f_N, for the respective ears. The signal modulation was applied with AM frequency f_m, modulation depth m, and starting phases ϕ_L and ϕ_R. When two amplitude modulators are to be applied to a carrier (e. g., a masker, N, and a signal modulator, S), they can be added together and applied as a common modulator ($1+S+N$) or applied in

series as separate modulators $(1 + S)(1 + N)$. The additive approach can result in overmodulation if either the signal or the masker has a large negative amplitude (i. e., $S+N < -1$). The multiplicative approach, used in this study, avoids overmodulation as long as $S > -1$ and $N > -1$, which allows for signal modulation depths (m) close to 0 dB (see also Houtgast, 1989). However, by multiplying the two modulators, additional spectral side-bands are created, which can complicate analysis of the data, as discussed in section 2.4.3.

The design of the stimuli was based on Ewert et al. (2002). Each stimulus had an overall duration of 600 ms, windowed with 50 ms $\cos^2$ onset and offset ramps. The AM signal was applied to the middle 500 ms of the carrier, gated with 50 ms $\cos^2$ onset and offset ramps, leaving 400 ms with the desired signal AM depth. Measurements were made with $f_m = 4$, 8 and 32 Hz. In order to avoid monaural cues, the experiment was designed as a discrimination task, so all three intervals in a trial (signal and two reference) had an applied signal modulation with the same modulation depth in each interval. The start phase of the signal modulation in the left ear ϕ_L was chosen randomly for each trial interval over the range [0, 2π] with a uniform probability distribution. In the two reference intervals, the modulation start phase in the right ear was set equal to ϕ_L (AM_0), while in the signal interval, ϕ_R was set equal to $\phi_L + \pi$ (AM_π). With the randomized modulation phase and equal modulation depth on all intervals in a trial, successful discrimination could only be performed by combining information from the two ears, not based on one ear's analysis alone.

The masker modulations had a fixed bandwidth of 1.4, 2.8 and 11.1 Hz for the $f_m = 4$, 8 and 32 Hz, respectively, corresponding to one half-octave centered at f_m. The masker center frequencies, f_N, were at octave steps from f_m over a range from -4 to +4 octaves, but with the additional limitation that f_N could not be below 2 Hz or above 128 Hz. This hard frequency limit was put in place because the envelope of the window function itself could interfere with detection below 2 Hz, and the modulation side bands could be resolvable above 128 Hz. In Ewert et al. (2002), the masker modulation was placed over a range from -2 to +2 octaves with a 2/3-octave step size. The larger range and step size were chosen here in expectation of broader tuning after the results from Experiment 1 (see Sec. 2.3). A new masker modulator ($N_{L/R}$) was created for each presentation interval by generating a 10 s Gaussian white noise in the

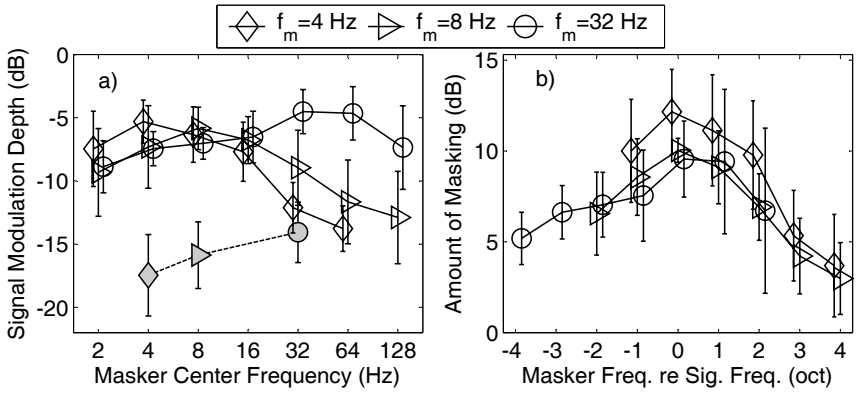

Figure 2.4: Panel a: Modulation depths required for discrimination of interaurally antiphasic AM from homophasic AM imposed on a pure-tone carrier in unmasked (shaded symbols, dashed line) and masked (open symbols, solid lines) conditions. Measurements were made with a fixed signal AM frequency of 4 Hz (diamonds), 8 Hz (triangles) and 32 Hz (squares) with interaurally uncorrelated narrowband noise maskers with a fixed power and bandwidth for a range of masker center frequencies. The data points for each curve are offset to more clearly show the error bars. Panel b: The same data from the left panel, but normalized for the unmasked threshold and signal frequency. The error bars in panel b show the standard deviation of the mean threshold difference across listeners.

time domain, setting all frequency components outside the passband to zero, and then scaling the variance to 0.1 (-10 dB re 1). The resulting noise was then added to a DC component ($1 + N_{L/R}$) and applied to the carrier as in Eq. 2.5. At this masker level, there was a small probability (less than 0.08% of samples, or less than 0.5 ms per presentation, on average) of overmodulation (i. e., $1 + N_{L/R} < 0$). This small occurrence was assumed to not have a significant effect on the results.

2.4.2 Results

Figure 2.4a shows the mean and standard deviation of the masked threshold patterns measured with a pure-tone carrier. The signal modulation depth in dB ($20 \log m$) is plotted as a function of the masker center frequency, with the signal modulation frequency as the parameter. In addition, the modulation depth required for discrimination without a masker present is plotted as a function of the signal modulation

frequency (dashed line, shaded symbols). Note that the three masked curves and their respective unmasked points have been offset slightly around the exact frequencies so that the error bars are more visible. In Fig. 2.4b, the same curves are replotted as an amount of masking, defined as the difference between the masked and unmasked thresholds for each signal modulation frequency at each masker center frequency, normalized by the signal modulation frequency in octaves. The error bars in panel b show the standard deviation of the mean threshold difference across listeners. In both panels, the symbols (diamond, triangle and circle) represent the 4, 8, and 32 Hz signal modulation frequencies, respectively.

The masking patterns (Fig. 2.4b) for the three signal frequencies are very similar in shape and amount of masking. All three curves show the highest amount of masking (approx. 10 dB) for the on-frequency condition and a monotonic decrease in masking as the spectral distance between signal and masker center frequency increases. The decrease in masking is greater when the masker center frequency is above the signal frequency than with lower masker frequencies. When the masker is 4 oct. above the signal, there is only about 3 dB of masking, however there is still about 6 dB of masking with the masker 4 oct. below the signal.

2.4.3 Discussion

The masking patterns obtained in the AM_π-AM_0 discrimination task with a narrow-band noise modulator masker imposed in series with a sinusoidal signal modulator on a pure-tone carrier showed consistency in shape and amount of masking for the three measured signal AM frequencies. The mean values of the three masking curves at each relative masker frequency are replotted in Fig. 2.5 (circles) along with a typical masking curve (squares) from the monaural masked AM detection experiments of Ewert et al. (2002) (adapted from their Fig. 2; 5.5 kHz carrier, 64 Hz AM signal, Q=1.25). Both curves show a maximum amount of masking when the masker is centered at the signal frequency, although the monaural curve shows a clearly higher masking value (about 17 dB) than the binaural curve. The slopes are similarly asymmetric for both curves, with a slightly steeper slope on the high-frequency side. For masker frequencies above the signal frequency, the monaural curve rolls off more rapidly than the binaural curve, so that the monaural curve already shows less masking

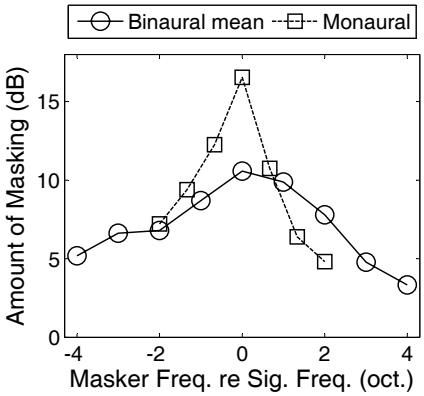

Figure 2.5: The mean of the three masking curves from panel Fig. 2.4b (circles, solid line) is shown with a typical monaural AM masking curve (dashed line, squares, adapted from Ewert et al., 2002, from their Fig. 2, 5.5 kHz carrier, 64 Hz AM signal).

than the binaural curve for maskers centered 1 oct. above the signal modulation frequency. This is consistent with the idea that monaural envelope processing has a sharper tuning than binaural processing of dynamic ILDs.

Multiplication of the signal and masker modulators creates additional side-bands in the stimulus through spectral convolution. This is represented in a sketch of the envelope spectra of two idealized stimuli in Fig. 2.6. A stimulus with only an applied noise masker AM would show an envelope spectrum with a DC component ($f = 0$) and a band of noise centered at f_N. Multiplication of the masking modulator with the signal modulator (tonal component at f_m) results in the two side-bands shown with dashed lines in Fig. 2.6 centered at $f_m \pm f_N$ (note that only positive frequencies are shown in the sketch). In an AM *detection* experiment, as in the monaural experiments from Houtgast (1989) and Ewert et al. (2002), where the listener's task is to distinguish between presentation intervals with only a masker modulator and a target interval with masker and signal modulators, the side-bands are only present in the target interval. Therefore, they can serve to enhance the detectability of the signal AM. However, with an AM *discrimination* experiment, like the one here, all stimuli have the same

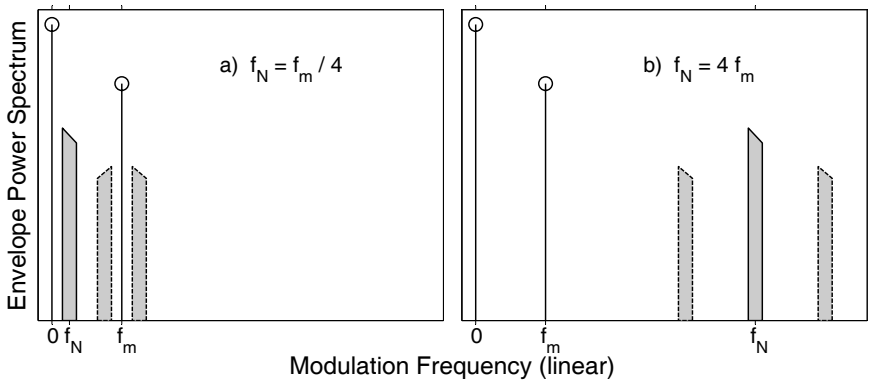

Figure 2.6: Theoretical envelope power spectra resulting from the application of a band-pass noise masker modulator and a tonal signal modulator. The DC-component ($f = 0$) and tonal component ($f = f_m$) are plotted with circles and the spectrum of the noise masker is shown with a bar with a solid line centered at f_N. The bars shown with the dashed lines show the result of applying the masker and signal modulators in series with the multiplicative approach described in the text. The left panel shows a case where $f_N < f_m$ and the right panel shows a case where $f_N > f_m$. Note that only the DC-component and positive frequencies are shown.

modulation and the same side-bands. In this case, the side-bands do not provide any cues for signal detection, and may actually hamper signal detection.

The amplitude of the side-bands is determined by the amplitudes of the masker and signal AM components. In this experiment, with a fixed masker energy, the side-bands' energy scales with the signal energy at a fixed ratio (-10 dB). The effect of these side bands should be considered when designing a model to account for the measured masking patterns. For example, a model could be designed with a symmetric band-pass modulation filter centered at the signal's modulation frequency, and a certain signal-to-noise ratio (SNR) required after the filter for detection of the signal modulation. With this model, the side-bands' energy would create a noise floor at a SNR that depends on the signal and masker frequencies. When $f_N \ll f_m$, the side-bands will be very close, spectrally, to the signal (Fig. 2.6a), and the side-bands' energy will be passed through the filter with little attenuation, creating a relatively high noise floor. As f_N increases, the side-bands move away from the signal in frequency, becoming more attenuated by the filter and reducing the noise floor. The side-bands

are only centered at frequencies larger than f_m if $f_N > 2f_m$. The result could be an assymetric masking pattern, even though the filter was assumed to be symmetric around the signal modulation frequency.

2.5 Implications for binaural models

The experimental data presented above suggest that a binaural model should include an array of ILD modulation band-pass filters to simulate human performance. Some preliminary simulations were made using the binaural model from Breebaart et al. (2001a) as an artificial observer in the experiments described in section 2.4. These simulations were performed with the original model, which uses a sliding integrator (low-pass filter) to limit its temporal resolution.

The Breebaart model was designed for static binaural conditions, such as for predicting binaural masking level differences (BMLD), and is quite successful at predicting human performance under many experimental conditions (see also Breebaart et al., 2001b,c, for more details). Breebaart et al. (2001c) focused on temporal parameters, including a simulation based on an experiment from Grantham (1984), where the listener's task was to discriminate between interaurally antiphasic and homophasic AM imposed on uncorrelated broadband noise carriers. Grantham's data showed a large variance between test subjects, but Breebaart's model was able to capture the general trend of the results. Since their model was able to simulate experimental results similar to those described above in section 2.3, it was chosen as a basis for testing with the new experiments and for possible future development.

The model starts with two parallel peripheral processing stages (one for each ear, see schematic in Fig. 2.7), based on the monaural processing model from Dau et al. (1996), which did not include a modulation filterbank. The first stages are an outer- and middle-ear transfer function, basilar-membrane filtering, consisting of an array of gammatone filters, inner hair cell transduction, modeled with half-wave rectification and a low-pass filter with a cutoff frequency of 770 Hz, and finally a series of five adaptation loops, which enable the simulation of forward masking. The output from each pair (Right/Left) of peripheral channels is then passed to an array of excitation-inhibition (EI) elements, which calculate the difference in the corresponding channels

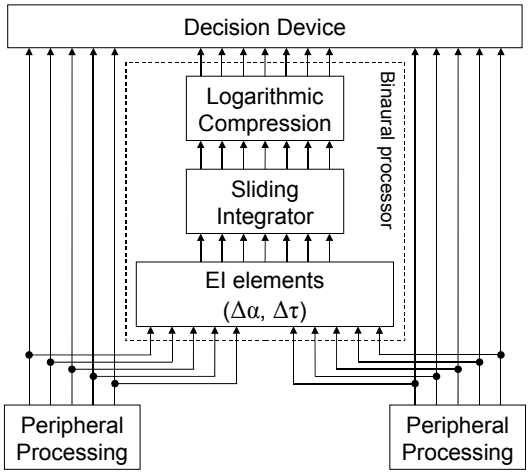

Figure 2.7: Schematic of the binaural model from Breebaart et al. (2001a). The model consists of two parallel monaural peripheral channels, including a gammatone filterbank, a half-wave rectification and low-pass filter inner hair cell model, and adaptation loops. The two monaural signals are combined in the binaural processor through an array of excitation-inhibition (EI) elements, which calculate the difference of the two signals for a range of applied interaural gains and delays. The resulting signals are smoothed with a sliding integrator and compressed with a logarithmic compression. Finally, an optimal detector tries to find a signal based on all monaural and binaural inputs.

for a range of characteristic interaural gains and delays. This is similar in concept to the equalization-cancellation (EC) model from Durlach (1963), which finds the optimal gain and delay before calculating the channel difference. The EI concept is based on neurons that receive excitatory input from the ipsilateral side and inhibitory input from the contralateral side, effectively calculating a difference between the two auditory signals. The output from each EI element is then smoothed with a sliding integrator, consisting of a symmetric double-sided exponential window with time constants of 30 ms. This sliding integrator acts as a low-pass filter with a cutoff frequency of about 5.3 Hz. A compressive (logarithmic) non-linearity is applied to the smoothed signal. The resolution of the system is limited by the addition of an internal noise. Finally, an optimal detector, with inputs from all monaural and binaural channels, is used as the decision device.

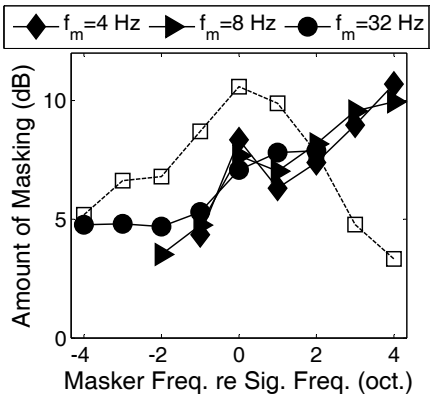

Figure 2.8: Masked tuning curves predicted by the model from Breebaart et al. (2001a) when used as an artificial listener in the experiments described in section 2.4. The mean tuning curves from the human listeners are shown with the dashed line and open square symbols (mean of the three curves from Fig. 2.4b). The simulation was made with $f_m = 4$ Hz (diamonds), 8 Hz (triangles) and 32Hz (circles).

The model does not track perceived motion or predict spatial perception of the sound source, but rather looks at the energy in each EI channel (i. e., for a fixed ILD and ITD combination) in order to detect a signal. Diotic signals have no energy in the ILD=0, ITD=0 channel (perfect cancellation, internal noise is added later) and any interaural decorrelation will result in an increase in energy in this channel. Therefore, the addition of an antiphasic tone to a diotic noise (N_0S_π) will result in a much larger increase in energy than the addition of a homophasic tone (N_0S_0), demonstrating the classic BMLD (see, e. g., Licklider, 1948; Hirsh, 1948).

The experimental conditions described in section 2.4 were simulated using the model. The simulation results are summarized as masking curves in Fig. 2.8, like those shown in Fig. 2.4b. It is clear from a comparison of the human listeners' (open symbols) and the model's results (filled symbols) that the tuning described in section 2.4 is not captured by the model. The model does show a small peak at the signal frequency, but then the masking level increases with higher relative masker center frequencies, while the human listeners show a decrease in masking level with higher masker frequencies. The reason for the increase in masking in the model with

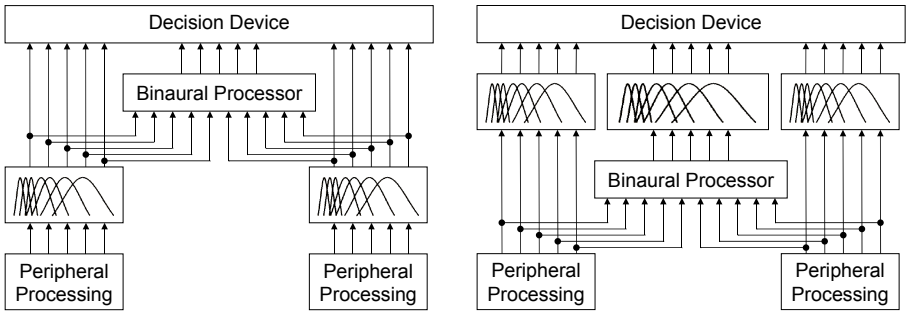

Figure 2.9: Possible concepts for the inclusion of modulation frequency selectivity in binaural processing of ILD fluctuations. In the left panel, the modulation filterbank (MFB) from Dau et al. (1997a) is applied to each monaural peripheral channel before the input to the binaural processor. The binaural system would thereby inherit its ILD modulation frequency tuning from the monaural AM processing. The second concept, in the right panel, takes the inputs to the binaural processor from before the monaural MFBs. This then requires an additional binaural ILD modulation filterbank to add frequency selectivity in the processing of interaural level fluctuations.

high masker center frequencies is that the sliding integrator smooths out the interaural fluctuations from the masker, thereby removing any locations with good cancellation and increasing the energy at the output of the EI channels. Adding the diotic AM to the reference intervals further increases the energy so that there is less of a difference between the signal and reference intervals, making the discrimination task harder to perform. Part of this effect stems from the fact that the model was not designed to look at temporal differences and only compares the total energy at the output of the EI channels. However, if there was an ILD modulation frequency filtering as a part of the binaural processing instead of the low-pass filter, the resulting masking patterns might be more similar to those obtained in the results presented in section 2.4.

In order for this binaural model to be able to predict thresholds with fluctuating stimuli, it requires frequency selectivity in the processing of monaural and interaural level fluctuations. The monaural modulation filterbank (MFB) from Dau et al. (1997a) could be added to the peripheral channels, but then a question arises as to the sequence of model stages: should the taps for the EI array come from before or after this filterbank? Two possible design concepts are shown in Fig. 2.9. The sequence of the stages would not be important except for the non-linearities in both

the monaural MFB and the EI process. The monaural MFB has a non-linear reduction of modulation phase information for frequencies above 10 Hz. Without the modulation phase information, there would be no interaural differences with an AM_π signal, and the model would not be able to discriminate between AM_π and AM_0. If the EI inputs were to come from after the monaural modulation filters, but before this non-linearity (left panel of Fig. 2.9), then the sharpness of tuning would be preserved through the output of the EI elements. That sharpness might be reduced to fit the measured data by adding additional noise and/or interaural differences in the processing, but the effect of this additional noise on other experiments would have to be investigated. Another option would be to take the EI inputs from before the monaural modulation filterbank (right panel of Fig. 2.9). In this manner, the interaural modulation phase differences would be preserved going into the binaural processor. However, an additional model step would then be required, namely an ILD modulation filterbank at the output of the EI system. This filterbank would replace the sliding integrator from the original Breebaart model. There is also a compression stage after the sliding integrator in the original model. The optimal sequence for the linear and non-linear model stages should be investigated in further simulations.

2.6 Summary and Conclusions

The first experiment showed that interaurally correlated and uncorrelated narrowband noise carriers have a significant effect on the discriminability of modulated ILDs (AM_π) from diotic AM (AM_0), particularly for modulation frequencies below the bandwidth of the carrier. This suggested that the binaural system shows broad band-pass modulation frequency tuning in processing of ILD fluctuations. A comparison of the results obtained with diotically modulated and unmodulated references underscored the importance of eliminating monaural cues in the design of binaural detection tasks because the signal detection will be based on monaural detection if the monaural cues are more salient than the binaural cues.

This modulation frequency tuning was further explored in the second experiment with AM_π discrimination in the presence of masking narrowband noise modulators.

The masking patterns also showed band-pass tuning, but with a broader tuning than that shown in similar monaural experiments (e. g., Houtgast, 1989; Ewert et al., 2002).

An analysis with an existing binaural model (from Breebaart et al., 2001a) showed that the model, which uses a low-pass filter to limit its temporal resolution in the processing of fluctuating interaural differences, and not a modulation filterbank, cannot predict the thresholds or the masking patterns measured with human listeners.

Further experiments should be performed to investigate the effect of diotic level fluctuations on the perception of ILD fluctuations through additional psychoacoustic tests as well as modeling. In addition, a binaural model should be developed that can predict the frequency selectivity shown here in the processing of interaural level fluctuations.

Chapter 3

A lack of spatial release from amplitude modulation masking

Abstract

Two experiments were performed to measure the effect of a perceived spatial separation on masked amplitude modulation (AM) detection thresholds. Two temporally-interleaved transposed stimuli were used as carriers for a narrowband-noise modulation masker and a 16-Hz sinusoidal modulation probe. With these stimuli, the interaural time difference (ITD) of the masker and probe carriers could be adjusted independently. In the first experiment, the listeners adjusted the interaural level difference of a pointer stimulus to be aligned with the perceived lateral position of either the masker or the probe stimulus, as a function of the masker and probe ITDs. The results showed that the listeners could lateralize the two stimuli separately and robustly. The second experiment measured masked AM detection thresholds as a function of the masker modulation frequency and masker ITD using a diotic 16-Hz AM probe. These results showed modulation frequency tuning without a spatial release from modulation masking even though the masker and probe were perceived to have a spatial separation. Implications of these results for models of binaural processing are discussed.

3.1 Introduction

In many masked signal detection experiments, there can be a significant improvement (often around 15 to 20 dB) in the detection thresholds when there is a perceived spatial separation between the target and the masker, created using interaural time differences (ITD). This has been demonstrated in the masked detection of pure-tones (e. g., Jeffress et al., 1962; Breebaart et al., 1998; van der Heijden and Trahiotis, 1999) and with temporally fluctuating sounds, e. g., in the detection of click trains (Saberi et al., 1991; Gilkey and Good, 1995), pulsed 1/3-octave noise bands (Zurek et al., 2004), and with sinusoidally-amplitude-modulated broadband noise (Kopčo and Shinn-Cunningham, 2008). These studies all dealt with the ability to hear a sound presented simultaneously with a masking noise, but were not concerned with the ability to hear details about the target sound itself, e. g., its pitch trajectory or the shape of its amplitude envelope. The present study sought to determine whether there is a similar spatial release from masking in the detection of sinusoidal amplitude modulation (AM) imposed on a suprathreshold carrier in the presence of an amplitude-modulated masker.

Previous studies of masked AM detection have focused on monaural or diotic listening (e. g., Bacon and Grantham, 1989; Houtgast, 1989; Ewert and Dau, 2000). In those studies, a sinusoidal AM target was imposed together with an AM masker (either sinusoidal or narrowband noise) on a broadband-noise or pure-tone carrier, and the minimum target modulation depth was measured as a function of the modulation-frequency spectrum of the masker. These experiments typically show modulation-frequency tuning in that the thresholds are highest when the modulation frequency of an AM target is close to or within the modulation-frequency spectrum of the AM masker, and lower when the target and masker are separated in their modulation spectra.

Other studies have investigated AM detection and binaural interactions in terms of binaural modulation detection interference (MDI; Bacon and Opie, 1994; Sheft and Yost, 1997). MDI is a form of masked modulation detection where the target AM and masker AM are imposed on carriers in different audio-frequency regions. For example, when the target AM was applied to a 1-kHz carrier, and a masker AM

was applied to a 4-kHz carrier, and both were presented monaurally in the same ear, then thresholds increased by about 7 to 8 dB relative to the unmasked case. When the masker was moved to the other ear, then thresholds were only about 2 to 3 dB higher than in the unmasked case (Bacon and Opie, 1994), showing about a 5 dB release from masking. The thresholds measured in such MDI experiments also show modulation-frequency-specific tuning, similar to that seen with masked AM detection, discussed above (Bacon and Opie, 1994; Moore et al., 1995). When the target and masker modulation were presented with different ITDs, so that there was target and masker energy in both ears (Sheft and Yost, 1997), there was an improvement of about 2 dB in the thresholds relative to the diotic condition, showing a small, but significant, release from MDI. These studies suggest that there is a small interaction between AM processing and binaural (spatial) processing in the auditory system, at least across audio-frequency channels. MDI may be based largely on auditory grouping, in that it can be difficult to distinguish details between sounds that are perceptually grouped together (Yost et al., 1989; Moore and Shailer, 1992; Oxenham and Dau, 2001; Gockel et al., 2002). It could be that the interaural differences between the masker and the target prevented them from being grouped together, thereby causing the release from MDI, and that it was not actually an interaction between AM and binaural processing.

The release from masking in signal detection due to the target and masker having different ITDs is often modeled with an equalization-cancellation (EC) mechanism (Durlach, 1963), in which the difference is taken between the two ears' signals at an optimal interaural delay (see, e. g., van der Heijden and Trahiotis, 1999; Zurek et al., 2004). The delay is selected so that the energy of the masker is minimized when the difference is calculated, thereby increasing the effective signal-to-noise ratio (SNR). In theory, the same mechanism could be applied with masked AM detection, canceling the masking modulator and improving the AM detection thresholds.

The goal of the experiments in the present study was to investigate the interactions between AM and binaural processing by measuring lateralization of AM carriers and masked AM detection thresholds when the probe and masker AMs were imposed on carriers within the same spectral region (similar to the experiments from Houtgast, 1989 and Ewert and Dau, 2000), but whose ITDs could be controlled independently.

The results of these experiments and some implications for binaural models will be discussed in the following sections.

3.2 General methods

The two experiments performed in the present study measured different aspects of perception using the same stimuli. The first experiment was designed to measure whether two AM carriers with the same spectral content could be lateralized separately using ITDs, and was a precondition for the second experiment. Given that the two carriers could be lateralized separately, then the second experiment measured the AM detection thresholds as a function of the masker ITD and modulation-frequency content. The common aspects of the experiments are presented here, and the specific details will be presented in the subsequent sections.

3.2.1 Stimuli

The main requirement for the stimuli was to create a perception of two temporally and spectrally overlapping carriers coming from different lateral positions. A sound can be lateralized by creating an ILD and/or an ITD. By definition, an ILD changes a signal's energy in at least one ear, thereby creating different SNRs in each ear when the target and masker have different ILDs. In many cases, signal detection thresholds with ILD-based lateralization can simply be predicted by considering the ear with the higher SNR, or the 'better-ear advantage' (e. g., Kopčo and Shinn-Cunningham, 2008). Therefore, it was desired to only use ITDs on the stimuli to create the lateralized percept for the present study.

In order to have separate ITDs on the probe and masker carriers without changes in level from interference effects resulting from adding two similar carriers together with slightly different phase alignments, temporally-interleaved transposed stimuli were used as the carriers. Transposed stimuli were originally designed to create firing patterns in high-frequency auditory-nerve fibers, based on the stimulus envelope, that are similar to those of low-frequency nerves, based on the stimulus fine-structure (van de Par and Kohlrausch, 1997). There is much greater phase-locking to the envelope

in auditory-nerve firing patterns with transposed stimuli than with sinusoidally-amplitude-modulated (SAM) stimuli (Dreyer and Delgutte, 2006), as well as much greater ITD sensitivity and extent of laterality (Bernstein and Trahiotis, 2002, 2003). This ITD sensitivity and the ability to temporally interleave two transposed stimuli without interference made them ideal for use in the present study.

The transposed stimuli used in the present study were created with a method similar to that from van de Par and Kohlrausch (1997). The results of the signal processing steps are shown in Fig. 3.1. All stimuli were digitally generated using MATLAB® (The MathWorks) at a sampling rate of 97 656 Hz (nominal 98 kHz for the TDT system 3). First, a 250-Hz sinusoid was generated (panel a), and half-wave rectified (panel b). This signal was then multiplied by a 125-Hz square-wave to select every other half-wave (panel c). The first three steps are equivalent to convolving the positive half-wave of a 250-Hz sinusoid with a 125-Hz click train. In order to restrict the bandwidth of the signal, it was low-pass filtered with a 1 250 Hz cut-off frequency (panel d). The signal was then multiplied onto a 5-kHz sinusoidal carrier (panel e). A second stimulus was created using the same method, only with a time delay that was randomly drawn from a distribution given by:

$$4 \pm \mathrm{B}(4,2)[\mathrm{ms}], \tag{3.1}$$

where B(4,2) is a beta distribution, and there was equal probability of plus or minus. This distribution of delays is bimodal with a range of [3, 5] ms and a minimum at 4 ms. It was selected so that the two stimuli would have no temporal overlap in either ear with any combination of ITDs up to ± 1 ms. It also has only a small probability of a 4 ms delay, which would recreate a 250-Hz transposed stimulus. It was expected that a 4 ms delay might enhance grouping of the two stimuli, but this assumption was not tested. The ITD of each stimulus was then modified independently, and an AM was applied separately to each stimulus, depending on the measurement condition. Note that the transposition already imposes an AM on the tonal carrier, so the AM imposed on a transposed-stimulus carrier is a second-order AM. The result of adding two stimuli together with a 3.3 ms delay is shown in Fig. 3.1f, with one stimulus in black and the other in gray in the left panel, and the magnitude of the combined spectrum

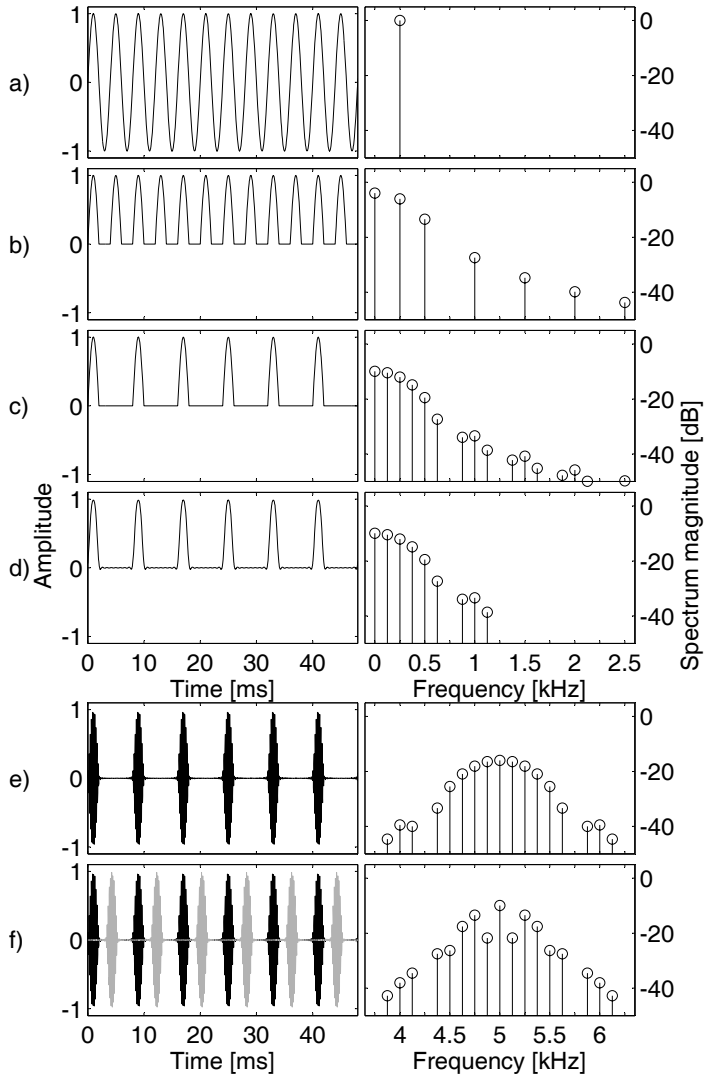

Figure 3.1: Steps for creation of the interleaved transposed stimuli. The left column shows the signal amplitudes, and the right column shows the spectrum magnitudes for each step. a) Generate a 250-Hz sinusoid. b) Half-wave rectify. c) Multiply by a 125-Hz square-wave. d) Low-pass filter with 1250-Hz cutoff frequency. e) Multiply by a 5-kHz sinusoidal carrier. f) Add a second signal with a specific time delay. The amplitude in (f) shows the two signals in black and gray, and the spectrum magnitude shows the combined spectrum for the particular temporal alignment.

shown in the right panel. One of the stimuli served as the carrier for the masker modulation, and the other stimulus served as the carrier for the probe modulation. In the following, the two transposed stimuli will be referred to as the masker stimulus and the probe stimulus.

In some of the measured conditions, an AM was applied to the masker and probe stimuli. The transposed stimulus effectively samples the modulator at a rate of 125 Hz, so the frequency content of the modulator is limited to half this rate, or 62.5 Hz, to avoid aliasing. The amplitude modulator on the probe stimulus was sinusoidal, so the probe stimulus was defined as a function of time t as:

$$y_{probe}(t) = [1 + m\sin(2\pi f_m t + \phi)]x_{probe}(t), \tag{3.2}$$

where m is the modulation depth, typically expressed in dB as $20\log m$, f_m is the modulation frequency, always 16 Hz in the present study, ϕ was a random starting phase, selected from a uniform distribution over the interval $[0, 2\pi]$, and $x_{probe}(t)$ is the transposed-stimulus carrier, described above. The masker modulator was a 5.6-Hz-wide (1/2 octave centered at 16 Hz) Gaussian-noise masker with a variance of 0.1 (power of -10 dB re the DC component) and was applied as:

$$y_{mask}(t) = [1 + n(t)]x_{mask}(t), \tag{3.3}$$

where $n(t)$ is the masker modulator and $x_{mask}(t)$ is the transposed-stimulus carrier described above. The center frequency of the masker noise was an experimental parameter, as described below. The masker modulator was generated by creating a two-second-long Gaussian noise and setting the magnitudes of the frequency components outside the desired passband to zero.

In order to further enhance the perceptual separability of the two stimuli, they were also gated separately. Both stimuli were gated with 10 ms $\cos^2$ ramps at the onset and at the end. The masker stimuli had an overall duration of 600 ms, and the probe stimuli had a 200 ms gating delay, relative to the onset of the masker stimulus, and an overall duration of 300 ms. An excerpt from an example of a combined stimulus is shown in Fig. 3.2. Here, the masker stimulus is shown in gray, and the probe stimulus is shown in black.

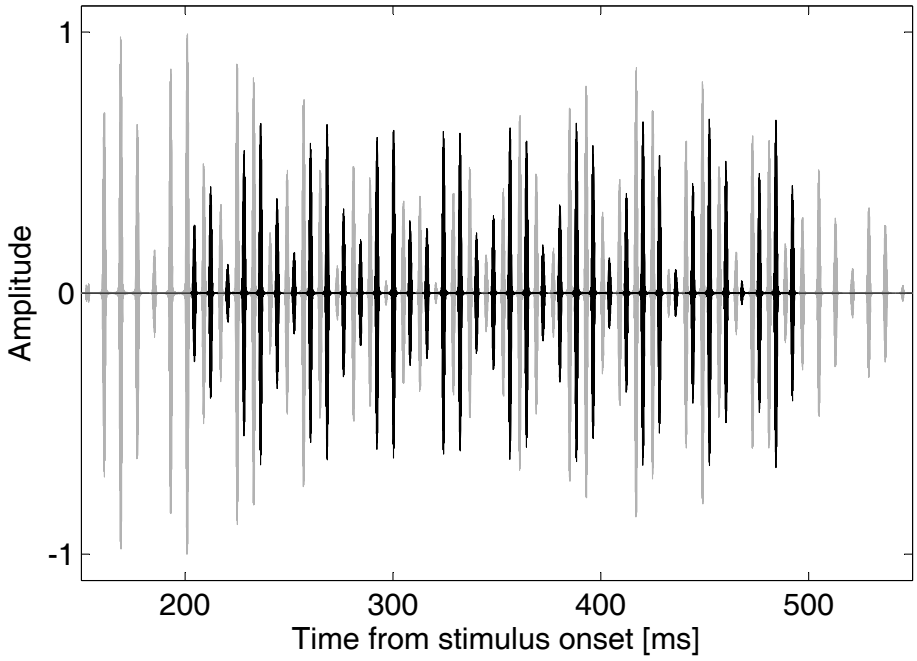

Figure 3.2: Example combined stimulus for one ear. The masker stimulus is in gray and the probe stimulus in black.

Next, an ITD was applied to the masker and probe stimuli by zero-padding the beginning and end of each stimulus as required for each ear. Then, the levels of the two stimuli were equalized, they were added together to form the complete stimulus, and the overall level was set to 65 dB SPL in each ear.

3.2.2 Equipment

The stimuli were generated on a PC running MATLAB® (The MathWorks), transmitted to a Tucker-Davis Technologies (TDT) System 3 (RP2.1 real-time processor, HB7 headphone buffer), and played through Etymotic Research ER-1 insert earphones. The participants sat in a sound-insulated booth with a computer monitor for instructions and a response interface, and a keyboard and mouse for response input.

3.2.3 Test subjects

Five test subjects (three female, two male, aged 18 to 32) participated in the experiments. All had pure-tone audiometric thresholds of 15 dB HL or better for octave frequencies from 125 Hz to 8 kHz. Three had little or no prior experience with psychoacoustic measurements and were paid for their participation, and the other two were members of the research center, including the author of this work. All subjects gave written informed consent (as approved by the Boston University Charles River Campus Institutional Review Board) before participating in the study.

3.3 Experiment I: ITD-based lateralization of the masker and probe stimuli

3.3.1 Procedure

The lateralization experiment was conducted to determine whether the listeners would be able to lateralize the masker and the probe stimuli separately, or whether the two temporally-interleaved stimuli could be heard at two separate locations. The listeners' task was to adjust the ILD of a pointer sound until it was perceived to come from the same lateral position within the head as the target stimulus, which was either the masker or the probe stimulus, depending on the condition. The ILD-pointer method was similar to ones used in previous studies to measure the extent of laterality of a stimulus (Bernstein and Trahiotis, 1985; Trahiotis and Stern, 1989; Bernstein and Trahiotis, 2003, 2005). There were six conditions measured: (1) the alignment target was the probe stimulus, presented alone, (2) the target was the masker stimulus, presented alone, (3) the target was the probe stimulus with a fixed ITD of 0 ms, presented in the combined stimulus, measured as a function of the masker ITD, (4) the target was the masker stimulus, presented in the combined stimulus, measured as a function of the masker ITD when the probe had a fixed ITD of 0 ms, (5) the target was the probe stimulus, presented in the combined stimulus, measured as a function of the probe ITD when the masker had a fixed ITD of 1 ms, and (6) the target was the masker with a fixed ITD of 1 ms, presented in the combined stimulus,

measured as a function of the probe ITD (the conditions are also summarized in the legend of Fig. 3.3). Note that the stimuli were identical in conditions 3 and 4, and in conditions 5 and 6, but the target for alignment changed between the conditions. These six conditions were measured as a function of either the masker or probe ITD τ for $\tau = \{-1, -0.5, 0, 0.5, 1\}$ ms, with a positive ITD indicating right ear leading. For this experiment, the masker AM was always centered at 16 Hz, the probe AM frequency.

The ILD pointer was a 5 kHz tone, fully amplitude modulated at 250 Hz. The pointer had a duration of 600 ms, and was gated with 10 ms $\cos^2$ ramps. Each measurement started with a random pointer ILD selected from a uniform distribution of integers from -8 to +8 dB, with a positive ILD indicating a higher level in the right ear. The probe, masker or combined stimulus was presented first, followed by a 200 ms pause and the pointer stimulus. The listener could choose to replay the two stimuli, move the ILD pointer to the left by a large or a small step, or to move the ILD pointer to the right by a large or a small step. A large step was ± 3 dB, and a small step was ± 1 dB, adding half the step-size to the right ear and subtracting half the step-size from the left ear. The listeners were encouraged to replay the signals repeatedly when unsure, and to move the pointer past the perceived location of the stimulus in order to get a better idea of the center of the stimulus. When they were certain of their alignment, they pressed a 'done' button to store the data and continue to the next stimulus. The stimuli were presented in blocks by the condition, with conditions 3 and 5 blocked together and conditions 4 and 6 blocked together, with the five ITDs for each condition presented in random sequence within each block. Each test subject completed five repetitions for each data point.

3.3.2 Results

Figure 3.3 shows the normalized pointer ILD required to align with the lateral position of the stimuli in the six measured conditions as a function of the ITD parameter τ. The conditions are summarized in the legend in Fig. 3.3, showing whether the probe or masker was the target for alignment, and the corresponding probe and masker ITDs. There was a large variance between test subjects in the pointer ILD required for alignment. One test subject, for example, required 15 dB ILD to align

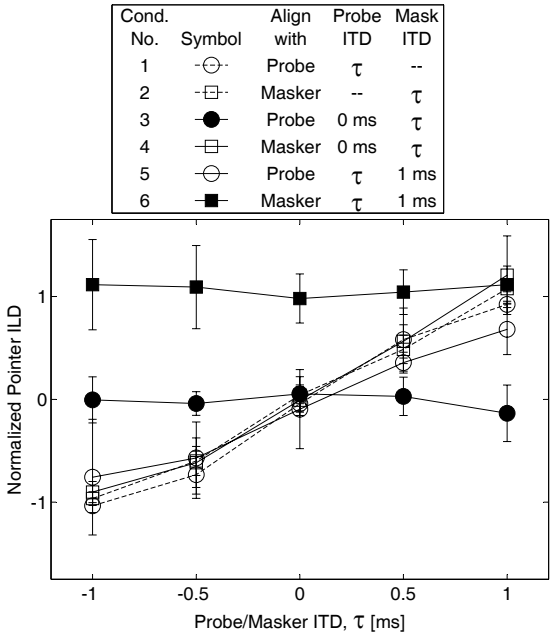

Figure 3.3: Normalized ILD pointers as a function of masker or probe ITD. The circles indicate alignment with the probe (P) stimulus, and the squares indicate alignment with the masker (M) stimulus. The data points connected with dashed lines were measured with one stimulus only, and those connected with solid lines were measured as part of the combined stimulus. The solid black markers indicate that the points were measured for a fixed ITD (see legend) as a function of the ITD (τ) of the other part of the combined stimulus.

with a 1-ms ITD, while another test subject required only 7 dB ILD. Similar variance between subjects has been reported in other studies (e. g., Heller and Trahiotis, 1996), and may reflect a different relative sensitivity to ILDs and ITDs between subjects. Therefore, each test subject's data were normalized for plotting and for the analysis using their mean data from the lateralizations when the two stimuli were presented alone (conditions 1 and 2). First, any ILD offset at 0-ms ITD was subtracted from all data points, then the data measured with positive ITDs were divided by the mean value of the two conditions at 1-ms ITD, and the data measured with negative ITDs were divided by the mean value of the two conditions at -1-ms ITD. The results shown in Fig. 3.3 are the mean normalized data across subjects with error bars representing

one standard deviation of the mean. The data were analyzed in the following using analyses of variance with repeated measures (RM-ANOVA) with a level of 0.05 required for significance.

The two conditions with a fixed ITD on the alignment target (conditions 3 and 6, filled symbols in Fig. 3.3), where the probe lateralization was measured as a function of the masker ITD and vice versa, show little change with the ITD of the other stimulus. An RM-ANOVA was performed on these two condition's data with a main factor of ITD. In condition 3 (filled circles in Fig. 3.3), the ITD of the probe stimulus was fixed at 0 ms, and there was no significant change in the normalized pointer ILD as the ITD of the masker stimulus was changed from -1 to 1 ms ($F(4, 16) = 0.56$, $p = 0.70$). In condition 6 (filled squares in Fig. 3.3), the ITD of the masker stimulus was fixed at 1 ms, and there was also no significant change in the normalized pointer ILD with the ITD of the probe stimulus ($F(4, 16) = 0.88$, $p = 0.50$).

The other four conditions, for which the alignment target's ITD was the experimental parameter, all show a monotonic increase in normalized pointer ILD as the ITDs of the target stimuli were increased from -1 to 1 ms. An RM-ANOVA across these four conditions with main factors of condition and ITD found a significant effect of ITD ($F(4, 16) = 159$, $p \ll 0.0001$), but no significant effect of condition ($F(3, 12) = 0.88$, $p = 0.48$). There was, however, a significant interaction ($F(12, 48) = 2.15$, $p < 0.05$), which a post hoc analysis showed to be a small repulsion effect in condition 5 (open circles connected with a solid line in Fig. 3.3) for positive ITDs. This repulsion was only seen in the data of two listeners.

3.3.3 Discussion

The results of the lateralization experiment show that the two stimuli (masker and probe) can be perceived as coming from two separate locations within the head. This suggests that they may be perceived as two separate auditory objects.

The two stimuli were gated separately, so the masker stimulus began 200 ms before the onset of the probe stimulus. Previous studies have shown robust lateralization with 100 ms and 200 ms long stimuli (Buell et al., 1991), so it is likely that the listeners could lateralize the masker stimulus before the probe stimulus even started. On the other hand, the probe stimulus started after the masker stimulus and ended before

the masker stimulus ended, so it could only have been lateralized separately from the masker stimulus if it was perceived to be a separate auditory object from the masker stimulus.

When two sounds are presented simultaneously with similar interaural parameters, but are not perceptually grouped together, there can be a repulsion effect, or pushing (Suzuki et al., 1993; Braasch and Hartung, 2002). This effect is seen in the data shown in Fig. 3.3 for condition 5 (open circles connected with a solid line) with positive ITDs. In that condition, the masker stimulus had an ITD of 1 ms, and some of the listeners required smaller pointer ILDs to align with the probe stimulus than were required for the same ITD in the other conditions (see conditions 1, 2 and 4 in Fig. 3.3), indicating that the perceived position was pushed towards the mid-line. The pushing effect has been reported most frequently in prior studies in conditions when a masker is gated on before the alignment target, as was the case with condition 5 in the present study. Previous studies (e. g., Canévet and Meunier, 1996; Carlile et al., 2001) have explained this effect by assuming that there are neurons tuned to certain spatial locations that show adaptation to an on-going stimulus. As a result, the spatial perception of later sounds is shifted away from the perceived position of the earlier sound, particularly for small perceptual distances between the sounds. Since there was no significant pushing effect seen with a masker ITD of 1 ms and a probe ITD of 0 ms (diotic), it was assumed that the perceived distance between the masker and probe could be used to investigate a spatial release from modulation masking in a modulation-detection experiment.

3.4 Experiment II: Masked amplitude-modulation detection

3.4.1 Procedure

The same temporally-interleaved transposed stimuli were used as carriers in the modulation detection experiment to test the hypothesis that a perceived spatial separation of a probe and a masker would improve the detectability of AM applied to the probe. Similar methods to those used in previous masked modulation detection

studies were used (e. g., Houtgast, 1989; Ewert and Dau, 2000; Thompson and Dau, 2008), in which a narrowband-noise modulator was used to interfere with detection of a sinusoidal modulator. In those previous studies, the masker and probe modulators were applied in series to a common carrier. In the present study, in order to achieve a spatial separation of the masker and the probe, the modulators were applied to separate carriers.

The minimum modulation depth m required to detect a sinusoidal AM imposed on the probe stimulus was measured as a function of the center frequency of the masker modulator, and of the masker ITD, either 0 ms (diotic) or 1 ms. For this experiment, the probe stimulus always had a 0-ms ITD. The center frequencies of the masker were 6.3, 10.1, 16, 25.4 and 40.3 Hz (-4/3, -2/3, 0, 2/3, 4/3 octaves from 16 Hz). In addition, two control measurements were made with no AM imposed on the masker transposed stimulus, both with 0-ms ITD and with 1-ms masker ITD.

A 3-interval, 3-alternative, forced-choice (3-AFC) design was used, with one randomly selected signal interval and two reference intervals. There was no AM imposed on the probe stimulus in the reference intervals, and a sinusoidal AM was imposed on the probe stimulus in the signal interval. The modulation depth was adaptively tracked, following a 2-down, 1-up rule (Levitt, 1971). Each track started with a modulation depth of -5 dB, and a step size of 4 dB. After the second and fourth change of step direction, the step size was halved, and the track continued for eight further reversals at the final step size of 1 dB. The threshold was defined as the mean of the modulation depths (in dB) of the last eight reversals in each track. Each test subject completed four tracks for each data point.

3.4.2 Results

Figure 3.4 shows the results of the modulation detection experiment with a plot of the mean detection thresholds across subjects and standard deviations of the subject means. The minimum modulation depths required to detect a sinusoidal AM imposed on the probe stimulus are shown in Fig. 3.4a for four masker conditions: (1) with no AM imposed on the masker stimulus with the masker carrier at 0-ms ITD (filled circles), (2) with no AM imposed on the masker stimulus with the masker carrier at 1-ms ITD (filled squares), (3) with a noise AM imposed on the masker stimulus

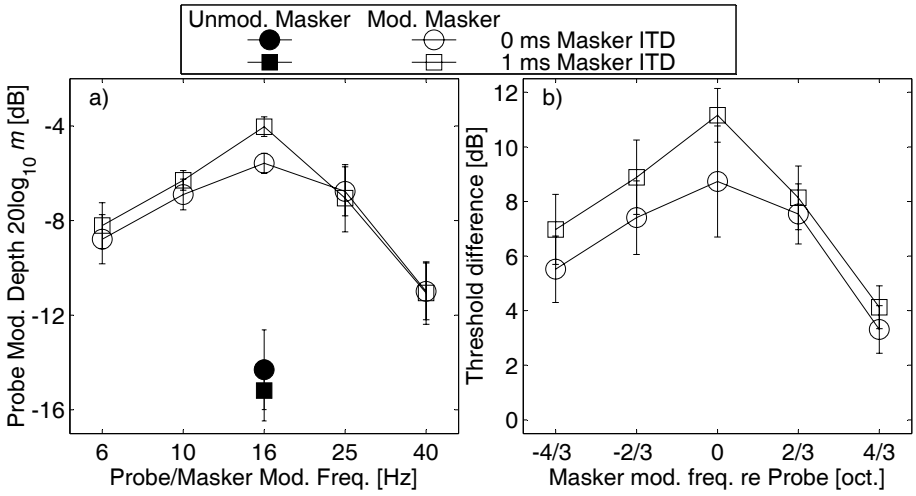

Figure 3.4: a) Thresholds in dB for detection of AM imposed on the probe stimulus with no AM imposed on the masker stimulus (filled symbols) and with a masker AM (open symbols), for a masker with 0-ms ITD (circles) and 1-ms ITD (squares). The probe stimulus always had 0-ms ITD. b) The difference between the thresholds measured with a modulated masker and with no masker modulation as a function of masker modulator center frequency and masker ITD.

with a masker ITD of 0 ms (open circles), and (4) with a noise AM imposed on the masker stimulus with a masker ITD of 1 ms (open squares). Figure 3.4b shows the same data, but replotted as the difference in threshold between the conditions with the modulated masker and the unmodulated masker for corresponding masker ITDs. This can be interpreted as masking in the modulation domain caused by the noise modulator imposed on the masker stimulus.

Statistical analyses were performed on the data using RM-ANOVA with a significance level of 0.05. The thresholds measured with a modulated masker were analyzed with a two-way RM-ANOVA with main effects of masker ITD and masker center frequency. The effect of masker ITD was significant ($\mathrm{F}(1,4) = 10.5$, $p < 0.05$), as was the effect of the masker center frequency ($\mathrm{F}(4,16) = 10.3$, $p < 0.001$). The interaction was not significant ($\mathrm{F}(4,16) = 2.04$, $p = 0.14$). Post-hoc analyses showed that the thresholds measured with the masker at 0-ms ITD were lower than those measured with the masker at 1-ms ITD. The same analysis was made on the

threshold difference data. With these data, the effect of masker ITD was no longer significant ($F(1,4) = 2.7$, $p = 0.18$), and the effect of masker center frequency and the interaction terms are the same as with the previous analysis. The data from the unmodulated-masker conditions were compared with the masker ITD as the main effect. The result showed no significant effect of the masker ITD on the thresholds when there was no AM imposed on the masker stimulus ($F(1,4) = 1.54$, $p = 0.28$).

3.4.3 Discussion

The data do not show any spatial release from the modulation masking. Even though the masker and probe could be lateralized separately, indicating a perceptual segregation, the detection thresholds did not improve. In fact, if there was any significant difference between the measured data points for different masker ITDs, then there was a small *increase* in the modulation detection thresholds when the masker was at a different perceived lateral position than the probe. However, this did not translate into a statistically significant increase in the modulation masking levels. These results were unexpected from the original hypothesis that there would be a spatial release from the modulation masking, in which case, the detection thresholds with a masker ITD of 1 ms would have been lower (less masked) than the detection thresholds measured with a diotic masker.

The measurement data show similar bandpass tuning to that reported in the previous masked modulation detection studies, even though the stimuli used in the present study are very different from those used in previous masked modulation detection studies (e. g., Bacon and Grantham, 1989; Houtgast, 1989; Ewert and Dau, 2000). Houtgast (1989) and Ewert and Dau (2000) imposed the masker and probe modulators in series on a common carrier, either a broadband noise or a pure-tone, while Bacon and Grantham (1989) added the masker and probe modulators together before imposing them on a carrier. In the present study, separate carriers were used for the masker and the probe. The magnitude spectra of the carriers were identical, with only a frequency-dependent phase change related to the temporal delay. This means that the modulation spectra of the masker and probe modulators were added together, as in the study from Bacon and Grantham (1989), instead of convolved, as they are when imposed in series. These procedural differences make a quantitative comparison

difficult, but, qualitatively, the prior and present studies show similar bandpass tuning in the modulation domain.

Other studies have shown interaction between binaural listening and modulation processing. For example, in one dichotic MDI experiment (Bacon and Opie, 1994), there was a small increase in monaural modulation detection thresholds when the masker was played in the opposite ear, although much less than when both target and masker were played in the same ear. Sheft and Yost (1997) showed that ITDs in the masker slightly reduced the amount of masking in another dichotic MDI experiment. These experiments were quite different from the present study, since the target and masker modulators in those studies were imposed on carriers with a large audio-frequency difference, as opposed to the present study where they were applied to carriers with the same audio-frequency content. The MDI studies require interactions across audio-frequency channels, and probably grouping, to explain their results. In the present study, the modulation masking occurs within one audio-frequency channel, and probably occurs more peripherally. It is possible that the binaural interaction affects the higher-level cross-channel processing from the MDI experiments, but creates no release from masking in the within-channel masking from the present experiments.

3.5 Implications for modeling

It was expected that a binaural EC-type model (Durlach, 1963) would predict a spatial release from modulation masking by canceling the masker, in the same way that it would predict a spatial release from masking for a tone-in-noise experiment. In order to test this, the model based on the work from Breebaart et al. (2001a) was used to process the stimuli used in the experiments. This model has been successful at predicting numerous thresholds, including just-noticeable differences for static interaural differences and binaural masking level differences (BMLD) for many combinations of interaural phase and correlation for the signal and noise (see also Breebaart et al., 2001b,c). However, this model was designed and tested for predictions of signal detection thresholds and not explicitly for the perceived lateral position of a sound.

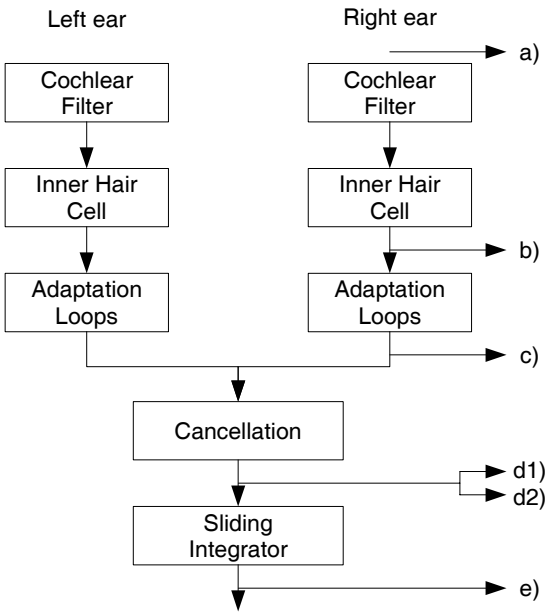

Figure 3.5: Overview of the model structure. Examples of the signals at the stages indicated by the arrows pointing to the right are shown in Fig. 3.6 in the panels corresponding to the letters indicated at the end of the arrows.

A schematic of the model used in the present study to process the signals is shown in Fig. 3.5. Signals were extracted from several stages within the model, as depicted by the arrows pointing to the right, and are plotted in the panels in Fig. 3.6 corresponding to the letter at the end of the arrows. The model consisted of a linear fourth-order gammatone filter (Hohmann, 2002), centered at 5 kHz, to simulate cochlear filtering, followed by half-wave rectification and a fifth-order low-pass filter with a cut-off frequency of 770 Hz, to simulate inner-hair-cell transduction. The inner-hair-cell stage was followed by a chain of five adaptation loops (Dau et al., 1996, 1997a), which was a non-linear stage that simulates properties of neural adaptation. The signals from each monaural channel were then subtracted from each other in the cancellation stage at discrete interaural delays, and its instantaneous energy was calculated. This output was smoothed with a sliding integrator window, which was a double-sided

exponential window with a time constant of 30 ms, acting as a low-pass filter with a cut-off frequency of 5.3 Hz.

In Fig. 3.6, panel (a) shows the amplitude of one ear's input stimulus, which is a combined transposed stimulus, as described above, with an on-frequency masker. For the purposes of demonstrating the binaural stages, the same stimulus was used with a 1 ms masker ITD and a 0 ms probe ITD. The probe stimulus is gated on at 200 ms and off at 500 ms. The internal representations after the various model stages shown in panels b-e (corresponding to the locations shown in Fig. 3.5) are plotted in arbitrary model units. Panel (b) shows the internal representation after the inner-hair-cell stage. This shows that the model has extracted the envelope (including the transposed stimulus) from the input signal. Panel (c) shows the internal representation after the adaptation loops stage. Two internal representations from after the cancellation stage are shown in panels (d1) and (d2), with the internal representation calculated with a 0-ms interaural delay shown in panel (d1), and calculated with a 1-ms interaural delay in panel (d2). These two delays were chosen to demonstrate the cancellation of the probe stimulus and the masker stimulus, respectively. Panel (d1) shows how the probe was canceled, leaving an internal representation of the masker stimulus (note the single density of the pulses between 200 and 500 ms compared to the double density of pulses in panels a-c), although its envelope has been modified as a result of the adaptation loops. Panel (d2) shows how the masker stimulus was canceled, leaving an internal representation of the probe stimulus with no energy before 200 ms or after 500 ms. This internal representation, which clearly shows the original 16 Hz modulations, is almost identical to the internal representation at this stage when the masker is unmodulated, indicating a spatial release from masking. However, the cancellation stage is followed by the sliding integrator in the model, and this smoothed output is shown in panel (e). The output from the 0-ms interaural delay channel is shown with the solid line, and the 1-ms interaural delay channel with the dashed line. These plots show how the sliding integrator smooths over relatively rapid fluctuations, attenuating the 16-Hz probe modulations and rendering the binaural outputs ineffective for predicting modulation detection thresholds. This means that the predictions from the model are consistent with the experimental data, in that a spatial release from modulation masking would not be predicted for these stimuli. The model

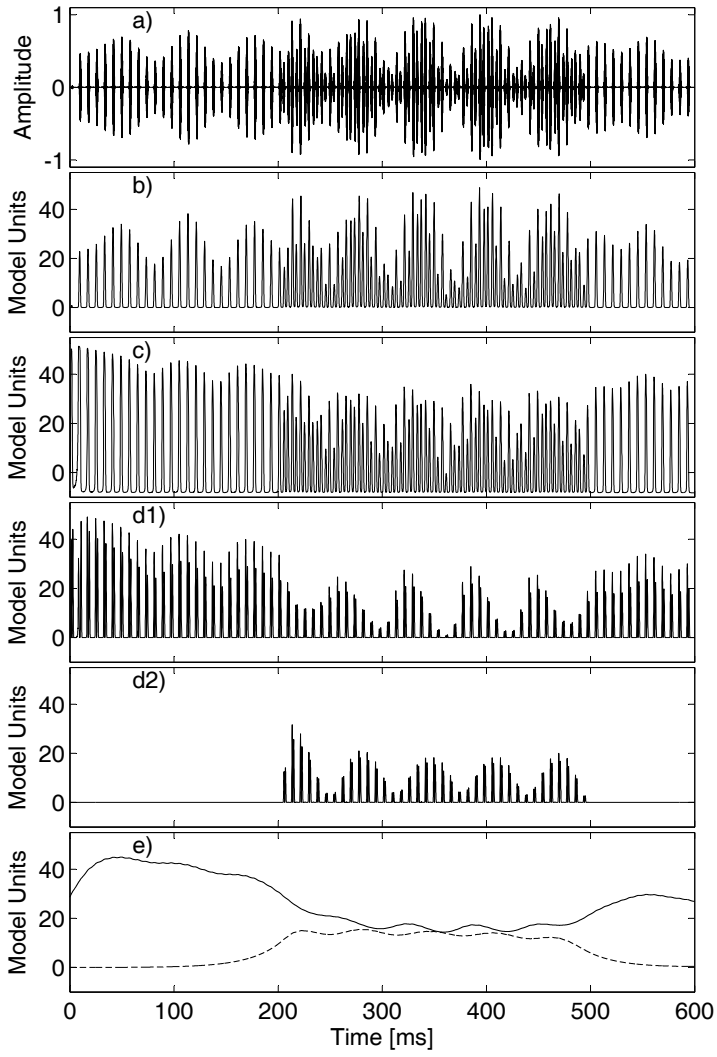

Figure 3.6: Stimulus and internal representations after several stages of the model. a) One ear's stimulus amplitude. For the later binaural calculations, the masker had a 1-ms ITD and the probe had a 0-ms ITD. The model internal representations are plotted in arbitrary model units. b) Internal representation from after the inner hair cell stage. c) Internal representation from after the adaptation loops stage. d1) Internal representation at the output of the 0-ms-ITD channel of the cancellation stage. d2) Internal representation at the output of the 1-ms-ITD channel of the cancellation stage. e) Internal representation after the sliding integrator stage for the 0-ms-ITD (solid line) and the 1-ms-ITD (dashed line) channels.

would predict a spatial release from modulation masking for slow modulations, e. g., around 4 Hz. The experimental data to test this prediction, needs to be collected in further investigations.

The model, as shown in Fig. 3.5 or the original from (Breebaart et al., 2001a), does not predict the bandpass modulation tuning seen in the data. In order to predict this tuning, a bandpass modulation filter, or filterbank, would have to be added after the adaptation loops, as in the model from Dau et al. (1997a).

The Breebaart model was not originally designed to predict the perceived lateral position of sounds, but there is information in the internal representations of the model that could be used to predict lateralization. Previous models of lateralization have used peaks in the distributions of minima with an EC-type model, or maxima in a coincidence-detection-based model (Jeffress, 1948) to predict the perceived lateral position of a sound for static ITD conditions (e. g., Stern and Colburn, 1978; Lindemann, 1986; Park et al., 2005). The minimum across the ITD-channel outputs in the present model as a function of time show where there is the best cancellation, or where the two ears' signals are most similar, which may be used as a cue for lateralization. Before the sliding integrator (Figs. 3.6d1 and d2), the masker and probe are canceled in the 1 and 0-ms-ITD channels, respectively. A histogram of the minima across the ITD channels over time shows a peak at 1 ms and another peak at 0 ms. These peaks may be used to predict the perceived lateral position of the stimuli. However, after the sliding integrator, the distribution of minima is unimodal with only one peak at a position between 0 and 1 ms ITD when both masker and probe stimuli are on. This would indicate that the model would need to predict the perceived lateral position of the stimuli based on internal representations from stages before the sliding integrator is applied. Further development is required with this model in order to predict the perceived lateral position of a sound.

3.6 Summary and conclusions

Two experiments were performed using two temporally-interleaved transposed stimuli as carriers for an AM masker and for an AM probe. The lateralization experiment showed that the two stimuli could be lateralized separately using ITDs. There was a

small amount of repulsion, or pushing, when the masker and probe stimuli had similar ITDs, indicating that the two stimuli were perceptually segregated.

In the masked modulation detection experiment, bandpass modulation tuning was seen in the data, but there was no release from modulation masking on the AM probe when the AM masker was lateralized with a 1-ms ITD. The measured tuning corresponds to the tuning reported in previous monaural studies of masked modulation detection.

Simulations with an existing binaural model showed that the model predictions are consistent with a lack of spatial release from modulation masking as the result of a low-pass filter at the output of the binaural stage. This model, however, would predict a spatial release from modulation masking with very slow (i. e., <5 Hz) modulations. Therefore, a further could investigate whether human listeners gain any benefit in AM detection from a spatial separation between slowly fluctuating stimuli.

Acknowledgments

The research for this article was performed during the first author's extended research stay at the Hearing Research Center at Boston University, which was supported by travel grants from the Denmark-America Foundation and the Idella Foundation. The research was also supported in part by a grant from the National Institutes of Deafness and Communication Disorders (NIDCD) to Barbara G. Shinn-Cunningham (R01 DC05778-02).

Chapter 4

Monaural and binaural subjective modulation transfer functions in reverberation

This chapter is based on Thompson and Dau (2009)

Abstract

The speech transmission index (STI) uses the magnitude of the modulation transfer function (MTF) in a single channel to predict speech intelligibility in rooms. This method often underestimates speech intelligibility when binaural listening is possible. The two ears often have large differences in the MTF, including magnitude and phase differences. Interaural modulation phase differences can create perceivable interaural intensity fluctuations, which can be used to improve detection of intensity modulations. Modulation detection measurements were made monaurally and binaurally with three dichotic impulse responses (IRs) and anechoically at modulation frequencies between 6 and 24 Hz. The first IR consisted of the direct sound and a single ideal reflection arriving at a different time in each ear, and the other IRs were from a simulation of a classroom and a recording from a concert hall. Monaurally, the thresholds measured with two of the IRs could be predicted within about 2 dB based on the anechoic condition's threshold and the magnitude of the MTF. However, with each of the IRs, there was a significant improvement in the modulation detection thresholds when listening binaurally over listening monaurally at a modulation frequency where

there was a large interaural modulation phase difference. The results suggest a way for the STI to incorporate binaural cues.

4.1 Introduction

Several methods exist for objectively predicting the intelligibility of speech in rooms. The two most widely used methods are the speech intelligibility index (SII – ANSI S3.5:1997, 2007) and the speech transmission index (STI – IEC 60268-16, 2003). The standardized versions of these methods are single channel (monaural) and may underestimate speech intelligibility in situations where there is a binaural listening advantage, e. g., when a speech source and an interfering noise source are in different spatial locations (Houtgast and Steeneken, 1973). Two recent studies have attempted to account for binaural processing with these prediction models. An equalization-cancellation (EC) model (Durlach, 1963) was tested with the SII (Beutelmann and Brand, 2006), and an interaural cross-correlation model was tested with the STI (van Wijngaarden and Drullman, 2008). Both studies showed improvements over the original models for some conditions, but both were focused on static binaural parameters, e. g., a fixed ITD or fixed spatial separation. In the present study, the effects of fluctuating binaural parameters, which result from the reverberation itself, were investigated to determine their effect on modulation detection thresholds. A binaural improvement in these thresholds would suggest that these cues could be included in the STI calculations for prediction of binaural speech intelligibility.

Of the two indices, the STI makes more use of the temporal aspects of the signal, assuming that the changes in a speech signal with time must be accurately transmitted in order for the speech to be intelligible, whereas the SII predicts the speech intelligibility based on long-term signal-to-noise ratios (SNR). Since this study was interested in temporally varying signals, the STI was chosen as a starting point.

The STI is based on the concept of the modulation transfer function (MTF). When a signal is transmitted through a channel (e. g., a room), the temporal intensity envelope of the signal is changed. The MTF measures the amount of attenuation imposed on the signal's intensity envelope as a function of the modulation frequency. In general, reverberation in a room acts as a low-pass modulation filter, attenuating

rapid intensity fluctuations and introducing a modulation-phase shift. Additive steady-state broadband noise uniformly attenuates modulation frequencies (Houtgast and Steeneken, 1985). For the STI, the MTF is measured for modulation frequencies in 1/3-octave intervals from 0.63 Hz to 12.5 Hz in audio-frequency bands in octave intervals from 125 Hz to 8 kHz. The values of the MTF are converted to SNRs and averaged across modulation frequencies within each audio-frequency band. These averaged values are then weighted, averaged across audio-frequency bands, and scaled to produce a single number between 0.0 and 1.0, with 0.0 predicting complete unintelligibility and 1.0 predicting complete intelligibility. The standardized STI is a single-channel measurement, and only considers the magnitude of the MTF, discarding the modulation-phase information. However, the physical separation of the ears by the head can create large differences in the MTF to each ear. Those interaural MTF differences may be able to be exploited by the auditory system to restore some of the envelope information obscured by the reverberation. The aim of this study was to investigate the circumstances under which a binaural advantage can be found in the detection of intensity fluctuations in reverberation.

The complex MTF in a room can be derived from the impulse response (IR) of the channel (Houtgast and Steeneken, 1973; Schroeder, 1981), assuming linearity. Since uncorrelated signals add linearly on an intensity basis, it is the transmission of the *intensity* envelope of the signal and not the *amplitude* envelope that is described by the MTF. Therefore, the MTF is derived by simply calculating the Fourier transform of the squared (amplitude) impulse response and normalizing by its total energy:

$$\mathrm{MTF}(f_m) = \frac{\int_0^\infty h^2(t) e^{-j2\pi f_m t} dt}{\int_0^\infty h^2(t) dt}, \tag{4.1}$$

where f_m is the modulation frequency and $h^2(t)$ is the squared impulse response. If a signal with a sinusoidal intensity envelope is transmitted through a room, with the input intensity given by:

$$i_0[1 + \sin(2\pi f_m t)], \tag{4.2}$$

where i_0 is the mean intensity, then, the measured output signal will have an intensity:

$$i_r[1 + m\sin(2\pi f_m t + \phi)] \tag{4.3}$$

where i_r is the mean output intensity, m is the output modulation depth, and ϕ is the output modulation phase. The output intensity i_r depends on the distance from the source and the reverberation level in the room. Since the input signal was fully modulated with zero starting phase, m is also the magnitude of the MTF at f_m, and ϕ is the phase-shift of the MTF at f_m resulting from the reverberation pattern.

The reverberation pattern in a room usually consists of a series of several discrete reflections arriving relatively soon after the direct sound, followed by a noise-like, exponentially-decaying reverberation tail (Kuttruff, 2000, ch. 4.2). Speech intelligibility is often enhanced by the arrival of early reflections within about 50 ms of the direct sound (Cremer and Müller, 1982; Bradley et al., 2003; Yang and Bradley, 2009). Those early reflections are assumed to be integrated with the direct sound, thereby increasing the energy in the perceived speech signal and improving the SNR (Haas, 1951; Lochner and Burger, 1964; Bradley et al., 2003). Several metrics have been defined in the room acoustics literature based on the relative energy levels in the early and late part of the room impulse response. For example, Deutlichkeit (definition), D_{50}, is the ratio of early energy (within the first 50 ms after the direct sound) and the total energy in the impulse response (Thiele, 1953), and the useful-to-detrimental ratio, U_{50}, is the ratio of the early-arriving speech energy and the sum of the late-arriving speech energy and the background noise (Lochner and Burger, 1964). These objective metrics, and others, have also been used as simple predictors of speech intelligibility in rooms (see, e. g., Bradley, 1986).

Early reflections can often be approximated as discrete, ideal reflections (assumed here to be Dirac delta pulses) arriving at time t_k relative to the arrival of the direct sound, with amplitude a_k relative to the amplitude of the direct sound. A single reflection (with $k = 1$) will create an MTF given by:

$$\text{MTF}\,(f_m) = \frac{1 + a_1^2 e^{-j2\pi f_m t_1}}{1 + a_1^2}, \tag{4.4}$$

where f_m is the modulation frequency. This MTF acts as a comb filter with magnitude:

$$|\text{MTF}\,(f_m)\,| = \frac{\sqrt{1+2a_1^2\cos(2\pi f_m t_1)+a_1^4}}{1+a_1^2} = \frac{\sqrt{(1+a_1^2)^2-4a_1^2\sin^2(\pi f_m t_1)}}{1+a_1^2}, \tag{4.5}$$

which has minima at $f_m = n/2t_1$, for $n = 1, 3, 5, \ldots$, and maxima at $f_m = n/t_1$, for $n = 0, 1, 2, \ldots$ (Houtgast and Steeneken, 1973). The phase of the MTF in Eq. 4.4 is:

$$\angle\text{MTF}\,(f_m) = \tan^{-1}\left(\frac{-a_1^2\sin(2\pi f_m t_1)}{1+a_1^2\cos(2\pi f_m t_1)}\right). \tag{4.6}$$

Therefore, a single reflection arriving within 50 ms of the direct sound would create a minimum in the MTF at $f_m > 10$ Hz, and later reflections are required to have a minimum in the MTF at lower modulation frequencies. The STI only considers modulation frequencies less than 12.5 Hz. This range of modulation frequencies is most affected by reflections arriving more than 40 ms after the direct sound, which corresponds quite well with the underlying concept of D_{50} and U_{50} that the late reflections are most detrimental for speech.

When there are n reflections occuring at times t_k for $k = 1 \ldots n$ with amplitudes a_k for $k = 1 \ldots n$ relative to the direct sound, the magnitude of the MTF is:

$$|\text{MTF}(f_m)| = \frac{\sqrt{E_h^2 - 4\sum_{k=1}^{n} a_k^2\sin^2(\pi f_m t_k) - 2\sum_{k=1}^{n}\sum_{i=1}^{n} a_k^2 a_i^2\sin^2[\pi f_m(t_k - t_i)]}}{E_h}, \tag{4.7}$$

where $E_h = 1 + \sum_{k=1}^{n} a_k^2$ is the total energy in the IR. The magnitude of the MTF is affected by the relative timing of each reflection to the direct sound and to every other reflection. The phase response is:

$$\angle\text{MTF}(f_m) = \tan^{-1}\left(-\frac{\sum_{k=1}^{n} a_k^2\sin(2\pi f_m t_k)}{1+\sum_{k=1}^{n} a_k^2\cos(2\pi f_m t_k)}\right). \tag{4.8}$$

The timing of all reflections contribute to the phase shift in the MTF at each modulation frequency.

An extreme example of an IR with many reflections is white noise with an exponential decrease in intensity with time, such as is often used as a model for the reverberation tail in a room. The MTF for such an IR, with decay time constant τ, is given by (from Schroeder, 1981):

$$\mathrm{MTF}\,(f_m) = \left(1 + j\frac{2\pi f_m \tau}{2}\right)^{-1} = \left(1 + j\frac{2\pi f_m T_{60}}{13.8}\right)^{-1}. \tag{4.9}$$

Equation 4.9 makes use of the relation between the time constant τ and the reverberation time T_{60} from the definition of T_{60}:

$$e^{\frac{-\tau}{T_{60}}} = 10^{-\frac{60}{20}} \qquad \Rightarrow \qquad \tau = \frac{T_{60}}{6.9}. \tag{4.10}$$

The magnitude of the MTF in eq. 4.9 is:

$$|\mathrm{MTF}\,(f_m)\,| = \left[1 + \left(\frac{2\pi f_m T_{60}}{13.8}\right)^2\right]^{-1/2}, \tag{4.11}$$

which acts as a low-pass filter, and the phases are random, due to the white-noise carrier. The cut-off frequency of the filter is:

$$f_c = \frac{13.8}{2\pi T_{60}}, \tag{4.12}$$

so higher reverberation times result in lower cut-off frequencies in the low-pass MTF. For a typical IR, the MTF will be the complex sum of the comb filters from the early reflections (Eqs. 4.7 and 4.8) and the low-pass filter from the noise-like reverberation tail (Eq. 4.9).

When calculating the STI, the MTF is typically measured with an omni-directional microphone. This is a convenient method, but it discards all spatial information from the impulse response and any benefit that might be derived from it. An extension of the STI approach could be to investigate the effect of binaural processing on the subjective MTF (Houtgast and Steeneken, 1973), i. e., the difference

between modulation detection thresholds in anechoic and reverberant environments:

$$\text{MTF}_{\psi}(f_m) = \theta_{an}(f_m) - \theta_{rev}(f_m), \tag{4.13}$$

where MTF_{ψ} is the subjective MTF, and θ_{an} and θ_{rev} are the modulation detection thresholds in anechoic and reverberant conditions, respectively, expressed in dB. An implicit assumption in the standardized STI is that the modulation detection thresholds in a reverberant environment can be predicted from the thresholds in an anechoic environment and the magnitude of the physical MTF, or that the physical and subjective MTFs have the same magnitude. Miyata et al. (1991) measured the monaural and binaural subjective MTF at 2.8 Hz and speech intelligibility in a reverberation chamber and compared them to RASTI (RApid STI, Houtgast and Steeneken, 1984) measurements made with an omni-directional microphone as well as with a KEMAR® (G.R.A.S. Sound & Vibration) manikin head. In their highly reverberant environment, the subjective MTF measured binaurally was similar to that measured with the better of the two ears and was also similar to the physical MTF. The measured speech intelligibility scores were also best predicted by the physical MTF measured at the better of KEMAR's two ears. Therefore, they concluded that the benefit of listening with two ears in that very reverberant environment was from 'selective better-ear listening,' where the best available information about the source signal from either ear is used for the binaural detection. However, it is unclear whether the results from that study in an extremely reverberant environment can be generalized to more normal listening environments.

The hypothesis of the present study was that an interaural modulation phase difference (IMPD) in the MTF can lead to a binaural advantage in a subjective MTF beyond 'better-ear listening.' An IMPD creates a fluctuating interaural intensity difference (IID):

$$\text{IID}(f_m, t) = 10 \log \frac{i_{r,L}\left[1 + m_L \sin\left(2\pi f_m t + \phi_L\right)\right]}{i_{r,R}\left[1 + m_R \sin\left(2\pi f_m t + \phi_R\right)\right]} \tag{4.14}$$

using Eq. 4.3 with the L and R subscripts indicating the left and right ears, respectively. This fluctuating IID may be detectable as a fluctuation in the perceived lateral

position of the sound (Grantham, 1984; Thompson and Dau, 2008), which may be an additional cue to detect the modulation signal. A simulation of an environment similar to that used in the study from Miyata et al. (1991) resulted in an MTF with only small interaural differences at the modulation frequency used in their measurements. Therefore, experiments were designed for the present study to test whether there can be a binaural advantage in a subjective MTF with less reverberant IRs and at modulation frequencies where large interaural differences in the MTF phase were calculated. The thresholds of detectability for intensity modulation imposed on a broadband noise carrier were measured under anechoic conditions and when convolved with dichotic IRs. The first IR consisted of the direct sound at time 0 and an ideal reflection in each ear with an interaural arrival time difference. The second IR was from a simulation of a classroom, and the third IR was from a binaural recording in a concert hall.

4.2 Methods

The minimum modulation depth required to detect a sinusoidal intensity modulation imposed on a broadband pink-noise carrier was measured. The thresholds were measured monaurally and binaurally via headphones with the stimuli presented alone (anechoic) or convolved with a dichotic IR. In contrast to many previous modulation detection studies (e. g., Viemeister, 1979; Dau et al., 1997a; Kohlrausch et al., 2000) that have used sinusoidal *amplitude* modulation (AM) signals, sinusoidal *intensity* modulation (IM) signals were used in this study in order to fit with the MTF concept. At small modulation depths ($m \ll 1$), the amplitude of a sinusoidal IM is approximately the same as the amplitude of a sinusoidal AM with half the modulation depth (i. e., $m_{int} \approx m_{amp}/2$). This means that IM detection thresholds should be about 6 dB higher than AM detection thresholds, assuming all other experimental parameters are held constant.

4.2.1 Stimuli

An independent Gaussian-pink-noise (0.1-5 kHz) carrier was generated for each interval in each trial. The target stimuli were designed to have a sinusoidal IM, so the amplitude of the stimuli was defined as follows:

$$x(t) = [1 + m\sin(2\pi f_m t + \phi)]^{\frac{1}{2}}\, n(t), \tag{4.15}$$

where m is the modulation depth, f_m is the modulation frequency, ϕ is the starting phase of the modulation, and $n(t)$ is the noise carrier. The starting phase ϕ was randomly selected from a uniform distribution over $[0, 2\pi]$ for each target interval. The modulation depth is often described on a dB-scale as $20\log_{10} m$ (e. g., Viemeister, 1979; Dau et al., 1997a), and will be presented here in the same fashion. Each stimulus in the non-anechoic tracks was convolved with a dichotic IR and the levels of the two ears' stimuli were scaled together so that the louder ear was presented at 65 dB SPL. Then a raised-cosine window with a 400 ms equivalent rectangular duration (ERD) and 100 ms ramps was applied to each stimulus. The three intervals were separated by silent gaps of 250 ms.

The measurements were made with three IRs. The first IR was artificially created and consisted of a Dirac impulse with an amplitude of 1 at $t = 0$ and a single, ideal reflection with an amplitude of 0.7 (-3 dB re direct sound) and a different arrival time in each ear. Since the stimuli were presented over headphones, the reflections could be controlled exactly, and there was no crossover of the reflections from one ear to the other. The reflection arrived in the left ear 55.6 ms after the direct sound, and the reflection in the right ear arrived 41.7 ms after the direct sound (see Fig. 4.1, label 'IR 9/12'). Note that a 13.9 ms interaural arrival time difference is not realistic for a single reflection, but these reflection arrival times were chosen to create minima in the magnitudes of the MTF at 9 Hz and 12 Hz in the left and right ears, respectively, as well as an IMPD in the MTF. The magnitudes of the MTFs for the different IRs are shown in Fig. 4.2, with the left ear's MTF plotted with a dashed line and the right ear's MTF plotted with a dotted line using the left ordinate. In addition, the IMPD is plotted in the same figure with a solid line, using the right ordinate. The IR described above, and the corresponding measurement results, will be referred to with the label '9/12'.

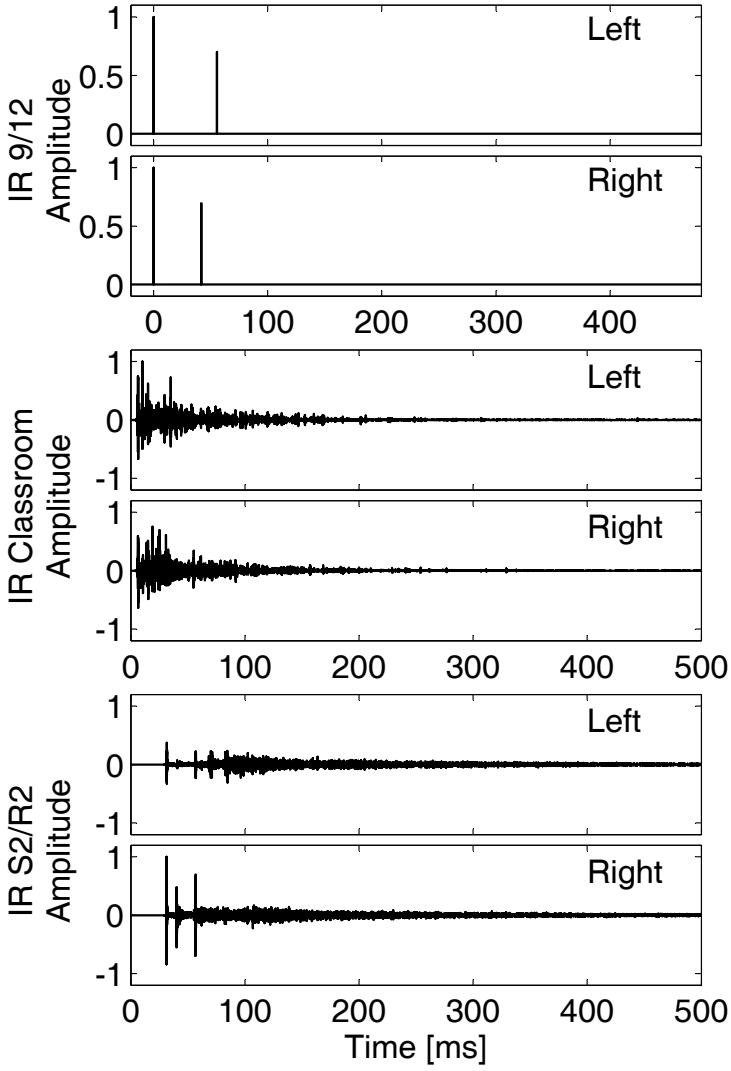

Figure 4.1: The impulse responses used in the first measurements. The top two panels show the left and right ear IRs for '9/12'. The middle two panels show the left and right ear IRs from the classroom. The bottom two panels show the left and right ear IRs from the concert hall, position S2/R2.

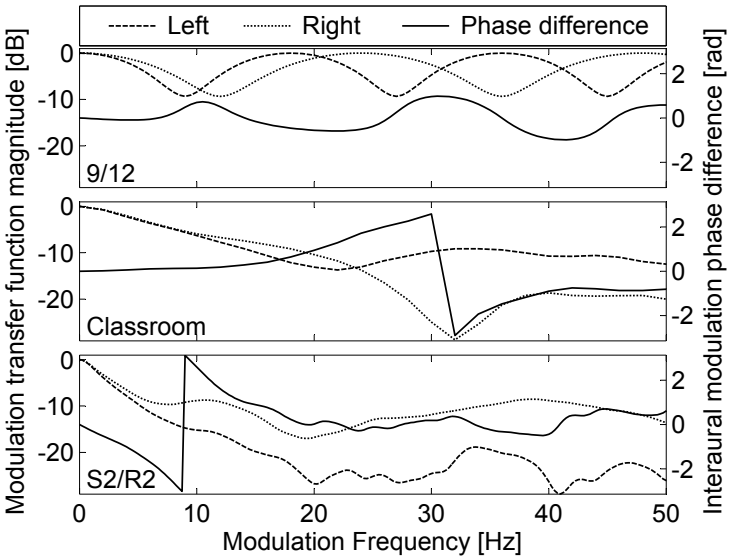

Figure 4.2: The MTFs derived from the IRs used in the first measurements (see Fig. 4.1). The dashed and dotted lines are the magnitudes of the MTFs for the left and right ear IRs, respectively, and refer to the left ordinate. The solid lines show the interaural modulation phase differences (IMPD) between the left and right ear MTFs, referring to the right ordinate. The top panel is for IR '9/12', the middle panel for IR 'Classroom' and the bottom panel for IR 'S2/R2'.

The second IR was from a simulation of a classroom made with the Odeon room acoustics simulation package (Christensen, 2005). The room has a reverberation time (T_{30}) of about 0.8 s in the mid-frequencies. The IR and corresponding MTF are shown in the middle panels of Figs. 4.1 and 4.2, respectively, with the label 'Classroom'. The third IR was a binaural recording from the Pori Promenadikeskus concert hall in Finland (Merimaa et al., 2005b), recorded in 2002 using a manikin head (Brüel & Kjær HATS) and diffuse field equalized to remove the filtering of the manikin's ear canals. The concert hall has a T_{30} of about 2.2 s. This IR is labeled as 'S2/R2' (label from Merimaa et al., 2005a) referring to source position 2 (on-stage, 5.2 m from the front of the stage and 4 m to the right of midline) and receiver position 2 (7^{th} row seat, 6 seats right of midline), and is shown in the bottom panel of Fig. 4.1 with its MTF shown in Fig. 4.2.

At first glance, it is perhaps surprising that there can be an IMPD of about π at modulation frequencies less than 10 Hz (see Fig. 4.2, S2/R2) in a 'real' environment, since a π-phase shift requires a 50 ms temporal shift at 10 Hz, which is much longer than the maximum ITD created by the physical distance between the ears. However, this can be understood by looking at the reflection patterns in the S2/R2 IR in Fig. 4.1 and considering Eq. 4.8. The S2/R2 IRs show two strong reflections in the right ear within about 30 ms of the direct sound, and a lot of energy from the left ear arriving much later. Since the MTF phase at a particular modulation frequency is affected by the timing and energy of every reflection, these differences in temporal energy distribution between the two ears results in a large IMPD at low modulation frequencies.

With each IR, modulation frequencies were selected for which the IMPD was large, but where the magnitude of the MTF was neither too small nor too large. If the MTF magnitude was too large (i. e., $|\text{MTF}| > -6$ dB), then monaural detection would be too close to the anechoic condition's thresholds, and there would be little room to show a binaural effect (if there was one). On the other hand, if the MTF magnitude was too small (i. e., $|\text{MTF}| < -14$ dB), then the intensity fluctuations of the output stimulus might be too small to generate any perceivable IID fluctuations.

4.2.2 Procedure

The experiment used a 3-interval, 3-alternative forced choice design with two reference intervals that had no IM and a target interval, randomly selected in each trial, that had an imposed IM. The modulation depth m was determined for each trial using an adaptive 1-up, 2-down tracking rule (Levitt, 1971). The initial modulation depth was -4 dB ($20 \log_{10} m$), and the initial step size was 4 dB. After every second change of direction, the step size was halved until the final step size of 1 dB was reached. The track continued for six further reversals, and the threshold was defined as the mean of the modulation depths at the last six reversals. The tracks were presented in blocks of listening condition (monaural and binaural) and IR, and the blocks were presented in random order. Within each block, the tracks for each modulation frequency were presented in random order. Each test subject completed four repetitions for each data point. The test subjects could stop the experiment after the completion of any track

and continue from the same point at a later time. Typically, the sessions lasted about thirty minutes before a break.

4.2.3 Test subjects

Six test subjects participated in this experiment. Five measured data with IR '9/12', four of those five measured with the 'Classroom' IR, and three of those four plus one additional test subject measured with the 'S2/R2' IR. They were not paid directly for their participation, but were directly affiliated with the research center, and included the first author of this article. All had pure-tone audiometric thresholds of 15 dB HL or better for octave frequencies from 125 Hz to 8 kHz, and were experienced in psychoacoustic tests, including other AM detection experiments. The experiments were performed with the approval of the ethics committee of the Technical University of Denmark.

4.2.4 Equipment

The stimuli were generated and presented using the AFC-Toolbox for MATLAB® (The MathWorks), developed at the University of Oldenburg, Germany and the Technical University of Denmark. The sounds were presented at a sampling rate of 48 kHz via a high-quality sound card (RME DIGI 96/8 PAD) and headphones (Sennheiser HD-580). During the experiment, the test subjects were seated in a sound-insulated test booth with a computer monitor, which displayed instructions and visual feedback, and a keyboard for response input.

4.3 Results

With the '9/12' IR, measurements were made at modulation frequencies of 6, 9, 10.5, 12 and 15 Hz in order to sample across the first minima in the left (dashed curve in Fig. 4.2, panel '9/12') and right (dotted curve) ears' MTFs at 9 and 12 Hz, respectively, as well as the first maximum in the IMPD (solid curve) at 10.5 Hz. The results are shown in Fig. 4.3a. The anechoic results (circles) are approximately constant at -14 dB for all measured modulation frequencies. A one-way analysis of variance with

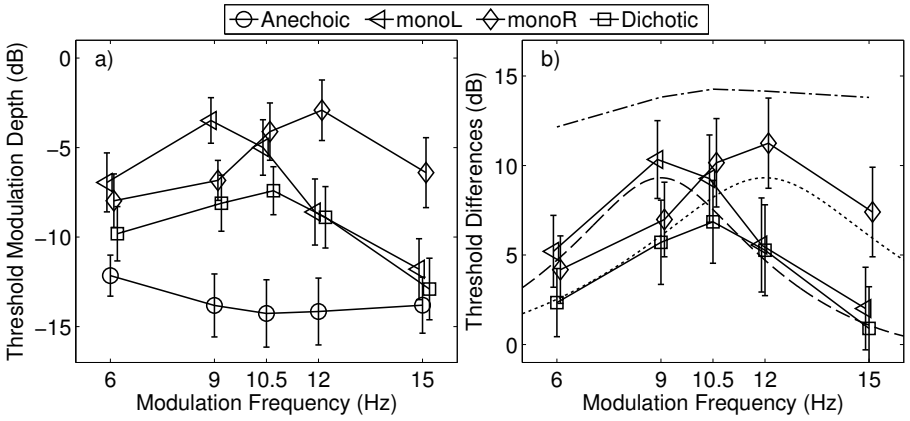

Figure 4.3: Panel a) shows the minimum modulation depth in dB ($20\log_{10} m$) required for detection of a sinusoidal intensity-modulation signal imposed on a broadband noise carrier as a function of the modulation frequency, measured with IR '9/12'. The measured conditions were anechoic (circles), left ear only (triangles), right ear only (diamonds) and dichotically (squares). Panel b) shows the difference between the thresholds obtained in the presence of a reflection and the thresholds in the anechoic condition. The auxiliary lines show the inverse of the theoretical MTF derived from the BRIR (dashed and dotted lines for the left and right ears, respectively) and the maximum possible threshold difference (i. e., 0 dB - anechoic condition's threshold, plotted with the dash-dot line). Note that the data points have been offset slightly around the measured frequencies to enhance visibility of the error bars.

repeated measures (RM-ANOVA) on the subjects' mean thresholds in the anechoic conditions showed a significant effect of modulation frequency (F(4,16)=4.6, $p < 0.05$). A post-hoc Tukey's HSD test showed that the mean threshold at 6 Hz was significantly higher ($p < 0.05$) than the mean threshold at 10.5 and 12 Hz. The elevated threshold at 6 Hz is similar to the increased modulation detection thresholds reported by Viemeister (1979) with gated carriers for modulation frequencies below about 8 Hz. This increase could be the result of modulation masking from the windowing function used, or that the listener was only presented with two cycles of the modulation when the windowing function was at its maximum level.

When the stimuli were convolved with the dichotic IR, the thresholds increased significantly. The left ear's threshold curve (triangles in Fig. 4.3) shows a maximum at 9 Hz, and the right ear's threshold curve (diamonds) shows a maximum at 12 Hz. The binaural (dichotic) threshold curve (squares) has a maximum at 10.5 Hz. A two-

way RM-ANOVA with main factors of listening condition (mono-left, mono-right and binaural) and modulation frequency showed significant effects of both factors and a significant interaction ($p < 0.001$ for each analysis). A post-hoc Tukey's HSD test showed significant differences ($p < 0.05$) between the binaural and right ear thresholds at all but the 9 Hz modulation frequency, and significant differences ($p < 0.05$) between the binaural and left ear thresholds for the 6, 9 and 10.5 Hz modulation frequencies.

Thresholds were obtained at only one modulation frequency, 24 Hz, with the 'Classroom' IR. At this modulation frequency, the magnitude of either ears' MTF is just greater than -14 dB (-13.6 and -13.0 dB). Therefore, assuming a threshold in the anechoic condition of about -14 dB, it was expected that the signal modulation should be just detectable with full input modulation (i. e., $m = 0$ dB). In addition, there is a large IMPD (1.7 rad) in the MTFs at 24 Hz. Therefore, it was expected that there would be a significant binaural advantage in the modulation detection experiment at this modulation frequency. The results of the measurement are shown in Fig. 4.4a. The anechoic condition's threshold is about -14 dB, similar to those reported at lower modulation frequencies with IR '9/12' (see Fig. 4.3a). Monaurally, the thresholds could not be measured reliably by any of the subjects with either ear alone (indicated by the asterisk at 0.5 dB), but none of the subjects had difficulty with the binaural detection, showing a mean threshold of -2.6 dB. The mean threshold difference between the reverberant and anechoic condition was 11.2 dB, which was significantly ($p < 0.01$) less than the model prediction of the smaller absolute value of the two ears' MTF magnitude (13.0 dB).

For the 'S2/R2' IR, measurements were made with modulation frequencies of 6, 8, and 10 Hz. The thresholds measured in the anechoic condition during the 'S2/R2' IR block were similar to those measured during the '9/12' IR block, with the threshold for 6 Hz slightly higher than the thresholds at higher frequencies (see Fig. 4.5a). A one-way RM-ANOVA showed a significant main effect of modulation frequency on the anechoic condition's thresholds (F(2,6)=26.5, $p < 0.01$), with a post-hoc analysis showing that the threshold at 6 Hz was significantly higher than the other two measured thresholds. The thresholds were unmeasurable at 8 and 10 Hz with the left ear (indicated by the asterisks at 0.5 dB) when convolved with

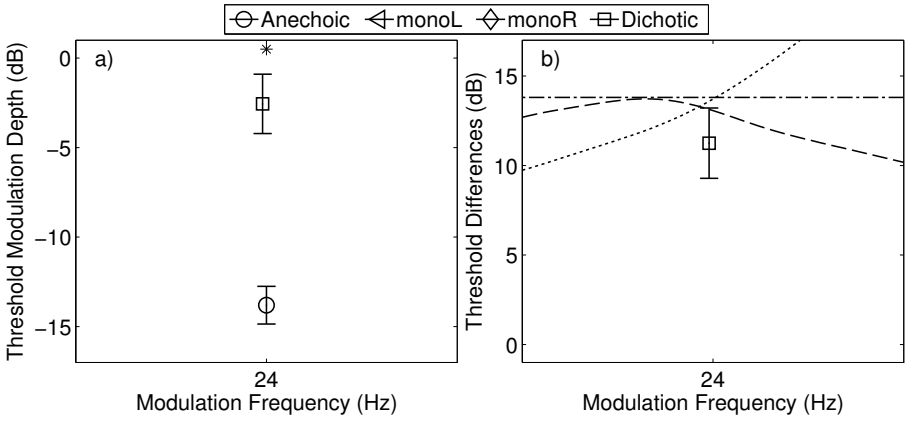

Figure 4.4: Panel a) shows the minimum modulation depth in dB ($20 \log_{10} m$) required for detection of a sinusoidal intensity-modulation signal imposed on a broadband noise carrier at a modulation frequency of 24 Hz, measured with IR 'Classroom'. Thresholds were measured under anechoic conditions (circle), and with the stimuli convolved with a binaural room impulse response (square). The thresholds could not be obtained monaurally (indicated by the asterisk at 0.5 dB). Panel b) shows the difference between the threshold obtained with the BRIR and the thresholds from the anechoic condition. The auxiliary lines show the inverse of the theoretical MTF derived from the BRIR (dashed and dotted lines for the left and right ears, respectively) and the maximum possible threshold difference (i. e., 0 dB - threshold in the anechoic condition, plotted with the dash-dot line).

the 'S2/R2' IR, because the modulation signal was not reliably detectable with full input modulation depth ($m = 0$ dB). Therefore, the left-ear data were excluded from the following ANOVA. A two-way RM-ANOVA with main effects of modulation frequency and listening condition on the right-ear and dichotic thresholds showed no significant effect of listening condition (F(1,3)=1.5, $p = 0.31$) or of modulation frequency (F(2,6)=1.8, $p = 0.25$), but there was a significant interaction (F(2,6)=25.7, $p < 0.01$). Post-hoc analysis with Tukey's HSD showed that the threshold measured with the dichotic stimulus with a 6 Hz modulation was significantly lower than the threshold measured with the right ear alone, but there were no significant differences at the other modulation frequencies.

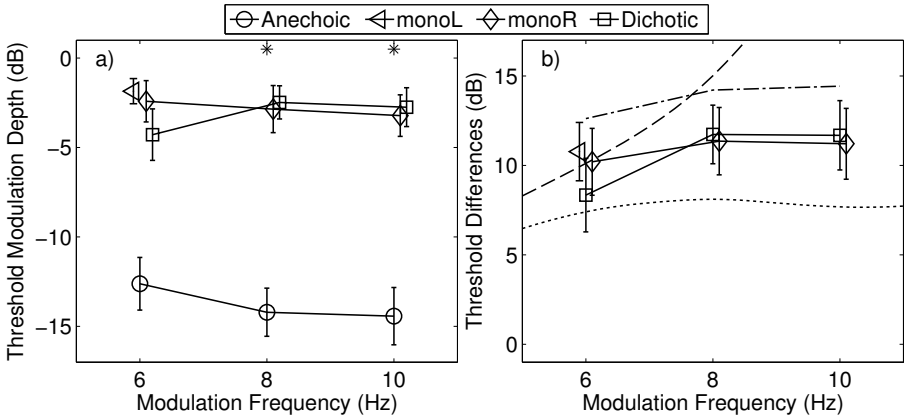

Figure 4.5: Panel a) shows the minimum modulation depth in dB ($20 \log_{10} m$) required for detection of a sinusoidal intensity-modulation signal imposed on a broadband noise carrier at a modulation frequency of 24 Hz, measured with IR 'S2/R2'. Thresholds were measured under anechoic conditions (circle), left ear only (triangles), right ear only (diamonds) and dichotically (squares). The thresholds could not be obtained monaurally (indicated by the asterisks at 0.5 dB) for the left ear at 8 or 10 Hz. Panel b) shows the difference between the threshold obtained with the BRIR and the thresholds from the anechoic condition. The auxiliary lines show the inverse of the theoretical MTF derived from the BRIR (dashed and dotted lines for the left and right ears, respectively) and the maximum possible threshold difference (i. e., 0 dB - threshold from the anechoic condition, plotted with the dash-dot line).

4.4 Discussion

The 'better-ear' MTF hypothesis would predict the binaural reverberant conditions' thresholds using the anechoic condition's thresholds and the maximum of the two ears' MTFs (i. e., least modulation attenuation) for a given f_m. This hypothesis holds for much of the data measured with the '9/12' IR, since the binaural detection thresholds are not significantly different from the minimum of the two monaural curves, except at a modulation frequency of 10.5 Hz. At that modulation rate, the binaural threshold was significantly better than either of the monaural thresholds and than the 'better-ear' model predictions. This is also the modulation frequency for which the IMPD is at a local maximum for this IR, and at which the largest interaural level fluctuations would be found. This lends support to the original hypothesis that an IMPD could create a binaural cue for IM detection.

The predictions from the 'better-ear' hypothesis were tested by replotting the threshold differences from Fig. 4.3a in Fig. 4.3b. These points were calculated by taking the mean threshold for each condition and modulation frequency for each listener and subtracting the mean threshold from the anechoic condition for that listener. The points are plotted as the mean of the threshold difference across listeners plus/minus one standard deviation. The negative of the MTF curves for the left and right ears are plotted with dashed and dotted lines, respectively, which should serve as the model for $\theta_{rev}(f_m) - \theta_{an}(f_m)$ (cf. Eq. 4.13). The dash-dotted curve in the plot is the negative of the anechoic condition's threshold. This represents where the threshold difference would lie if the reverberant threshold were at the maximum possible modulation depth of 0 dB (to avoid overmodulation). For modulation frequencies where the -MTF curve is greater than this curve, the prediction would be that the reverberant threshold would be unmeasurable, i. e., the modulation signal would be undetectable even with a fully modulated input stimulus.

Qualitatively, the MTF is a good predictor of the reverberant results. The two monaural data curves are always close to, but greater than the model, and the binaural data curve is very close to the prediction. The models can account for 73, 48 and 86% of the variance of the threshold differences for the monaural left, right and binaural data curves, respectively. An analysis of the remaining variance showed no significant effect of modulation frequency (F(4,16)=0.238, $p = 0.91$), but still a significant effect of listening condition (F(2,8)=8.2, $p < 0.05$) and a significant modulation frequency/listening condition interaction (F(8,32)=2.4, $p < 0.05$). Post-hoc analysis suggested that the fit of the model to the right ear's data was significantly worse than to the binaural data.

Some of the binaural advantage can be explained with signal detection theory and two independent observers (ears) (see, e. g., Green and Swets, 1966). In signal detection theory, it is assumed that an observer converts a physical stimulus into a perceived quantity of some internal variable through a noisy process. The internal noise level of this conversion limits the sensitivity of the detector, often called d'. If two independent observers are presented with a stimulus, then each will measure a value of the internal variable, and each will have a sensitivity d' to the signal. Their joint sensitivity to the stimulus is the square-root of the sum of the squared sensitivities

(i. e., $d'^2_{1,2} = d'^2_1 + d'^2_2$). For two equally sensitive observers, this would predict an increase in overall sensitivity of the joint observation by a factor of $\sqrt{2}$, or 3 dB.

In the analysis of threshold differences, it was assumed that the left and right ears have equal sensitivities to the modulation signal. This allows a comparison of the anechoic data, which was measured diotically, with the monaural data. However, since the model fit for the right ears' data was worse than for the left ears', it is possible that the left ears were more sensitive than the right ears. In this case, the anechoic condition's thresholds would be mostly determined by the left ears, thereby predicting a relatively good fit to the MTF model. If the right ears were less sensitive than the left ears, then the monaural-right thresholds from the anechoic condition would be slightly higher than the left-ear thresholds, which would result in an improved fit of the right ears' data by the MTF model. There is no *a priori* reason to assume that the right ears are less sensitive to modulation than the left ears, but this should be tested in further experiments by comparing diotic modulation detection thresholds with monotic thresholds for each ear.

The thresholds measured with the 'Classroom' IR (Fig. 4.4a) are interesting because a threshold was measurable in the dichotic condition even though the thresholds could not be measured in either of the monaural conditions. In addition, the difference between the dichotically measured threshold and the anechoic condition's threshold was significantly smaller than the better-ear MTF model. This shows that there can be a binaural advantage over monaural listening in the detection of modulation at a modulation frequency where there is a large IMPD, and it suggests that there may be a need to extend the STI to account for this advantage. Further testing with other IRs will be required to conclusively determine whether there is a causal relationship between the IMPD and this binaural advantage.

The 'S2/R2' IR presents a slightly different picture in the results than the other two IRs. With the first two IRs (Figs. 4.3b and 4.4b), the dichotic threshold differences were very close to, or slightly below, the prediction from the MTF, but with this IR (Fig. 4.5b), the MTF model is only a good fit for the dichotic data at 6 Hz. At 8 and 10 Hz, the dichotic threshold difference is significantly worse than the MTF model and is better predicted by the monaural-right data. The monaural data follow the trend established with the '9/12' data that the right ears' threshold differences are almost

3 dB larger than the MTF model predicts, and the left ears' mean difference at 6 Hz is close to the MTF model.

The two room IRs present some challenges for the data analysis. The MTF analysis that was done for this study was based on the broadband IR. The STI, and presumably the human ear, analyzes the signal in narrow frequency bands – octave bands for the STI, and cochlear filters for the human ear (see, e. g., Moore, 2003). The MTF and interaural MTF differences will be different for the different bands, and therefore the IID fluctuations will also vary across bands. The '9/12' IR did not have this problem, because the reflections were Dirac pulses, so the MTFs were the same in any frequency band. With different sensitivities in each audio-frequency band, the broadband modulation detection threshold could be based on the threshold of the most sensitive band, or the sensitivities of the different audio bands could be combined with the square-root of the sum of the squares of the sensitivities. Either method would predict that the multi-band signal detection is at least as good as the most sensitive frequency band. However, there could also be cross-channel interference, where the broadband modulation detection threshold is higher (less sensitive) than the threshold of the most sensitive detection channel. Further experiments are needed to determine how modulation detection thresholds are combined across frequency channels.

4.5 Conclusions

Monaurally, the thresholds for detecting a sinusoidal intensity-modulation signal in a reverberant environment can be predicted reasonably well by the difference of the thresholds measured in an anechoic environment and the MTF of the reverberant environment. The difference between thresholds measured in anechoic conditions and monaurally after convolution with the IR were close to, but always larger than, the magnitude of the MTF at the measured modulation frequencies. For most modulation frequencies in the conditions tested, the binaural threshold could be predicted by the minimum of the two ears' thresholds, i. e., the 'better-ear' model, but there were several instances where the binaural threshold was significantly better than either ear alone. The data presented here suggest that a binaural advantage beyond just 'better-ear' listening may be gained due to a fluctuating IID caused by interaural modulation

phase differences (IMPDs). Further testing is required to determine whether the IMPD is the cause of the binaural advantage, and whether the IID fluctuation can provide any information that is useful for decoding speech.

Chapter 5

Monaural and binaural consonant identification in reverberation

Portions of this chapter were presented at Acoustics'08 Paris, the 2nd joint conference of the Acoustical Society of America and the European Acoustics Association, 29 June-4 July, 2008.

Abstract

Consonant identifications were obtained monaurally and binaurally using vowel-consonant-vowel (VCV) stimuli convolved with three different impulse responses (IRs). The IRs were binaural recordings from a concert hall with a reverberation time of about 2.5 s in the mid-frequency range, and all had large interaural differences in their modulation transfer functions. The binaural percent correct identifications were significantly higher than either monaural condition, and the percent correct identifications were significantly lower for one IR than for the other two, despite having similar speech transmission indices (STI). Not every stimulus that was correctly identified monaurally was also correctly identified binaurally, indicating binaural interference. 12% of the stimuli that were correctly identified binaurally were not correctly identified with either monaural condition, which shows a binaural advantage beyond simple 'better-ear' listening. The most frequent errors were voicing confusions, contrary to findings from previous studies, which have found voicing to be relatively robust against reverberation.

5.1 Introduction

Many previous studies have shown a significant increase in speech intelligibility when listening binaurally, with a dichotic presentation, as compared to monaural or diotic listening. This binaural advantage has been demonstrated in speech-in-noise tasks (e. g., Dirks and Wilson, 1969; Bronkhorst and Plomp, 1988; Dubno et al., 2008), speech-on-speech tasks (e. g., Cherry, 1953; Freyman et al., 2001; Edmonds and Culling, 2006), and in investigations of the effects of reverberation on speech intelligibility (e. g., Moncur and Dirks, 1967; Gelfand and Hochberg, 1976; Nábělek and Mason, 1981; Nábělek and Robinson, 1982). Some of this binaural advantage can be attributed to 'better-ear' listening, where one ear has a better signal-to-noise ratio as a result of the head-shadow effect (Edmonds and Culling, 2006), or where the useful-to-detrimental energy ratio is higher in one ear than the other (Lochner and Burger, 1964; Bradley et al., 2003). Other gains in intelligibility come from binaural interactions, where interaural differences in the received signal may be used to improve the effective signal-to-noise ratio, similar to the binaural masking level difference (BMLD; Durlach, 1963), or to reduce the degradation of the speech signal by reverberation (Koenig, 1950; Nábělek and Pickett, 1974b). The goal of the present study was to investigate monaural and binaural speech intelligibility in reverberation, without interfering noise, to shed some light on binaural mechanisms that may be used to improve intelligibility in rooms.

An acoustic speech signal is characterized by fluctuations in amplitude and pitch with time, and it is assumed that those fluctuations carry information that is critical for understanding the speech signal (Houtgast and Steeneken, 1985; Drullman et al., 1994; Greenberg et al., 1996; van der Horst et al., 1999; Apoux and Bacon, 2008). Amplitude fluctuations may be particularly important for the perception of consonants, since they can be distinguished from non-consonants in that they have relatively fast changes in overall amplitude with time (Stevens, 1980; Rosen, 1992). Reverberation, such as is created by a room, generally acts as a low-pass filter on these fluctuations, attenuating fast fluctuations and letting slow fluctuations pass through relatively unaffected (Houtgast and Steeneken, 1973). As the reverberation time increases, the cutoff frequency of the low-pass filter decreases (Schroeder, 1981), and

the intelligibility of speech also decreases (see, e. g., Nábělek and Pickett, 1974a). The attenuation of the intensity fluctuations as a function of the modulation frequency is called the modulation transfer function (MTF). One way of modeling the decrease in speech intelligibility with increasing reverberation times is the speech transmission index (STI; Houtgast and Steeneken, 1973; IEC 60268-16, 2003), which is based on the MTF. The STI is calculated as a weighted average of a measured or calculated MTF across modulation frequencies and audio frequencies for a single receiver (e. g., microphone or ear). There can, however, be large interaural differences in the MTF magnitude and phase, which may create perceivable interaural level fluctuations. Those fluctuations can lead to a binaural advantage in modulation detection thresholds (see Chap. 4, or Thompson and Dau, 2009), and may be useful for enhancing speech intelligibility, particularly consonant identification.

Some of the previous studies of speech intelligibility in reverberation have either discussed consonant identification and errors as part of a word-identification study (e. g., Nábělek and Mason, 1981), or have focused explicitly on consonant reception (e. g., Nábělek and Pickett, 1974b; Gelfand and Silman, 1979; Helfer, 1994). These studies used different materials, most frequently the modified rhyme test (MRT) or nonsense syllables, and a range of reverberation times from 0.1 to 1.6 s. The MRT uses primarily mono-syllabic real words of the form consonant-vowel-consonant (CVC), and the nonsense syllables were either CV or VC tokens. These studies compared word-initial to word-final consonants and generally found that final consonants had worse performance than initial consonants in reverberation. Final consonants are affected by 'overlap-masking', where the energy from the preceding vowel is temporally smeared through the consonant sound by the reverberation, while initial consonants are only affected by 'self-masking', assuming that there is no preceding word, where the consonant's own energy is temporally smeared (Nábělek et al., 1989; Libbey and Rogers, 2004). None of these studies have measured with inter-vocalic consonants (i. e., VCV), which may show effects of both initial and final consonants.

It is common to analyze consonant errors in terms of articulatory features, such as place of articulation (e. g., front, middle, back), manner of articulation (e. g., stop, fricative, nasal) and voicing (voiced, unvoiced) (see, e. g., Miller and Nicely, 1955). Some more recent studies have proposed analyses based on acoustic features in the

Table 5.1: Articulatory consonant features used in the discussion of consonant identification errors.

Consonant	Place	Voicing	Manner
b	front	voiced	stop
d	middle	voiced	stop
f	front	unvoiced	fricative
g	back	voiced	stop
k	back	unvoiced	stop
m	front	voiced	nasal
n	middle	voiced	nasal
p	front	unvoiced	stop
s	middle	unvoiced	fricative
t	middle	unvoiced	stop
v	front	voiced	fricative
z	middle	voiced	fricative

speech signal, reasoning that these features are what is actually heard by the listener and are, therefore, more relevant for perception (e. g., Soli and Arabie, 1979; Gordon-Salant, 1985; Allen, 2005; Phatak et al., 2008). In the present study, the consonant errors will be discussed in terms of articulatory features, as shown in Table 5.1, for comparison with the prior studies from Nábělek and Pickett (1974b), Gelfand and Silman (1979), and Helfer (1994). Those studies found that the articulatory feature that was most affected by reverberation was the place of articulation. For example, a place of articulation error would be a 't' response, an unvoiced middle stop, when /k/, an unvoiced back stop was presented. After place of articulation, manner was next most affected by reverberation, and voicing was least affected by reverberation. Nábělek and Pickett (1974b) reported the largest binaural improvement for the manner and voicing features. Helfer (1994) also reported the largest binaural improvement in reverberation for manner features, although the identification of the nasals /n/ and /m/ was already quite good monaurally.

There have been a number of studies that have investigated the relation between the modulation spectrum of speech and the transmission of information about the articulatory features. Rosen (1992) proposed that temporal features in the range of 2 to 50 Hz are most critical for manner, the voicing feature relied mostly on temporal

fluctuations in the 50 to 500 Hz range, although there was also an influence of the slower modulation rates on voicing, and place of articulation is communicated through high-frequency fluctuations in the 0.6 to 10 kHz range (temporal fine structure). Christiansen and Greenberg (2005) found that manner of articulation was most affected when modulation frequencies above 12 Hz were filtered out, voicing relied most on modulation frequencies in the range of 3 to 6 Hz, and place of articulation required broadband (audio-spectrum) integration, and modulation rates above 6 Hz. The question as to what modulation rates are critical for which speech features remains unanswered. It is possible that the distinction between certain pairs of consonants is a difference in the temporal energy pattern, or in the modulation spectra of the consonants (Gallun et al., 2008). So if binaural listening can restore information about those temporal pattern differences that had been obscured by reverberation, for example, through the use of interaural differences in the MTF, then this could result in a reduction in the number of confusions reported between those consonants.

The confusion matrices generated in consonant identification studies are often quite asymmetric. For example, there may be many 'p' responses when /t/ was presented, but relatively few 't' responses when /p/ was presented. This response bias is often not considered in the discussion of the results, or is simply averaged out by analyzing the data based on the means of the rows and columns of the confusion matrix (e. g., van der Horst et al., 1999; Allen, 2005). In signal detection theory, a response bias leading to an increase in the false identifications of a signal will also result in an increase in the correct identifications of the signal (see, e. g., Green and Swets, 1966; Wickens, 2002). Applying this concept to the consonant identification case, this would mean that, in the example above, given that the false /p/ identifications were disproportionately high, the percent correct /p/ identifications were probably higher than they would have been with unbiased responses. The method of making the confusion matrices symmetric through an averaging of the off-diagonal elements removes a bias from the errors, but does not remove any bias from the correct responses. The asymmetries may reveal interesting insights into the perception of the consonants by pointing out how, for example, a /t/ in one environment can sound like a /k/, but not vice versa. Therefore, in the present study, the confusion matrices will be presented and discussed with their asymmetries.

The goal of the present study was to investigate the identification of consonants in a reverberant environment without background noise, and to compare the patterns of errors made monaurally, with each ear alone, and binaurally. The impulse responses (IR) were selected to have a large interaural differences in MTF magnitude and phase for modulation frequencies less than 25 Hz, which are considered to be important for decoding speech (Christiansen et al., 2007; Apoux and Bacon, 2008), in order to test whether a large binaural advantage could be measured in these environments.

5.2 Methods

Listeners were presented with vowel-consonant-vowel (VCV) speech tokens convolved with dichotic IRs and were asked to identify the consonant sound presented. Presentations were made monaurally and binaurally through headphones, and consonant confusion matrices were generated for each listening condition and IR.

5.2.1 Stimuli

The speech material was a subset of the corpus of VCV tokens from Cooke and Scharenborg (2008) using four talkers (two males, two females), twelve consonants (/b/, /d/, /f/, /g/, /k/, /m/, /n/, /p/, /s/, /t/, /v/, /z/), three vowel contexts (/aCa/, /iCi/, /uCu/) and two stress patterns (first vowel stressed, second vowel stressed). Four of the tokens were rejected during pilot testing due to recording artifacts and three tokens were rejected after the experiment due to high error rates in the anechoic condition, so there were a total of 281 VCV tokens used in the experiment.

The VCV tokens were convolved with one of three binaural room impulse responses (BRIR) recorded from the Promenadikeskus concert hall in Pori, Finland (Merimaa et al., 2005b). The BRIRs were recorded using a manikin head (Brüel & Kjær HATS), and were diffuse field compensated to remove the effect of the manikin's ear canals for playback over headphones. The BRIRs had reverberation times (T_{30}) of about 2.5 s in the mid-frequencies (see Merimaa et al., 2005a, for more details on the BRIRs), but only the first 1 s of the BRIRs was used for the experiment. Three source/receiver position combinations were used in the current study, and the IRs and

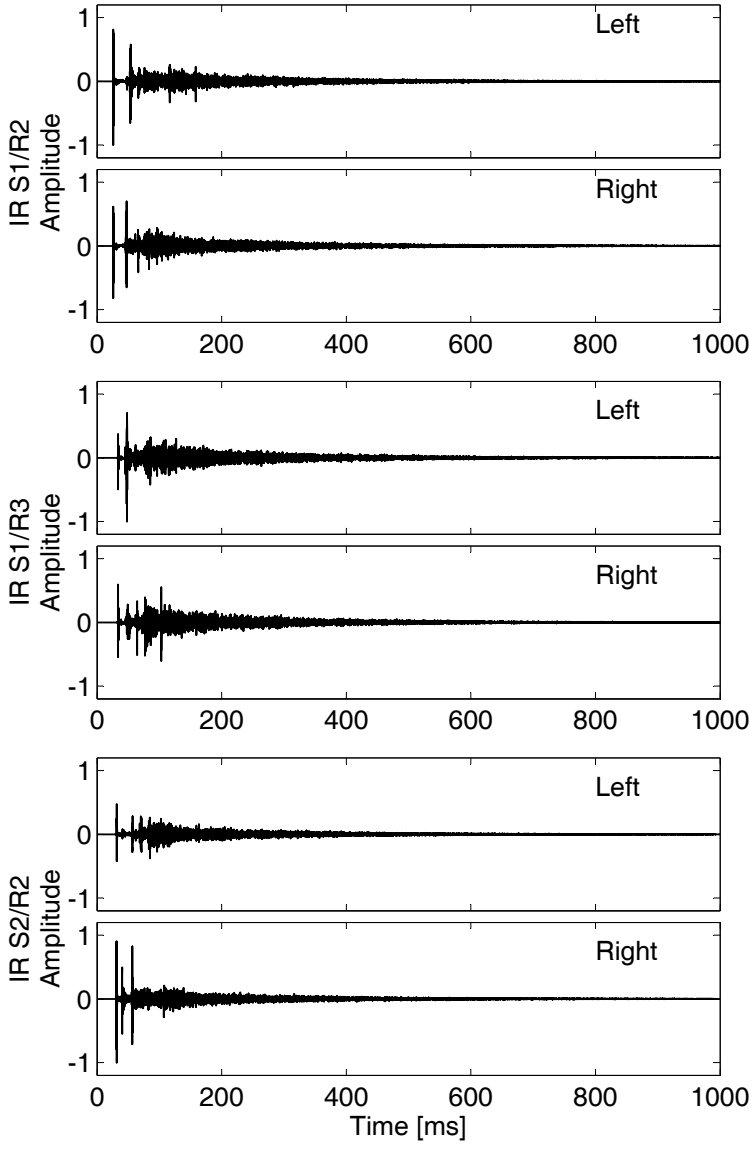

Figure 5.1: The impulse responses (IRs) used in the experiment. The left and right ears' IRs are shown for the S1/R2 source/receiver position (top two panels), for the S1/R3 position (middle two panels) and for the S2/R2 position (bottom two panels).

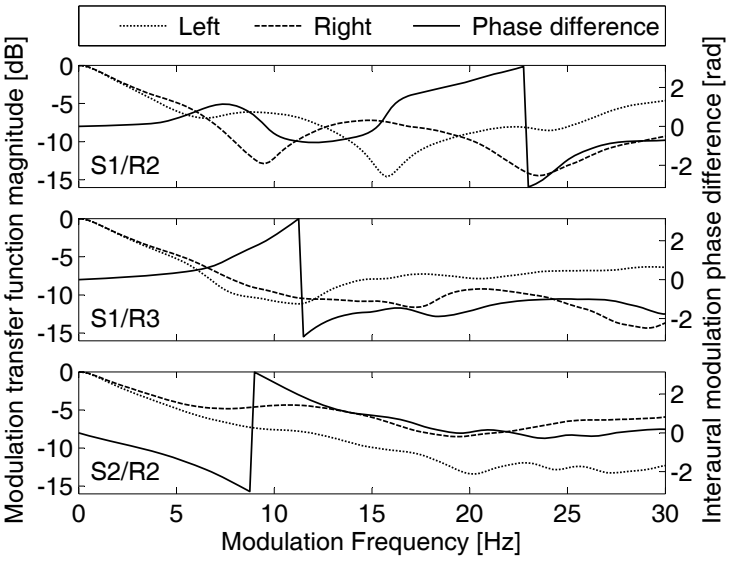

Figure 5.2: The modulation transfer functions (MTFs) calculated from the IRs shown in Fig. 5.1. The magnitudes of the MTFs for the left and right ears are shown with dotted and dashed lines, respectively, and refer to the left ordinate. The interaural modulation phase differences (IMPD) are plotted for each IR with solid lines, and refer to the right ordinate.

their respective data are labeled with the source (S) and receiver (R) position numbers from Merimaa et al. (2005b). IRs from two source positions (S1 and S2) and two receiver positions (R2 and R3) were used. Both source positions were on the stage, with S1 2.0 m from the front of the stage and 1.6 m left of center, and S2 5.2 m from the front of the stage and 4.0 m right of center. Both receiver positions were in the audience, with R2 in the seventh row and six seats right of center, and R3 in the eleventh row and five seats left of center. The BRIRs were recorded with the manikin 'sitting' in a seat with its nose pointed toward the stage in a direction perpendicular to the back of the seat, meaning that the direct sound did not necessarily come from 0° azimuth. The recorded IRs from S1/R2, S1/R3, and S2/R2 are shown in Fig. 5.1. The MTFs calculated from these IRs are shown in Fig. 5.2, with the magnitudes of the left and right ears' MTFs plotted with dotted and dashed lines, respectively, referring to the left ordinate, and the interaural modulation phase differences (IMPD) plotted

Table 5.2: Calculated speech transmission indices (STI).

S/R location	Left	Right	Better ear
S1/R2	0.47	0.47	0.49
S1/R3	0.48	0.48	0.50
S2/R2	0.49	0.49	0.51

with the solid line, referring to the right ordinate. The STI was calculated for each IR using the method from Houtgast and Steeneken (1985) for the left and right ears, and a 'better-ear' STI was calculated using the maximum of the two ears' MTFs for each modulation frequency and audio band. These calculated STIs are shown in table 5.2.

5.2.2 Equipment

The VCV tokens were convolved with the IRs using MATLAB® (The MathWorks) running on a PC. The stimuli were presented at a sampling rate of 25 kHz through Sennheiser HD-580 headphones. During the experiment, the test subjects were seated in a sound-insulated listening booth with a computer monitor, which displayed instructions and the response interface, and a keyboard and mouse for response input.

5.2.3 Procedure

The three IRs (S1/R2, S1/R3 and S2/R2) and three listening conditions (monaural-left, monaural-right and binaural) were presented to the nine test subjects in a latin-square design to avoid sequential effects in presentation to the listeners. Monaural conditions were chosen instead of presenting the monaural signals diotically to better reflect half-binaural hearing, and because previous studies have found insignificant or only slight differences in consonant ID in reverberation between monaural and diotic presentations (e. g., Moncur and Dirks, 1967; Helfer, 1994). In the following, the monaural conditions will be referred to as L and R for the left and right ears, respectively. The test subjects were presented with every VCV token in random sequence within each IR/listening-condition block before continuing to the next block. After completion of the nine blocks, each test subject completed one anechoic block as a control condition.

The experiment was set up with a 12-alternative, forced-choice design. Each token was presented once within each block, and the test subject had to provide a response before moving on to the next token. The subjects were instructed to respond with the consonant sound (phoneme) that they heard, not necessarily with the consonant that they would use to write the 'word.' The computer graphical user interface (GUI) showed twelve buttons for response, with each consonant represented once in a layout that mimicked their location on a standard QWERTY-keyboard. The subjects could either respond by clicking on the button in the GUI or by typing the consonant on the computer keyboard. After entering their response, they pressed a 'ready' button on the GUI or the space-bar on the keyboard to continue. The keyboard-entry method proved to be the fastest method and was preferred by the test subjects. No feedback was given on the responses.

5.2.4 Test subjects

Nine test subjects (five female, four male) participated in the experiments. Six test subjects were paid for their participation and the other three were affiliated with the research center and were not paid directly for their participation. All were native speakers of American English and had normal hearing, defined here as having audiometric thresholds of 15 dB HL or less. All subjects gave written informed consent (as approved by the Boston University Charles River Campus Institutional Review Board) before participating in the study. One test subject's data were excluded from the analysis due to a very low score ($< 90\%$ correct) in the anechoic condition, and a replacement test subject was used instead. The final nine test subjects each had at least 97% correct in the anechoic condition with a mean score of 98.7% correct (the confusion matrix generated in anechoic conditions, pooled across listeners, can be found in the Appendix, Table A.1).

5.3 Results

The overall mean percent correct consonant identification scores across listeners are shown in table 5.3 for each IR and listening condition, as well as for the anechoic

Table 5.3: Mean percent correct consonant identification scores for each source/receiver (S/R) position and listening condition, and the anechoic control condition.

Condition	S1/R2	S1/R3	S2/R2	Anechoic
Left	72.6	69.6	71.6	
Right	72.9	69.5	74.5	98.7
Binaural	82.5	77.2	82.8	

control condition. The confusion matrices for each listening condition and IR, pooled across listeners, can be found in the Appendix. A two-way analysis of variance with repeated measures (RM-ANOVA) on arcsine-transformed percent correct data with main factors of listening condition and IR showed a significant effect of listening condition (F(2,16)=99.0, $p < 0.0001$) and of IR (F(2,16)=4.6, $p < 0.05$), but no significant interaction (F(4,32)=1.7, $p = 0.17$). A post-hoc analysis showed that the binaural condition had significantly higher percent-correct scores than either monaural listening condition. Also, the S1/R3 IR had significantly lower scores than the other IRs, even though there was very little difference in the STIs calculated for the IRs.

In the following, L will represent the monaural-left response, R the monaural-right response, B the binaural response, and S the source. The pooled responses were analyzed to see how often both L and R were correct P($L = S \wedge R = S$), how often either L or R were correct P($L = S \vee R = S$), and then how often B was correct in each situation. The results of the analysis for each IR are shown in the panels of Fig. 5.3. In each panel, the left-hatched bar shows how often the L response was correct P($L = S$), and the right-hatched bar shows how often the R response was correct P($R = S$). The unhatched bar shows how often both the L *and* R responses were correct P($L = S \wedge R = S$), and the square-hatched bar shows how often either the L *or* the R (or both) response was correct P($L = S \vee R = S$). The horizontally-hatched bar shows the binaural percent correct P($B = S$). The right-most bar in each panel shows the binaural percent correct divided into how often B, L and R were correct P($L = S \wedge R = S \wedge B = S$) (cross-hatch), how often B and L, but not R, were correct P($L = S \wedge R \neq S \wedge B = S$) (left-hatch), how often B and R, but not L, were correct P($L \neq S \wedge R = S \wedge B = S$) (right-hatch), and how often B was correct and both L and R were incorrect, P($L \neq S \wedge R \neq S \wedge B = S$).

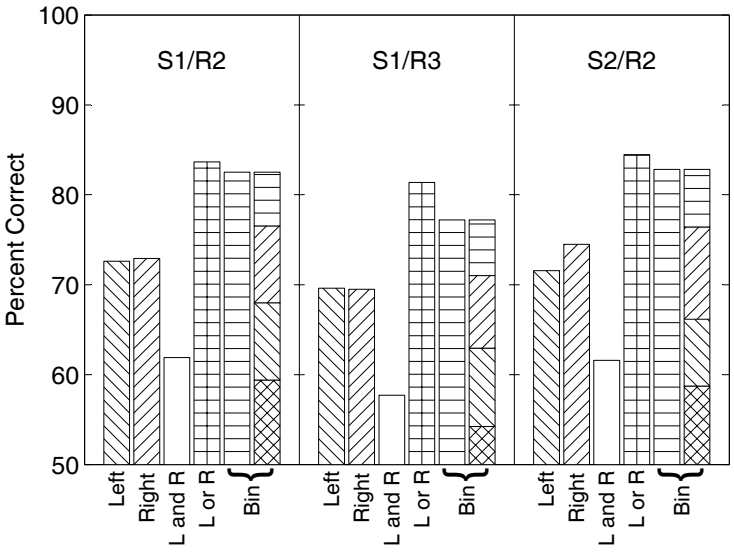

Figure 5.3: Percent correct consonant identifications by listening condition for each impulse response, pooled across listeners. The first five bars from left-to-right in each panel show P($L = S$) (left-hatching), P($R = S$) (right-hatching), P($L = S \wedge R = S$) (unhatched), P($L = S \vee R = S$) (square-hatch) and P($B = S$) (horizontal-hatch). The rightmost bar in each panel shows P($B = S$) broken into P($L = S \wedge R = S \wedge B = S$) (cross-hatch), P($L = S \wedge R \neq S \wedge B = S$) (left-hatch), P($L \neq S \wedge R = S \wedge B = S$) (right-hatch) and P($L \neq S \wedge R \neq S \wedge B = S$) (horizontal-hatch). See the text for more details.

When listening monaurally, the listeners had about 70% correct responses, across the three IRs. In about 60% of the consonant presentations, the correct response was given in both monaural conditions P($L = S \wedge R = S$). For 95% of the stimuli where the L and R responses were correct, the B response was also correct. In over 80% of the stimulus presentations with all three IRs, either the L or R response was correct, but the B response was only correct for 90% of those presentations. When only one of the L or R responses was correct, i. e., exclusive or, only 75% of the B responses were correct. Of the correct B responses, only 88% were for presentations that also had either correct L or R responses. The additional 12% of the correct B responses cannot be attributed to a correct monaural response.

More details in the percent correct data can be seen by looking at the percent correct for each consonant presented (see Fig. 5.4). In the L and R conditions, as a

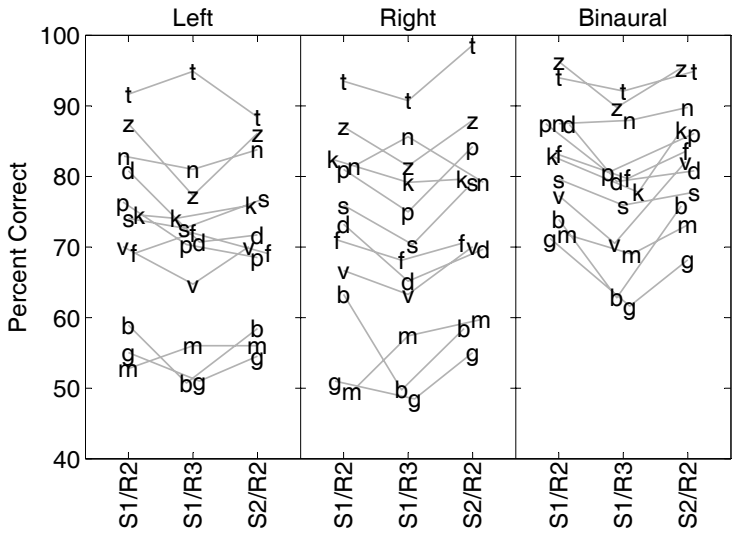

Figure 5.4: Percent correct consonant identifications by consonant presented, pooled across listeners.

group, unvoiced stops (/p/, /t/, /k/) were correctly identified more often than voiced stops (/b/, /d/, /g/), with the middle stops (/t/ and /d/) highest within their respective groups. With fricatives, the voiced fricative /z/ has a much higher percent correct than the unvoiced fricatives (/f/, /s/), and the voiced fricative /v/ is generally the lowest of the group. This shows that there is nothing inherent with these IRs that unvoiced consonants are always easier to identify than voiced consonants. The two nasal consonants (/m/, /n/) have very different correct identification scores, with /m/ around 55% in the monaural conditions and /n/ around 80% correct. As was seen in the overall percent correct data, the binaural percent correct identifications are generally higher than either of the monaural scores. For example, the correct identification of /m/ improves from about 55% in the monaural conditions to around 70% in the binaural condition. The three worst performers monaurally (/b/, /g/, /m/) all show large improvements of 10-20 percentage points in the binaural condition, but it is not only the worst performers that have large binaural improvements. There is an improvement of almost 10 percentage points from the monaural to binaural condition

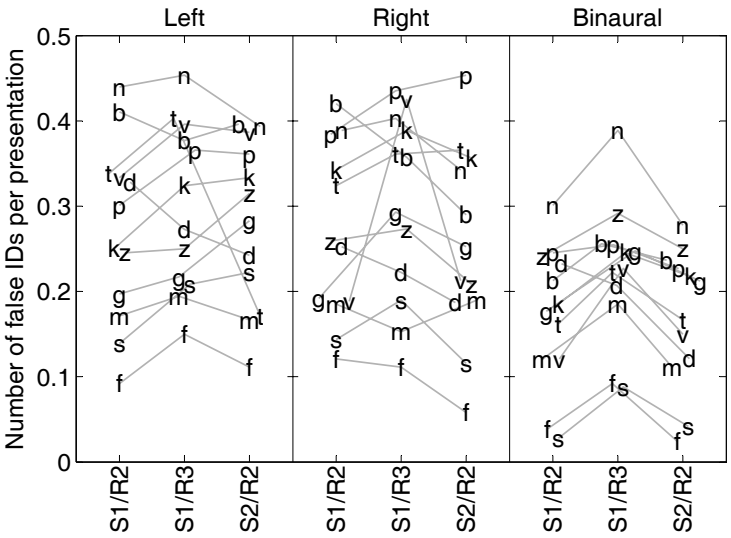

Figure 5.5: The number of false identifications divided by the number of presentations for each consonant for each IR, grouped by listening condition.

for /z/, which had 78-88% correct in the monaural conditions. The consonants with the largest binaural gain in percent correct (defined here as the percentage point difference between the binaural percent correct and the largest of the monaural percent correct for each consonant) were /b/, /f/, /g/ and /m/.

As discussed above (Sec. 5.1), percent correct reflects only part of the picture. Another part is the response bias, which is shown by including information about the false identification rate. For example, if the respondents always pressed the 't' button whenever they were unsure, then the percentage of correct identifications of /t/ would be high, but there would also be a lot of false /t/ identifications. The false identifications are presented in Fig. 5.5 as the ratio of false IDs of a consonant to the number of presentations of that consonant. These data show that there were a lot of incorrect 'n' responses, so the high percent correct number may not actually reflect reliable identification of the consonant sound, but rather a bias toward 'n' responses. There was a similar pattern for /t/ with a high percent correct in all conditions, and a high number of false IDs, except in the L-S2/R2 condition, where there were less

than 0.2 false IDs per presentation of /t/. On the other hand, /b/ had a lot of incorrect IDs and also a very low percent correct score. There was not a significant correlation between the percent correct and the false IDs ratio for any of the conditions, so good performance overall was not solely driven by response bias. The largest binaural reduction in false IDs (defined here as the minimum of the monaural false IDs minus the binaural false IDs for each consonant) was seen with the consonants /b/, /p/, /s/, /t/ and /v/. The false IDs of /v/ present an interesting pattern in that the percent correct scores are approximately equal across the IRs within each condition, but the false IDs show a very large variation, particularly with right-ear listening. An investigation into the acoustic differences between the IRs, perhaps in their MTFs, may reveal why there were fewer 'v' responses with R-S1/R2 and R-S2/R2, which may show what acoustic features are critical for the perception of /v/.

The analysis can drill down even further into the false ID data by considering which consonant responses were given for each consonant stimulus. This is just a graphical representation of the off-diagonal elements of the confusion matrix. The error plots for the stimuli /d/ and /t/ are shown in Fig. 5.6 (left panels), where the labels (/C/) in the binaural sub-panel show the presented consonant and the letters within the plots show the responses. Comparing the /d/ and /t/ panels clearly shows the bias towards 't' responses in the /d/-/t/ pair, when more than 10% of the responses when /d/ was presented were 't' (upper-left panel), but generally less than 5% of the responses when /t/ was presented were 'd' (lower-left panel). There is also only a small improvement from the monaural conditions to the binaural for either 'd' for /t/ errors or 't' for /d/ errors. /d/ and /t/ are both middle-articulated stop consonants (see the articulatory features in Table 5.1), so /d/-/t/ errors are voicing errors. A similar bias is seen with the /m/-/n/ pair shown in the right panels of Fig. 5.6. Close to 30% of the responses when /m/ was presented were 'n' (upper-right panel), but only around 10% of the responses when /n/ were presented were 'm' (lower-right panel). /m/ and /n/ are both voiced nasal consonants, so /m/-/n/ errors are place of articulation errors. When switching from monaural to binaural listening, there were only small reductions in the number of 'n' responses for /m/ presentations and 'm' responses for /n/ presentations. Monaurally, there were many 'b' responses for /m/ presentations (a manner of articulation error), with around 10% of the responses monaurally, but

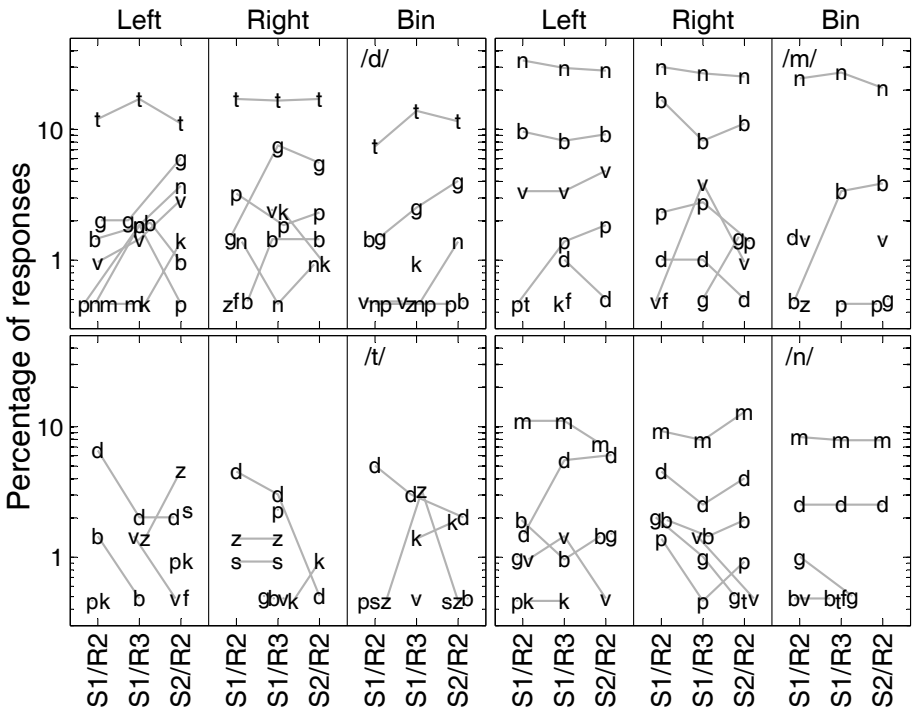

Figure 5.6: The percent false IDs of a consonant when the consonants /d/, /m/, /n/ and /t/ were presented (panel labels in the binaural sub-panel show the presented consonant).

there was a binaural reduction in the number of 'b' responses with only 4% or less with binaural /m/ presentations. The response biases in the /d/-/t/ and /m/-/n/ pairs probably contribute to the relatively high percent-correct scores for /t/ and /n/.

The errors made when a fricative consonant was presented are shown in Fig. 5.7. The fricatives /s/ and /z/ (left panels) are primarily confused with each other, with a slight bias toward /z/ in some of the conditions. When changing from monaural to binaural listening, there was little reduction in the number of 'z' responses when /s/ was presented, but there was a comparatively large reduction in 's' responses when /z/ was presented. /s/ and /z/ presentations had errors with only a few consonants, with primarily 'z' and 't' responses for /s/ presentations, and primarily 's', 't' and 'd'

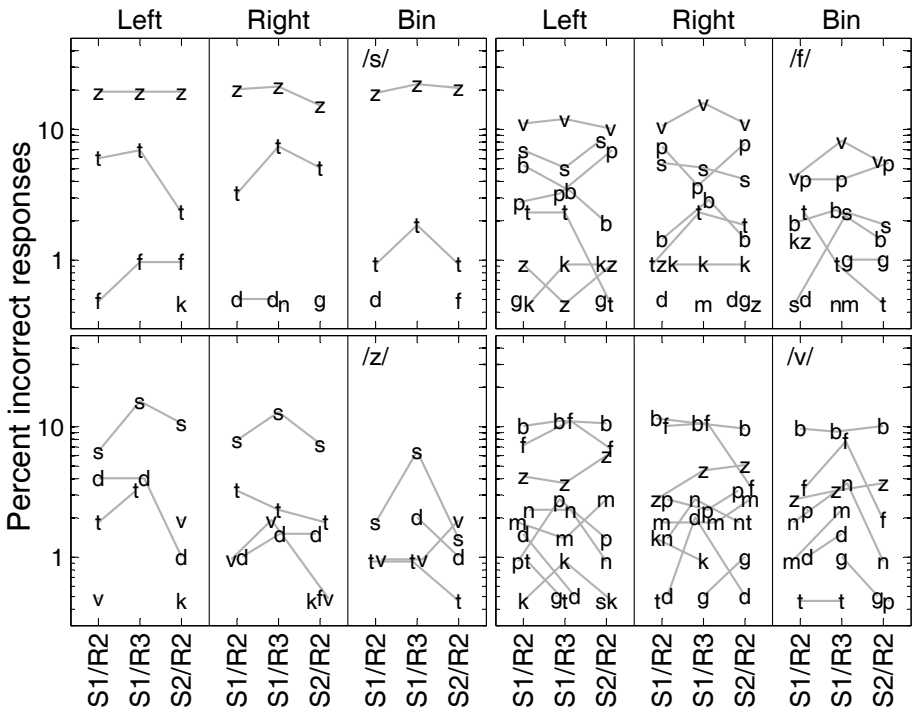

Figure 5.7: The percent false IDs of a consonant when the fricative consonants /f/, /s/, /v/ and /z/ were presented (panel labels in the binaural sub-panel show the presented consonant).

responses for /z/ presentations. /f/ and /v/ presentations (right panels), had errors with reports of almost every other consonant. The most common incorrect response when /v/ was presented was ‘b’, which was the only place in the confusion matrices that the most common error was a manner of articulation error. The second most common error when /v/ was presented was an ‘f’ response (an error of voicing), and the most common incorrect response was ‘v’ when /f/ was presented, but there were also a lot of ‘s’ (place error) and ‘p’ (manner error) responses.

The remaining four stop consonants’ (/b/, /g/, /k/, /p/) confusion data are shown in Fig. 5.8. The most common errors for these four consonant sounds were voicing errors (/b/-/p/ and /g/-/k/), with a bias in each pair toward the unvoiced consonants /p/

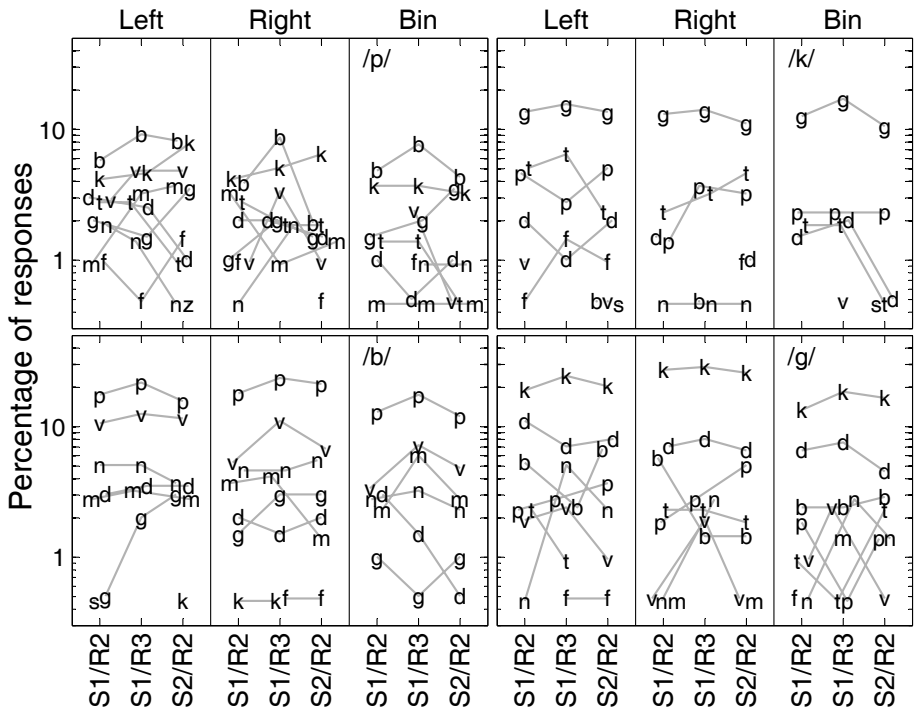

Figure 5.8: The percent false IDs of a consonant when the stop consonants /b/, /g/, /k/ and /p/ were presented (panel labels in the binaural sub-panel show the presented consonant).

and /k/. There were also a lot of 'v' responses for /b/ (manner error), and 'd' for /g/ (place error). There was very little binaural reduction in the number of 'g' responses for /k/, but there was a binaural reduction in the number of 'k' responses for /g/ and of 'p' responses for /b/.

5.4 Discussion

The overall percent-correct results showed a significant binaural advantage in consonant identifications over monaural listening, similar to reports from previous studies. The proportion of binaural correct responses was about the same magnitude of the

proportion of presentations for which there was either a correct monaural-right or monaural-left response. However, the binaural response was not always correct for the presentations that had a correct monaural response, and there were many correct binaural responses for which neither monaural response had been correct. This indicates that there is some interference between the two monaural signals that makes the binaural ID more difficult for some consonant presentations. This also shows that there is a binaural advantage that goes beyond just 'better-ear' listening that makes the binaural ID easier for some presentations. This analysis is, however, based on the assumption that the monaural and binaural consonant IDs were repeatable, since they were not made, and cannot be made, simultaneously. This repeatability assumption needs to be tested by presenting the same stimuli to the same listeners over several sessions. This data will show a probability of responses for each presentation. For example, for a certain test subject and a certain monaural presentation of /b/, there may be a 40% likelihood of responding with 'b', a 30% likelihood of responding with 'p', a 20% likelihood of responding with 'v', and small percentages of other responses. For the same test subject and stimulus in the other ear, the response probabilities may be 20% 'b', 40% 'p' and 30% 'v'. Then the question is whether a model can be created to predict the binaural response based on a combination of these monaural probabilities.

The most common errors made in this study were errors of voicing (i. e., /t/-/d/, /k/-/g/, /p/-/b/, /s/-/z/, /f/-/v/). This is in contrast to previous studies of monaural and binaural consonant identification in reverberation (Nábělek and Pickett, 1974b; Gelfand and Silman, 1979; Helfer, 1994) that found that the place of articulation was more degraded by reverberation than manner, and voicing was the least affected by reverberation. There are two main differences between this study and those previous studies: the reverberation times and the speech material used. The reverberation time used in this study was about 2.5 s in the mid-frequency range, while the longest reverberation time used in those prior studies was 1.6 s. The previous studies only looked at word-initial and word-final consonants, where the consonants in the present study were intervocalic. It is unknown at this point which of these differences caused the change in effect on the articulatory features.

The binaural gain in consonant identifications has been attributed previously to improvements in the manner and voicing features, because these features rely on low-frequency information, while place relies on high-frequency information (Nábělek and Pickett, 1974b). Of the four consonants with the largest binaural improvement in percent correct identifications in the present study (/b/, /f/, /g/ and /m/), three had large reductions in voicing errors ('p' for /b/, 'v' for /f/, and 'k' for /g/). With /m/ presentations, the most common errors were 'n' responses, a place error, but the largest binaural reduction in errors when /m/ was presented were 'b' and 'v' responses, both manner errors. For these consonants, the improvements are consistent with the binaural improvements reported in the previous study from Nábělek and Pickett (1974b). However, the improvements are quite asymmetric. For example, the number of 'k' responses for /g/ presentations decreased from monaural to binaural presentations by about 6 percentage points, but there was little, if any, reduction in the number of 'g' responses for /k/ presentations. A similar asymmetry is seen with the /s/ and /z/ presentations. There was a large decrease in the number of 's' responses for /z/, but there was little decrease in the number of 'z' responses for /s/. The largest binaural improvement in correct identifications of /s/ came through a reduction in the number of 't' responses. It should also be noted that the asymmetry in the binaural improvement with /s/ and /z/ *increases* the response bias in this pair of consonants.

These asymmetries in binaural improvements may reveal a weakness of relying on articulatory features for the analysis of perceptual errors. If there was a perceptual feature of a consonant sound that determined whether the consonant was voiced or not, then it could be expected that the binaural improvement in this voicing feature would be symmetrical. An analysis of the acoustic features of the consonants, as has been performed for some consonants obscured by noise (Allen, 2005; Phatak et al., 2008; Régnier and Allen, 2008), may provide more insight into what features are being restored through binaural listening. This analysis should take into account the temporal features, e. g., amplitude modulations, of the speech tokens. This may reveal whether the hypothesis that the interaural differences in the MTF are, at least partially, responsible for binaural gains in consonant identifications. Using an example of articulatory features from previous studies, Nábělek and Pickett (1974b) proposed that the binaural gain was largest with manner and voicing features because these are

based on low-audio-frequency information, and Christiansen and Greenberg (2005) proposed that manner features needed amplitude modulations above 12 Hz to be successfully transmitted. In the present study, the MTFs for the S1/R2 and S1/R3 IRs showed large IMPDs at modulation frequencies around 12 and 24 Hz, respectively (see Fig. 5.2). Therefore, if the original hypothesis in the present study is true, then these IRs should show a bigger binaural reduction in manner errors than the S2/R2 IR. There were no significant differences in the articulatory-feature error patterns between the IRs, only in some differences in the errors between specific consonant pairs. It may be more appropriate to investigate the MTFs in audio-frequency bands, as is done with the STI, in order to draw specific conclusions about the binaural improvement. This will require also gathering more consonant ID data with more IRs to develop and test a model for the monaural and binaural perception of consonants in reverberation.

5.5 Conclusions

Consonant sounds that were corrupted by reverberation were identified by test subjects in monaural and binaural presentations. There was a significant improvement in the overall percent correct identifications made with binaural presentations over monaural presentations. This binaural advantage was not just the result of using the 'better-ear' identification, since many binaural correct IDs were made with presentations for which neither monaural ID was correct. The most frequent errors made were errors of the articulatory feature voicing, contrary to previous studies that have found voicing to be more robust to reverberation than either manner or place of articulation (Nábělek and Pickett, 1974b; Gelfand and Silman, 1979; Helfer, 1994).

There was a large response bias in some pairs of consonants (e. g., /d/-/t/, /m/-/n/), which was illustrated by a high percentage of correct responses and false identifications for one of the consonants and a low percentage for the other. Methods of analyzing consonant confusions that rely on symmetric confusion matrices should take into account that both the percent correct and the false IDs are affected by the response bias when forcing symmetry on the matrices.

There was a greater binaural reduction in voicing and manner errors than place of articulation errors, as has been found in previous studies (e. g., Nábělek and Pickett,

1974b). The reduction in errors was asymmetric between consonants, with, for example, a large reduction in the number of ‘s’ responses for /z/ presentations, but little reduction in ‘z’ responses for /s/ presentations. This asymmetry in binaural improvement may argue for an analysis based on acoustic features instead of articulatory features.

The data did not show large differences between the error patterns in terms of articulatory features between the three IRs used in the present study. There were some differences in the error patterns between specific pairs of consonants, which may help to show what acoustic features are critical for identifying specific consonants. However, more consonant ID data must be gathered with more IRs with different MTFs before concrete conclusions can be drawn on whether interaural differences in the MTFs contribute to the binaural advantage in consonant identifications. This will also determine whether a binaural calculation should be added to the STI model.

Acknowledgments

The research for this study was carried out while co-author Eric Thompson was on an extended research stay at Boston University, which was funded in part by grants from the Denmark-America Foundation and the Idella Foundation. The research was also funded in part by an NSF grant to the Center of Excellence for Learning in Education, Science and Technology (CELEST), Thrust 2.

Chapter 6

Overall summary and discussion

The data presented in this thesis provide interesting insights into the interactions between amplitude modulation (AM) and binaural processing in the human auditory system. Two main AM/binaural interactions were investigated: the processing of interaural level fluctuations caused by interaural modulation phase differences, and the release from AM masking caused by a perceived spatial separation of an AM target and an AM masker. Chapters 2 and 3 presented basic psychoacoustic experiments aimed at characterizing human performance regarding these AM/binaural interactions. The experiments presented in Chaps. 4 and 5 investigated situations in which interaural level fluctuations resulting from modulation phase differences could provide a binaural cue to enhance modulation signal detection thresholds or speech intelligibility as compared to monaural performance.

When there is an interaural modulation phase difference, a fluctuating interaural level difference is created. This is a cue that is only available in the auditory system by comparing the two ears' signals. The results of the experiments in Chap. 2 showed that listeners are able to switch between using monaural level fluctuations ('up-down') and binaural ILD fluctuations ('right-left') for the detection of AM imposed on a carrier, depending on which cue is most salient. The listeners were able to discriminate between interaurally antiphasic and homophasic AM at rates well above 100 Hz, indicating that modulation phase information must be preserved in the auditory system at least up to the level of binaural interaction. Differences in the modulation depths required for AM interaural phase discrimination for pure-tone, diotic noise or dichotic noise carriers suggested some modulation frequency tuning in the processing of

modulated ILDs. The modulation discrimination thresholds measured in the presence of interaurally uncorrelated narrowband noise AM maskers showed a bandpass shape as a function of the masker noise center frequency. This suggests that an array of bandpass modulation filters may be required for a binaural model to predict the thresholds reported in the data. Simulations with an existing binaural model that uses a low-pass 'binaural sluggishness' filter to limit binaural temporal resolution showed that the model was unable to predict the bandpass shape observed in the listeners' threshold patterns.

Chapter 3 presented two experiments related to a different aspect of binaural hearing, namely the influence of a perceived difference in lateral position of a target and masker in the detection of amplitude modulation ('up-down'). The experiments both used temporally interleaved transposed-stimulus carriers for a target AM and a masker AM. The transposed stimuli allowed the interaural time difference (ITD) of each carrier to be controlled independently. In the first experiment, the listeners adjusted the ILD of a pointer stimulus so that it was perceived to come from the same lateral position as either the target or masker carrier as a function of the ITD. The data showed that the perceived location of target and masker carriers were almost unaffected by the other carrier's presence, indicating that the target and masker were perceived as separate auditory objects. However, when modulation detection thresholds were measured as a function of the modulation frequency and ITD of the masker, the data showed very little difference between the thresholds measured when both masker and target were diotic and when the target was diotic, but the masker had a 1 ms ITD. This indicates that there is no spatial release from modulation masking when the spatial difference is based on ITDs.

The data presented in Chaps. 2 and 3 showed that binaural processing makes use of rapid interaural fluctuations. It was possible for the listeners to discriminate interaural modulation phase differences in AM stimuli at modulation rates above 100 Hz. The listeners could also lateralize two sounds separately for which the stimulus ITD is changing between 0 and 1 ms at a rate of 125 Hz. This means that a model of binaural processing must preserve this timing information at least up to the point of binaural interaction and extraction of lateral position from the signals. Many existing binaural models use a low-pass 'binaural sluggishness' filter to limit the

temporal resolution of binaural processing. Such models cannot predict the bandpass tuning shown in the data in Chap. 2, and cannot predict the separate lateralization of the two carriers in Chap. 3. However, this low-pass filter, by eliminating the lateralization of the carriers, also eliminates any spatial release from modulation masking with the stimuli from Chap. 3, thereby corresponding with the listeners' data. The challenge remains to develop a model of binaural processing and perception that predicts all of the thresholds reported in this thesis.

The hypothesis that an interaural modulation phase difference in the modulation transfer function in a reverberant environment could help to improve modulation detection thresholds was tested in Chap. 4. For a simple impulse response for which the modulation transfer function is constant across audio-frequency bands, the data showed a significant binaural improvement in modulation detection thresholds over thresholds obtained with either ear alone. The monaural modulation detection thresholds could be predicted reasonably well from the modulation detection thresholds obtained in the anechoic condition and the modulation transfer function at that modulation frequency, adding support to the model underlying the speech transmission index (STI). However, this model does not make use of interaural differences, and cannot predict the binaural advantage seen in the data. For a more complex, realistic room impulse response, for which the modulation transfer function varies across audio frequency bands, the STI model predicted much better performance than was seen in the data, and there was little, if any, binaural advantage at expected modulation frequencies (i. e., those modulation frequencies at which there was a large interaural modulation phase difference). It is as yet unknown how monaural modulation detection is performed across audio frequency bands that have different modulation depths and different modulation phases. Further studies should be performed to see if detection performance is based on the most sensitive band, or if sensitivities are combined across bands using, e. g., sum-of-squares, with synergy, or with interference.

The experiment from Chap. 5 investigated the effect of interaural modulation phase differences on a different task, namely the identification of consonant sounds in reverberant environments. There was a significant binaural improvement in the percentage of correct consonant identifications for the three impulse responses used

in the experiment as compared to monaural identifications. The most common errors made were confusions regarding the articulatory feature voicing, e. g., reporting 'd' when /t/ was spoken. Voicing and manner were the articulatory features that showed the greatest improvement from monaural to binaural listening. Even though the three impulse responses used in this experiment showed large differences in their modulation transfer functions, the resulting error patterns were largely similar. There were some differences seen in error patterns for particular consonants between the three impulse responses. A more detailed acoustic analysis of the stimuli in the three conditions may shed light on what is required to decode those particular consonant sounds.

Chapters 4 and 5 started with the hypothesis that interaural level fluctuations resulting from interaural modulation phase differences can be used as a cue to enhance the detection of amplitude (or intensity) modulation and the identification of consonants in reverberant environments, respectively. The results of each experiment can neither confirm nor disprove the hypothesis, but both suggest paths that further research can take to test the hypothesis. A model that can explain how modulation information is combined across audio-frequency bands, and then across ears, may be able to be used as an enhanced STI for the prediction of speech intelligibility in rooms. First, further testing with consonant identification should be done with impulse responses that are specifically designed to test the effect of certain modulation frequency ranges on specific consonant groups. When a robust effect can be demonstrated, then tests can be made to determine whether an interaural modulation phase difference can improve performance again.

One of the goals coming out of this research is to develop a computational model of the auditory system that can predict performance in a wide range of experiments for human listeners with normal and elevated audiological thresholds. This model would build on the work presented in this thesis, as well as ongoing work from others using a common modeling platform. The other work includes investigations into non-linearities in monaural hearing from Jepsen et al. (2008) and across-audio-channel modulation processing from Piechowiak et al. (2007). The combined model should be able to predict signal detection and discrimination thresholds in experiments involving simultaneous and non-simultaneous masking, across-channel and across-ear

processing in complex listening environments. This could also be usable as a front-end for predicting speech intelligibility scores and for automatic speech recognition systems, and as a test bed for new hearing appliances and algorithms.

Bibliography

Akeroyd, M. A. and Summerfield, A. Q. (1999). A binaural analog of gap detection. *J. Acoust. Soc. Am.*, **105**(5), 2807–2820.

Allen, J. B. (2005). Consonant recognition and the articulation index. *J. Acoust. Soc. Am.*, **117**(4), 2212–2223.

ANSI S3.5:1997 (R2007). *Methods for the calculation of the Speech Intelligibility Index*. New York: American National Standards Institute.

Apoux, F. and Bacon, S. P. (2008). Selectivity of modulation interference for consonant identification in normal-hearing listeners. *J. Acoust. Soc. Am.*, **123**(3), 1665–1672.

Bacon, S. P. and Grantham, D. W. (1989). Modulation masking: Effects of modulation frequency, depth, and phase. *J. Acoust. Soc. Am.*, **85**(6), 2575–2580.

Bacon, S. P. and Opie, J. M. (1994). Monotic and dichotic modulation detection interference in practiced and unpracticed subjects. *J. Acoust. Soc. Am.*, **95**(5), 2637–2641.

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), 1868–1880.

Bernstein, L. R. and Trahiotis, C. (1994). Detection of interaural delay in high-frequency sinusoidally amplitude-modulated tones, two-tone complexes, and bands of noise. *J. Acoust. Soc. Am.*, **95**(6), 3561–3567.

Bernstein, L. R. and Trahiotis, C. (2002). Enhancing sensitivity to interaural delays at high frequencies by using 'transposed stimuli' . *J. Acoust. Soc. Am.*, **112**(3), 1026–1036.

Bernstein, L. R. and Trahiotis, C. (2003). Enhancing interaural-delay-based extents of laterality at high frequencies by using 'transposed stimuli' . *J. Acoust. Soc. Am.*, **113**(6), 3335–3347.

Bernstein, L. R. and Trahiotis, C. (2005). Measures of extents of laterality for high-frequency 'transposed' stimuli under conditions of binaural interference. *J. Acoust. Soc. Am.*, **118**(3), 1626–1635.

Beutelmann, R. and Brand, T. (2006). Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.*, **120**(1), 331–342.

Blauert, J. (1972). On the lag of lateralization caused by interaural time and intensity differences. *Audiology*, **11**(5), 265–270.

Braasch, J. and Hartung, K. (2002). Localization in the presence of a distracter and reverberation in the frontal horizontal plane. I. Psychoacoustical data. *Acta Acust. United Ac.*, **88**(6), 942–955.

Bradley, J. S. (1986). Predictors of speech intelligibility in rooms. *J. Acoust. Soc. Am.*, **80**(3), 837–845.

Bradley, J. S., Sato, H., and Picard, M. (2003). On the importance of early reflections for speech in rooms. *J. Acoust. Soc. Am.*, **113**(6), 3233–3244.

Breebaart, J., van de Par, S., and Kohlrausch, A. (1998). Binaural signal detection with phase-shifted and time-delayed noise maskers. *J. Acoust. Soc. Am.*, **103**(5), 2079–2083.

Breebaart, J., van de Par, S., and Kohlrausch, A. (2001a). Binaural processing model based on contralateral inhibition. I. Model structure. *J. Acoust. Soc. Am.*, **110**(2), 1074–1088.

Breebaart, J., van de Par, S., and Kohlrausch, A. (2001b). Binaural processing model based on contralateral inhibition. II. Dependence on spectral parameters. *J. Acoust. Soc. Am.*, **110**(2), 1089–1104.

Breebaart, J., van de Par, S., and Kohlrausch, A. (2001c). Binaural processing model based on contralateral inhibition. III. Dependence on temporal parameters. *J. Acoust. Soc. Am.*, **110**(2), 1105–1117.

Bronkhorst, A. W. and Plomp, R. (1988). The effect of head-induced interaural time and level differences on speech intelligibility in noise. *J Acoust Soc Am*, **83**(4), 1508–1516.

Buell, T. N., Trahiotis, C., and Bernstein, L. R. (1991). Lateralization of low-frequency tones: Relative potency of gating and ongoing interaural delays. *J. Acoust. Soc. Am.*, **90**(6), 3077–3085.

Canévet, G. and Meunier, S. (1996). Effect of adaptation on auditory localization and lateralization. *Acta Acust. United Ac.*, **82**(1), 149–157.

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), 416–424.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.*, **25**(5), 975–979.

Christensen, C. L. (2005). *ODEON room acoustics program, version 8.0, user manual*. Lyngby, Denmark: Odeon A/S.

Christiansen, T. U., Dau, T., and Greenberg, S. (2007). Spectro-temporal processing of speech – An information-theoretic framework. In B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. Verhey (Eds.), *Hearing – From sensory processing to perception* (pp. 515–523). Springer Verlag.

Christiansen, T. U. and Greenberg, S. (2005). Frequency selective filtering of the modulation spectrum and its impact on consonant identification. In Rasmussen,

A. N., Poulsen, T., Andersen, T., and Larsen, C. B. (Eds.), *21st Danavox Symposium, Hearing Aid Fitting*, (pp. 585–599).

Cooke, M. and Scharenborg, O. (2008). The interspeech 2008 consonant challenge. In *Interspeech'08, Brisbane, Australia*.

Cremer, L. and Müller, H. (1982). *Principles and applications of room acoustics*, volume 1. London: Applied Science. [Translated by T.J. Schultz].

Dau, T. (1996). *Modeling auditory processing of amplitude modulation*. PhD thesis, Universität Oldenburg.

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

Dau, T., Kollmeier, B., and Kohlrausch, A. (1997b). Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration. *J. Acoust. Soc. Am.*, **102**(5), 2906–2919.

Dau, T., Püschel, D., and Kohlrausch, A. (1996). A quantitative model of the 'effective' signal processing in the auditory system. I. Model structure. *J. Acoust. Soc. Am.*, **99**(6), 3615–3622.

Dirks, D. D. and Wilson, R. A. (1969). Binaural hearing of speech for aided and unaided conditions. *J. Speech Hear. Res.*, **12**(3), 650–664.

Dreyer, A. and Delgutte, B. (2006). Phase locking of auditory-nerve fibers to the envelopes of high-frequency sounds: Implications for sound localization. *J. Neurophysiol.*, **96**(5), 2327–2341.

Drullman, R., Festen, J. M., and Plomp, R. (1994). Effect of temporal envelope smearing on speech reception. *J. Acoust. Soc. Am.*, **95**(2), 1053–1064.

Dubno, J. R., Ahlstrom, J. B., and Horwitz, A. R. (2008). Binaural advantage for younger and older adults with normal hearing. *J. Speech Lang. Hear. Res.*, **51**(2), 539–556.

Durlach, N. I. (1963). Equalization and cancellation theory of binaural masking-level differences. *J. Acoust. Soc. Am.*, **35**(8), 1206–1218.

Edmonds, B. A. and Culling, J. F. (2006). The spatial unmasking of speech: Evidence for better-ear listening. *J. Acoust. Soc. Am.*, **120**(3), 1539–1545.

Ewert, S. D. and Dau, T. (2000). Characterizing frequency selectivity for envelope fluctuations. *J. Acoust. Soc. Am.*, **108**(3), 1181–1196.

Ewert, S. D., Verhey, J. L., and Dau, T. (2002). Spectro-temporal processing in the envelope-frequency domain. *J. Acoust. Soc. Am.*, **112**(6), 2921–2931.

Feddersen, W. E., Sandel, T. T., Teas, D. C., and Jeffress, L. A. (1957). Localization of high-frequency tones. *J. Acoust. Soc. Am.*, **29**(9), 988–991.

Fleischer, H. (1982). Modulationsschwellen von Schmalbandrauschen [Modulation thresholds of narrow-band noise]. *Acustica*, **51**, 154–161.

Formby, C. and Muir, K. (1988). Modulation and gap detection for broadband and filtered noise signals. *J. Acoust. Soc. Am.*, **84**(2), 545–550.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *J. Acoust. Soc. Am.*, **109**(5), 2112–2122.

Gallun, F. J., Durlach, N. I., Colburn, H. S., Shinn-Cunningham, B. G., Best, V., Mason, C. R., and Kidd, G. (2008). The extent to which a position-based explanation accounts for binaural release from informational masking. *J. Acoust. Soc. Am.*, **124**(1), 439–449.

Gelfand, S. A. and Hochberg, I. (1976). Binaural and monaural speech discrimination under reverberation. *Audiology*, **15**(1), 72–84.

Gelfand, S. A. and Silman, S. (1979). Effects of small room reverberation upon the recognition of some consonant features. *J. Acoust. Soc. Am.*, **66**(1), 22–29.

Gilkey, R. H. and Good, M. D. (1995). Effects of frequency on free-field masking. *Hum. Factors*, **37**(4), 835–843.

Gockel, H., Carlyon, R. P., and Deeks, J. M. (2002). Effect of modulator asynchrony of sinusoidal and noise modulators on frequency and amplitude modulation detection interference. *J. Acoust. Soc. Am.*, **112**(6), 2975–2984.

Gordon-Salant, S. (1985). Phoneme feature perception in noise by normal-hearing and hearing-impaired subjects. *J. Speech Hear. Res.*, **28**(1), 87–95.

Grantham, D. W. (1982). Detectability of time-varying interaural correlation in narrow-band noise stimuli. *J. Acoust. Soc. Am.*, **72**(4), 1178–1184.

Grantham, D. W. (1984). Discrimination of dynamic interaural intensity differences. *J. Acoust. Soc. Am.*, **76**(1), 71–76.

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

Grantham, D. W. and Wightman, F. L. (1978). Detectability of varying interaural temporal differences. *J. Acoust. Soc. Am.*, **63**(2), 511–523.

Green, D. M. and Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: John Wiley & Sons, Inc.

Greenberg, S., Hollenback, J., and Ellis, D. (1996). Insights into spoken language gleaned from phonetic transcriptions of the Switchboard corpus. In *Proc. Int. Conf. Spoken Language Processing (ICSLP)*, Philadelphia, PA.

Griesinger, D. (1997). The psychoacoustics of apparent source width, spaciousness and envelopment in performance spaces. *Acustica*, **83**(4), 721–731.

Haas, H. (1951). Über den Einfluss eines Einfachechos auf die Hörsamkeit von Sprache [On the influence of a single echo on the audibility of speech]. *Acustica*, **1**(2), 49–58.

van der Heijden, M. and Trahiotis, C. (1999). Masking with interaurally delayed stimuli: The use of 'internal' delays in binaural detection. *J. Acoust. Soc. Am.*, **105**(1), 388–399.

Helfer, K. S. (1994). Binaural cues and consonant perception in reverberation and noise. *J. Speech Hear. Res.*, **37**(2), 429–438.

Heller, L. M. and Trahiotis, C. (1996). Extents of laterality and binaural interference effects. *J. Acoust. Soc. Am.*, **99**(6), 3632–3637.

Henning, G. B. (1974). Detectability of interaural delay in high-frequency complex waveforms. *J. Acoust. Soc. Am.*, **55**(1), 84–90.

Hirsh, I. J. (1948). The influence of interaural phase on interaural summation and inhibition. *J. Acoust. Soc. Am.*, **20**(4), 536–544.

Hohmann, V. (2002). Frequency analysis and synthesis using a gammatone filterbank. *Acta Acust. United Ac.*, **88**(3), 433–442.

van der Horst, R., Leeuw, A. R., and Dreschler, W. A. (1999). Importance of temporal-envelope cues in consonant recognition. *J. Acoust. Soc. Am.*, **105**(3), 1801–1809.

Houtgast, T. (1989). Frequency selectivity in amplitude-modulation detection. *J. Acoust. Soc. Am.*, **85**(4), 1676–1680.

Houtgast, T. and Steeneken, H. J. M. (1973). The modulation transfer function in room acoustics as a predictor of speech intelligibility. *Acustica*, **28**(1), 66–73.

Houtgast, T. and Steeneken, H. J. M. (1984). A multi-language evaluation of the RASTI-method for estimating speech-intelligibility in auditoria. *Acustica*, **54**(4), 185–199.

Houtgast, T. and Steeneken, H. J. M. (1985). A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *J. Acoust. Soc. Am.*, **77**(3), 1069–1077.

IEC 60268-16 (2003). *Objective rating of speech intelligibility by speech transmission index* (3rd ed.). Geneva, Switzerland: International Electrotechnical Commission.

Jeffress, L. A. (1948). A place theory of sound localization. *J. Comp. Physiol. Psychol.*, **41**(1), 35–39.

Jeffress, L. A., Blodgett, H. C., and Deatherage, B. H. (1962). Masking and interaural phase. II. 167 cycles. *J. Acoust. Soc. Am.*, **34**(8), 1124–1126.

Jepsen, M. L., Ewert, S. D., and Dau, T. (2008). A computational model of human auditory signal processing and perception. *J. Acoust. Soc. Am.*, **124**(1), 422–438.

Klumpp, R. G. and Eady, H. R. (1956). Some measurements of interaural time difference thresholds. *J. Acoust. Soc. Am.*, **28**(5), 859–860.

Koenig, W. (1950). Subjective effects in binaural hearing. *J. Acoust. Soc. Am.*, **22**(1), 61–62.

Kohlrausch, A., Fassel, R., and Dau, T. (2000). The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers. *J. Acoust. Soc. Am.*, **108**(2), 723–734.

Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., Oxenham, A. J., and Püschel, D. (1997). Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations. *Acustica*, **83**(4), 659–669.

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), 2236–2250.

Kuttruff, H. (2000). *Room acoustics* (4$^{\text{th}}$ ed.). London: Spon Press.

Lawson, J. L. and Uhlenbeck, G. E. (1950). *Threshold signals*, volume 24 of *Radiation Laboratory Series*. New York: McGraw-Hill.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.*, **49**(2), 467–477.

Libbey, B. and Rogers, P. H. (2004). The effect of overlap-masking on binaural reverberant word intelligibility. *J. Acoust. Soc. Am.*, **116**(5), 3141–3151.

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), 150–159.

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), 1608–1622.

Lochner, J. P. A. and Burger, J. F. (1964). Influence of reflections on auditorium acoustics. *J. Sound Vib.*, **1**(4), 426–454.

Merimaa, J., Peltonen, T., and Lokki, T. (2005a). Concert hall impulse responses - Pori, Finland: Analysis results. Available online at: `http://www.acoustics.hut.fi/projects/poririrs` [Last viewed 25-Jul-2008].

Merimaa, J., Peltonen, T., and Lokki, T. (2005b). Concert hall impulse responses - Pori, Finland: Reference. Available online at: `http://www.acoustics.hut.fi/projects/poririrs` [Last viewed 25-Jul-2008].

Miller, G. A. and Nicely, P. E. (1955). An analysis of perceptual confusions among some english consonants. *J. Acoust. Soc. Am.*, **27**(2), 338–352.

Mills, A. W. (1960). Lateralization of high-frequency tones. *J. Acoust. Soc. Am.*, **32**(1), 132–134.

Miyata, H., Nomura, H., and Houtgast, T. (1991). Speech-intelligibility and subjective MTF under diotic and dichotic listening conditions in reverberant sound fields. *Acustica*, **73**(4), 200–207.

Moncur, J. P. and Dirks, D. (1967). Binaural and monaural speech intelligibility in reverberation. *J. Speech Hear. Res.*, **10**(2), 186–195.

Moore, B. C. J. (2003). *An introduction to the psychology of hearing* (5$^{\text{th}}$ ed.). London: Academic Press.

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (1993). Detection of temporal gaps in sinusoids: Effects of frequency and level. *J. Acoust. Soc. Am.*, **93**(3), 1563–1570.

Moore, B. C. J., Sek, A., and Shailer, M. J. (1995). Modulation discrimination interference for narrow-band noise modulators. *J. Acoust. Soc. Am.*, **97**(4), 2493–2497.

Moore, B. C. J. and Shailer, M. J. (1992). Modulation discrimination interference and auditory grouping. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **336**(1278), 339–346.

Nábělek, A. K., Letowski, T. R., and Tucker, F. M. (1989). Reverberant overlap- and self-masking in consonant identification. *J. Acoust. Soc. Am.*, **86**(4), 1259–1265.

Nábělek, A. K. and Mason, D. (1981). Effect of noise and reverberation on binaural and monaural word identification by subjects with various audiograms. *J. Speech Hear. Res.*, **24**(3), 375–383.

Nábělek, A. K. and Pickett, J. M. (1974a). Monaural and binaural speech perception through hearing aids under noise and reverberation with normal and hearing-impaired listeners. *J. Speech Hear. Res.*, **17**(4), 724–739.

Nábělek, A. K. and Pickett, J. M. (1974b). Reception of consonants in a classroom as affected by monaural and binaural listening, noise, reverberation, and hearing aids. *J. Acoust. Soc. Am.*, **56**(2), 628–639.

Nábělek, A. K. and Robinson, P. K. (1982). Monaural and binaural speech perception in reverberation for listeners of various ages. *J. Acoust. Soc. Am.*, **71**(5), 1242–1248.

Nuetzel, J. M. and Hafter, E. R. (1981). Discrimination of interaural delays in complex waveforms: Spectral effects. *J. Acoust. Soc. Am.*, **69**(4), 1112–1118.

Oxenham, A. J. and Dau, T. (2001). Modulation detection interference: Effects of concurrent and sequential streaming. *J. Acoust. Soc. Am.*, **110**(1), 402–408.

Park, M., Nelson, P. A., and Kim, Y. (2005). An auditory process model for sound localization. In *Proceedings IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (ASPAA)*, (pp. 122–125)., New Paltz, NY.

van de Par, S. and Kohlrausch, A. (1997). A new approach to comparing binaural masking level differences at low and high frequencies. *J. Acoust. Soc. Am.*, **101**(3), 1671–1680.

Phatak, S. A., Lovitt, A., and Allen, J. B. (2008). Consonant confusions in white noise. *J. Acoust. Soc. Am.*, **124**(2), 1220–1233.

Piechowiak, T., Ewert, S. D., and Dau, T. (2007). Modeling comodulation masking release using an equalization-cancellation mechanism. *J. Acoust. Soc. Am.*, **121**(4), 2111–2126.

Plomp, R. (1964). Rate of decay of auditory sensation. *J. Acoust. Soc. Am.*, **36**(2), 277–282.

Price, R. (1955). A note on the envelope and phase-modulated components of narrow-band gaussian noise. *IRE Trans. Inform. Theory*, **1**(2), 9–13.

Pumplin, J. (1985). Low-noise noise. *J. Acoust. Soc. Am.*, **78**(1), 100–104.

Régnier, M. S. and Allen, J. B. (2008). A method to identify noise-robust perceptual features: application for consonant /t/. *J. Acoust. Soc. Am.*, **123**(5), 2801–2814.

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos. Trans. R. Soc. London, Ser. B*, **336**(1278), 367–373.

Saberi, K., Dostal, L., Sadralodabai, T., Bull, V., and Perrott, D. R. (1991). Free-field release from masking. *J. Acoust. Soc. Am.*, **90**(3), 1355–1370.

Schroeder, M. R. (1981). Modulation transfer-functions: Definition and measurement. *Acustica*, **49**(3), 179–182.

Sheft, S. and Yost, W. A. (1997). Binaural modulation detection interference. *J. Acoust. Soc. Am.*, **102**(3), 1791–1798.

Sheft, S. and Yost, W. A. (2007). Discrimination of starting phase with sinusoidal envelope modulation. *J. Acoust. Soc. Am.*, **121**(2), EL84–EL89.

Soli, S. D. and Arabie, P. (1979). Auditory versus phonetic accounts of observed confusions between consonant phonemes. *J. Acoust. Soc. Am.*, **66**(1), 46–59.

Stellmack, M. A., Viemeister, N. F., and Byrne, A. J. (2005). Monaural and interaural temporal modulation transfer functions measured with 5-kHz carriers. *J. Acoust. Soc. Am.*, **118**(4), 2507–2518.

Stern, R. M. and Colburn, H. S. (1978). Theory of binaural interaction based in auditory-nerve data. IV. A model for subjective lateral position. *J. Acoust. Soc. Am.*, **64**(1), 127–140.

Stevens, K. N. (1980). Acoustic correlates of some phonetic categories. *J. Acoust. Soc. Am.*, **68**(3), 836–842.

Suzuki, Y., Yokoyama, T., and Sone, T. (1993). Influence of interfering noise on the sound localization of a pure tone. *J. Acoust. Soc. Jpn. (E)*, **14**(5), 327–339.

Terhardt, E. (1968). Über die durch amplitudenmodulierte Sinustöne hervorgerufene Hörempfindung. [The auditory sensation produced by amplitude modulated tones]. *Acustica*, **20**(4), 210–214.

Thiele, R. (1953). Richtungsverteilung und Zeitfolge der Schallrückwürfe in Räumen [Direction and time dependence of sound reflections in rooms]. *Acustica*, **3**, 291–302.

Thompson, E. R. and Dau, T. (2008). Binaural processing of modulated interaural level differences. *J. Acoust. Soc. Am.*, **123**(2), 1017–1029.

Thompson, E. R. and Dau, T. (2009). Monaural and binaural subjective modulation transfer functions in reverberation. *J. Acoust. Soc. Am.* [submitted].

Trahiotis, C. and Stern, R. M. (1989). Lateralization of bands of noise: Effects of bandwidth and differences of interaural time and phase. *J. Acoust. Soc. Am.*, **86**(4), 1285–1293.

Viemeister, N. F. (1979). Temporal modulation transfer functions based upon modulation thresholds. *J. Acoust. Soc. Am.*, **66**(5), 1364–1380.

Wickens, T. D. (2002). *Elementary signal detection theory*. Oxford University Press.

van Wijngaarden, S. J. and Drullman, R. (2008). Binaural intelligibility prediction based on the speech transmission index. *J. Acoust. Soc. Am.*, **123**(6), 4514–4523.

Witton, C., Green, G. G., Rees, A., and Henning, G. B. (2000). Monaural and binaural detection of sinusoidal phase modulation of a 500-Hz tone. *J. Acoust. Soc. Am.*, **108**(4), 1826–1833.

Yang, W. and Bradley, J. S. (2009). Effects of room acoustics on the intelligibility of speech in classrooms for young children. *J. Acoust. Soc. Am.*, **125**(2), 922–933.

Yost, W. A., Sheft, S., and Opie, J. (1989). Modulation interference in detection and discrimination of amplitude modulation. *J. Acoust. Soc. Am.*, **86**(6), 2138–2147.

Zurek, P. M., Freyman, R. L., and Balakrishnan, U. (2004). Auditory target detection in reverberation. *J. Acoust. Soc. Am.*, **115**(4), 1609–1620.

Appendix A

Confusion matrices from the consonant identification experiment

The following tables contain the confusion matrices from the consonant identification experiment, pooled across listeners, for the different listening conditions and impulse responses.

Table A.1: Confusion matrix for the anechoic condition.

		Response												
		p	t	k	b	d	g	m	n	f	s	v	z	Total
Presentation	p	216	0	0	0	0	0	0	0	0	0	0	0	216
	t	0	216	0	0	0	0	0	0	0	0	0	0	216
	k	0	0	215	0	0	1	0	0	0	0	0	0	216
	b	4	0	0	203	0	0	0	0	0	0	0	0	207
	d	0	0	1	0	197	0	0	0	0	0	0	0	198
	g	0	0	3	0	0	195	0	0	0	0	0	0	198
	m	0	0	0	0	0	0	215	1	0	0	0	0	216
	n	0	0	0	1	1	0	0	214	0	0	0	0	216
	f	0	0	0	0	0	0	0	0	205	0	2	0	207
	s	0	0	1	0	0	0	0	0	0	205	0	10	216
	v	0	0	0	1	0	0	0	0	1	0	204	1	207
	z	0	0	0	0	0	0	0	0	0	4	0	212	216
Total		220	216	220	205	198	196	215	215	206	209	206	223	2529

Table A.2: Confusion matrix with left-ear listening and the S1/R2 impulse response

		Response												
		p	t	k	b	d	g	m	n	f	s	v	z	Total
Presentation	p	165	6	9	12	6	4	2	4	2	0	6	0	216
	t	1	198	1	3	13	0	0	0	0	0	0	0	216
	k	10	11	161	0	4	27	0	0	1	0	2	0	216
	b	38	0	0	122	6	1	6	11	0	1	22	0	207
	d	1	26	0	3	160	4	1	1	0	0	2	0	198
	g	5	5	41	11	22	109	0	1	0	0	4	0	198
	m	1	1	0	20	0	0	114	73	0	0	7	0	216
	n	1	0	1	4	3	2	24	179	0	0	2	0	216
	f	6	5	1	11	0	1	0	0	143	15	23	2	207
	s	0	13	0	0	0	0	0	0	1	160	0	42	216
	v	2	2	1	21	3	0	4	5	15	0	145	9	207
	z	0	4	0	0	8	0	0	0	0	14	1	189	216
		230	271	215	207	225	148	151	274	162	190	214	242	2529

Table A.3: Confusion matrix with right-ear listening and the S1/R2 impulse response

		Response												
		p	t	k	b	d	g	m	n	f	s	v	z	Total
Presentation	p	175	6	9	8	4	2	7	1	2	0	2	0	216
	t	0	202	0	0	9	0	0	0	0	2	0	3	216
	k	3	5	178	0	3	26	0	1	0	0	0	0	216
	b	39	0	1	131	4	3	8	10	0	0	11	0	207
	d	7	37	0	1	145	3	0	3	1	0	0	1	198
	g	4	5	59	12	14	101	1	1	0	0	1	0	198
	m	5	0	0	35	2	0	107	65	1	0	1	0	216
	n	3	0	0	4	9	4	20	176	0	0	0	0	216
	f	16	2	2	3	1	0	0	0	147	12	22	2	207
	s	0	7	0	0	1	0	0	0	0	164	0	44	216
	v	6	1	3	24	1	0	4	3	21	0	138	6	207
	z	0	7	0	0	2	0	0	0	0	17	2	188	216
Total		258	272	252	218	195	139	147	260	172	195	177	244	2529

Table A.4: Confusion matrix with binaural listening and the S1/R2 impulse response

		Response												
		p	t	k	b	d	g	m	n	f	s	v	z	Total
Presentation	p	189	3	8	10	2	3	1	0	0	0	0	0	216
	t	1	203	0	0	10	0	0	0	0	1	0	1	216
	k	5	4	179	0	3	25	0	0	0	0	0	0	216
	b	28	0	0	153	6	2	5	6	0	0	7	0	207
	d	1	16	0	3	173	3	0	1	0	0	1	0	198
	g	4	2	29	5	13	141	0	1	1	0	2	0	198
	m	0	0	0	1	3	0	155	53	0	0	3	1	216
	n	0	0	0	1	5	2	18	189	0	0	1	0	216
	f	9	5	3	4	1	0	0	0	172	1	9	3	207
	s	0	2	0	0	1	0	0	0	0	172	0	41	216
	v	5	1	0	20	2	0	2	4	7	0	160	6	207
	z	0	2	0	0	0	0	0	0	0	4	2	208	216
Total		242	238	219	197	219	176	181	254	180	178	185	260	2529

Table A.5: Confusion matrix with left-ear listening and the S1/R3 impulse response

		Response												
		p	t	k	b	d	g	m	n	f	s	v	z	Total
Presentation	p	152	6	10	19	5	3	7	3	1	0	10	0	216
	t	0	205	0	1	4	0	0	0	0	0	3	3	216
	k	6	14	160	0	2	31	0	0	3	0	0	0	216
	b	47	0	0	105	7	4	7	11	0	0	26	0	207
	d	4	37	1	4	140	4	1	4	0	0	3	0	198
	g	6	2	53	5	14	101	0	11	1	0	5	0	198
	m	3	0	1	17	2	0	121	64	1	0	7	0	216
	n	0	0	1	2	11	0	24	175	0	0	3	0	216
	f	7	5	2	7	0	0	0	0	149	11	25	1	207
	s	0	15	0	0	0	0	0	0	2	157	0	42	216
	v	6	1	2	23	1	1	3	5	23	0	134	8	207
	z	0	7	0	0	8	0	0	0	0	34	0	167	216
Total		231	292	230	183	194	144	163	273	180	202	216	221	2529

Table A.6: Confusion matrix with right-ear listening and the S1/R3 impulse response

		Response												
		p	t	k	b	d	g	m	n	f	s	v	z	Total
Presentation	p	162	4	11	18	4	4	2	4	0	0	7	0	216
	t	5	196	1	1	6	1	0	0	0	2	1	3	216
	k	8	7	171	1	0	28	0	1	0	0	0	0	216
	b	51	0	1	103	3	6	9	10	1	0	23	0	207
	d	4	36	5	3	129	15	0	1	0	0	5	0	198
	g	6	5	62	3	16	96	0	6	0	0	4	0	198
	m	6	0	0	17	2	1	124	58	0	0	8	0	216
	n	1	0	0	3	5	2	17	185	0	0	3	0	216
	f	8	5	2	6	0	0	1	0	141	11	33	0	207
	s	0	16	0	0	1	0	0	1	0	152	0	46	216
	v	5	0	2	22	4	1	4	6	22	0	131	10	207
	z	0	5	0	0	3	0	0	0	0	28	4	176	216
Total		256	274	255	177	173	154	157	272	164	193	219	235	2529

Table A.7: Confusion matrix with binaural listening and the S1/R3 impulse response

		Response												
		p	t	k	b	d	g	m	n	f	s	v	z	Total
Presentation	p	174	3	8	16	1	4	1	2	2	0	5	0	216
	t	0	199	3	0	6	0	0	0	0	0	1	7	216
	k	5	4	168	0	4	34	0	0	0	0	1	0	216
	b	38	0	0	130	3	1	13	7	0	0	15	0	207
	d	1	30	2	0	157	5	0	1	0	0	1	1	198
	g	1	1	40	5	15	122	3	6	0	0	5	0	198
	m	1	0	0	7	0	0	149	59	0	0	0	0	216
	n	0	1	0	1	5	1	17	190	1	0	0	0	216
	f	9	2	0	5	0	2	1	1	165	5	17	0	207
	s	0	4	0	0	0	0	0	0	0	164	0	48	216
	v	0	1	0	19	3	2	5	8	16	0	146	7	207
	z	0	2	0	0	4	0	0	0	0	14	2	194	216
Total		229	247	221	183	198	171	189	274	184	183	193	257	2529

Table A.8: Confusion matrix with left-ear listening and the S2/R2 impulse response

		Response												
		p	t	k	b	d	g	m	n	f	s	v	z	Total
Presentation	p	148	2	17	17	2	7	8	1	3	0	10	1	216
	t	2	191	2	0	4	0	0	0	1	5	1	10	216
	k	11	5	164	1	4	27	0	0	2	1	1	0	216
	b	34	0	1	121	7	6	6	8	0	0	24	0	207
	d	1	24	3	2	142	12	0	8	0	0	6	0	198
	g	8	0	44	14	16	108	0	5	1	0	2	0	198
	m	4	0	0	19	1	0	121	61	0	0	10	0	216
	n	0	0	0	3	12	3	16	181	0	0	1	0	216
	f	15	1	2	4	0	1	0	0	143	18	21	2	207
	s	0	5	1	0	0	0	0	0	2	166	0	42	216
	v	3	0	1	22	0	0	6	2	14	1	145	13	207
	z	0	0	1	0	2	0	0	0	0	23	4	186	216
Total		226	228	236	203	190	164	157	266	166	214	225	254	2529

Table A.9: Confusion matrix with right-ear listening and the S2/R2 impulse response

		Response												
		p	t	k	b	d	g	m	n	f	s	v	z	Total
Presentation	p	182	4	14	4	3	3	3	0	1	0	2	0	216
	t	0	213	2	0	1	0	0	0	0	0	0	0	216
	k	7	10	172	0	2	22	0	1	2	0	0	0	216
	b	46	0	0	121	4	6	3	12	1	0	14	0	207
	d	5	37	2	3	138	11	0	2	0	0	0	0	198
	g	11	4	56	3	13	109	1	0	0	0	1	0	198
	m	3	0	0	23	1	3	129	55	0	0	2	0	216
	n	2	1	0	4	8	1	28	171	0	0	1	0	216
	f	17	4	2	3	1	1	0	0	146	9	23	1	207
	s	0	11	0	0	0	1	0	0	0	171	0	33	216
	v	7	4	0	20	1	2	6	4	7	0	145	11	207
	z	0	4	1	0	3	0	0	0	1	16	1	190	216
Total		280	292	249	181	175	159	170	245	158	196	189	235	2529

Table A.10: Confusion matrix with binaural listening and the S2/R2 impulse response

		Response												
		p	t	k	b	d	g	m	n	f	s	v	z	Total
Presentation	p	186	1	7	9	2	7	1	2	0	0	1	0	216
	t	0	205	4	1	4	0	0	0	0	1	0	1	216
	k	5	1	187	0	1	21	0	0	0	1	0	0	216
	b	26	0	0	157	1	2	6	5	0	0	10	0	207
	d	1	25	0	1	160	8	0	3	0	0	0	0	198
	g	3	5	36	6	9	135	0	3	0	0	1	0	198
	m	1	0	0	8	0	1	158	45	0	0	3	0	216
	n	0	0	0	0	5	0	17	194	0	0	0	0	216
	f	12	1	0	3	0	2	0	0	173	4	12	0	207
	s	0	2	0	0	0	0	0	0	1	168	0	45	216
	v	1	0	0	21	0	1	0	2	4	0	170	8	207
	z	0	1	0	0	2	0	0	0	0	3	4	206	216
Total		235	241	234	206	184	177	182	254	178	177	201	260	2529

Contributions to Hearing Research

Vol. 1: *Gilles Pigasse*, Deriving cochlear delays in humans using otoacoustic emissions and auditory evoked potentials, July, 2008.

Vol. 2: *Olaf Strelcyk*, Peripheral auditory processing and speech reception in impaired hearing, April, 2009.

Vol. 3: *Eric R. Thompson*, Characterizing binaural processing of amplitude-modulated sounds, June, 2009.