Binaural auditory processing and temporal periodicity - Experiments and models

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

von Dipl. Phys.
Martin Julius Christoph Klein-Hennig
geboren am 18. Januar 1984
in Oldenburg

Gutachter: Prof. Dr. Volker Hohmann
Zweitgutachter: Prof. Dr. Dr. Birger Kollmeier
Tag der Dissertation: 12. 12. 2014

Abstract

The ability of the human auditory system to function under unfavorable acoustic conditions, for example focusing on a single talker in a multi-talker environment with background noise and reverberation, is well known (e.g., Cherry, 1953; Kaiser and David, 1960). The theory of auditory scene analysis (ASA, Bregman, 1994) states that the auditory system actively and passively groups several signal features into internal representations of sound sources. This separation helps in tasks like understanding speech in noise or tracking the movement of a sound source. Two of the important signal features are binaural cues derived from a joint processing of the signals received at both ears and pitch or harmonicity, as, e.g., generated by voiced speech sounds (Darwin and Carlyon, 1995). Binaural cues give information about the location and movement of a sound source, while harmonicity processing enables grouping of signal energy according to a fundamental frequency (F0), which helps for example in the separation of different voices. As the signals of interest in auditory scene analysis (voiced speech, animal calls, music) are often periodical (e.g., Fletcher, 1992), the aim of this thesis is to provide psychophysical data and develop and evaluate models for binaural and harmonic processing of periodic signals.

The first part of this thesis deals with the processing of interaural time differences (ITDs) in signals with a periodic envelope. Here, a custom periodic envelope is constructed consisting of several segments. The influence of each envelope segment on the sensitivity to ITDs is measured in psychophysical experiments. The ability of an established model for ITD sensitivity (Bernstein and Trahiotis, 2002) to predict the data is evaluated and the model is extended by principles of neuronal adaptation to improve model performance. This is of particular interest as it offers valuable hints on the binaural processing of voiced speech sounds at high frequencies and possible coding strategies for binaural hearing aids or cochlear implants.

The second and third part deal with combined processing of harmonicity and ITDs. The second part establishes the method of measuring detection thresholds in harmonicity research. In psychophysical experiments, the ability of subjects to detect a single target component embedded in a tone complex is tested, while the target tone is in harmonic or mis-

tuned relationship to the masking tone complex. Based on this method, the third part of this thesis reports on combination experiments, where the target tone is additionally presented with an interaural phase difference (IPD). An auditory computer model based on amplitude modulation processing (Dau et al., 1997) and equalization-cancellation (Durlach, 1963) is used to predict the psychophysical results. Various hypotheses on the combination of binaural and harmonicity information are tested against the human and model data. The results of both studies shed light on the combined processing of two important signal features in ASA and are of high relevance for the development of ASA models.

The findings presented here give valuable insights into the joint processing of binaural and periodicity cues and lead to a better understanding of the ASA process in humans. The psychophysical and model results offer hints for the development of models of ASA in the field of computational auditory scene analysis (CASA) and new coding strategies for binaural hearing aids and cochlear implants.

Zusammenfassung

Die Fähigkeit des menschlichen auditorischen Systems auch unter schwierigen akustischen Bedingungen zu funktionieren, zum Beispiel bei der gezielten Wahrnehmung eines einzelnen Sprechers in einer Umgebung mit anderen Sprechern, Hall und Hintergrundgeräuschen, ist bekannt (z.B. Cherry, 1953; Kaiser and David, 1960). Die Theorie der auditorischen Szenenanalyse (ASA, Bregman, 1994) besagt, dass das auditorische System in aktiven und passiven Gruppierungsmechanismen eine Vielzahl von Signaleigenschaften in interne Representationen von Schallquellen zusammenfasst. Diese Separation von Signalen erleichtert das Verstehen von Sprache in Hintergrundgeräuschen oder das Verfolgen einer beweglichen Schallquelle. Zwei der wichtigen Signaleigenschaften sind binaurale Merkmale, die aus einer gemeinsamen Verarbeitung der Signale an beiden Ohren extrahiert werden, und die Tonhöhe oder Harmonizität eines Signals, wie z.B. in stimmhafter Sprache (Darwin and Carlyon, 1995). Die binauralen Merkmale geben Aufschluss über den Ort und die Bewegung einer Schallquelle im Raum, während durch die Verarbeitung der Harmonizität eines Signals die Signalenergie einer bestimmten Grundfrequenz (F0) zugeordnet werden kann, was die Unterscheidung von Stimmen ermöglicht. Die in der auditorischen Szenenanalyse relevanten Zielsignale (z.B.Sprache, Tierrufe, Musik) sind meist periodischer Natur (z.B. Fletcher, 1992). Das Ziel dieser Arbeit ist deshalb der Gewinn psychophysischer Daten und die Entwicklung und Evaluation von Modellen für die binaurale und harmonizitätsbezogene Verarbeitung von periodischen Signalen.

Der erste Teil der Arbeit beschäftigt sich mit der Verarbeitung von interauralen Laufzeitunterschieden (engl. interaural time difference, ITD) in Signalen mit einer periodischen Einhüllenden. Hier wird eine spezifisch angepasste Einhüllendenform entwickelt, die aus verschiedenen Segmenten besteht. Der Einfluss jedes Segments auf die Sensitivität für ITDs wird in psychophysischen Experimenten gemessen. Ein etabliertes Modell zur Vorhersage von ITD-Sensitivität (Bernstein and Trahiotis, 2002) wird anhand der gewonnen Daten evaluiert und zur Verbesserung der Vorhersagen um Prinzipien der neuronalen Adaptation erweitert. Die Ergebnisse sind besonders relevant für die binaurale Verarbeitung von stimmhafter Sprache bei hohen Frequenzen und mögliche Codierungsstrategien für binau-

rale Hörgeräte oder Cochleaimplantate.

Der zweite und dritte Teil der Arbeit untersuchen die kombinierte Verarbeitung von Harmonizität und ITDs. In der zweiten Studie wird eine Methode zur Messung von Detektionsschwellen für die Untersuchung von Harmonizitätsverarbeitung etabliert. Hier wird in psychophysischen Experimenten die Fähigkeit der Versuchspersonen, eine einzelne Zielkomponente aus einem Tonkomplex herauszuhören, getestet, wobei das Zielsignal in harmonischer oder verstimmter Beziehung zum maskierenden Tonkomplex ist. Ausgehend von dieser Methode wird in der dritten Studie von Kombinationsexperimenten berichtet, in denen das Zielsignal zusätzlich mit einem interauralen Phasenunterschied (engl. interaural phase difference, IPD) versehen wird. In der Arbeit wird ein auditorisch motiviertes Computermodell mit Modulationsverarbeitung (Dau et al., 1997) und einem "equalization-cancellation" Ansatz (Angleichung und Auslöschung der Signale an beiden Ohren, siehe Durlach, 1963) zur Vorhersage der psychophysischen Daten entwickelt. Mit dem Modell und den Versuchsdaten werden mehrere Hypothesen zur Kombination von binauralen und harmonischen Signalmerkmalen im auditorischen System überprüft. Die Ergebnisse der beiden Studien liefern neue Hinweise zur kombinierten Verarbeitung von zwei wichtigen Signalmerkmalen der auditorischen Szenenanalyse, die von großem Nutzen für Entwicklung von ASA-Modellen sind.

Die in dieser Arbeit vorgestellten Ergebnisse geben wertvolle Einsichten in die gemeinsame Verarbeitung von binauralen und periodizitätsbasierten Signalmerkmalen und führen zu einem besseren Verständnis der auditorischen Szenenanalyse im Menschen. Die psychophysischen Daten und Modellergebnisse geben Hinweise für die Entwicklung von Modellen im Feld der rechnergestützten auditorischen Szenenanalyse (engl. computational auditory scene analysis, CASA), sowie für neue Codierungsstrategien in binauralen Hörgeräten und Cochleaimplantaten.

Contents

1 General introduction **11**

1.1 Auditory scene analysis 11
1.2 Localization of periodic signals 12
1.3 Pitch, periodicity and harmonicity 13
1.4 Harmonicity and localization 16
1.5 Consequences for binaural hearing aids and CASA 17

2 The influence of envelope segments on sensitivity to interaural time delays **19**

2.1 Introduction 20
2.2 Methods 23
2.2.1 Subjects 23
2.2.2 Apparatus and stimuli 23
2.2.3 Procedure 26
2.2.4 Models 26
2.3 Experimental results 30
2.3.1 Experiment 1: Attack duration 30
2.3.2 Experiment 2: Hold duration 32
2.3.3 Experiment 3: Decay duration 34
2.3.4 Experiment 4: Pause duration 35
2.3.5 Experiment 5: Level 37
2.3.6 Experiment 6: Modulation frequency 39
2.3.7 Experiment 7: Direct current offset 41
2.3.8 Experiment 8: Temporal asymmetry 42
2.3.9 Experiment 9: Transposed tone 43
2.4 Discussion 45
2.4.1 Influence of isolated envelope segments 45
2.4.2 Influence of analytical envelope parameters 47
2.4.3 Relation between data and NCC model 49

2.4.4 Relation between data and model with adaptation mechanism . . . 51
2.4.5 Implications for future modeling . . . 53
2.5 Conclusions . . . 54

3 Effect of mistuning on the detection of a tone masked by a tone complex 57
3.1 Introduction . . . 58
3.2 Materials and Methods . . . 59
3.2.1 Ethics statement . . . 59
3.2.2 Subjects . . . 59
3.2.3 Stimuli . . . 59
3.2.4 Procedure . . . 61
3.3 Results . . . 61
3.3.1 Experiment 1: resolved harmonics, F0 = 160 Hz . . . 61
3.3.2 Experiment 2: unresolved harmonics, F0 = 40 Hz . . . 62
3.3.3 Experiment 3: unresolved harmonics, F0 = 160 Hz . . . 63
3.4 Discussion . . . 64
3.5 Conclusions . . . 66

4 Combination of binaural and harmonic masking release effects 67
4.1 Introduction . . . 68
4.2 Methods . . . 71
4.2.1 Subjects . . . 71
4.2.2 Apparatus and stimuli . . . 71
4.2.3 Procedure . . . 73
4.2.4 Models . . . 73
4.3 Experimental results . . . 76
4.3.1 Experiment 1: F0 = 160 Hz, f_t = 800 Hz . . . 76
4.3.2 Experiment 2: F0 = 40 Hz, f_t = 800 Hz . . . 77
4.3.3 Experiment 3: F0 = 40 Hz, f_t = 800 Hz, broadband . . . 77
4.4 Model results . . . 78
4.4.1 Binaural processing before modulation processing . . . 78
4.4.2 Binaural processing after modulation processing . . . 79
4.4.3 Parallel processing . . . 79
4.5 Discussion . . . 80
4.5.1 Psychophysical results . . . 80
4.5.2 Model results . . . 81

4.6 Conclusions 82

5 General conclusions **85**

Bibliography **87**

Chapter 1

General introduction

1.1 Auditory scene analysis

The human auditory system performs well at extracting desired acoustical information from challenging listening environments, such as reverberant rooms, the presence of multiple, distracting sound sources and generally difficult signal-to-noise ratios (SNRs). Being able to listen to a talker in such an environment is commonly referred to as the "cocktail-party" effect (e.g., Cherry, 1953; Kaiser and David, 1960). Several simple mechanisms responsible for the cocktail-party effect have been proposed. These involve the use of binaural cues such as interaural time differences (ITDs) and interaural level differences (ILDs) for sound source localization (e.g., Kaiser and David, 1960; Carhart et al., 1967; Mitchell et al., 1971). Some studies, however, showed that there are more cues that are used to separate sound sources, like for example fundamental frequencies (F0s) of talkers (e.g., Parsons, 1976), common temporal onsets over frequency ranges (Dannenbring and Bregman, 1978), frequency modulation statistics (McAdams, 1989) and spectral regularity (Roberts and Bailey, 1996).

A widely accepted theory about the combined processing of these signal features into so-called "auditory objects" is called auditory scene analysis (Bregman, 1994). Auditory objects combine several cues like those mentioned above into an internal representation of an external sound source. The process of assigning signal features to auditory objects is called auditory grouping. Still, little is understood about the way auditory grouping works on a physiological level. Signal processing approaches using multiple signal features, as in auditory grouping, have proven successful for example in noise reduction systems in hearing aids or speech recognition. But these approaches are mostly based on complex and computationally elaborate signal processing that is unlikely to occur in the auditory system. Auditory motivated models would provide a means of studying the influence of certain hearing aid processing schemes on the ability of auditory scene analysis. The performance

of such schemes at ASA could be evaluated without the need for large-scale psychophysical studies.

The current development of models of auditory scene analysis (computational auditory scene analysis, CASA) is still far away from a comprehensive model, as there are many signal features and combination possibilities to investigate. Apart from sound source localization, which is undoubtedly the most important mechanism in auditory scene analysis, a promising research direction is its combination with the processing of periodic sounds, as the most important acoustic signals in every-day life are periodic: (voiced) speech, animal calls, music (e.g., Fletcher, 1992; Ladefoged, 1996).

Thus, the goal of this work is to provide psychophysical data and model approaches on the processing of binaural, periodic signals. The first part of this work focuses on the processing of binaural signal features in the envelope of sounds, while the second and third part deal with the combined processing of harmonicity and binaural cues in periodic signals.

1.2 Localization of periodic signals

The most simple periodic signals are "pure tones", i.e. sine waves. The localization of pure-tones has been studied extensively in the last 100 years, since the generation of pure-tones does not involve elaborate signal processing equipment. One of the important findings on sound source localization is the "duplex theory" by Rayleigh (1907), which states that the human auditory system uses the interaural time difference (ITD), which occurs due to the distance between left and right ear, for frequencies up to about 1500 Hz. Above 1500 Hz, the interaural level difference caused by sound shadowing of the head is used for localization. This frequency limitation for the use of ITDs is believed to be caused by the lack of phase-locking in auditory nerve cells at such high frequencies. The cells are no longer able to fire synchronously with the frequency of the sound signal due to physicochemical limitations (Palmer and Russell, 1986).

With more sophisticated methods of signal generation, an interesting problem not described by the duplex theory could be studied: amplitude modulated signals with a high carrier frequency (e.g., Leakey et al., 1958; Henning, 1974). These studies found that the binaural system in such cases is able to exploit interaural time differences in the envelope only, as the carrier ITD is not accessible due to its high frequency. With the discovery, development and improvement of cochlear implants, the use of localization cues in the envelope of sounds has gained interest, as the electrical signals generated by the cochlear implant encode most information in the envelope.

The localization of sinusoidally amplitude modulated (SAM) sounds has been studied in depth by, e.g., McFadden and Pasanen (1976), Bernstein and Trahiotis (1985) and Bernstein and Trahiotis (1994), with studies investigating the influence of modulation frequency and depth, and more recently using different waveforms such as transposed tones (van de Par and Kohlrausch, 1997; Bernstein and Trahiotis, 2002) or so-called "raised sines" (Bernstein and Trahiotis, 2009), where the exponent of the modulator is increased. These "analytical" waveforms have a disadvantage, though: modifying a parameter such as modulation frequency intrinsically changes secondary envelope parameters such as the steepness of its slopes, or in the case of raised sines the duration of zero modulation energy segments in the envelope cycle with increased exponent of the modulator. The isolated influence of these secondary parameters is worth investigating, as it could give insight about the most influential parts for localization of an amplitude modulated signal. This would provide valuable information about binaural processing in the auditory system, as well as hints for possible coding strategies in cochlear implants.

The second chapter of this work investigates the influence of such secondary parameters like the "attack" flank duration and the "pause" time between two envelope cycles on the sensitivity to interaural time delays. The psychophysical experiments involved normal-hearing test subjects. In the modeling part of the study, the ability of an established lateralization model based on cross-correlation between left and right ear signals (e.g., Bernstein and Trahiotis, 2002) to predict the experimental results, was evaluated.

1.3 Pitch, periodicity and harmonicity

The role of temporal periodicity in pitch perception becomes clear by looking at the definition of pitch itself. According to the American National Standards Institute (ANSI), pitch is "that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high. Pitch depends primarily on the frequency content of the sound stimulus, on the sound pressure and the waveform of the stimulus." (American National Standards Institute, 1994). This definition involves both spectral and temporal coding in pitch perception. Spectral information is thought to be extracted by place coding, meaning that the places of excitation on the basilar membrane are evaluated. The place theory of pitch goes back to Helmholtz (von Helmholtz, 1863), and is still, in an evolved form, used to explain certain pitch phenomena (e.g., Goldstein, 1973; Moore, 1993). Temporal coding means that the temporal characteristics of the signal are transducted from mechanical excitation on the basilar membrane into a temporal code by auditory nerve cells. Most

modern pitch perception models depend on temporal coding (e.g., Patterson et al., 1992; de Cheveigné, 1998).

Both place and time coding mechanisms have limitations, but complement each other. Time coding is, as mentioned above in the binaural system, limited by the phase-locking capability of auditory nerve cells transmitting the signal information. For monaural cells involved in pitch processing, this limit is thought to lie between 4-5 kHz (Moore, 1973; Sek and Moore, 1995). In this range, place coding can still be used to extract information about the frequencies contained in the signal. Place coding, however, is limited by the mechanical properties of the basilar membrane. The single frequencies of multiple signal components can not be extracted if they lie within a certain range, corresponding to the characteristic bandwidth of the place of excitation. In a spectrogram, these single frequencies would fall into the same frequency channel, making it impossible to tell them apart. This information, however, can still be extracted from the temporal periodicity of the signal. A common modeling approach for periodicity pitch is the calculation of the autocorrelation of the signal. The peaks of the autocorrelation function yield the period durations contained in the signal, which in turn can be transformed into frequencies of the components. Place coding can be implemented using more elaborate, cell-based models that simulate the excitation pattern caused by the signal on the basilar membrane (for an overview of pitch models see de Cheveigné, 2005).

The models mentioned above mimic early, peripheral stages of the auditory system. In higher stages, the frequencies extracted by these models have to be combined into a single pitch percept, as elicited for example by voiced speech or a musical note played by an instrument. These higher processing stages are probably more complex, as pitch perception is quite robust against missing information. Voiced speech signals such as vowels, for example, contain only few harmonics (so-called “formants”) spectrally distant from the fundamental frequency. Still, the formant signal energy is perceived as belonging to the rest of the speech signal it is contained in (e.g., Darwin and Sutherland, 1984). This robustness makes pitch processing a key example of an auditory grouping mechanism: Signal energy with a common fundamental frequency is grouped into a single auditory object representing the speech source. This grouping allows the signal to stand out of other, undesired signals such as background noise or, in the case of speech, allows the separation of voices due to their different fundamental frequencies.

Studies on pitch perception often work with artificial harmonic tone complexes that are generated by the addition of multiple pure-tones with the desired frequency relationships (Moore et al., 1985, 1986; Hartmann et al., 1990; Hartmann and Doty, 1996). It has been

shown that the grouping mechanism underlying the perception of harmonic complexes can be disturbed by introducing discrepancies into one or more harmonic components. Presenting a single harmonic shortly before or after the rest of the tone complex, for example, leads to the perception of that single component as a second auditory object, apart from the tone complex (Moore et al., 1985; Hartmann et al., 1990). Harmonic grouping can be disrupted by mistuning, i.e. manipulating the harmonicity of the tone complex by changing the frequency relationship of a single component to the harmonic complex as a whole. Humans are quite sensitive to mistuning. In Moore et al. (1985), human test subjects could distinguish a mistuned tone complex from its harmonic counterpart for mistuning frequencies as low as 1.1%, depending on parameters such as stimulus duration, fundamental frequency and number of components in the complex.

In mistuning detection experiments by Moore et al. (1985), subjects reported that they were able to "hear out" a mistuned component, perceiving it as a second auditory object apart from rest of the tone complex. Darwin (1981) observed a similar effect for speech-like stimuli consisting of narrow bands of noise that simulate formants. Most mistuning studies measure the mistuning detection performance of the subjects by letting them compare a mistuned and a harmonic tone complex, with the degree of mistuning being the experimental variable (e.g., Moore et al., 1985).

A simpler approach which is especially suited for combination studies with, e.g., binaural information, is the measurement of detection thresholds. The detection threshold of a target tone, which is a component of a tone complex, is measured by adaptively decreasing its level until the stimulus can no longer be distinguished from a tone complex lacking the target tone (i.e., the reference stimulus). Klinge et al. (2011) and Oh and Lutfi (2000) employed such a method. Both studies observed that the detection threshold for a target in a harmonic relationship to the rest of the tone complex (the masker) has a higher detection threshold than a mistuned target. Depending on the resolvability, this can be attributed to the ability to "hear out" the mistuned target tone as mentioned above. In Klinge et al. (2011) however, the components of the tone complexes were added up in sine phase, meaning that the reference and target intervals clearly differed in their envelope structure. This enabled the test subjects to identify the target interval by comparing the stimulus to a template that was learned in the course of the experiments. Oh and Lutfi (2000) put the study focus on informational masking and randomized the distribution of component frequencies in a wide range.

To test the suitability of the detection threshold method for the investigation of diotic harmonicity processing, a study was conducted using headphones in a soundproof booth, with tone complexes of varying resolvability, deterministic frequency settings and randomized

phase relations in every interval. The design, execution and results are reported in Chapter 3.

1.4 Harmonicity and localization

As mentioned above, harmonicity can be a strong grouping cue employed in auditory scene analysis. This means that harmonicity information is likely used in combination with binaural information to detect and track sound sources. Although combination of multiple cues is the core of auditory scene analysis, still little is known about the combined processing of harmonicity and temporal binaural cues. Several studies have investigated harmonicity or modulation and binaural processing in combination, but come to different and partially contradiction conclusions.

Krumbholz et al. (2009) studied the ability of subjects to perform modulation detection tasks or musical interval recognition tasks above threshold or in binaurally unmasked conditions (i.e. the stimuli could on be detected to to interaural differences in the masker). In binaurally unmasked conditions, the subjects were not able to determine the musical interval in the stimuli and had a worse modulation detection performance as compared to diotic conditions. Thus, Krumbholz et al. (2009) conclude that binaural processing precedes temporal modulation (i.e. periodicity) processing, with an integration stage that degrades temporal modulation information. Klinge et al. (2011) studied the influence of mistuning and localization in the free field by presenting a mistuned or harmonic target component spatially separated with a different loudspeaker than the rest of the complex. The detection thresholds for the target component decreased with mistuning and spatial separation. They observe a linear additivity of both effects, showing that the subjects profited from both harmonicity and binaural information. This is not in line with the results of Krumbholz et al. (2009), probably due to Klinge et al. (2011) measuring in the free field, where control of binaural cues available to the subjects is difficult. The linear additivity of modulation based and binaural cues, however, was also observed by Epp and Verhey (2009a), who found a linear combination of comodulation masking release (Hall et al., 1990) and binaural masking level difference (BMLD, e.g., Jeffress et al., 1956). Nitschmann and Verhey (2012) measured BMLDs with varying spectral distance between target and masker signals and conclude that their observed threshold decreases can be explained by a processing scheme where the binaural path has only limited or no access to modulation information.

Chapter 4 reports on a study that was performed to test these hypotheses with further psychophysical measurements and auditory modeling. For full control over the available

binaural information, the experiments were performed with headphones in a sound-proof booth. The same stimuli as in the previous study from Chapter 3 were used, additionally applying an interaural phase difference to the target component for a combination of periodicity and binaural information. An auditory model based on amplitude modulation processing (Dau et al., 1997) and equalization-cancellation (Durlach, 1963) is used to predict the psychophysical results. The model results as well as the human data are evaluated against the above-mentioned processing order hypotheses.

1.5 Consequences for binaural hearing aids and CASA

Hearing aid and cochlear implant users often have trouble to understand talkers in multi-talker situations. A major improvement is the use of binaurally linked hearing aids that are able to provide more precise interaural time and level differences than unlinked, independent devices in both ears. Due to their small form factor, however, hearing devices have a limited battery capacity, requiring hearing aid algorithms with a small computational cost. Thus, knowledge about the important parts of a binaural signal is invaluable for the development of binaural hearing aid algorithms and helps understanding mechanisms that enable ASA in complex environments. Such knowledge is gained in Chapter 2, where the influence of different parts of a periodic envelope on ITD sensitivity is investigated.

The goal of CASA is to develop algorithms that mimic human performance at ASA. Like ASA, these algorithms are used to enhance the signal-to-noise ratio (SNR) of a desired signal (e.g. a talker) or to track its location, in acoustically complex environments. CASA approaches that perform a joint processing of periodicity and binaural information (e.g., Ma et al., 2007; Christensen et al., 2009) could profit from the outcomes of Chapters 3 and 4, as they contribute knowledge of the possible processing order of these cues in the human auditory system, enabling the creation of algorithms closer to real auditory processing. CASA algorithms that closely model auditory processes are especially useful for the objective evaluation of hearing aid processing schemes, giving insight into the ability of hearing aid algorithms to improve or restore the ability of auditory scene analysis in hearing impaired persons.

Chapter 2

The influence of different segments of the ongoing envelope on sensitivity to interaural time delays

Abstract The auditory system is sensitive to interaural timing disparities in the fine structure and the envelope of sounds, each contributing important cues for lateralization. In this study, psychophysical measurements were conducted with customized envelope waveforms in order to investigate the isolated effect of different segments of a periodic, ongoing envelope on lateralization. One envelope cycle was composed of the four segments attack flank, hold duration, decay flank, and pause duration, which were independently varied to customize the envelope waveform. The envelope waveforms were applied to a 4-kHz sinusoidal carrier, and just noticeable envelope interaural time differences were measured in six normal hearing subjects. The results indicate that attack durations and pause durations prior to the attack are the most important stimulus characteristics for processing envelope timing disparities. The results were compared to predictions of three binaural lateralization models based on the normalized cross correlation coefficient. Two of the models included an additional stage to mimic neural adaptation prior to binaural interaction, involving either a single short time constant (5 ms) or a combination of five time constants up to 500 ms. It was shown that the model with the single short time constant accounted best for the data.

This chapter is a reformatted reprint of "The influence of different segments of the ongoing envelope on sensitivity to interaural time delays", M. Klein-Hennig, M. Dietz, V. Hohmann, and S. D. Ewert, J. Acoust. Soc. Am. 129, 3856. The original article can be found at `http://dx.doi.org/10.1121/1.3585847`.

2.1 Introduction

In contrast to visual perception, which enables us to observe the world in front of us, hearing allows us to capture sound events from all possible directions. Accurate localization of a sound source in the horizontal plane is strongly facilitated by two-ear interaction (binaural hearing) in humans. In order to determine the position of a sound source, the interaural disparities between the signals that arrive at each ear are evaluated and form binaural cues for localizing sounds. The binaural cues used for analyzing the azimuthal position of a sound source are the interaural time difference (ITD) and the interaural level difference (ILD).

Rayleigh (1907) hypothesized that the ITD is used to localize sounds for frequencies up to about 1500 Hz, and the ILD dominates localization at frequencies above about 1500 Hz. However, the auditory system can exploit timing disparities in the envelope (envelope ITDs) of the signal in the high-frequency region and many psychoacoustic studies have used envelope waveforms that provide binaural cues of differing salience. Henning (1974), for example, studied the influence of sinusoidal amplitude modulation (SAM) on lateralization, and Hafter and Buell (1990) investigated the lateralization of clicks with a Gaussian envelope and found that just-noticeable differences (JNDs) in envelope ITDs were lower than for SAM tones. Bernstein and Trahiotis (2002) measured the JND of transposed tones as introduced by van de Par and Kohlrausch (1997), which have, unlike SAM tones, a segment of silence ("pause") in every cycle and have steeper flanks than SAM tones at a similar rate. The same applies to the filtered impulse trains as used by Hafter and Dye (1983). The use of both the transposed tones and the filtered impulse trains resulted in generally lower JNDs than observed for SAM tones. Recently, Bernstein and Trahiotis (2009) determined JNDs for so-called "raised sine" stimuli. In these stimuli, the pause duration and flank steepness could be controlled by the exponent of a sinusoidal modulator. An increased exponent and therefore longer pause durations and steeper flanks in the stimuli led to lower JNDs. The same study demonstrated that a reduced modulation depth caused higher JNDs.

While all of the above-mentioned studies revealed the influence of analytical envelope parameters such as sine exponent, modulation frequency, or modulation depth, the observed changes in perceptual sensitivity were often discussed in terms of envelope properties such as the steepness of the flanks or the existence of a pause. Unfortunately, the existing analytical envelope manipulations presented in the literature usually lead to a co-variation of these properties. For instance, transposed tones and raised sine tones have both steeper flanks as well as a more pronounced modulation depth when compared to SAM tones. Likewise,

changing the analytical parameters similarly affected the attack and the decay flanks. Therefore the isolated influence of different envelope segments or features on binaural sensitivity is still unknown.

Given that temporal envelope cues are based on rapid monaural level fluctuations, it is plausible that monaural adaptation mechanisms, which alter the internal representation of the signal's envelope, play a role in the processing of envelope ITD if they take place prior to binaural processing. This was confirmed by Hafter et al. (1988), who found a monaural adaptation mechanism "at a location peripheral to binaural interaction" (p. 663) that affects binaural thresholds. Several parts of the auditory system exhibit adaptive behavior, with some adaptive behavior likely to occur prior to binaural processing. For example, an auditory nerve fiber can have a maximal discharge rate at the onset of a stimulus response and a gradual decrease with the ongoing stimulus: Right after the onset of a stimulus, sensitivity is reduced. This behavior was used by Smith (1979) to explain forward masking. The gain and loss of sensitivity with onset and ongoing excitation are parameters that differ from cell to cell and multiple cells with different types of firing patterns have been found in the auditory system (e.g., Young, 1988). By investigating post-stimulus time histograms of SAM and transposed tones, physiological studies (e.g., Griffin et al., 2005; Dreyer and Delgutte, 2006) have found that neural responses for transposed tones are more synchronized to the stimulus envelope than for SAM tones. Bernstein and Trahiotis (2009) have also related the lower psychoacoustic JNDs achieved with transposed tones to the higher neural synchronization. Neuronal adaptation is well established in monaural auditory models (Dau et al., 1996a, 1997; Meddis and O'Mard, 2005). Most binaural models, however, do not include any form of adaptation prior to binaural interaction (Jeffress, 1948; Sayers and Cherry, 1957; Durlach, 1963; Colburn, 1977; Lindemann, 1986). An exception is the binaural processing model of Breebaart et al. (2001a,b,c) . It combines the monaural model of Dau et al. (1996a), including adaptation, with subsequent binaural processing. However, this model is not able to predict lateralization measurements in its present form. Other models which can predict lateralization, such as the normalized 4 th-moment model (Dye et al., 1994), the normalized cross correlation coefficient model (Bernstein and Trahiotis, 2002), the position-variable model (Stern and Shear, 1996) and the two-channel interaural phase difference (IPD) model (Dietz et al., 2009) do not include neuronal adaptation.

The aim of the current study was to clarify the role of different envelope features occurring in earlier studies that used SAM tones, transposed tones, or "raised sine" stimuli on binaural sensitivity and to study the influence of adaptation. A custom periodic envelope was created and applied as a modulator to a 4-kHz pure tone carrier. Each cycle of the

envelope waveform consisted of four segments: Attack, hold, decay, and pause (see Figure 2.1). The attack segment (or flank) had a duration defined by the time taken for the initial increase of the envelope from minimum to peak amplitude. The hold duration specified the time of constant peak amplitude. The decay duration or flank was the time of decrease from peak to minimum amplitude after the hold duration. Finally, the pause duration specified the time of constant minimum amplitude at the end of the cycle. The isolated influence of these four specific envelope segments on binaural envelope ITD sensitivity was examined: For stimuli with envelope ITDs in these specific segments of an ongoing, periodic envelope, the just-noticeable envelope interaural time difference (referred to as JND in the following) was determined as a function of the duration of each segment. Experiments 1-4 addressed the isolated influence of the four envelope segments, while Experiments 5-9 studied the effect of secondary envelope parameters such as level, amplitude offset, and modulation frequency. A summary of the experiments and their parameters can be found in Table 2.1.

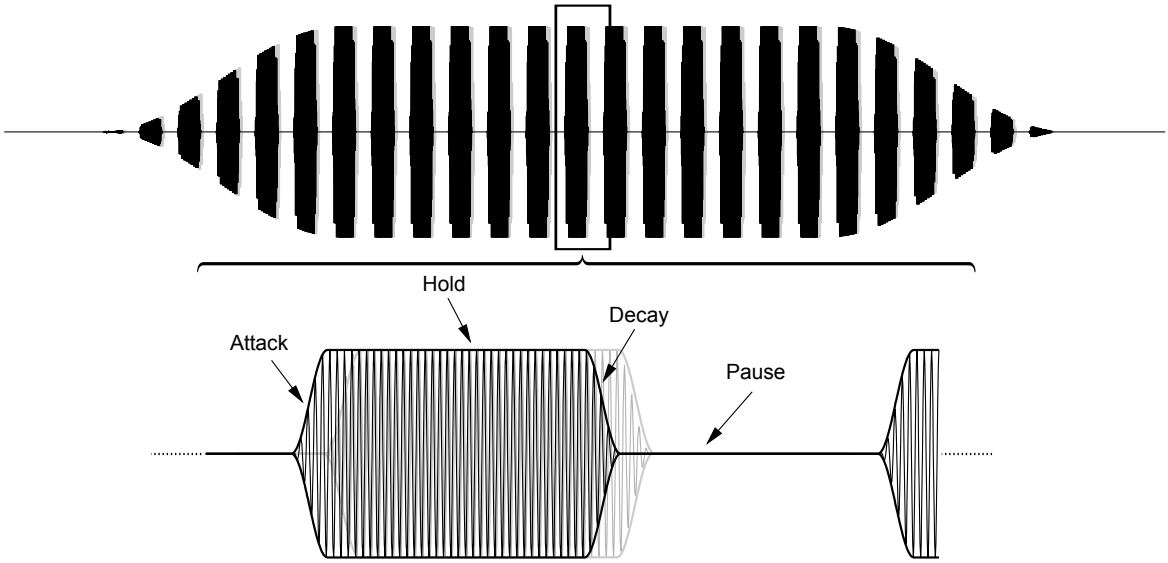

Figure 2.1: Illustration of a typical stimulus used in the experiments. The upper trace shows the whole amplitude-modulated stimulus (4-kHz sine carrier) with the on- and off-gating and the periodic variation of the ongoing envelope. The lower trace shows a close-up of a single envelope segment. The envelope parameters attack, hold, decay, and pause are indicated by pointers. The gray waveform in the background indicates the stimulus in the other ear if an envelope ITD was applied.

2.2 Methods

2.2.1 Subjects

Six normal-hearing listeners (three female, three male) aged 24-30 years participated in the experiments. Before data acquisition, all subjects took part in 5 h of training with stimuli similar to those used in the final experiment. Three of the subjects received compensation on an hourly basis for taking part in the experiment. The other subjects were lab members, including two of the authors. The experiments were approved by the ethics committee of the Universität Oldenburg.

2.2.2 Apparatus and stimuli

Subjects were seated in a double-walled, sound-attenuating booth. The stimuli were generated at runtime on a personal computer using MATLAB. The AFC software package developed at Universität Oldenburg was used for presentation and experiment control[1] . The digitally generated stimuli had a sampling frequency of $f_s = 48$ kHz and were converted to analog signals by an external RME ADI-8 PRO D/A converter connected to a 24-bit RME DIGI96/8 PAD sound card. A Tucker Davis (Alachua, FL) HB7 headphone buffer was used to drive Sennheiser (Wedemark-Wennebostel, Germany) HD 580 headphones. The subjects used a computer keyboard or mouse to indicate their response and received visual feedback on a computer monitor.

The stimuli used in the JND measurements were periodically amplitude-modulated pure tones with a carrier frequency of 4 kHz. This frequency was selected because at lower frequencies, fine-structure cues become increasingly salient and the decreasing bandwidth of the auditory filters limits the applicability of short attack and decay durations, but at higher frequencies, sensitivity to interaural envelope cues decreases (Bernstein and Trahiotis, 2002). The envelope had a customized waveform for each experimental condition. A single envelope period consisted of four parts (see Figure 2.1): Attack, hold, decay, and pause. For the attack and decay flanks, squared-sine functions were used. The hold duration specified the duration of constant modulator amplitude of 1. During the pause, the modulator amplitude was set to zero, except for Experiment 7 where the amplitude offset was the parameter of interest. Setting hold and pause durations to 0 ms resulted in a waveform mathematically identical to SAM. To limit spectral broadening of the stimuli, the attack and

[1] AFC: A psychophysical-measurement package for MathWorks MATLAB, developed by Stephan D. Ewert, Universität Oldenburg and Centre for Applied Hearing Research, e-mail: stephan.ewert@uni-oldenburg.de

Experiment	Attack (ms)	Hold (ms)	Decay (ms)	Pause (ms)	Mod. rate (Hz)	Energy portion within filter	Level (dB SPL)	JND (µs)	Relative std. dev. (%)	NCC (µs)	NCC1A (µs)	NCC5A (µs)	RMSE NCC (µs)	RMSE NCC1A (µs)	RMSE NCC5A (µs)
1: Attack duration	1.25	8.75	1.25	8.75	50	0.99	61	129	45	208	167	208	1.62	1.31	1.56
	2.5	8.75	1.25	8.75	47	0.99	61	136	45	250	208	250			
	5	8.75	1.25	8.75	42	1.00	61	211	59	333	250	313			
	10	8.75	1.25	8.75	35	1.00	61	341	47	479	375	417			
2: Hold duration	1.25	0	1.25	13.125	64	0.81	52	101	33	83	83	125	1.43	1.35	1.56
	1.25	4.375	1.25	13.125	50	0.98	59	104	45	125	125	167			
	1.25	13.125	1.25	13.125	35	0.99	61	107	39	188	167	188			
3: Decay duration	1.25	8.75	1.25	8.75	50	0.99	61	(1059)	N/A	250	313	479	N/A	N/A	N/A
	1.25	8.75	2.5	8.75	47	0.99	61	(1958)	N/A	292	333	521			
	1.25	8.75	5	8.75	42	1.00	61	(1666)	N/A	354	396	625			
	1.25	8.75	10	8.75	35	1.00	61	(2169)	N/A	500	521	833			
4: Pause duration	1.25	17.5	1.25	0	50	0.99	64	479	59	479	750	417	1.34	1.37	1.43
	1.25	13.125	1.25	4.375	50	0.99	63	150	60	188	188	167			
	1.25	8.75	1.25	8.75	50	0.99	61	98	55	167	146	167			
	1.25	4.375	1.25	13.125	50	0.98	59	104	45	125	125	167			
	1.25	0	1.25	17.5	50	0.81	51	105	34	83	83	146			
5: Level	10	0	10	0	50	1.00	36	789	65	875	479	354	1.13	1.34	1.61
	10	0	10	0	50	1.00	48	438	47	521	354	313			
	10	0	10	0	50	1.00	60	282	43	292	292	292			
	10	0	10	0	50	1.00	66	200	110	229	250	292			
6: Modulation freq.	10	0	10	0	50	1.00	60	282	43	292	313	250	1.41	1.32	1.42
	5	0	5	0	100	1.00	60	181	42	188	208	167			
	1.25	13.125	1.25	13.125	35	0.99	61	107	39	188	167	188			
	1.25	8.75	1.25	8.75	50	0.99	61	98	55	167	146	167			
	1.25	3.75	1.25	3.75	100	0.98	61	122	40	125	125	125			
7: dc offset	10	0	10	0	50	1.00	60	282	43	292	313	250	1.58	1.36	1.97
	10	0	10	0	50	1.00	66	1676	107	875	1104	646			
	1.25	8.75	1.25	8.75	50	0.99	61	98	55	167	146	167			
	1.25	8.75	1.25	8.75	50	1.00	66	251	51	500	667	417			
8: Temporal asymmetry	1.25	0	18.75	10	33	0.99	58	114	34	208	146	188	1.74	1.40	1.45
	18.75	0	1.25	10	33	0.99	58	377	54	229	250	313			
9: Transposed tone	10	0	10	0	50	1.00	60	282	43	292	313	250	1.08	1.11	1.28
	N/A	N/A	N/A	N/A	50	1.00	58	150	28	167	167	208			
Attack duration (full waveform shift)	1.25	8.75	1.25	8.75	50	0.99	61	98	55	167	146	167	1.49	1.36	1.51
	2.5	8.75	1.25	8.75	47	0.99	61	106	56	188	167	188			
	5	8.75	1.25	8.75	42	1.00	61	187	70	208	188	229			
	10	8.75	1.25	8.75	35	1.00	61	227	91	250	250	271			

Table 2.1: Parameters and properties of the stimuli used in the experiments along with the psychoacoustic results and model predictions.

decay durations were always greater than or equal to 1.25 ms, resulting in at least 98% of the total energy (see Table 2.1) being within the equivalent rectangular bandwidth (ERB) of the auditory filter at 4 kHz (Moore and Glasberg, 1996). Two conditions with a broader spectrum, where 81% of the total energy was within the ERB, are marked as exceptions where they appear and are analyzed in the discussion. Details on the envelope parameter configurations are given in the respective experiment sections and in Table 2.1.

In several experiments, it was necessary to test the shortest possible attack and decay durations of 1.25 ms each. In that case, the total cycle duration of usually 20 ms resulted in a nearly square shape of the amplitude modulation. It is therefore referred to as “pseudo-

square-wave (PSW) modulation"[2]. The inverse of the period duration, which was the sum of the attack, hold, decay, and pause durations resulted in a modulation frequency f_m ranging from 33 to 100 Hz across conditions.

The stimuli had a total duration of 500 ms, including stimulus on- and offset gating ramps of 125 ms each. The envelope ITD was applied to the complete 500-ms envelope of the right-ear stimulus. Exceptions were the attack and decay duration experiments, where the envelope ITD was applied to the respective flanks only. The left-ear stimulus was never modified. The gating ramps were synchronously applied to both left and right stimulus after the application of the ITD. The relatively long gating ramps were used in order to minimize the influence of the gating on the perception of envelope ITDs in the ongoing periodic envelope[3].

In order to prevent subjects from exploiting information in the low-frequency domain potentially produced by nonlinear distortion, a low-pass noise was added to the stimulus. This low-pass noise had a root-mean-square (rms) level of 45 dB sound pressure level (SPL) and a flat spectrum up to 200 Hz. Above 200 Hz its spectral density decreased with a slope of -3 dB/octave, and was additionally filtered with a fifth-order low-pass filter with a cutoff frequency of 1000 Hz. The low-pass noise was uncorrelated between both ears. The noise duration was 600 ms, gated with 50-ms raised cosine ramps. The stimulus was temporally centered in the noise.

As the flank steepness of the stimuli was an experimental parameter, it was essential to control the maximum amplitude of the stimuli. Any rms equalization across the conditions with different envelope waveforms would have led to different maximum amplitudes and thus to an unintentional modification of flank steepness. Thus, if not stated otherwise, all conditions had the same maximum of 1, which corresponded to a rms level of 60 dB SPL for a sinusoidal amplitude modulation (c.f. Griffin et al., 2005). The resulting rms levels in dB SPL are given in Table 2.1.

[2] A true square-wave stimulus with infinitely steep slopes passed through an ERB filter at 4 kHz would have similar limitations of the rise/fall times.

[3] In preliminary experiments with longer gating ramp durations, 125 ms turned out to be sufficiently long to exclude effects of the on- and off-gating ramps on the data. In these experiments, the JND of two conditions was obtained for different gating durations. In the first condition, the waveform started with the pause segment of the envelope, whereas in the second condition the waveform started with the attack flank of the envelope. The gating duration for which both conditions yielded the same JND was used for the final experiments.

2.2.3 Procedure

The JNDs were determined using an adaptive 2-interval, 2-alternative forced-choice procedure. A 1-up, 3-down tracking rule was used estimating the 79.4%-correct point on the psychometric function (Levitt, 1971). A reference stimulus without envelope ITD and the test stimulus with adaptively varied envelope ITD were presented in random order. The test subject had to indicate whether the second sound was toward the left or right of the first sound. The envelope ITD started at 2 ms and was initially varied by a factor of 2.0. After the second reversal, the factor was reduced to 1.4 and after the fourth reversal it was further reduced to 1.1. The adaptive run was terminated after a total of ten reversals.

The envelope waveform delay was applied in the time domain by rounding to and shifting by an integer number of samples. Based on the sampling frequency, the smallest possible envelope ITD change of a single sample was $\mathrm{ITD}_{\mathrm{min}} = \frac{1}{f_\mathrm{s}} = 20.83\mu\mathrm{s}$. This minimum ITD change was small enough to reliably measure the JND in all conditions. The JND was defined as the geometric mean of the envelope ITD values at the last six reversals. For each test subject, five runs were measured. The individual mean JND was calculated on the basis of the last four valid runs by geometrical averaging. A run was considered valid when the geometric standard deviation of the last six reversals was less than 20% of the JND and when the JND did not exceed 8 ms. In the very few cases in which only three of the last four runs were valid, an additional run was conducted. If there were still less than four valid runs, the condition was marked as failed and will be discussed separately. The mean data shown here are geometric means and standard deviations across the individual JNDs of all test subjects.

2.2.4 Models

Three models were used. All models had identical peripheral preprocessing that was effectively identical to the preprocessing employed by Bernstein and Trahiotis (2002), with minor difference in the rectification and compression stages. A 4-kHz fourth-order gammatone filter (Patterson et al., 1987; Hohmann, 2002) was used as the auditory filter. Cochlear compression and part of the inner hair cell transduction process were modeled by a half-wave, square-law rectification followed by a power-law compression with an exponent of $n = 0.23$, leading to an effective exponent of $n = 0.46$. The transduction process of the inner hair cells was further modeled by a 425-Hz fourth-order low-pass filter (Weiss and Rose, 1988). A 150-Hz first-order low-pass filter was used to account for modulation rate limitation (Kohlrausch et al., 2000; Ewert and Dau, 2000; Bernstein and Trahiotis, 2002).

The first model was a re-implementation of the normalized cross-correlation coefficient (NCC) model by Bernstein and Trahiotis (2002). After the monaural preprocessing, the normalized cross-correlation coefficient ρ was determined according to Equation 2.1 in Bernstein and Trahiotis (1996):

$$\rho = \frac{\sum x(t)y(t)}{\sqrt{\sum x(t)^2}\sqrt{\sum y(t)^2}} \tag{2.1}$$

ρ was calculated using a single envelope cycle from the steady-state part of the left, $x(t)$, and right, $y(t)$, preprocessed stimuli for envelope ITDs in the range of 0-3000 μs in $\mathrm{ITD}_{\mathrm{min}}$ steps. Due to the normalization, ρ is independent of the stimulus level. The calculation of ρ led to a function assigning every envelope ITD in this range a value of ρ. With increasing envelope ITD, ρ decreased monotonically. The JND was therefore found by using a threshold value for ρ as criterion. A criterion value of $\rho_{\mathrm{crit}} = 0.9993$ was used for all experimental conditions, and the envelope ITD at which the correlation coefficient dropped below this threshold was used as the model prediction. Figure 2.2 illustrates the method for the 50-Hz PSW stimulus of the modulation-frequency experiment, in the ITD range from 0 to 700 μs. The criterion value $\rho_{\mathrm{crit}} = 0.9993$ is plotted as a dashed line. The intersection of ρ and ρ_{crit} determines the simulated JND (indicated by the downward-pointing arrow). ITDs with $\rho > \rho_{\mathrm{crit}}$ are not detected by the model. The criterion value ρ_{crit} was determined by minimizing the square root of the mean of the squared deviations of the model predictions for all experiments from the mean data of all experiments [root-mean-square error (RMSE)]. The supplemental attack and decay experiments (see Section 2.4.1) were not included for the determination of the threshold value. Before calculating the RMSE, the base-ten logarithm of the model predictions and mean data was taken, to fit the model on a log-scale JND axis comparable to the way the data are shown in the plots and according to the geometric mean and log-symmetric errors. The resolution for variation of ρ during the fitting process was 0.00001. In addition to this first model, two modified models were generated by combination of the first model with a monaural adaptation stage. In the second model, adaptation loops (Dau et al., 1996a) were included prior to calculating ρ in order to model monaural adaptation, as in the binaural processing model by Breebaart et al. (2001a). The output signals of the preprocessing stage were passed through an array of five adaptation loops with time constants of 5, 50, 129, 253, and 500 ms[4]. Then, the normalized cross correlation co-

[4]The function of the adaptation loops requires a restriction of the input range to a minimum value in order to avoid division by zero. In the published models (Dau et al., 1996a; Breebaart et al., 2001a), this minimum value also serves to define the absolute detection threshold. As the current model is not used for pure-tone detection threshold predictions, the choice of the minimum value became a parameter to adjust the

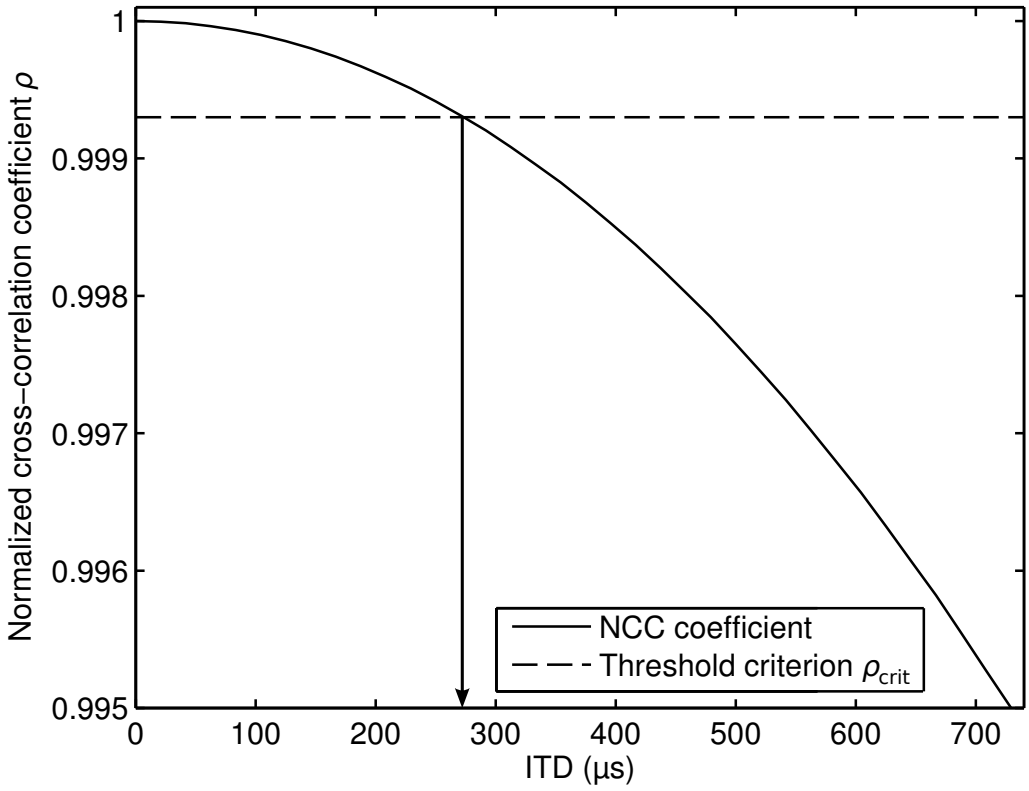

Figure 2.2: The normalized cross coefficient ρ for the 50-Hz PSW stimulus of the modulation-frequency experiment, calculated for stimuli in the ITD range from 0 to 700 μs. The criterion value $\rho_{\text{crit}} = 0.9993$ that simulates the JND is plotted as a dashed line. The assumption is that ITDs with $\rho > \rho_{\text{crit}}$ cannot be discriminated from a zero ITD by the model. The resulting simulated JND is indicated by the downward arrow.

efficient was calculated as described earlier. The detection threshold criterion was derived in the same manner as before, resulting in a criterion value of $\rho_{\text{crit}} = 0.99735$. In the third model, only the first adaptation loop with the smallest time constant of $\tau_1 = 5$ms was used to predict the data. For this model, a criterion value of $\rho_{\text{crit}} = 0.99942$ was used. The two modified models are referred to as NCC5A (all five adaptation loops) and NCC1A (only the first adaptation loop), respectively.

The general behavior of the adaptation loops is shown in Figure 2.3 for the 50-Hz PSW stimulus of the modulation-frequency experiment. The internal envelopes without adapta-

characteristic of the adaptation. Two values were tested, the "original" value of 10^{-5} and $(10^{-5})^{0.46}$, to adjust for the compression exponent as used in the preprocessing. As in the literature, the input signals were scaled in such a way that a peak value of 1 represents a peak-equivalent value of 100 dB sound pressure level. The NCC5A model predictions using a value of 10^{-5} are shown in the current study because they showed smaller deviations from the data than the model results using $(10^{-5})^{0.46}$.

tion (upper trace), at the output of the first adaptation loop (middle trace), and at the output of all five adaptation loops (lower trace) are shown. In each modulation cycle a pronounced attack followed by an adaptively decaying region is obvious for the single and the five adaptation loops. In case of the five adaptation loops (lower trace) an additional "overshoot" at the initial onset of the whole stimulus can be observed[5].

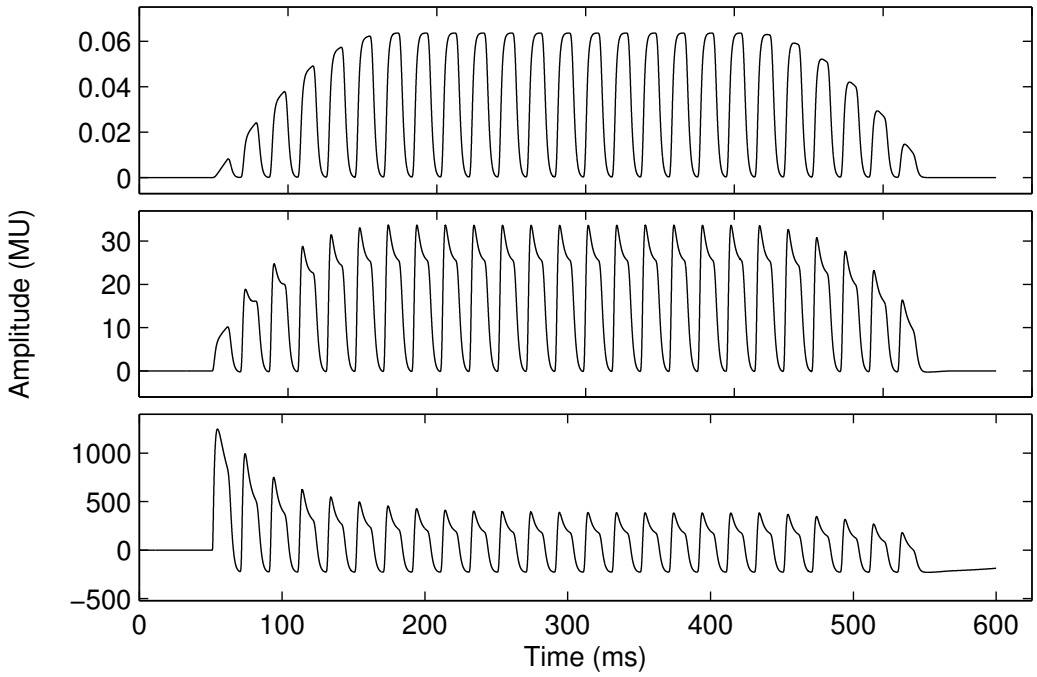

Figure 2.3: The internal signal representation in the model at the output of the preprocessing stage is shown in the upper trace. Middle trace: Internal representation with a single adaptation loop ($\tau = 5$ms) after preprocessing (as employed for the NCC1A model). Bottom trace: Internal representation with five adaptation loops after preprocessing as used in the NCC5A model. The input stimulus was the 50-Hz PSW condition of the modulation frequency experiment (Experiment 6).

[5]The pronounced overshoot at the overall signal onset as generated by the five adaptation loops does not reflect the lack of onset sensitivity achieved with the 125-ms gating employed in the psychoacoustic experiments. To disregard possible effects of the overshoot in the model predictions, a single modulation cycle from the center of the stimulus after adaptation processing was used as input to the correlation model.

2.3 Experimental results

Here, the results along with the rationale and the parameter configuration of each experiment are given. For convenience, the durations of the envelope segments are rounded to one decimal place in the text and figures. The exact values are given in Table 2.1, along with the values of the JNDs and the RMSE of the model predictions.

2.3.1 Experiment 1: Attack duration

Rationale

The attack flank of an envelope period marks its beginning and, if a time of silence preceded the attack flank, it is the segment of a signal that is processed by auditory neurons with maximal sensitivity after their quiescent period. A high sensitivity to envelope ITD in the attack segment would indicate a strong contribution of the attack to the overall lateralization of a sound source. In order to investigate how envelope ITDs in the attack flank of an envelope period contribute to lateralization, the JNDs of four different attack durations were measured.

Conditions

Four conditions with attack durations of 1.3, 2.5, 5, and 10 ms were measured. The decay duration was 1.3 ms, the hold and pause durations were 8.8 ms (the small insets above the x axis in Figure 2.4 and the following figures illustrate a 30-ms portion of the stimulus envelopes). To prevent a possible influence of the decay flank, the envelope ITD was applied to the attack flank only by extending the pause duration and shortening the hold duration in the right-ear envelope. This resulted in an ILD (particularly for larger envelope ITDs), because the right-ear signal had an overall lower rms level due to a shorter hold duration. In the range of the expected JNDs, however, the resulting ILD was smaller than 0.2 dB. Given that this is far below the ILD detection threshold of 1-2 dB (Grantham, 1984), it was expected to have little or no influence on the results. The method of applying the ITD shift only to the attack segment is different than the commonly employed procedure of a constant ITD. Therefore, a supplementary experiment was also performed where the ITD was applied to all segments (i.e., a complete waveform shift).

Increasing the attack duration resulted in longer modulation periods and consequently in modulation frequencies of 50, 47, 42, and 35 Hz. The modulation periods were always

identical in both ears (in Experiment 6, a variation of the modulation frequencies in this range was found to have no significant effect on the JND).

Results

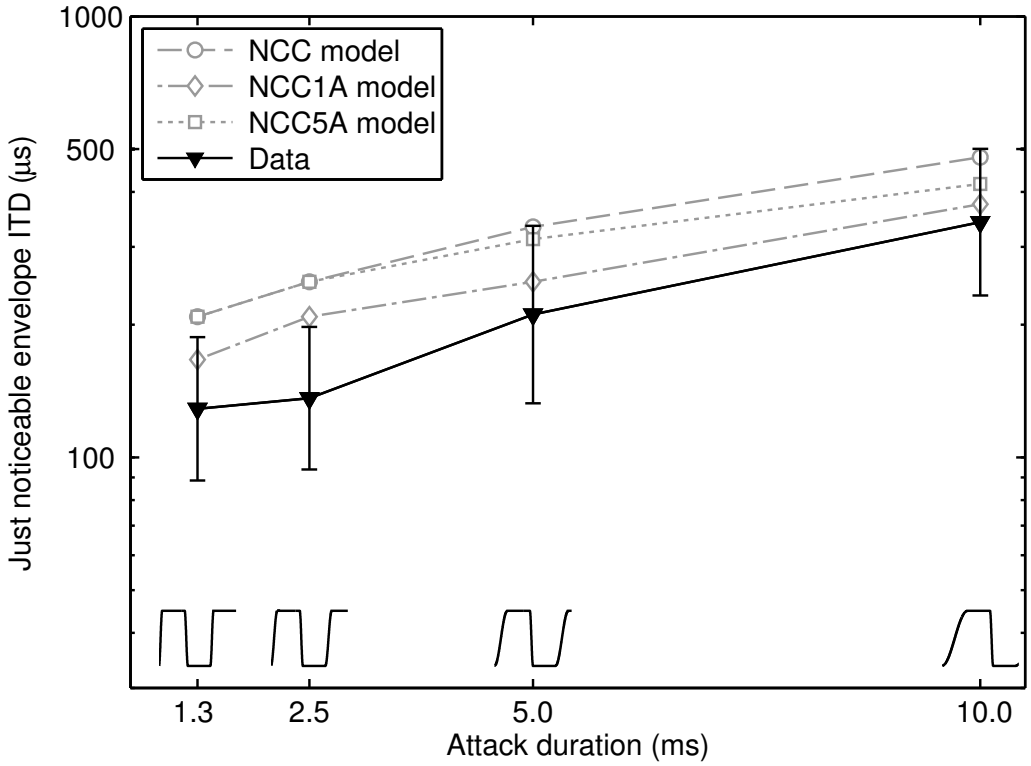

Figure 2.4: Results for Experiment 1: Just noticeable envelope ITDs (geometric mean, error bars: geometric standard deviation across subjects) in microseconds for envelopes with attack durations of 1.3, 2.5, 5, and 10 ms, shown as black triangles. The predictions of the normalized cross-correlation model (NCC) are shown as gray circles with dashed lines. The predictions of the normalized cross-correlation model with monaural adaptation using five adaptation loops (NCC5A) and using only the first adaptation loop (NCC1A) are shown as gray squares with dotted lines and gray diamonds with dash-dotted lines, respectively. The small insets above the x axis show 30-ms portions of the stimulus envelopes for the respective conditions.

The results are shown in Figure 2.4 as filled black triangles (geometric mean, error bars: geometric standard deviation across subjects). The JND monotonically increased with increasing attack duration, rising by a factor of 3 from 129 μs for an attack duration of 1.3 ms to 341 μs for an attack duration of 10 ms. A repeated-measures analysis of variance

(ANOVA) showed a highly significant main effect of the attack duration on JND: $F(3,15) = 22.0$, $p < 0.001$. Post hoc pairwise comparisons (Bonferroni) indicated that the 5- and 10-ms conditions were significantly different ($p < 0.05$) from the other conditions with 1.3- and 2.5-ms attack duration, which were not significantly different from one another.

The stimulus configurations and results of the supplementary experiment with a complete waveform shift are shown in the last four rows of Table 2.1. The supplementary experiment led to lower JNDs but followed the same trend as those obtained when only the attack flank was shifted.

The predictions of the NCC model are shown as gray circles with dashed lines, and the predictions of the NCC1A and NCC5A models are indicated by gray diamonds with dashed-dotted lines and gray squares with dotted lines, respectively. All model predictions follow the general trend observed in the data, although they all show higher JND values. The predictions exhibited slightly less variation as a function of attack duration, particularly between 2.5- and 5-ms attack duration. The NCC5A model predictions differed by less than 50 μs from the NCC results. The NCC1A model predictions had overall smaller JNDs compared to the other two models, leading to a better agreement with the data. The values of the RMSE demonstrated that the NCC1A model performed best in accounting for these data (see Table 2.1).

2.3.2 Experiment 2: Hold duration

Rationale

The hold duration is the portion of the stimuli with the maximal modulator amplitude and constitutes the steady state portion of the signal. Smith (1979) has shown that during the steady state portion of the signal the neuronal sensitivity to an ongoing stimulus declines, resulting from a decrease in firing rate of auditory-nerve fibers with increasing stimulus duration. This could hamper the interaural difference detection. A parametrical variation of the hold duration was used here to assess its effect on the JND.

Conditions

The experiment used three PSW stimuli with three different hold durations of 0, 4.4, and 13.1 ms. All other envelope segments were kept constant: The attack and decay durations were 1.3 ms, and the pause duration was 13.1 ms. Varying only the hold duration led to longer modulation periods and thus to stimuli with modulation frequencies of 64, 50, and 35 Hz. The stimulus with a zero hold duration is an exception from the rule that 98% of

the energy falls within the ERB around the 4-kHz carrier frequency. The complete absence of a constant nonzero envelope segment results in spectral broadening and only 80% of the energy fell within the ERB around 4 kHz.

Results

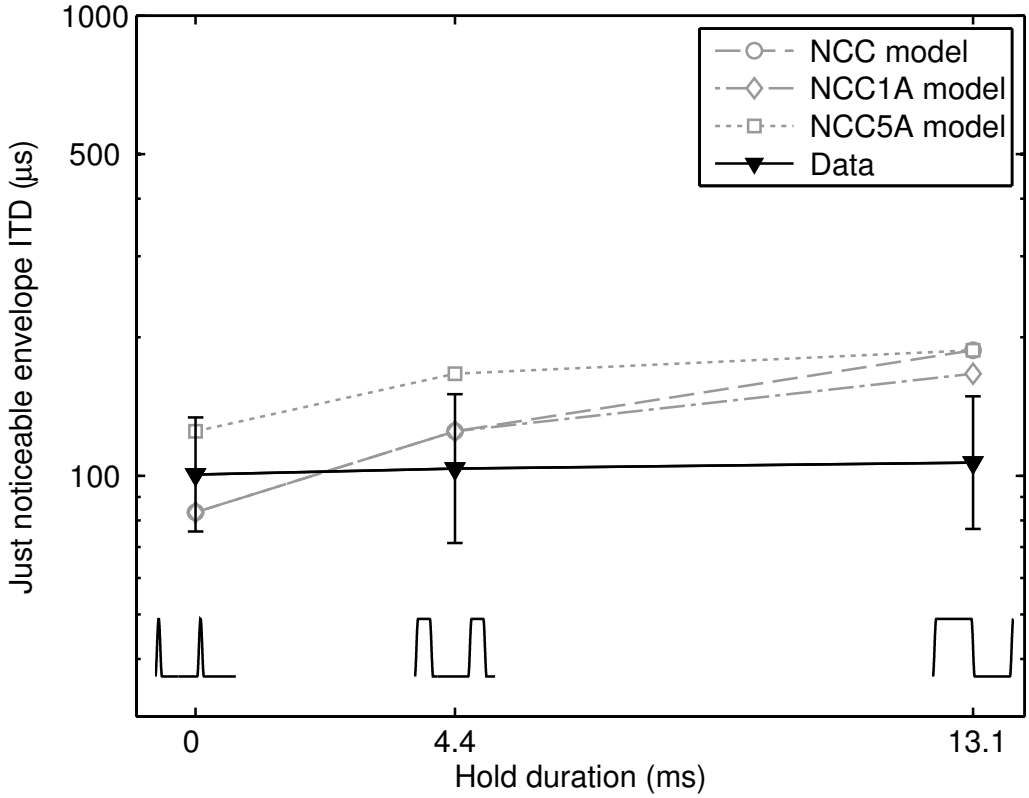

Figure 2.5: Results of Experiment 2: Just noticeable envelope ITDs in microseconds for envelopes with hold durations of 0.0, 4.4, and 13.1 ms, shown as black triangles. The attack and decay durations were fixed at 1.3 ms, the pause duration was fixed at 13.1 ms. Conventions are as in Figure 2.4.

Figure 2.5 shows the JND results of the hold duration experiment. The plotting conventions are the same as in Figure 2.4. It can be seen that the JND did not depend on the hold duration of the envelope. The JNDs were in the range between 101 and 107 μs. A repeated-measures ANOVA showed no significant effect of the hold duration on the JND: $F(2,10) = 0.13$, $p = 0.88$.

The gray symbols indicate the model predictions. All three models predicted a reasonable agreement with the data for 0-ms hold duration but produced an increase in JND as soon

as a nonzero hold duration was applied, so overestimating the data. The NCC (circles) and NCC1A (diamonds) model predictions deviated from the results for a hold duration of 0 and 4.4 ms by the minimal ITD step size of 20.83 μs. The NCC1A model showed slightly less increase for the longest hold duration. The NCC5A (squares) model also had a weaker dependence on hold duration than the other two models. All NCC5A predictions were above the measured JNDs. The NCC1A model showed the smallest prediction error.

2.3.3 Experiment 3: Decay duration

Rationale

The decay segment of a signal marks its end and likely a region of reduced neuronal sensitivity resulting from preceding activity. A reduced sensitivity could hamper the processing of interaural time differences. In order to investigate this potential effect, four conditions with different decay durations were examined.

Conditions

The decay durations were 1.3, 2.5, 5.0, and 10.0 ms, leading to modulation frequencies of 50, 47, 42, and 35 Hz at both ears. The attack, pause, and hold durations of the left-ear signal were fixed to 1.3, 8.8, and 8.8 ms, respectively. To prevent any influence of the attack flank, it was synchronously applied to both channels. As in Experiment 1, this shift of only one flank was achieved by extending the pause duration and shortening the hold duration in the right-ear envelope.

Results

Only two subjects were able to complete the experiment successfully. The other subjects were usually able to lateralize the stimuli before reaching the 8-ms maximum ITD (in 82% of the runs), but their intra-run standard deviation was quite often greater than 20% of the JND. Even though up to six measurements (instead of five) were performed for this reason, there were usually less than four valid runs. It was therefore not possible to show quantitative results with error bars for these conditions. The mean results of the successfully completed runs were 1059, 1958, 1666, and 2169 μs for the increasing decay durations of 1.3, 2.5, 5.0, and 10.0 ms, respectively.

All models predicted an increase in JND with increasing decay duration, but with values considerable below 2 ms, in contrast to the experimental data. The NCC model predictions

were in the range between 250 and 500 μs for the 1.3- and 10-ms decay flank condition, whereas the NCC1A model predicted JNDs in the range between 313 and 521 μs. The NCC5A model predictions increased from 479 to 833 μs, showing reasonable agreement with the data for the successfully completed runs. Given the overall poor predictions and the lack of valid data, a goodness of fit in terms of RMSE was not derived for this experiment.

2.3.4 Experiment 4: Pause duration

Rationale

Given that the sensitivity of auditory neurons recovers during the pause before a signal onset, it could be expected that the duration of the pause has an influence on the JND. In contrast, if the neuron sensitivity is lowered by preceding stimuli and does not recover during brief pauses, the signal onset might not be well represented in the neural code and interaural differences could become more difficult to detect. To assess the effect of the pause duration, five conditions with different pause durations were used.

Conditions

Pause durations of 0.0, 4.4, 8.8, 13.1, and 17.5 ms were tested. As before, PSW tones with attack and decay durations of 1.3 ms were used. The envelope ITD was applied to all envelope segments. Based on the results of Experiment 2, the influence of the hold duration can be neglected, and so hold duration was modified depending on the pause duration to maintain a fixed cycle duration of 20 ms. Therefore, all conditions had a fixed modulation frequency of 50 Hz. The stimulus with the longest pause duration of 17.5 ms resulted in having zero hold duration and consequently increased spectral broadening. In this condition, only 81% of the energy fell within the ERB around the 4-kHz carrier.

Results

The results are shown in Figure 2.6. The condition with zero pause duration led to the highest JND of about 480 μs. In this condition, the decay flank of one period was immediately followed by the attack flank of the next period, without a pause in between. Increasing the pause duration to 4.4 ms led to a dramatic drop in the JND by a factor of 3 to about 150 μs. For pause durations longer than 4.4 ms, the JND decreased only slightly with increasing pause duration. A repeated-measures analysis of variance revealed a highly significant main effect of the pause duration on the JND: $F(4,20) = 32.53$, $p < 0.001$. Post hoc pairwise

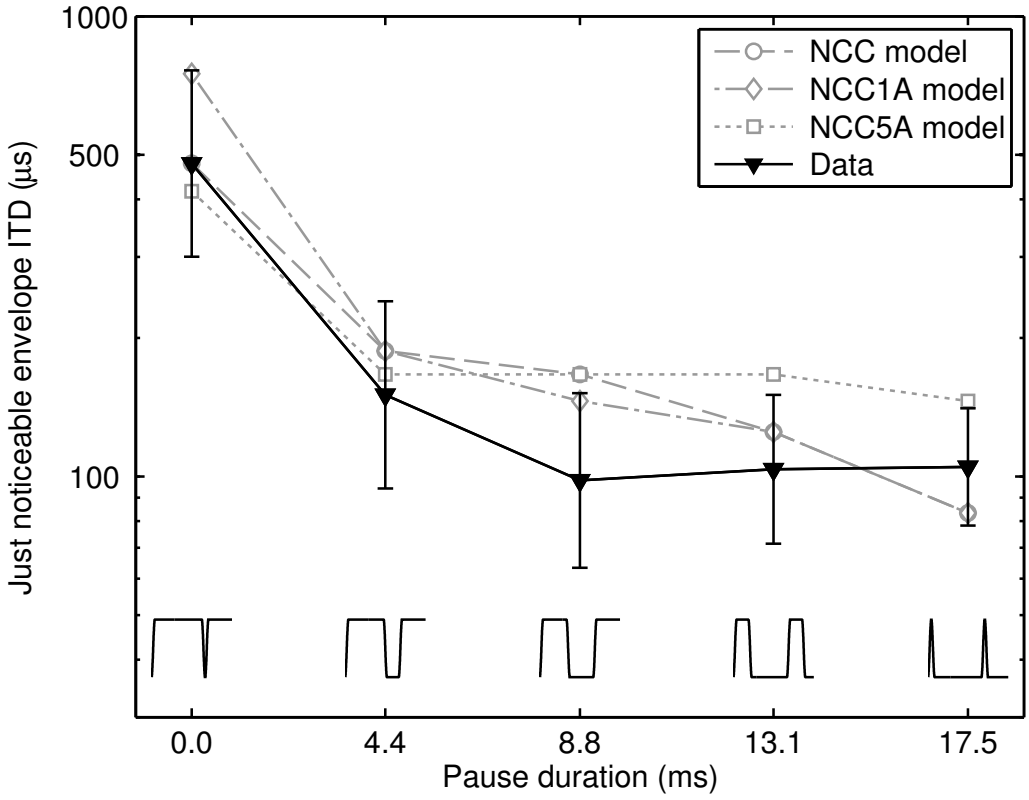

Figure 2.6: Results of Experiment 4: Just noticeable envelope ITDs in microseconds for envelopes with pause durations of 0.0, 4.4, 8.8, 13.1, and 17.5 ms, shown as black triangles. The attack and decay durations were fixed at 1.3 ms, the modulation frequency was 50 Hz. Conventions are as in Figure 2.4.

comparisons (Bonferroni) showed that the shortest pause condition was significantly different ($p < 0.05$) from all other conditions. No significant differences were found between the conditions with pause durations larger than 0 ms.

All model predictions (indicated in gray) follow the general pattern of the data, while overestimating JNDs in most conditions. In agreement with the data, a strong decrease in threshold was observed when the pause duration increased from 0 to 4.4 ms. The NCC and the NCC1A model (circles and diamonds) predicted a further decrease in JND with increasing pause duration, especially between 13.1 and 17.5 ms where no difference can be found in the data. In contrast, the NCC5A model (squares) predicted almost constant JNDs for pause times greater than 4.4 ms and overestimated the JNDs for longer pause durations. However, it followed the general trend of the data better than the other models. The RMSE was lowest for the NCC model.

2.3.5 Experiment 5: Level

Rationale

Previously published data on the level dependence of the JND is not particularly consistent. One study (Smoski and Trahiotis, 1986) found a very strong effect: When decreasing the level from 80 to 45 dB SPL the JND increased from 22-28 μs to about 600 μs. Two other studies (McFadden and Pasanen, 1976; Nuetzel and Hafter, 1976) found a moderate dependence of up to 40% increase of JND per 10 dB overall level. Dye and Hafter (1984) found that increasing the intensity of high-frequency click trains improved ITD detection performance. Dreyer and Oxenham (2008) as well as Bernstein and Trahiotis (2008) reported only a very small effect. It is still subject to debate if level dependence is caused by spread of excitation or by an increased within-channel salience (e.g., Bernstein and Trahiotis, 2010). However, the influence of level is important for the discussion of the other experiments, because a change in level automatically leads to a change of flank steepness in SAM tones but leaves the attack duration constant. In this experiment, JNDs were acquired for SAM tones from 36 to 66 dB SPL in order to investigate the influence of overall level on lateralization.

Conditions

The four conditions for the level experiment consisted of SAM tones with levels of 36, 48, 60, and 66 dB sound pressure level and a modulation frequency of 50 Hz. The SAM tones were generated by setting the pause and hold durations to 0 ms and the attack and decay durations to 10 ms, resulting in a modulation frequency of 50 Hz. The envelope ITD was applied to all segments of the envelope.

Results

The results are shown in Figure 2.7. It was found that the JND decreased with increasing level by a factor of about 4 from 800 μs for 36 dB to 200 μs for 66 dB. A repeated-measures ANOVA revealed a highly significant main effect of the level with $F(3,15) = 19.57$, $p < 0.001$. Post hoc pairwise comparisons (Bonferroni) indicated that the 36-dB level condition was significantly different ($p < 0.05$) from all other conditions. No significant differences were found between the conditions with levels 48, 60, and 66 dB SPL.

As a consequence of the normalization of the cross correlation and the single-channel framework, all models of this study are level independent. To overcome this issue, the model results were generated without normalization of the dot product, $\sum x(t)y(t)$, for this

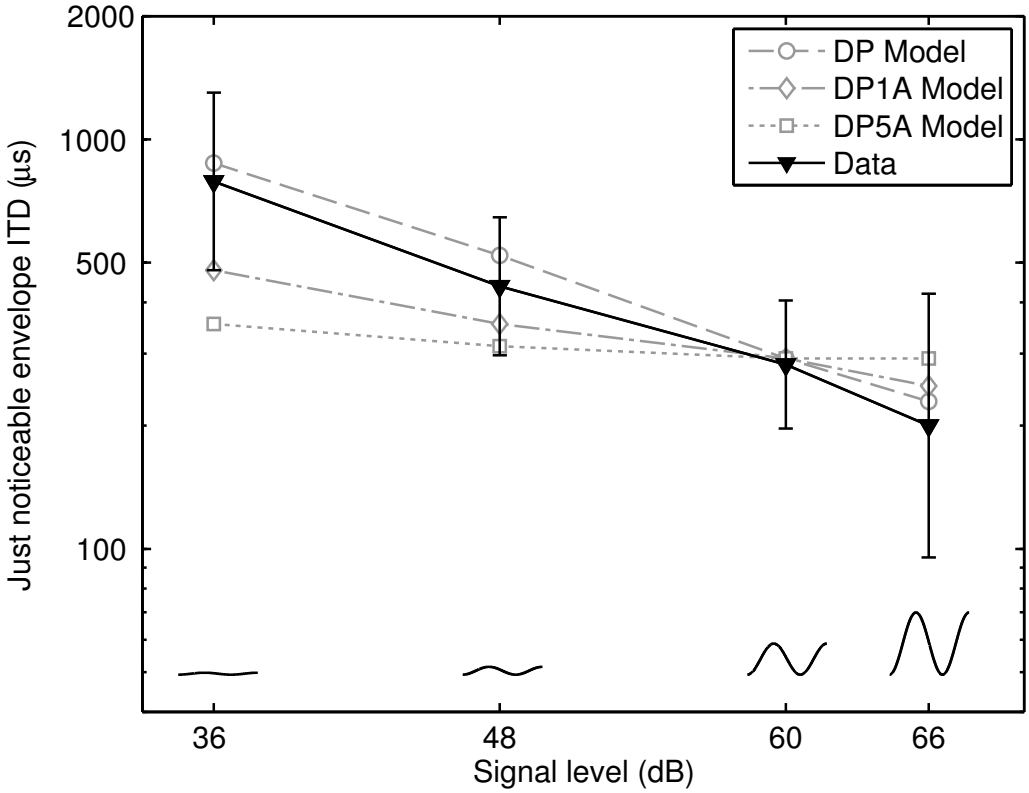

Figure 2.7: Results of Experiment 5: Just noticeable envelope ITDs in microseconds for SAM tones with rms levels of 36, 48, 60, and 66 dB SPL, shown as black triangles. Conventions are as in Figure 2.4, except that the model results were computed without normalizing the cross-correlation coefficient (see Section 2.3.5.

experiment [see Eq. (1)]. The preprocessing was otherwise identical to the general NCC model type. Due to the missing normalization of the dot product, the JND was simulated using the detection criterion $\Delta\rho$[6]. The NCC model without normalization is termed the "DP" model (dot product model), and 1 or 5 adaptation loops models without normalization "DP1A" and "DP5A". The DP model shows the least prediction errors and correctly predicts the trend of the data. The DP1A and DP5A models predicted decreasing JNDs, but with a too shallow slope. that is, too shallow. In terms of the RMSE as given in Table 2.1 (in the columns labeled as the respective NCC models), the DP model provided the best goodness

[6]To predict the JND of a specific condition, the dot-product dp(ITD=0) for the stimulus with an ITD of 0 μs was determined. The detection criterion was defined as dp(ITD=0)-δ, where δ was fitted for a correct prediction of the 50-Hz 60-dB SAM condition. Thus, it is assumed that, independent of the level, a fixed reduction of the dot-product is required to predict the JND. The values of δ were 0.03 for the DP model and 4451.9 for the DP1A and DP5A models. In the case of the DP model, the value δ of 0.03 can be calculated by multiplying the dot-product dp(ITD=0) and $(1-\rho_{\text{crit}})$, with $\rho_{\text{crit}} = 0.9997$, as described in Section 2.2.4.

of fit. Nevertheless, the DP models can only be employed for identical envelope waveforms in both channels, so they cannot account for Experiments 1 and 3. If the envelope waveforms are identical in both ears (except for a time shift) and if the level is kept constant, the DP and NCC models and their modified versions produce identical predictions.

2.3.6 Experiment 6: Modulation frequency

Rationale

Given that the stimuli used in the attack-duration and hold-duration experiments (Experiments 1 and 2) vary in their modulation periods, modulation frequency is a confounding covariant. The influence of modulation frequency on the JND should therefore be studied. Bernstein and Trahiotis (2002) found a decreasing JND in SAM tones for modulation frequencies from 32 to 128 Hz. Hafter and Dye (1983) found decreasing JNDs with increasing modulation frequency (or click rate) using click trains. They explained their results with a "multiple-look" approach assuming that more events reduce the error of the estimation proportionally to the square root of the number of observed clicks. For a constant number of clicks, however, Hafter and Dye (1983) found an increasing JND with increasing click rate. This effect can be accounted for by either peripheral adaptation or by binaural adaptation, although the latter does not play a significant role for modulation frequencies below 100 Hz. In the current study, the PSW tones can be used to vary modulation rate independent of flank steepness (fixed attack and decay durations) by alteration of only the pause and hold durations. By comparison with SAM tones, it can be examined to which degree a potential change in JND is caused by a change in flank steepness, by the effects of peripheral adaptation, and a change in modulation rate per se.

Conditions

For SAM and PSW tones, the influence of modulation frequency was investigated using five conditions. Two SAM conditions with modulation frequencies of 50 and 100 Hz and three PSW conditions with modulation frequencies of 35, 50, and 100 Hz were used in the experiment. The attack and decay durations of the PSW tones were fixed at 1.3 ms, the hold and pause durations were identical and set to 13.1, 8.8, and 3.8 ms to achieve the desired modulation frequencies. The envelope ITD was applied to all segments of the envelope.

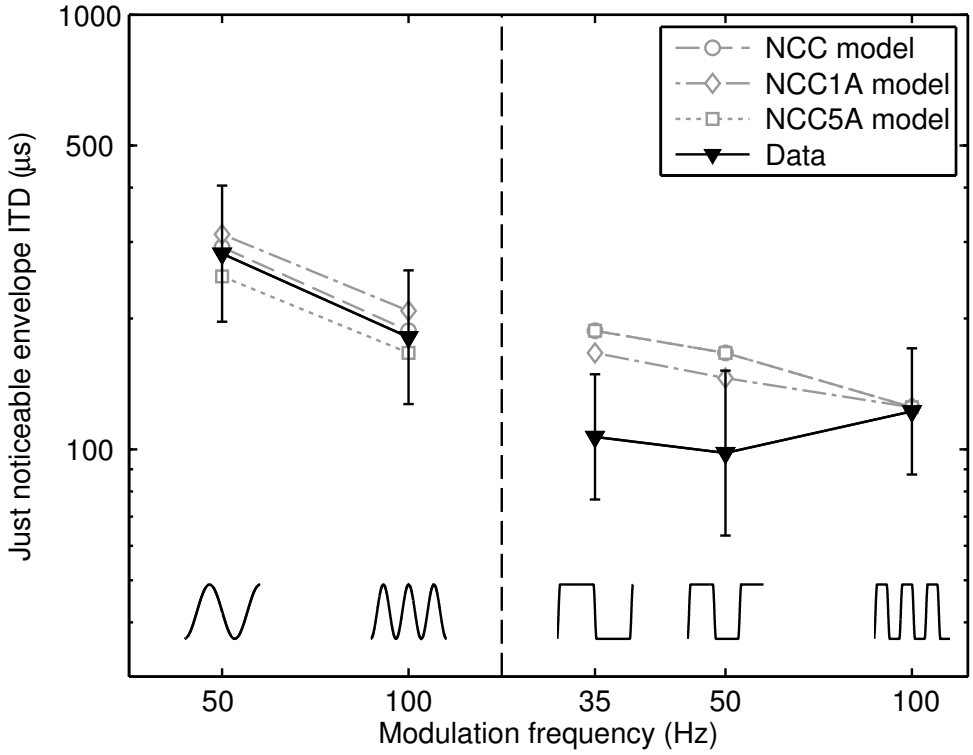

Figure 2.8: Results of Experiment 6: Just noticeable envelope ITDs in microseconds for SAM tones with $f_m = 50$ and 100 Hz (left panel), and PSW tones with $f_m = 35$, 50, and 100 Hz (right panel) shown as black triangles. The attack and decay durations were 1.3 ms. Conventions are as in Figure 2.4.

Results

Figure 2.8 shows the JNDs for SAM tones (left panel) and PSW tones (right panel). For SAM tones, the JND dropped from 282 μs for $f_m = 50$ Hz to 181 μs for $f_m = 100$ Hz. This behavior was not observed with PSW stimuli, where the JNDs for the stimuli with three different modulation frequencies exhibited hardly any variation with values of 107, 98, and 122 μs for $f_m = 35$, 50, and 100 Hz, respectively. A repeated-measures ANOVA indicated a significant main effect of the modulation frequency on the JND for SAM tones: F(1,5) = 9.2, $p < 0.05$. For PSW tones, no significant main effect was found [F(2,10) = 0.9, $p = 0.43$].

The open gray symbols are the model predictions. The predictions of all three models provided a good fit to the data for the SAM conditions. For the PSW conditions, however, the model predictions decreased with increasing modulation rate, contrary to the psychoacoustic results. The goodness of fit for the models showed that the NCC1A model performed best.

2.3.7 Experiment 7: Direct current offset

Rationale

To further investigate the influence of the pause duration on the JND, the SAM and PSW envelope was modified by adding a constant value to the envelope, resulting in an upward shift in amplitude of the envelope. The direct current (dc) offset created by this modification of the envelope replaces the pause segment (silence or no signal present) by a region of reduced, nonzero carrier intensity. This way, the resulting "pause" segment can be assumed to allow for less recovery in the sensitivity of auditory neurons than expected for a segment of silence. The dc offset modification employed here allows for a change of overall level of SAM and PSW tones without changing their attack or decay steepness. Furthermore, the modification also corresponds to a reduction of the modulation index of the signal. McFadden and Pasanen (1976), Nuetzel and Hafter (1981), Stellmack et al. (2005), and Bernstein and Trahiotis (2009) found increased JNDs for SAM tones with decreased modulation index.

Conditions

Two SAM and two PSW conditions were used. The envelope was shifted upward by adding a value of 0.67, creating a dc offset together with a level increase, but retaining constant flank steepness. For both the SAM and PSW stimuli, the condition without dc offset had a modulation index of 1, and the dc offset condition had a modulation index of 0.43. The modulation rate for all four conditions was 50 Hz, with the envelope ITD applied to all segments of the envelope.

Results

The JND results for stimuli with a dc offset are shown in Figure 2.9. For both stimulus types, the introduction of the dc offset clearly led to an increase in the JND. This effect was more pronounced in the SAM tone results, where the JND increased by a factor of more than 5 from 282 to 1676 μs. For PSW tones, the JND increased by a factor of 2.5 from 98 to 251 μs. A one-way repeated-measures ANOVA showed a significant main effect of the dc offset on the JND for both the SAM [$F(1,5) = 56.2$, $p < 0.01$] and the PSW [$F(1,5) = 15.5$, $p = 0.011$] conditions.

As in the previous figures, the gray symbols indicate the model predictions. All models predicted an increase in JND when an envelope dc offset was applied. The predictions

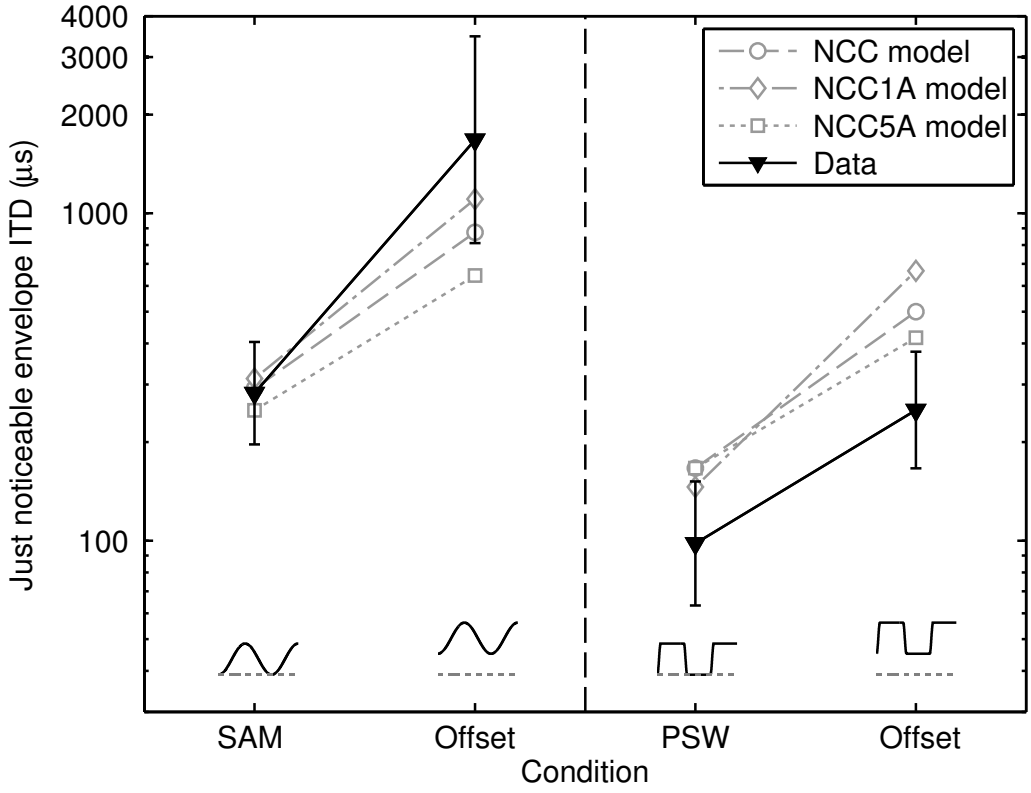

Figure 2.9: Results of Experiment 7: Just noticeable envelope ITDs in microseconds for SAM and PSW tones with and without dc offset, shown as black triangles. Conventions are as in Figure 2.4.

matched the JND range well for the SAM tones while they considerably overestimated JNDs in case of the PSW tones. For all models, the factor of the increase was too small for SAM tones and too large for PSW tones. The NCC1A model produced the smallest prediction error.

2.3.8 Experiment 8: Temporal asymmetry

Rationale

All three models are sensitive to differences in the envelope power spectrum between stimuli. In this experiment, the JND of two temporally asymmetric stimuli which are temporally reversed versions of each other (and thus have the same power spectrum) were determined. As a consequence of differences in the attack durations, the stimuli may yield different JNDs in the experimental data. Such behavior is not predictable using the NCC model. However, differences in the data for the time-reversed envelope versions would be in line

with monaural data by Akeroyd and Patterson (1997). They investigated the discrimination of a similar envelope shape and its time-reversed form as a function of modulation depth and modulation frequency. Their results indicate that the subjects did not use spectral cues to discriminate the two waveforms. In their study, envelope-spectrum-based models predicted no discrimination, contrary to the experimental results.

Conditions

The envelope of the first stimulus ("damped") had an attack duration of 1.3 ms and a decay duration of 18.8 ms. The second stimulus ("ramped") was a time-reversed version of the damped stimulus and consequently had an attack duration of 18.8 ms and a decay duration of 1.3 ms. Both stimuli had a pause duration of 10 ms and 0-ms hold duration. The modulation rate for both conditions was 33 Hz, with the envelope ITD applied to the whole envelope.

Results

The results are shown in the left panel of Figure 2.10. The first stimulus (left-hand side) with the short attack duration led to a JND of 114 μs. Temporal reversal and consequently swapping attack and decay durations resulted in an increase of JND by a factor of about 4 (right-hand side). A repeated-measures ANOVA revealed a highly significant main effect of the temporal reversal on the JND: $F(1,5) = 143.0$, $p < 0.01$.

The model predictions (open gray symbols) differed across models. The NCC model (circles) predicted no JND change, as expected. The NCC1A (diamonds) and NCC5A (squares) models showed larger JND changes, but still deviated from the data in at least one condition. The NCC1A model showed the smallest prediction error.

2.3.9 Experiment 9: Transposed tone

Rationale

Transposed tones (van de Par and Kohlrausch, 1997) have been used in many recent lateralization experiments (e.g., Bernstein and Trahiotis, 2002; Furukawa, 2008; Bernstein and Trahiotis, 2009). In these studies, transposed tones were generated by multiplying a high-frequency pure tone carrier with a half-wave rectified and 2000-Hz low-pass filtered, low-frequency pure tone. In order to provide a more complete reference for testing the current and future models, the comparison between transposed and SAM tone thresholds

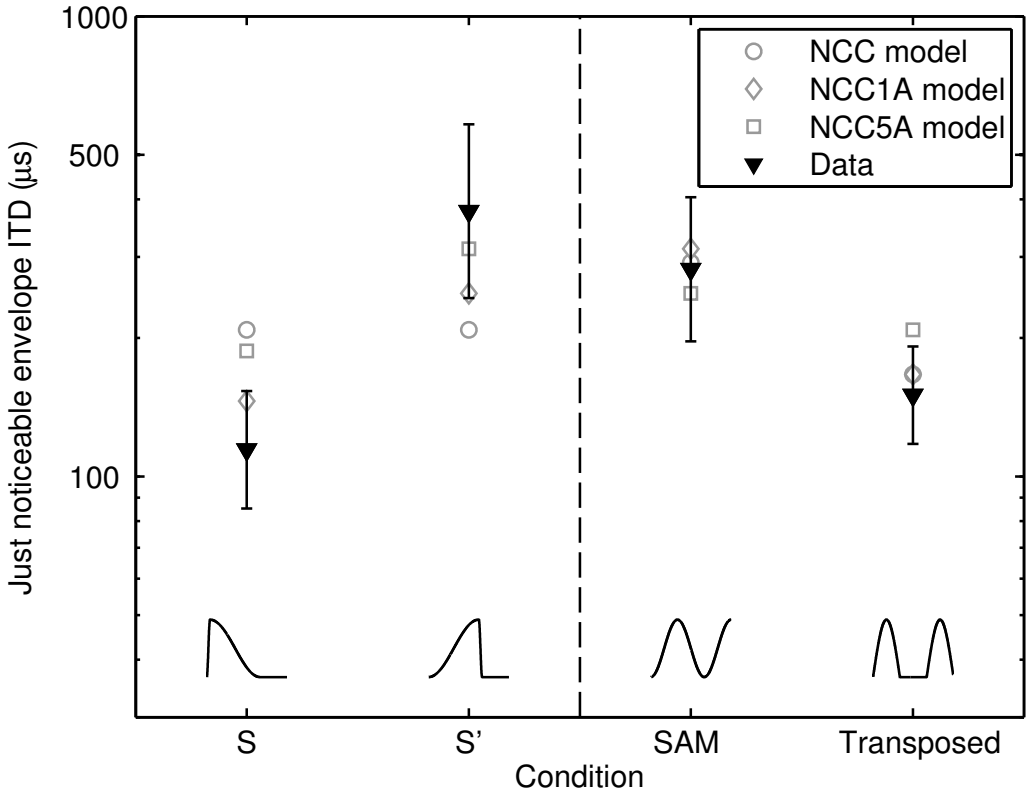

Figure 2.10: Left panel: Results of Experiment 8. Just noticeable envelope ITDs in microseconds for a condition with an attack duration of 1.3 ms, a decay duration of 18.8 ms, and a pause duration of 10 ms (S). Condition S' is the time-reversed version of condition S. Conventions are as in Figure 2.4, except that the insets show 60-ms portions of the stimulus envelopes. Right panel: Results of Experiment 9. Just noticeable envelope ITDs for a SAM tone and a transposed tone. Both stimuli had a modulation frequency of 50 Hz. Conventions are as in Figure 2.4.

from Bernstein and Trahiotis (2002) was repeated with the modulation frequency, stimulus parameters, and subjects of the current study.

Given that a transposed tone cannot be generated by defining specific attack, hold, decay, and pause durations, it was investigated separately by comparison of its JND to that of a SAM tone of the same modulation frequency. The results can be interpreted with regard to differences in the attack flank shapes.

Conditions

The transposed tone was generated in accordance with Bernstein and Trahiotis (2002) from a 50-Hz pure tone and a 4-kHz carrier. This results in an envelope waveform with a modulation frequency of 50 Hz. The right-hand waveform above the x axis in the right panel of Figure 2.10 shows a 30-ms portion of the stimulus envelope for the transposed tone. The left-hand waveform in the right panel indicates the envelope of the corresponding 50-Hz SAM tone.

Results

The right panel of Figure 2.10 shows the results of the transposed-tone experiment. Compared to the SAM tone, the transposed tone led to a JND decrease by a factor of about 2. A repeated-measures ANOVA showed a significant difference between the SAM and transposed tone JNDs: $F(1,5) = 34.4$, $p < 0.01$.

The NCC1A (diamonds) and NCC (circles) models provided good JND predictions for the transposed tone in comparison to the SAM tone. The predictions of the NCC5A model (squares) for the transposed tones were shifted to a higher JND, showing little JND decrease for the transposed tone compared to the SAM tone. The goodness of fit given as RMSE in Table 2.1 indicates that the NCC1A model performed best.

2.4 Discussion

In summary, the experimental results showed that the attack and pause envelope segments have the strongest influence on the JND, while changing the hold and decay duration did not alter the JND significantly. No single version of the tested models could predict the correct trend of the data in all experiments. The RMSE of the model predictions indicated that the model with a single adaptation loop and a time constant of 5 ms (NCC1A) had the best overall performance.

2.4.1 Influence of isolated envelope segments

Even though all four envelope segments could be changed independently, an isolated change of a single segment always resulted in a change of the modulation frequency which might have affected the data. However, Experiment 6, as well as published data (e.g., Hafter and Dye, 1983; Bernstein and Trahiotis, 2002), indicated that the influence of modulation

frequency per se is no larger than the square root of the ratio of the number of cycles during a given observation time, $\sqrt{N1}/\sqrt{N2}$, with N1 and N2 referring to the number of cycles in the conditions that are compared (for further details see Section 2.4.2). The data of Experiment 2 (see Figure 2.5), where constant JNDs were found for three hold durations, indicated that this parameter did not affect JNDs. It is therefore assumed that the co-variation of attack duration and either hold duration or modulation frequency did not have a considerable effect in the data.

Given that the response of most auditory neurons shows a maximum at onset, it appears plausible that sensitivity is most pronounced in the attack segment, making changes in the attack flank more salient than changes in the remaining parts of the ongoing envelope cycles. An increased salience of shorter rise time is in line with Smith and Brachman (1980), who found stronger onset responses in PSTHs of gerbil auditory nerve fibers for shorter rise (i.e., attack) times. It is also in line with Heil (2001), who found shorter first-spike latencies with shorter rise times in the primary auditory cortex of a cat. The small JND increase between the 1.25- and 2.5-ms conditions can be accounted for by the small increase in rise time of the "internal" stimuli after peripheral auditory preprocessing. Comparing the 20%-80% rise times of the "internal" and unprocessed "external" stimuli shows that the internal rise times of the 1.25- and 2.5-ms conditions differ by only 0.4 ms (see Table 2.2). This shows that the peripheral monaural preprocessing has a considerable effect on the current model predictions.

In Experiment 2, the hold duration was found to have no significant influence on the JND (see Figure 2.5), indicating that JNDs are not influenced by spectral narrowing caused by an increased hold time. This result also rules out an undesired confounding effect in experiments with a co-variation of the hold duration, such as Experiment 4 (influence of pause duration).

Attack duration (ms)	20%–80% unprocessed (ms)	20–80% preprocessed (ms)
1.25	0.5	3.1
2.5	1.0	3.5
5	2.0	4.6
10	4.0	7.3

Table 2.2: 20%-80% rise times of the unprocessed and preprocessed attack experiment stimuli

The results of the decay experiment (Experiment 3) revealed severe difficulty in successfully completing the experiment for many subjects. JNDs larger than 1000 μs were found, indicating that the decay segment of the envelope contributes minimally to lateralization. In some cases, the envelope ITD became so large that an additional, confounding ILD cue was introduced as a consequence of the reduction of the hold duration of the leading side envelope. The ILD cue was into the opposite direction of the envelope ITD cue, which could have led to confusion of the subjects as soon as a certain ITD of 1000-2000 μs was reached during the adaptive procedure[7].

In contrast to the decay experiment where ITD cues were only present in the decay flank, the results of the attack experiment with envelope-waveform shift showed an influence of the decay flank even at low envelope ITDs. Thus the decay flank might play a supporting role, improving the use of envelope ITDs as a cue in conditions employing a more "natural" full envelope-waveform shift. However, a lateralization of the stimuli solely based on an ITD in the decay flank was difficult.

Experiment 4 (influence of pause duration) indicated a higher sensitivity to envelope ITD with increasing pause duration (Figure 2.6). A large difference in JND was observed between the 4.4- and 0-ms pause condition. No significant difference was observed between the four conditions with pause durations $\geq$ 4.4 ms. This observation is in line with the assumption that the pause prior to the attack allows for a sensitivity recovery of the auditory nerve fibers and that the recovery has a short time constant, less than about 5-10 ms (Westerman and Smith, 1984; Wickesberg and Oertel, 1990).

2.4.2 Influence of analytical envelope parameters

Some manipulations, such as varying the modulation frequency (Experiment 6) or modulation depth of SAM tones (Experiment 7) were previously explored by McFadden and Pasanen (1976), Nuetzel and Hafter (1981), Stellmack et al. (2005), and Bernstein and Trahiotis (2009). The results of the current study are in line with their data. Bernstein and Trahiotis (2009) found an envelope ITD threshold of about 290 μs for a 64-Hz SAM and 120 μs for a 64-Hz transposed tone. These are comparable to the JNDs of 282 μs (50-Hz SAM) and 150 μs (50-Hz transposed tone) observed in Experiment 9 (Figure 2.10).

In Experiment 5, it was found that the JND decreased by a factor of 3.9 when increasing

[7]The counteracting ILD cue led to confusion in some of the subjects during the data collection and the adaptive procedure failed in some of those cases. A collection of the whole psychometric function could have provided more detailed insight but would have increased the measurement time substantially without noticeable improvement of the otherwise still clear cut results of the adaptive procedure in the context of this study.

the overall level by 30 dB (Figure 2.7). This effect of level was stronger than that reported in most other studies (e.g., Nuetzel and Hafter, 1976; Dreyer and Oxenham, 2008; Bernstein and Trahiotis, 2008) but weaker than in Smoski and Trahiotis (1986). Given that the envelope flanks in the SAM tones steepen with increasing level, the steepness of the attack flanks may be the dominating cue for envelope ITD discrimination. This is supported by a comparison between the 100-Hz, 60-dB SAM condition from Experiment 6 and the 50-Hz, 66-dB condition from Experiment 5. Both stimuli have the same maximum flank steepness and showed JNDs of 177 and 214 μs, respectively. A post hoc, pairwise comparison indicated no significant difference between the results of the two conditions. In Experiment 7, where the overall level was increased without varying the flank steepness, no decrease in JNDs was observed: Instead, increasing JNDs were observed, which are likely related to the absence of a segment of silence and reduced modulation index. This observation is in line with the results of McFadden and Pasanen (1976), where increased envelope ITD thresholds were found for a decrease in the modulation index.

The results of Experiment 6 show that the modulation frequency influenced the JND for SAM tones only (Figure 2.8). For PSW tones, no significant dependency could be observed for the range of modulation frequencies tested. This emphasizes the important role of the envelope flank steepness, which only changed in the SAM stimuli.

In the PSW conditions of Experiment 6, the pause time decreased with increasing modulation rate. From the pause experiment it can be estimated that the short pause duration (3.8 ms) of the 100-Hz condition should have lead to an increase of JND by about 50% while no effect on the JND would be expected for the 35- and 50-Hz conditions as observed in the data. However, the 100-Hz condition has a much larger number of cycles than the 35- and 50-Hz conditions, leading to more observable informative events (N). Hafter and Dye (1983) argued that given sufficiently long integration times, the available information increases by a factor of square root of N. A combination of the two effects explains the trends observed in the data of Figure 2.8: From 35 to 50 Hz, thresholds decrease slightly (16%) as a result of the higher N, while for 100 Hz, the thresholds increase again because now the reduction in pause duration starts to reduce sensitivity and counteracts a potential gain related to an increased number of observations.

The combined effect of flank steepness and pause duration was examined in Experiment 9 by means of transposed tones (see Figure 2.10). The results are in line with Bernstein and Trahiotis (2002, 2009), where the JND for a SAM tone and a transposed tone differed by a factor of about 2 at a modulation frequency of 64 Hz. The transposed tone has, in contrast to the SAM tone, a pause segment and a steeper attack and decay, presumably leading to

the high sensitivity to envelope ITD with transposed tones. In the current data, the JND of 150 μs for the transposed tone was higher than the JND of 98 μs for the comparable PSW tone with 8.8-ms pause duration (a post hoc pairwise comparison showed a significant difference, $p < 0.05$, between both conditions). Hafter and Buell (1990) used Gaussian click trains and found an average JND of 87 μs for a single click and around 50-70 μs for different click trains (estimated from their Figure 2.3). Given that their stimulus envelopes and the envelope of the PSW tone with 8.8-ms pause duration have a 5%-100% rise and decay-flank duration of about 1.1 ms, the current results are broadly in line with their data.

2.4.3 Relation between data and NCC model

The NCC model prediction for the transposed tone in Experiment 9 is in line with Bernstein and Trahiotis (2009), where the JND for a SAM tone and a transposed tone differed by a factor of about 2 and a normalized cross-correlation based model was able to correctly predict the data.

The underestimation of the increase in JND for the SAM condition with dc offset in Experiment 7, is in line with the trend observed in the model predictions in Bernstein and Trahiotis (2009) and Stellmack et al. (2005). In their studies, a cross-correlation coefficient based model was used to predict JNDs (or discriminability) for SAM tones with reduced modulation depth, showing thresholds lower than observed in their experimental data. For the PSW dc offset condition of the current study, all models predicted higher JNDs than observed in the psychoacoustic data. This is again in line with Bernstein and Trahiotis (2009), who tested raised sine stimuli with different exponents and reduced modulation depth (e.g., $m = 0.5$). The raised sine stimuli with an exponent of 8 are comparable to PSW stimuli regarding the pause, attack, and decay durations, except for the absence of a hold duration.

The level dependence observed in Experiment 5 cannot be modeled with a single-channel version of the normalized cross correlation coefficient model. Either the normalization has to be abandoned or the model would have to be extended to a multi-channel approach. In this study, the first approach was chosen and it was found that the level dependence could be modeled correctly, by using a static peripheral compression.

The inaccurate predictions in the hold duration (Experiment 2), pause duration (Experiment 4), and PSW modulation rate (Experiment 6) conditions are caused by the spectral sensitivity of the model. With the left-ear and right-ear envelopes of the stimuli being almost identical, the cross correlation of the envelope in both ears closely reflects the inverse Fourier transform of the envelope power spectrum in one ear. Thus, the decrease of the nor-

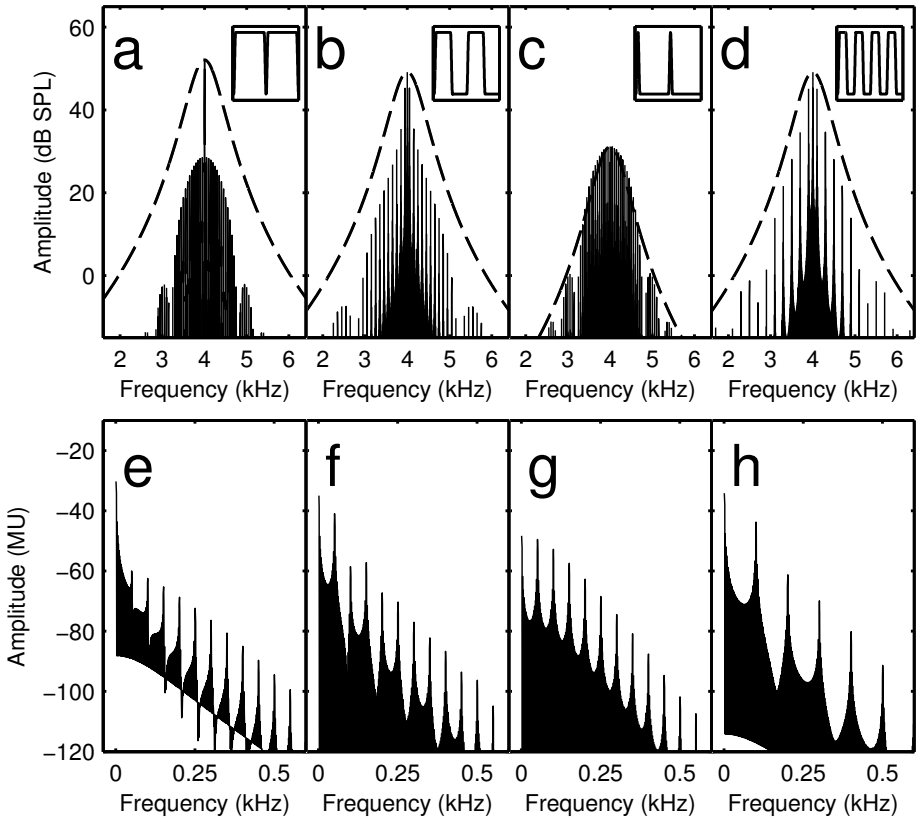

Figure 2.11: Top row: Spectra of the 0.0-, 8.8-, and 17.5-ms conditions of the pause experiment [panels (a)-(c)] and the 100-Hz condition of the modulation frequency experiment [panel (d)]. The spectra were obtained from the unprocessed stimuli and are given in dB sound pressure level. The small insets show 40 ms of the respective stimulus envelopes. The transfer function of a 4-kHz, fourth-order gammatone filter is indicated by the dashed line. Bottom row: Internal envelope spectra of the stimuli in the respective panels of the top row after auditory preprocessing as described in Section 2.2.4, given in model units (MU).

malized cross correlation coefficient with increasing envelope ITD is steeper for a stimulus with increasing spectral width than it is for a stimulus with a constant spectrum (Bernstein and Trahiotis, 2002). In Experiment 2, for example, the NCC predictions decreased with decreasing hold duration (see Figure 2.5), in contrast to the psychoacoustic data. Here, the increased width of the envelope spectrum related to shorter hold or longer pause durations, resulted in a faster decay of the correlation coefficient with increasing envelope ITD. Figure 2.11 shows the audio power spectra and the envelope power spectra after monaural preprocessing of three stimuli with a pause duration of 0.0, 8.8, and 17.5 ms taken from Experiment 4. The 0-ms pause condition has a strong peak at 4 kHz. With decreasing hold

time, this peak becomes less distinct and energy is increasingly spread across the spectrum. Such an increasing width is also observed in the envelope spectrum in the lower row with increasing energy at all modulation frequencies in relation to the dc component. Because of the increasing width of the envelope spectrum, the model predicted lower JNDs than observed in the data (see Figure 2.5 and 2.6).

In Experiment 1, the predicted trend of the JND as a function of the attack duration is slightly shallower than that in the data. These deviations are caused by the fact that in the stimuli for this experiment the envelope ITD was applied to the attack flank only and that the other identical envelope segments dominate the cross-correlation coefficient. The overall agreement between model and data was better (specified by RMSE in Table 2.1) in the supplementary attack experiment with full envelope-waveform shift.

Previous studies that have altered analytic parameters (e.g., Henning, 1974; Young and Carhart, 1974; Bernstein and Trahiotis, 2002, 2009; Dietz et al., 2009) did not test a potential effect of asymmetry between attack and decay and the NCC model could successfully model the data (Bernstein and Trahiotis, 2002, 2009). In contrast, the disparities in isolated envelope segments investigated in the current study demonstrated a stronger importance of attack flank and pause duration, which cannot be predicted by the NCC model in its present form. The envelope power spectral dependency of the NCC model becomes most obvious in Experiment 8, where the conditions are temporally reversed versions of each other and thus have identical power spectra. In this case, the NCC model predicts identical JNDs independent of the time reversal.

2.4.4 Relation between data and model with adaptation mechanism

Based on the average RMSE across Experiments 1, 2, 4, 6, 7, 8, and 9, performance of the NCC1A model was best (mean RMSE=1.32), followed by the NCC (mean RMSE=1.46), and then the NCC5A model (mean RMSE=1.52).

The NCC5A model produced the largest deviations from the data, caused by the five adaptation loops. With time constants ranging from 5 to 500 ms, the adaptation loops generate steeper attack flanks in the internal representation when compared to the other models. They contribute more strongly to the calculation of the normalized cross-correlation coefficient than the rest of the signal, leading to lower JND predictions for stimuli with long attack durations. This can be observed in the predictions of attack duration (see Figure 2.4). The NCC1A model predictions showed the same general trend, however. If the models were configured to correctly predict the results of the 1.3-ms attack condition by adjusting the ρ-criterium accordingly, all other predicted JNDs in the attack experiment would generally

be too low for the NCC5A model. The NCC1A model would predict three out of the four attack conditions fairly accurately, resulting from the weaker adaptation. The predictions of the NCC1A and NCC5A models in the supplementary attack experiment were similar to the ones for Experiment 1, again overestimating the JNDs and showing less dependency on the attack duration as observed in the data.

In Experiment 4, the NCC5A model overestimated JNDs for pause durations greater than 4.4 ms, in contrast to the NCC model. With the long time constants involved, the NCC5A model does not recover its sensitivity fast enough during the pause duration, leading to a reduced steepness of the attack flanks in the internal representation of the envelopes. Given that binaural processing takes place at early stages of the auditory pathway (Yin and Chan, 1990) and given that the five adaptation loops simulate the adaptation process along the complete monaural auditory pathway in Dau et al. (1996a,b), it appears reasonable to hypothesize that only the first fast adaptation stage or stages precede the binaural processor[8]

For Experiment 2, both the NCC5A and NCC1A models showed a decrease in sensitivity with increasing hold duration similar to the NCC model and in contrast to the data. In this case, the spectral changes between the conditions cannot be compensated by the adaptation stage.

A similar dependency of the model predictions on spectral changes was observed for the PSW stimuli in Experiment 6. Here, all the models predicted decreasing thresholds for increasing modulation frequency. In this case, the width of the internal envelope spectra increases with increasing modulation frequency as shown in Figure 2.11. In contrast, the data show constant thresholds for the PSW stimuli. Only for SAM tones, where the flank steepness increases with increasing modulation frequency, were decreasing thresholds were observed.

The NCC5A and NCC1A model predictions for conditions using SAM tones were similar to those of the NCC model, because the preprocessed SAM stimuli maintain their shape after passing through the adaptation loops. This led to almost identical predictions with the same trends for the NCC, NCC5A, and NCC1A models in the SAM modulation frequency and dc offset experiments.

In Experiment 8, the NCC5A and NCC1A models showed a larger JND for the time-reversed condition with shallow attack flank than for the condition with steep attack flank and shallow decay flank. However, the sensitivity difference introduced by the adaptation stage (a factor of 1.6) was much lower than the factor of 4 that was observed in the data.

[8] Adaptation time constants in this context do not relate to time constants for temporal binaural resolution (binaural sluggishness) (e.g., Culling and Summerfield, 1998; Akeroyd and Summerfield, 1999) as in the case of the NCC1A model.

With respect to Experiment 5, the adaptation stage used in the DP1A and DP5A models cannot maintain good agreement with the data as observed for the non-normalized dot product approach. The reason is that the adaptation loops compress the level range of the input stimulus. This is, however, just a specific aspect of the implemented adaptation loops and not a general drawback of adaptation. Other peripheral models that include adaptation (e.g., Sumner et al., 2002) include less compression.

Taken together, the overall performance of the NCC5A and NCC1A models was dominated by the envelope spectral sensitivity of the normalized cross-correlation coefficient. For pause/hold duration-related spectral changes, a single adaptation loop was able to better counteract the envelope spectral dependency than five adaptation loops. Using five adaptation loops with time constants up to 500 ms, the adaptation resulted in an overestimation of JNDs for stimuli with short pause durations, when compared to the psychoacoustic data. This effect was not found in the NCC1A model.

2.4.5 Implications for future modeling

In line with physiological findings in the cochlear nucleus (e.g. Wickesberg and Oertel, 1990), the modeling of the current data clearly supports relatively fast adaptation in the region of 5 ms. However, the current study cannot specify the exact characteristics of such adaptation. Further investigations could use a modified adaptation stage with an adjustable low-pass characteristic to better account for the data. In this work, the adaptation loops were taken "as is" from previously published monaural and binaural models (Dau et al., 1996a; Breebaart et al., 2001a) that differ in their front-end and back-end processes. Alternatively, a more detailed hair cell and auditory nerve model showing adaptation effects could be used (e.g., Meddis, 1986; Meddis et al., 1990; Sumner et al., 2002) or mechanisms from the model by Neubauer and Heil (2008) could be used to obtain a level and rise-time-dependent auditory nerve stage. These potential preprocessing stages, however, might still not sufficiently counteract the envelope-spectral dependency of the NCC detection process.

An alternative binaural processing scheme to the NCC model could be based on interaural differences, mimicking an excitatory-inhibitory (EI) neural circuit as used in the binaural model by Breebaart et al. (2001a,b,c). Again, the current data suggest that such an EI-type binaural stage should be preceded by fast-acting adaptation, contrary to Breebaart et al. (2001a,b,c) and Thompson and Dau (2008). A simple EI-type binaural stage was tested in Ewert et al. (2010) for a more limited set of data and could provide a good starting point for future investigations. Such an approach would also be in line with the findings by Joris and Yin (1995). They showed that the sensitivity to envelope disparities is based on the fact that

these cells are EI cells and that they can be characterized by a subtractive mechanism.

2.5 Conclusions

The role of four envelope segments, attack, hold, decay, and pause of ongoing, periodic envelope waveforms on the just-noticeable interaural envelope time difference (JND) was investigated. The stimuli had a carrier frequency of 4 kHz. Psychophysical data and model predictions were compared. The following was found:

1. The psychophysical data revealed differential effects of the individual features of the ongoing envelope waveform on lateralization: Increased attack steepness and increased pause duration prior to the attack resulted in the lowest JND. This observation is in line with the assumption of a neuronal adaptation mechanism.

2. For stimuli with pseudo-square-wave modulation with fixed attack and decay flanks, JNDs did not depend on modulation frequency in contrast to stimuli with sinusoidal amplitude modulation for the range of frequencies investigated (33-100 Hz). This result emphasizes the importance of flank steepness for JND rather than modulation frequency.

3. For stimuli with temporally asymmetric envelopes, JNDs changed with time reversal. Thus the data do not support the existence of a simple functional relation between the normalized cross-correlation coefficient (or envelope power spectrum) and the JND. A time-dependent preprocessing stage (e.g., an adaptation stage) prior to the calculation of the normalized cross-correlation coefficient appears suited to solve this problem.

4. The incorporation of a single adaptation stage with a short time constant of 5 ms prior to the calculation of the cross correlation coefficient accounted best for the data. This result indicates that it might not be necessary to include long adaptation time constants prior to the binaural interaction stage as done in, e.g., Breebaart et al. (2001a) and Thompson and Dau (2008).

Acknowledgments

This study was supported by the DFG (SFB/TRR31 “The Active Auditory System”). We would like to thank the Medical Physics group and Birger Kollmeier for constant support and fruitful discussions. Reviewer Leslie R. Bernstein, two anonymous

reviewers, and the associate editor Michael Akeroyd provided valuable input to improve the manuscript.

Chapter 3

Effect of mistuning on the detection of a tone masked by a harmonic tone complex

Abstract The human auditory system is sensitive in detecting "mistuned" components in a harmonic complex, which do not match the frequency pattern defined by the fundamental frequency of the complex. Depending on the frequency configuration, the mistuned component may be perceptually segregated from the complex. In other cases, mistuning can be detected using different signal features such as envelope fluctuations. In the context of a masking experiment, mistuning a single component decreases its masked threshold. In this study we propose to quantify the ability to detect a single component for fixed amounts of mistuning by adaptively varying its level. This method produces mistuning masking level differences that can be compared to those of other masking release effects. Detection thresholds were obtained for various frequency configurations where the target component was resolved or unresolved. The results from 6 normal-hearing listeners show a significant decrease of masked thresholds between harmonic and mistuned conditions in all configurations and provide evidence for the employment of different detection strategies for resolved and unresolved components. The data suggest that across-frequency processing is involved in the release from masking, showing that it is unlikely to be produced solely by a peripheral process. The results emphasize the ability of this method to assess integrative aspects of pitch and harmonicity perception.

This chapter is a reformatted reprint of "Effect of Mistuning on the Detection of a Tone Masked by a Harmonic Tone Complex", M. Klein-Hennig, M. Dietz, A. Klinge-Strahl, G. Klump, V. Hohmann, PLoS ONE 7(11): e48419 (2012). The original article can be found at `http://dx.doi.org/10.1371/journal.pone.0048419`.

3.1 Introduction

The harmonicity of a signal is an important feature in auditory grouping. It allows humans to group frequency components that have a common fundamental frequency (F0) into a single auditory object (e.g., Bregman, 1994). This helps, for example, in the segregation of concurrent speech from different talkers or speech from noise (e.g., Darwin and Carlyon, 1995).

In a harmonic complex it is difficult to detect or "hear out" single frequency components. To facilitate this task, additional cues are needed. The most commonly used cue is mistuning, i.e., shifting the frequency of a target component in such a way that it no longer matches the harmonic frequency pattern defined by the F0 of the complex (e.g., Moore et al., 1985, 1986; Hartmann et al., 1990; Hartmann and Doty, 1996). Mistuning effectively provides release from masking, enabling or facilitating the detection of a single component that would otherwise be masked by the rest of the harmonic complex. This masking release has been shown only indirectly in studies on mistuning detection (e.g., Moore et al., 1985; Hartmann et al., 1990), as they measured the just noticeable amount of mistuning using a paradigm in which the subjects compared mistuned complexes to harmonic complexes, while the amount of mistuning was adaptively varied. A direct investigation of the effect of mistuning on detection thresholds is possible in detection experiments, where a single component of a complex is the target signal that has to be detected, and the rest of the complex is regarded as the masker, effectively masking the component (Oh and Lutfi, 2000; Klinge et al., 2011). Here, the level of the target component is varied adaptively to obtain detection thresholds for various stimulus configurations. This method has the advantage that it generates detection thresholds that allow comparison to and possibly combination with other masking release effects such as comodulation masking release (e.g., Hall et al., 1984) or binaural unmasking (e.g., Licklider, 1948). These effects are also investigated by measuring and comparing detection thresholds, while the results of the mistuning detection studies mentioned above cannot be expressed in terms of masking level differences, rendering the comparison to other masking release effects difficult. The two studies that investigated single-component detection in harmonic complexes (Oh and Lutfi, 2000; Klinge et al., 2011) have several methodological specifics and limitations that make it difficult to derive a comprehensive picture of the auditory processing involved. Oh and Lutfi (2000) used non-deterministic frequency configurations in their stimuli, as they designed their experiment in the context of informational masking, whereas Klinge et al. (2011) used deterministic frequency configurations. Klinge et al. (2011) performed their measurements in free-field, which makes

control of the stimuli at the ear-level difficult. Both studies measured detection thresholds for one fixed percentage of mistuning only.

To shed further light on the effects involved in harmonicity processing, this study provides a data set of detection thresholds for a single sinusoidal target tone masked by harmonic and mistuned complexes for an extensive set of critical conditions. The stimuli consisted of a harmonic complex as masker and a single sinusoidal target component that was either harmonic or mistuned to the masker's fundamental frequency. The stimuli were presented in a controlled environment with headphones, with an adaptively varied target. In order to test the ability of the method to account for different strategies to detect the mistuned component, thresholds were obtained for frequency configurations in which the harmonics of the tone complex are either resolved or unresolved[1]. The possible involvement of across-frequency processes in the detection of a mistuned component is investigated by increasing the stimulus bandwidth.

3.2 Materials and Methods

3.2.1 Ethics statement

Written consent was obtained from each participant prior to the experiments. The experiments were approved by the local ethics committee of the University of Oldenburg.

3.2.2 Subjects

Six normal-hearing listeners (3 male, 3 female), aged 22 to 27, took part in the study. Pure tone audiograms were measured for all test subjects, showing no hearing loss ($>$ 15 dB HL) between 250 and 8000 Hz. Prior to data collection, all subjects completed a 3-hour training run with the same stimuli as used in the experiments.

3.2.3 Stimuli

The stimuli consisted of a harmonic complex used as masker and a pure tone target signal. The masker was generated by adding up eight pure tones of different frequencies in random phase for each stimulus presentation, to prevent subjects from learning spectral or temporal

[1] A harmonic is defined as resolved if it individually excites a place on the basilar membrane (e.g., Plack and Oxenham, 2005), i.e., it is the only harmonic that occurs within the equivalent rectangular bandwidth (ERB, see Moore and Glasberg, 1996) of an auditory filter centered around that harmonic's frequency. If multiple harmonics fall into the same ERB, they are defined as being unresolved.

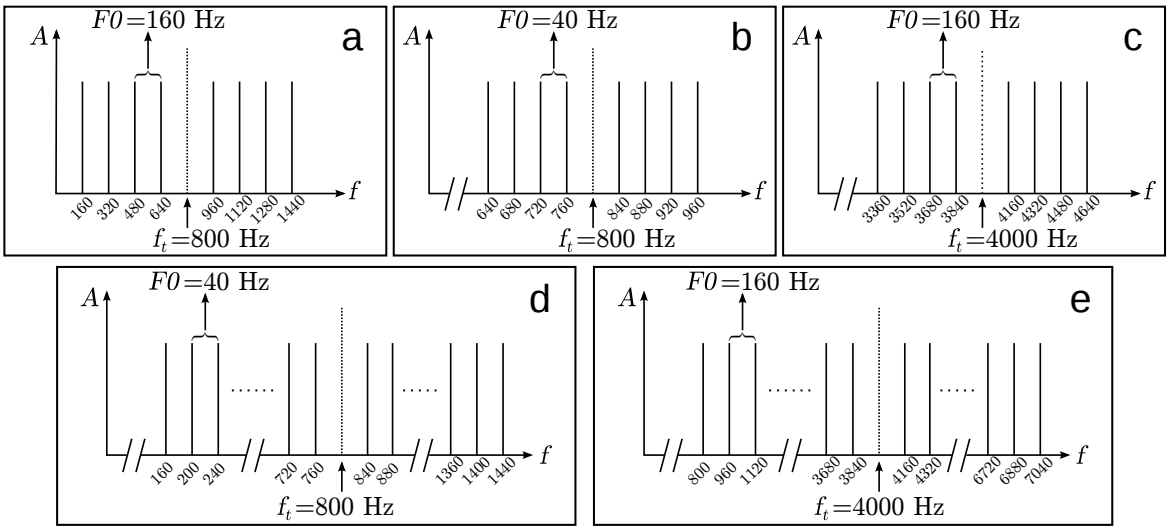

Figure 3.1: Pictogram of the spectra of the stimuli used in the experiments. Black lines indicate the harmonics of the tone complex used as masker. The harmonics are multiples of the masker's fundamental frequency F0. The lowest harmonic is not necessarily the fundamental frequency. The dotted line shows the frequency of the target component that had to be detected and was not part of the masker. Panels a to c show the frequency configurations of Experiments 1-3 in the harmonic condition. Panels d and e show the frequency configurations of the broad-band conditions with additional masker harmonics as used in Experiments 2 and 3.

templates and exploiting them in the detection task. The frequencies were integer multiples of the fundamental frequency F0 of the masker and were the four harmonics below and above the target signal frequency f_t . In Experiment 1, a fundamental frequency F0 = 160 Hz was used, with a target frequency of f_t = 800 Hz (see 3.1a). To generate unresolved harmonics in Experiment 2, F0 was set to 40 Hz, while keeping f_t at 800 Hz (see Figure 3.1b). Experiment 3 also had unresolved harmonics in a high frequency range, by shifting f_t to 4 kHz, while keeping F0 at 160 Hz (see Figure 3.1c).

Experiments 2 and 3 also contain "broad-band" conditions that were configured to have the same bandwidth-to-target-frequency ratio as Experiment 1. This was achieved by additional masker components below and above the target, leading to stimuli with 32 and 40 components in Experiments 2 and 3, respectively (see Figures 3.1d and 3.1e).

Each of the masker components had a level of 55 dB SPL, resulting in an overall masker level of 64 dB SPL. The broad-band conditions of Experiments 2 and 3 had overall masker levels of 70 and 71 dB SPL, respectively. Target and masker were gated and presented synchronously with 25 ms Hanning windows. The total stimulus duration was 400 ms.

In order to create a mistuned condition, the F0 of the masker was increased while keeping

the frequency of the target f_t constant, which effectively led to a downward mistuning of the target component. To have a comparable amount of mistuning in all experiments, the percentage of mistuning was chosen such that the fourth masker component, which is next to the target component, was shifted upwards by 10, 20 and 40 Hz in Experiments 1 and 2, as these values are proper divisors of the F0s of the stimuli. Higher frequency shifts of 20, 40, 80 and 160 Hz were used in Experiment 3 due to the high target frequency of f_t = 4 kHz. The fourth component was selected as a measure of mistuning since it is the most likely component to fall into the auditory filter centered on the target frequency after mistuning. A single-channel auditory model would use this frequency band as the only signal channel, because it has the highest target-to-masker ratio.

3.2.4 Procedure

The experiments were conducted in a double-walled, sound-attenuating booth. The stimuli were generated digitally with a sampling frequency of f_s = 48 kHz and presented via Sennheiser (Wedemark-Wennebostel, Germany) HD 650 headphones. The headphone was free-field calibrated on a Brüel&Kjær (Nærum, Denmark) 4135 artificial ear. The subjects responded using a computer keyboard and visual feedback was provided on a computer monitor.

A 3-interval 2-alternative forced-choice procedure was used to measure the detection thresholds. Using a 1-up 2-down tracking rule, the 70.7%-correct point on the psychometric function was estimated (see Levitt, 1971). The reference intervals did not contain the target signal. The first interval presented to the subjects was always a reference interval, and could not be selected. The target signal was first presented with a level of 65 dB SPL, which was initially varied in 5 dB steps. Stepsize was reduced to 2 and 1 dB after the second and fourth reversal, respectively.

Each adaptive run was terminated after eight reversals with 1 dB steps. The individual means were obtained by averaging over the last eight reversals of five experimental runs. The mean thresholds were obtained by averaging over the individual means of the subjects.

3.3 Results

3.3.1 Experiment 1: resolved harmonics, F0 = 160 Hz

In this experiment, the masker had a fundamental frequency of 160 Hz and the target signal frequency was f_t = 800 Hz (5^{th} harmonic of F0 = 160 Hz, see Figure 3.1a). This way, all

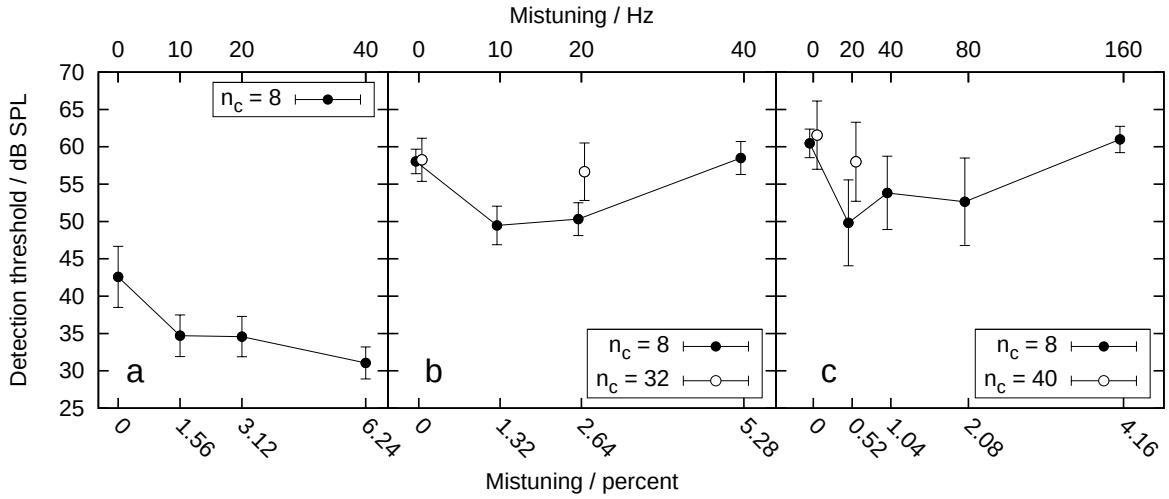

Figure 3.2: Detection thresholds of the single target component in dB SPL as a function of mistuning in percent (bottom axis) or in Hz (top axis). Filled circles indicate thresholds obtained with a masker comprised of eight components (i.e. $n_c = 8$). Open circles show the thresholds obtained with the broadband conditions, with $n_c = 32$ in Experiment 2, and $n_c = 40$ in Experiment 3. Panel a: Thresholds obtained in Experiment 1, with resolved harmonics, with a fundamental frequency F0 = 160 Hz and a target frequency $f_t = 800$ Hz. Panel b: Thresholds obtained in Experiment 2, with unresolved harmonics, with a fundamental frequency F0 = 40 Hz and a target frequency $f_t = 800$ Hz. Panel c: Thresholds obtained in Experiment 3, with unresolved harmonics, with a fundamental frequency F0 = 160 Hz and a target frequency $f_t = 4000$ Hz. The error bars indicate the standard deviation across six normal-hearing subjects.

harmonics were resolved. The results are shown in Figure 3.2a.

With increased mistuning, the masked threshold decreased from 43 dB SPL to 31 dB SPL. Thus, a 12 dB release from masking was found. A repeated-measures analysis of variance (ANOVA) showed a highly significant main effect of mistuning on the detection threshold: $F(3,15) = 30.4$, $p < 0.001$. Post-hoc pairwise comparisons (Bonferroni corrected) indicated that the 0.0% condition was significantly different from the mistuned conditions ($p < 0.001$). The mistuned conditions were not significantly different from each other (assuming $\alpha = 0.001$).

3.3.2 Experiment 2: unresolved harmonics, F0 = 40 Hz

In Experiment 2, F0 was set to 40 Hz, whereas f_t was kept at 800 Hz (20th harmonic of F0 = 40 Hz, see Figure 3.1b). With these settings, the components of the complex were unresolved. The results are shown in Figure 3.2b.

As in Experiment 1, a threshold decrease with increasing mistuning can be observed in the first three conditions with 8 masker harmonics. The 0.0% condition as well as the 5.28% condition yield a threshold of 58 dB SPL. For the 1.32% and 2.64% conditions, thresholds of 50 and 51 dB SPL are found, leading to a maximal difference in masked threshold of 8 dB. A repeated-measures ANOVA showed a highly significant main effect of mistuning on the detection threshold: $F(3,15) = 48.9$, $p < 0.001$. Post-hoc pairwise comparisons (Bonferroni corrected) indicated that the 0.0% condition and the 5.28% condition were significantly different from the other conditions ($p < 0.001$). There was no significant difference between the 1.32% and 2.64% mistuned conditions and no significant difference between the 0.0% and 5.28% condition.

For the broad-band condition, the thresholds for 0.0% and 2.64% mistuning were obtained. In these conditions there was no significant effect of mistuning on the thresholds: $F(1,5) = 2.8$, $p = 0.15$. Comparing the 2.64% conditions with 8 and 32 masker harmonics, it was found that the inclusion of additional harmonics in the masker had a significant effect on the mistuned thresholds: $F(1,5) = 42.59$, $p < 0.01$. The 0.0% thresholds were not significantly influenced by the increase of masker bandwidth: $F(1,5) = 0.05$, $p = 0.84$.

3.3.3 Experiment 3: unresolved harmonics, F0 = 160 Hz

Here, the masker fundamental frequency was again set to F0 = 160 Hz, but the target frequency was f_t = 4 kHz Hz (25^{th} harmonic of F0 = 160 Hz, see Figure 3.1c). In addition to the harmonic masker and the target signal, a continuous white noise, second-order low-pass filtered (butterworth) at 1.5-kHz, was presented throughout the experiment to interfere with low-frequency distortion products that could influence the detection. The noise had an overall level of 40 dB SPL. In Figure 3.2c, the results show a decrease in masked thresholds with increasing mistuning between the first two conditions with 8 masker harmonics. The maximal masking level difference of 10 dB can be found between the 0.0% and 0.52% conditions. The 0.0% and 4.16% conditions yield the same threshold. A repeated-measures ANOVA showed a highly significant main effect of mistuning on the detection threshold: $F(4,20) = 12.4$, $p < 0.001$. Post-hoc pairwise comparisons (Bonferroni corrected) indicated that the 0.0% condition and the 4.16% condition were significantly different from the other conditions ($p < 0.05$). There was no significant difference between the 0.52%, 1.04% and 2.08% mistuned conditions, and no significant difference between the 0.0% and 4.16% condition (assuming $\alpha = 0.05$).

For the broad-band condition, the thresholds for 0% and 0.52% mistuning were obtained. The detection threshold decreased from 61 to 58 dB when a mistuning of 0.52% was applied.

Between the two broad-band conditions, the effect of mistuning was significant: $F(1,5) = 23.7$, $p < 0.01$. Comparing the 0.52% conditions with 8 and 40 masker harmonics, it can be seen that the inclusion of additional harmonics in the masker had a significant effect on the thresholds: $F(1,5) = 16.21$, $p < 0.05$. The thresholds of the 0.0% conditions with 8 and 40 maskers were not significantly influenced by the increase of masker bandwidth: $F(1,5) = 0.77$, $p = 0.42$.

3.4 Discussion

The method of measuring single-component detection thresholds yielded reliable and statistically significant results that are in line with previously published data. A significant effect of mistuning on the masked threshold of the target component has been shown in all three experiments. Comparing the harmonic condition to the condition with the smallest mistuning, the results show masking level differences between 8 and 12 dB. This outcome is in line with Oh and Lutfi (2000) and Klinge et al. (2011). In a comparable condition with a target frequency of 1 kHz and 10 masker components, Oh and Lutfi (2000) measured a masking level difference of about 5 dB between harmonic and mistuned stimuli. Klinge et al. (2011) observed masking level differences of about 7 dB with a target frequency of 1 kHz, and 11 dB with a target frequency of 8 kHz.

In Experiments 2 and 3, where the harmonics were unresolved, detection thresholds of both harmonic and mistuned conditions were increased by up to 18 dB compared to Experiment 1. This is in line with Klinge et al. (2011), where an increase of 11 to 14 dB was observed comparing the thresholds of a 1-kHz target and an 8-kHz target. The increased thresholds occur due to the unresolvability of the target component. Additional energy in the target frequency filter, which is present in the unresolved conditions as multiple components falling into the same cochlear filter, makes it difficult to detect the target. Thus, the inability to perceptually segregate the mistuned harmonic, as investigated in Moore et al. (1986), is reflected in the overall increase of thresholds in Experiments 2 and 3. In Experiment 2, the 0.0% and the 5.28% conditions yield the same thresholds, because a mistuning value of 5.28% creates a harmonic condition, since the frequency of the fourth masker component is shifted from 760 to 800 Hz, coinciding with the target frequency. This is also the case for the 0.0% and 4.16% conditions in Experiment 3. These conditions were introduced to investigate if the additional energy contributed by the shifted fourth component in the ERB around the target frequency had an influence on the thresholds. This was not the case.

A large effect of mistuning can still be found in the unresolved conditions in this study

as well as in Klinge et al. (2011). The persistence of the effect with unresolved harmonics could be caused by signal features apart from mistuning that were used by the subjects to detect the target component. Klinge et al. (2011) hypothesized that the predominant cue in their unresolved condition was a change of the temporal envelope of the stimulus waveform caused by adding the target signal to the complex masker. As their harmonic complexes were generated by adding up pure tones in sine phase, their stimuli had a constant envelope waveform throughout the experiment. In this study, the phases of the components were randomized in each presentation interval, and a large release from masking could still be found. This observation does not rule out the possibility of temporal or spectral envelope cues being exploited, as hypothesized in, e.g., Moore et al. (1986) and Hartmann et al. (1990). However, it is incompatible with a detection mechanism based on a stored temporal template of the stimuli, as discussed by Klinge et al. (2011).

In the broad-band conditions of Experiments 2 and 3, no significant effects of mistuning were observed. Studies on comodulation masking release (e.g., Hall et al., 1990), which have a paradigm similar to this study, have shown that increasing masker bandwidth by adding energy in remote frequency ranges increases the release from masking. The contradicting result in this study could be explained by the additional lower harmonics that are added to the complex to increase bandwidth. As pitch perception is dominated by low-frequency resolved harmonics (e.g., Hartmann et al., 1990), the masker F0 is a too strong cue that disables or hampers the detection of the mistuned target tone. This would be in line with Houtsma and Smurzynski (1990), who found that F0 discrimination performance decreased with increasing lowest harmonic number of a tone complex.

The fact that an increased number of remote components hampers detection of the target component in the mistuned condition points towards the involvement of across-frequency processes. The data cannot be explained by signal changes in the on-frequency channel around the target component. Thus, an on-frequency single-channel auditory model would not be sufficient to predict the decrease of masking release in the broadband conditions.

The obtained masking level differences show that release from masking by mistuning has a magnitude with other effects such as binaural masking level differences in Metz et al. (1968), ranging from 7 to 16 dB for a target frequency of 1 kHz, depending on the bandwidth of the noise masker. For comodulation masking release, Verhey et al. (2003) report masking releases ranging from 3 to 11 dB when comparing an uncorrelated to a comodulated condition, depending on various stimulus statistics.

The overall similarity of magnitude of the investigated mistuning effect and other masking/unmasking effects is promising, as it facilitates combinational experiments (e.g. with

dichotic targets, as in Klinge et al. (2011) where the effects are employed simultaneously to investigate possible interactions (e.g. as in Epp and Verhey, 2009a, where comodulation masking release and binaural unmasking are combined).

3.5 Conclusions

We presented a method to investigate the influence of mistuning on component detection by measuring masking level differences of a single target component in harmonic complexes as a function of mistuning. The results show that the method is able to assess effects in harmonicity and pitch perception and can account for effects of resolvability. The measured detection thresholds allow for quantification of the masking release effect by mistuning and challenge current auditory processing models, in particular because of the observed across-frequency interaction.

Acknowledgments

This study was supported by the DFG (SFB/TRR31 “The Active Auditory System”). We would like to thank the Medical Physics group and Birger Kollmeier for constant support and fruitful discussions.

Chapter 4

Combination of binaural and harmonic masking release effects in the detection of a single component in complex tones

Abstract Harmonic and binaural signal features are relevant for auditory scene analysis, i.e., the segregation and grouping of sound sources in complex acoustic scenes. The way these features are combined in the auditory system, however, is still unclear. This study provides psychophysical data and model simulations to evaluate three possible combinations of auditory processing schemes suggested in literature. Detection thresholds for an 800-Hz tone masked by a diotic harmonic complex tone (fundamental frequency: 160 or 40 Hz) were measured in 6 normal-hearing subjects in resolved or unresolved conditions. The target tone was presented diotically or with an interaural phase difference (IPD) of 180° and in a harmonic or "mistuned" relationship to the diotic masker. Both mistuning and IPD provided release from masking in a non-additive way. A single-channel auditory model with different binaural processing schemes was used to predict the unresolved conditions. Experimental and model results hint at a parallel processing scheme with a binaural processor that has limited access to modulation information. The predictions of the monaural processor were in line with the experimental results and literature data. The modeling results form a basis for a subsequent investigation and modeling of combinatory effects in resolved harmonic complexes that require across-frequency processing.

A modified version of this chapter was submitted as "Combination of binaural and harmonic masking release effects in the detection of a single component in complex tones", M. Klein-Hennig, M. Dietz, V. Hohmann, to PLOS ONE on January 20, 2015.

4.1 Introduction

Auditory scene analysis (ASA) allows humans to detect, identify and track sound sources (e.g., talkers) in complex acoustic environments (Bregman, 1994). According to Bregman (1994), ASA partly relies on the grouping of auditory information that likely stems from the same sound source into single auditory objects. Important signal features for auditory grouping are binaural and harmonic features (e.g. Hukin and Darwin, 1995b; Darwin and Hukin, 1999). The binaural information (interaural time differences, ITD and interaural level differences, ILD) allow the azimuthal localization of a sound source. In turn, sound source location is an auditory grouping cue. Darwin and Hukin (1999) found that small differences in ITDs alone can be used to separate words in the absence of talker or fundamental frequency (F0) differences. Harmonicity is also a strong grouping cue that fuses individual components of a complex tone or speech formants with a common fundamental frequency (F0) into a single auditory object (e.g., Moore et al., 1985; Hukin and Darwin, 1995a). The human auditory system is sensitive to "mistuning", i.e., deviations from the harmonic frequency relationship between complex components. Moore et al. (1986) and Hartmann et al. (1990) have shown that deviations between 0.5-4% of the fundamental frequency can be detected and lead to the perception of a mistuned component as second auditory object in addition to the complex tone.

For a speech signal embedded in realistic acoustic environments, harmonicity and binaural features of the talker co-vary with its spatial position and with those of interfering signals. As a consequence, several computational models of ASA (e.g., Kepesi et al., 2007; Ma et al., 2007) assume combined processing of the two features for optimal information processing. This study aims at further investigating these combination mechanisms by contributing psychophysical data and auditory model simulations on the detection of single sinusoidal components masked by a harmonic complex tone masker. In particular, complex tones with unresolved harmonics were used. In this case, harmonicity information is mainly coded by temporal periodic amplitude modulation of the envelope in auditory frequency bands, which is known to be exploited for pitch analysis and mistuning detection (Moore et al., 1985; Hartmann et al., 1990; Lee and Green, 1994).

Several studies address auditory processing of combined binaural and harmonicity information: McDonald and Alain (2005) measured event-related brain potentials (ERPs) for harmonic and mistuned target tones in a ten-tone complex, while presenting the target tone either on the same or on a different loudspeaker than the masker (i.e., the other nine harmonics). Their behavioral and electrophysiological data showed that both harmonicity and

location are evaluated to separate sounds and that localization cues can be used to resolve ambiguity in harmonicity cues. They found some evidence that harmonicity-based segregation of sound sources occurred during active and passive listening, whereas location effects were only observed during active listening. Based on this finding, they conclude that the evaluation of localization information is more reliant on active top-down processing than harmonicity processing. Their setup, however, did not allow any conclusions as to which underlying binaural features and processing schemes are responsible for their findings, as the stimuli at ear-level are difficult to control in a free-field experiment. It is not clear if subjects used spectral localization cues, interaural time differences (ITDs), or interaural level differences (ILDs) for detection. Given the target tone frequency of 600 Hz, the dominant cue is most likely the ITD. Their findings therefore hint towards ITD being an additional grouping cue when offering location information to a complex tone (e.g. speech) that is already grouped by harmonicity. This is in contradiction to Culling and Summerfield (1995), who found that ITD is a weak grouping cue for simultaneous grouping in, e.g., formant-like noise bands, which suggests that additional localization cues may have been involved in the effects observed by McDonald and Alain (2005). Klinge et al. (2011) also investigated the combined influence of binaural and harmonic signal features in the free field by measuring detection thresholds of a sinusoidal target component in a harmonic complex-tone masker. The target component was either in a harmonic or in a mistuned relationship to the masker with fundamental frequency F0 and could additionally be presented on a separate loudspeaker located at 90° azimuth. They found an addition of the two masking release effects: Both mistuning and spatial separation of the target decreased its detection threshold. As in McDonald and Alain (2005), however, the free-field setup employed by Klinge et al. (2011) does not exclude the exploitation of spectral or level-based localization information. Furthermore, they presented all harmonic components in constant sine phase, which could have lead to subjects using template matching to identify the target interval based on its envelope shape. A diotic experiment by Klein-Hennig et al. (2012) with the same adaptive procedure and similar stimuli, but random phases in each interval, found slightly smaller masking release by mistuning. Further evidence regarding the joint processing of binaural and harmonic features was provided by Krumbholz et al. (2009), who used headphone experiments that allowed for a strict control of interaural parameters and could thus yield more precise findings on ITD processing. They found that subjects had difficulties to perform musical interval recognition (MIR) tasks with binaurally unmasked complex tones. With increasing fundamental frequency, they found decreasing MIR performance. Since their complex tones were unresolved, the authors suggested that the temporal envelope fluctuations that would

convey the required periodicity information were not accessible to the auditory system in binaurally unmasked conditions. According to Krumbholz et al. (2009), this hints at a processing scheme in which binaural processing precedes pitch processing, with a temporal integration step in between that leads to the observed MIR performance decrease.

Further studies directly investigated the relation between amplitude modulation in frequency subbands and binaural processing. Epp and Verhey (2009b) studied the combination of comodulation masking release (CMR, e.g., Hall et al., 1984, 1990) and binaural masking level differences (BMLD) in headphone experiments. Here, only interaural timing disparities were available as binaural cues. Epp and Verhey (2009b) found a linear addition of the two masking releases. They offer no conclusion regarding the processing order of envelope and binaural processing, but their results indicate a serial, rather than parallel processing scheme. Nitschmann and Verhey (2012) measured BMLDs as a function of frequency separation between masker and the pure tone target signal. They found that BMLDs decrease with increasing spectral distance between masker and target and state that the observed decrease in masking release could be caused by modulation information not being available to the binaural system, but to the monaural system only. Thompson and Dau (2008) found that modulation filters in binaural processing are probably broader than monaural modulation filters proposed in Dau et al. (1996a). They state that such broader filters could be employed either before or after binaural processing. Thus, the order of modulation and binaural processing remains unclear.

In summary, the results of the afore-mentioned studies on combined processing of binaural and temporal periodicity and modulation information led to model hypotheses that are partly inconsistent. Three general hypotheses can be extracted:

- Binaural processing precedes periodicity processing (Krumbholz et al., 2009),
- Periodicity processing precedes binaural processing (McDonald and Alain, 2005),
- Both processing stages work in parallel, and the binaural stage has no or reduced modulation selectivity compared to the monaural stage (Nitschmann and Verhey, 2012).

To evaluate these hypotheses, this study investigated the combined influence of harmonic and binaural signal features by psychophysically measuring detection thresholds of a single 800-Hz target tone in a resolved (fundamental frequency F0 = 160 Hz) or unresolved (F0 = 40 Hz) tone-complex masker. The target tone could either be harmonic or mistuned in relation to the masker, and was presented diotically (M0S0) or with an interaural phase difference of 180° (M0Sπ). For full control of the binaural stimulus parameters at ear-level, the measurements were performed with headphones. To gain information about the

order and type of processing of harmonic and binaural features, predictions from a binaural auditory model were compared to the psychophysical results.

4.2 Methods

4.2.1 Subjects

Six normal-hearing listeners (4 male, 2 female) aged 24 to 32 years participated in the experiments. Before data acquisition, all subjects took part in one hour of training with the same stimuli as used in the final experiment. Five of the subjects received compensation on an hourly basis for their participation. The other subject was an author of this study (MK). The experiments were approved by the ethics committee of the University of Oldenburg.

4.2.2 Apparatus and stimuli

The experiments were conducted in a double-walled, sound-attenuating booth. The stimuli were generated digitally with a sampling frequency of f_s = 48 kHz at runtime, on a PC using MATLAB. After conversion to analog signals by an external RME ADI-8 PRO D/A converter connected to a 24-bit RME DIGI96/8 PAD sound card, the stimuli were presented via Sennheiser HD 650 headphones. A Tucker Davis HB7 headphone buffer was used to drive the headphones. The subjects gave responses using a computer keyboard or mouse. Visual feedback was given on a computer monitor.

The stimulus configuration was identical to Klein-Hennig et al. (2012) and is illustrated in Figure 4.1. The target tone consisted of a pure-tone with a frequency f_t of 800 Hz. The masker was a harmonic complex consisting of four harmonics below and above the target tone. In Experiment 1, the masker had a fundamental frequency F0 of 160 Hz, such that the components had frequencies of 160, 320, 480, 640, 960, 1120, 1280 and 1440 Hz. In Experiment 2, the masker F0 was 40 Hz, leading to component frequencies of 640, 680, 720, 760, 840, 880, 920 and 960 Hz. Experiment 3 used the same configuration as in Experiment 2, with 12 additional masker harmonics below and above the target component frequency, leading to a broad-band masker with 32 components ranging from 160 to 1440 Hz. The tones were added up in the temporal domain, with random phases identical on both channels (diotic, M0S0). To achieve binaural masking release, a dichotic M0Sπ was created by applying an additional phase shift of 180° to the target tone on the right ear channel prior to the addition of the masker. Mistuning was applied by increasing the fundamental frequency F0 of the masker by a certain percentage, while keeping the target frequency f_t

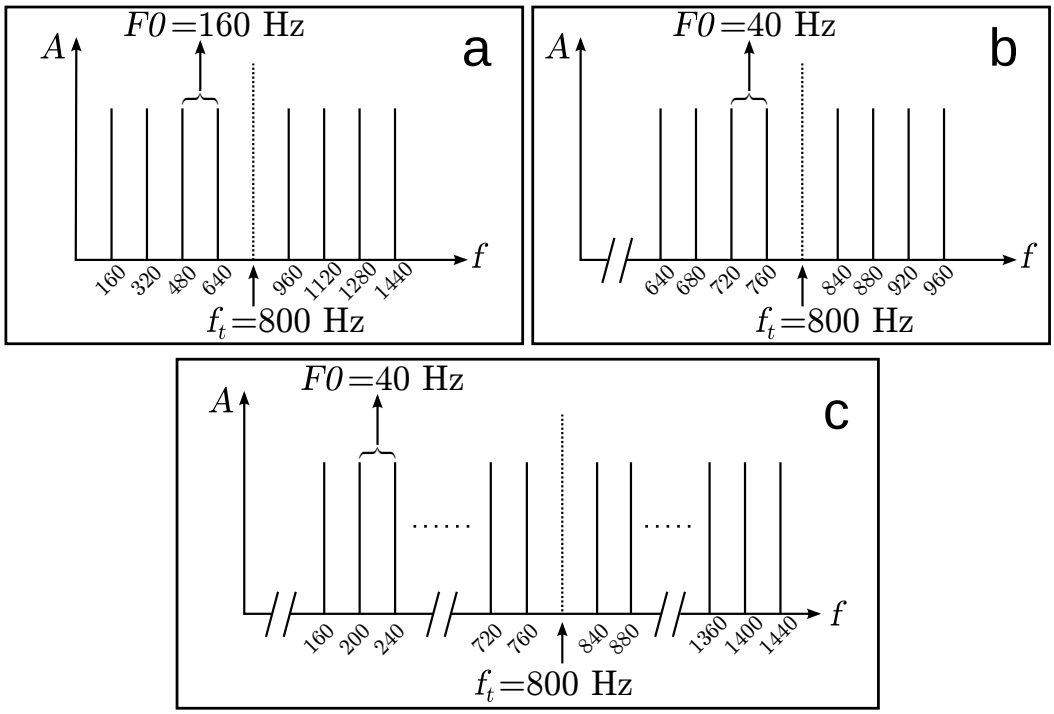

Figure 4.1: Pictograms of the frequency configurations for the stimuli used in the experiments. The masker harmonic complex is indicated by black lines, its respective fundamental frequency F0 is indicated in each panel. The dotted line represents the frequency f_t of the target tone that had to be detected. Panels a and b indicate the configurations for Experiments 1 and 2. Panel c shows the configuration of the broad-band stimuli with additional masker harmonics used in Experiment 3.

constant. The mistuning values were 3.12% in Experiment 1 and 2.64% in Experiments 2 and 3. These values were chosen because they produced the largest masking releases in Klein-Hennig et al. (2012). The percentages lead to an upward shift of 20 Hz of the 4th masker component, which is the one below the target component. In contrast to Klein-Hennig et al. (2012), all thresholds were measured while presenting a continuous, binaurally uncorrelated 380-Hz low-pass noise at a level of 45 dB SPL to prevent the possible influence of cochlear distortion products by pure-tone harmonics (e.g., Goldstein, 1967; Pressnitzer and Patterson, 2001).

The stimuli had a duration of 400 ms, including on- and off-gating using 25 ms Hanning windows.

4.2.3 Procedure

The detection thresholds were determined using an adaptive 3-interval, 2-alternative forced-choice procedure. A 1-up, 2-down tracking rule was used, estimating the 70.7%-correct point on the psychometric function (Levitt, 1971). The first interval always contained a reference stimulus and could not be selected as a response. The test subject had to indicate which interval contained the target tone. As in Klein-Hennig et al. (2012), the target signal was initially presented at a level of 65 dB SPL and was varied in 5 dB steps. The step size was reduced to 2 dB after the second and 1 dB after the fourth reversal.

After eight reversals with 1dB steps, each adaptive run was terminated. The individual mean thresholds were calculated by arithmetically averaging over the last eight reversals of five experimental runs. The mean thresholds reported in the results are arithmetic averages over the individual averages of all subjects.

4.2.4 Models

This section describes the single-channel model used for simulating the unresolved conditions from Experiments 2 and 3. Experiment 1 was excluded from modeling, as all masker components were outside the critical bandwidth of the on-target auditory filter. The development and evaluation of a multi-channel model with across-frequency processing to account for this resolved condition is beyond the scope of this study.

Peripheral processing

To simulate auditory processing in the cochlea, a single-channel auditory preprocessing stage was implemented, similar to Dietz et al. (2009). It consisted of

- Real-valued 4th-order gammatone filtering at the target frequency f_t = 800 Hz, with a bandwidth of one auditory filter (one ERB),
- Half-wave rectification,
- 5th-order 770-Hz lowpass filtering (Breebaart et al., 2001a).

Modulation and binaural processing stages

The modulation processing stage used in the model is based on the envelope power spectrum model (EPSM) by Ewert and Dau (2000). It is used to extract target-related envelope information and attenuate masker-related modulation components, leading to a larger signal-to-

masker ratio and thus facilitating detection. To extract the envelope of a stimulus, the time signal after peripheral processing was sampled down to a sampling frequency of 2000 Hz. Then, a single modulation filter (Dau et al., 1997) was used for envelope processing. The modulation filter frequency was set to the fundamental frequency F0 = 40 Hz in harmonic conditions and to 20 Hz in the mistuned conditions. This value corresponds to the upward frequency shift of the masker component below the target component, which generates a 20-Hz beating between both components that should be visible in the modulation spectrum and could be a detection cue. The 20-Hz modulation filter extracts this cue at the maximum SNR.

The model was divided into a monaural and a binaural pathway. For monaural processing, the absolute left and right ear output signals after modulation processing were combined into a single output signal by addition. The binaural processing stage is a simplification of the equalization-cancellation model as proposed by Durlach (1963) and calculates the absolute value of the difference between the left and right input signals, which depend on the processing order (see Section 4.2.4 for the different versions of the binaural pathway).

Decision stage

Depending on the stimulus condition (diotic or dichotic), the target and reference intervals were processed by the monaural or the binaural pathway. As a decision variable, the squared output signal of the employed pathway was integrated across the central 50% steady-state part of the interval, yielding a single value for each target and reference interval that establishes the decision variable and corresponds to its total energy E_{total}. Passing through the same adaptive procedure as the test subjects, the model chose the interval with the largest E_{total}. This approach establishes a classic signal detection approach that employed the respective most sensitive pathway for detection in the different configurations. Because experimental dichotic thresholds were at least 3 dB better than the diotic thresholds (see Section 4.3), a combination of information across monaural and binaural pathways, e.g., via a combination of sensitivity indices d', is not indicated. To limit detection accuracy, Gaussian noise was added to the decision variable in both pathways. The standard deviation was set to match the harmonic detection thresholds observed in the human data. This resulted in a fixed standard deviation of $\sigma_m = 0.01$ for the monaural pathway in all experiments. The standard deviation σ_b for the binaural pathway was also the same across experiments, but had to be adapted to the order of processing in each configuration of the binaural pathway. Respective values are given in Section 4.4 (model results).

Model configurations

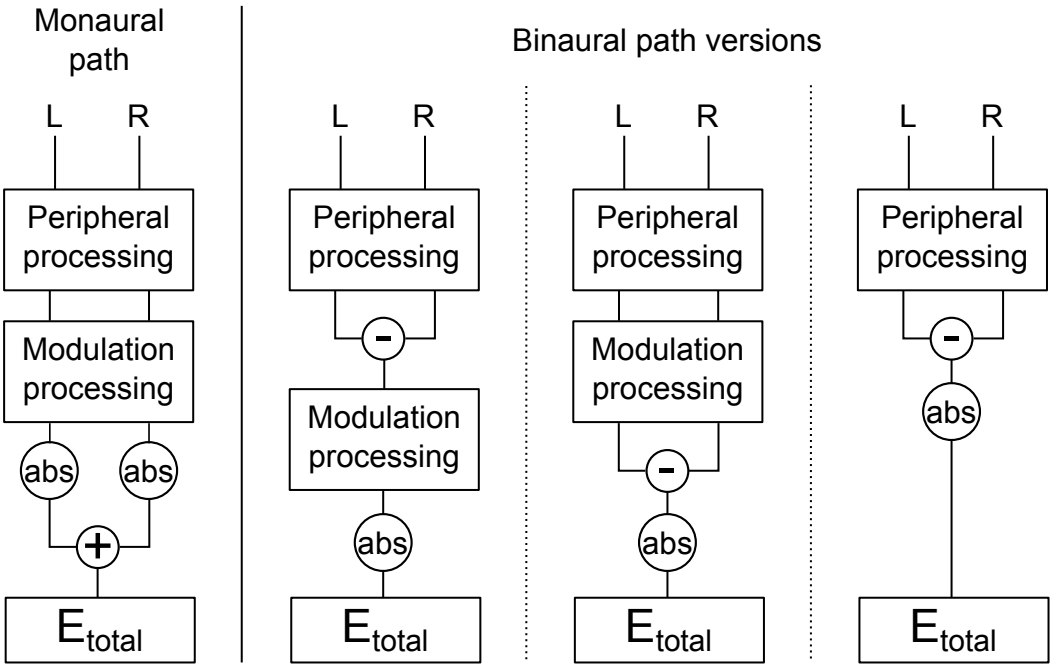

Figure 4.2: Schemes of the monaural pathway (leftmost panel) and the three tested binaural pathway configurations, employing different processing order. First panel: Monaural pathway employing modulation processing to predict diotic thresholds. Second panel: Binaural processing precedes modulation processing. Third panel: Modulation processing precedes binaural processing. Fourth panel: Binaural processing has no access to the output of the modulation processor. The peripheral processing as well as the calculation of the integrated power E_{total} of the respective output signal is identical in all configurations.

Figure 4.2 illustrates the model configurations used to predict the data. The left panel shows the monaural processing pathway which is employed to predict the diotic data. This pathway was identical in the three model configurations that were tested. The model configurations differ in the order of binaural and modulation processors in the binaural processing pathway. In the first model configuration (second panel of Figure 4.2), binaural processing precedes modulation processing, hence modulation information is extracted from the output of the binaural processor. The second configuration (third panel of Figure 4.2) reverses this order: a common modulation stage processes left and right ear signals, followed by a binaural processor that works on the output of the modulation stage. The third model configuration (fourth panel of Figure 4.2) does not include modulation processing, simulating a binaural pathway without access to modulation information. To reduce variation in the model predictions, random phase angles for the complex tone components were chosen and

kept constant throughout the model run. As mentioned above, component phase randomization was introduced to prevent detection based on envelope templates in the subjects. Since the models do not include methods for template building and matching, they do not gain a detection advantage over the subjects by using constant component phases.

4.3 Experimental results

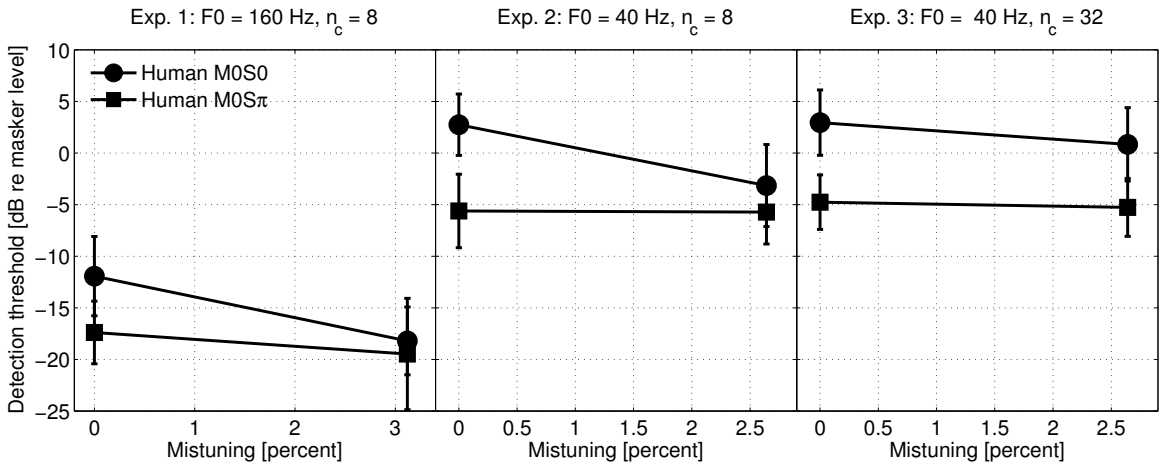

Figure 4.3: Detection thresholds for Experiments 1 to 3. Data points and error bars show mean and standard deviation across 6 subjects. Results for a diotic target tone (M0S0) are plotted with circles. Squares indicate thresholds for a target tone with an interaural phase difference of 180° (M0Sπ). The fundamental frequency F0 and total number of masker components n_c of the stimuli is indicated at the top of the panels.

4.3.1 Experiment 1: F0 = 160 Hz, f_t = 800 Hz

The results of Experiment 1 are shown in the left panel of Figure 4.3. Experiment 1 used a fundamental frequency F0 of 160 Hz and a target frequency f_t of 800 Hz (5th harmonic of F0=160 Hz). The number of masker components was n_c = 8. In the diotic M0S0 condition, a threshold decrease by 6.3 dB can be observed comparing the harmonic and the 3.12% mistuning condition. Presenting the target dichotically with an IPD of 180° lowered the thresholds from -12 to -17 dB in the harmonic (0% mistuning) condition, which corresponds to a BMLD of 5.5 dB. With a mistuning of 3.12%, the BMLD decreased from 5.5 to 0.7 dB. There was no significant difference between the dichotic 0% and 3.12% conditions: a repeated-measures analysis of variance (ANOVA) showed a significant main effect of

mistuning on the thresholds in the diotic case ($p < 0.05$, F(1,5) = 17.2), but not in the dichotic case ($p = 0.23$, F(1,5) = 1.7). A significant main effect of IPD in the harmonic case was found ($p < 0.05$, F(1,5) = 14.7). In the mistuned case there was no significant effect of IPD ($p < 0.47$, F(1,5) = 0.6). In a two-way repeated measures ANOVA, no significant interaction of mistuning and IPD was found ($p = 0.073$, F(1,5) = 3.579).

4.3.2 Experiment 2: F0 = 40 Hz, f_t = 800 Hz

In this experiment, F0 was set to 40 Hz while keeping the target frequency f_t at 800 Hz (20th harmonic of F0 = 40 Hz), to generate unresolved harmonics in the auditory filter centered around f_t. As in Experiment 1, the number of masker components was n_c = 8. The results are shown in the center panel of Figure 4.3. In the diotic condition, mistuning led to a decrease in detection threshold by 5.8 dB. In the harmonic 0% condition, a BMLD of 7 dB was found. At 2.64% mistuning, the BMLD decreased to 2 dB. Repeated-measures ANOVA showed a highly significant main effect of mistuning on the thresholds in the diotic case ($p < 0.001$, F(1,5) = 14.9), but not in the dichotic case ($p = 0.96$, F(1,5) = 0.003). Regarding the influence of the target IPD, a highly significant main effect of IPD in the harmonic case ($p < 0.001$, F(1,5) = 25.3) was found. In the mistuned case, no significant main effect of IPD was observed ($p = 0.17$, F(1,5) = 2.3). A two-way repeated measures ANOVA showed a significant interaction of mistuning and IPD ($p < 0.05$, F(1,5) = 5.9).

4.3.3 Experiment 3: F0 = 40 Hz, f_t = 800 Hz, broadband

This experiment used the same frequency configuration as Experiment 2 (F0 = 40 Hz, f_t=800 Hz), but increased the number of masker components to n_c = 32. This way, the stimuli had the same target-frequency-to-bandwidth-ratio as in Experiment 1. The results are shown in the right panel of Figure 4.3. In the diotic conditions, mistuning decreased the thresholds by 2.1 dB. In the dichotic conditions, a threshold decrease of 0.5 dB in the mistuned condition can be observed. Thus, the BMLD was 7 dB in the harmonic condition and 5 dB in the mistuned condition. No significant effect of mistuning could be found in both the diotic case ($p = 0.22$, F(1,5) = 1.8) and the dichotic case ($p = 0.82$, F(1,5) = 0.1). In the harmonic case, a highly significant effect of IPD was found ($p < 0.001$, F(1,5) = 35.2). In the mistuned case, a significant effect of IPD could be observed ($p < 0.05$, F(1,5) = 16.106). A two-way repeated measures ANOVA showed no significant interaction of IPD and mistuning ($p = 0.39$, F(1,5) = 0.771).

4.4 Model results

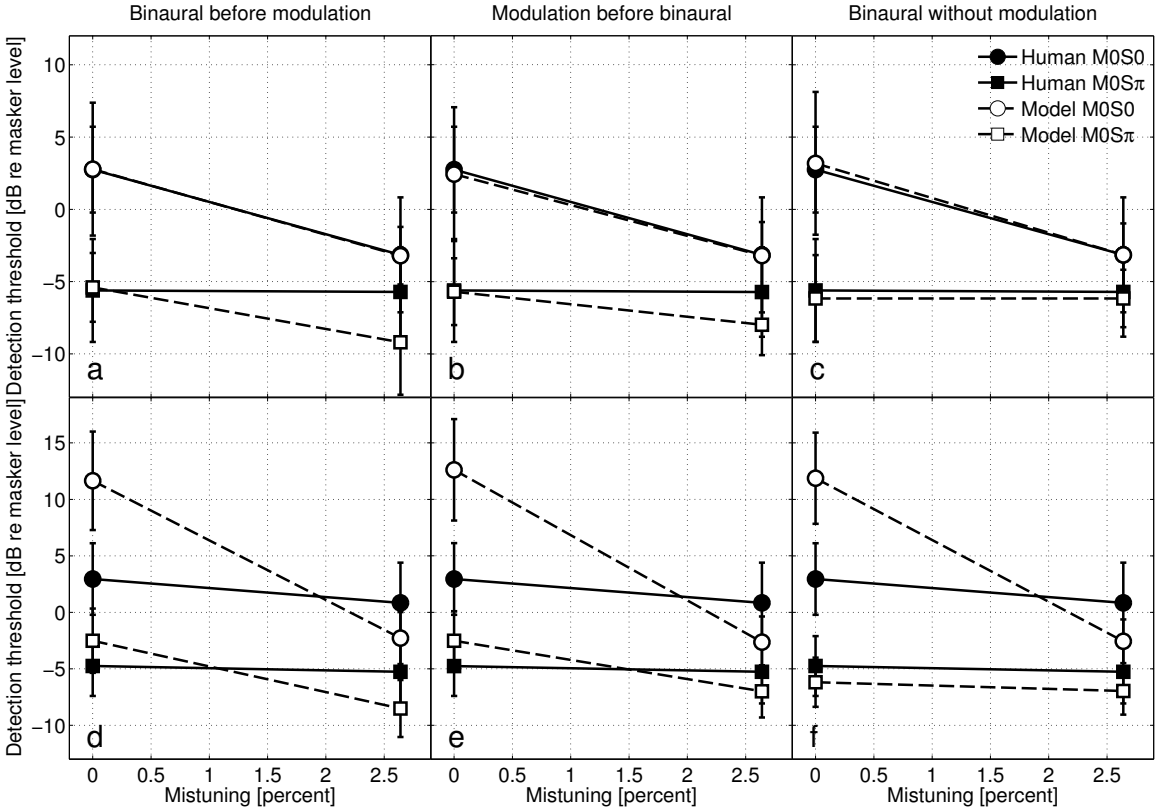

Figure 4.4: Human (filled symbols) and model thresholds (open symbols) for Experiment 2 (top row, panels a-c) and Experiment 3 (bottom row, panels d-f) for the three tested model configurations. Panels a and d: detection thresholds for a model in which binaural processing precedes modulation processing. Panels b and e: detection thresholds for a model configuration where modulation processing precedes binaural processing. Panels c and f: detection thresholds for a model where the binaural processor has no access to modulation information. Plotting conventions are as in Figure 4.3.

4.4.1 Binaural processing before modulation processing

In the processing chain of this model, the binaural equalization-cancellation stage processes left and right ear signals. The output is then analyzed by a modulation filter, as described in Section 4.2.4 (see 2nd panel of Figure 4.2). The standard deviation of the internal noise was $\sigma_b = 0.04$. The predictions of this model configuration are shown in the left panels (a and d) of Figure 4.4. For Experiment 2 (panel a), the model shows a sensitivity to mistuning in

the dichotic conditions, predicting a threshold decrease of 4.5 dB that is not observed in the human data. The diotic predictions are in line with the human data. In Experiment 3 (panel d), the model predictions for the diotic and dichotic harmonic conditions are by 10 dB and 3 dB too high, while the mistuned thresholds are by 5 dB too low compared to the human data.

Please note that the same model was used in all diotic conditions. Predictions slightly vary because models runs were repeated for each condition.

4.4.2 Binaural processing after modulation processing

This model configuration performs modulation processing after auditory preprocessing. The binaural difference is calculated on the output of the modulation stage (see 3rd panel of Figure 4.2). The standard deviation of the internal noise was $\sigma_b = 0.0076$. The predictions are shown in the center panels (b and e) of Figure 4.4. For the dichotic condition of Experiment 2 (panel b), the model predicts a mistuning effect of 3 dB on the thresholds, which was not observed in the human data. The diotic predictions are in line with the human data. For Experiment 3 (panel e), the model predicts higher harmonic thresholds for both IPD configurations, deviating by up to 10 dB from the human data. The predictions for the mistuned conditions are 2 to 5 dB lower than the human data.

4.4.3 Parallel processing

This model does not include a modulation filter in the binaural pathway (see 4th panel of Figure 4.2). The binaural processor works directly on the preprocessed signals. The model thresholds are shown in the right panels (c and f) of Figure 4.4. The standard deviation of the internal noise was $\sigma_b = 0.0161$. The model predicts no threshold decrease with mistuning for Experiment 2 in the dichotic condition (panel c), i.e., both the diotic and dichotic predictions are in line with the human data. For Experiment 3 (panel f), the diotic harmonic threshold is 10 dB larger than the experimental threshold, while the diotic mistuned threshold is 5 dB lower than observed in the human results. The dichotic model predictions deviate less than 1 dB from the human data, showing no difference between harmonic and mistuned conditions.

4.5 Discussion

4.5.1 Psychophysical results

In summary, the results show that both harmonic and binaural manipulation of the stimuli affected the detection thresholds.

The observed diotic thresholds in the resolved case (Experiment 1) are in line with Klinge et al. (2011) and Klein-Hennig et al. (2012), who found a mistuning effect of the same magnitude using similar methods and similar frequency configurations (identical in the case of Klein-Hennig et al., 2012). The diotic thresholds found for unresolved stimuli in Experiments 2 and 3 are in line with Klein-Hennig et al. (2012). Given that the thresholds in Klein-Hennig et al. (2012) were obtained without using a low-frequency masking noise as described in Section 4.2.2, cochlear distortion products caused by the pure-tone components appear to have little influence on the detection thresholds.

In all harmonic conditions, a significant BMLD of 5 to 7 dB was found. The resolved stimuli of Experiment 1 yielded a BMLD of 5 dB. The stimuli are comparable (with regard to resolvability and frequency range) to the 1000-Hz target-frequency conditions in Klinge et al. (2011), which yielded a BMLD of 5 dB as well. The unresolved conditions of Experiments 2 and 3 generated BMLDs of 7 dB, 2 dB more than the resolved condition in Experiment 1. With regard to resolvability, the results can be compared to the 8-kHz harmonic condition of Klinge et al. (2011), as a target frequency of 8 kHz and a fundamental frequency of 200 Hz leads to unresolved harmonics as well. Klinge et al. (2011), however, found a BMLD of 15 dB in that configuration, which is not in line with the results from Experiments 2 and 3. Since the stimuli in Klinge et al. (2011) were presented via loudspeakers, and the exploitation of binaural timing disparities at 8 kHz is highly limited, interaural level differences and spectral cues could have led to this large BMLD.

In the dichotic conditions of all experiments, no significant effect of mistuning could be observed. This is not in line with Klinge et al. (2011), who still found a significant effect of mistuning in "separated" stimulus conditions where the target tone was presented from a loudspeaker located at 90° azimuth and the masker from a loudspeaker in front of the test subjects.

As in Klein-Hennig et al. (2012), broad-band stimuli generated by additional masking components decreased the effect of mistuning to such an extent that no significant effect of mistuning could be observed. As the additional components are added outside of the auditory filter centered around the target frequency $f_t = 800$ Hz, this hints at the involvement of across-frequency processing. In the mistuned broad-band conditions, the same BMLD is

found as in the harmonic condition, the only threshold difference observed was in the diotic mistuned condition. This could mean that the modulation processor uses across-frequency information, not available to the binaural processor, and would be in favor of a processing scheme as proposed by Nitschmann and Verhey (2012), where the binaural pathway has only limited access to the periodicity information required for a threshold decrease by mistuning.

4.5.2 Model results

Overall, the energy-based detection approach predicted thresholds in the correct range for Experiment 2. The modulation filter in the monaural pathway of all model configurations was successful in achieving a masking release by mistuning in the diotic stimulus conditions similar to the human data. Although all models performed 100 adaptive runs for every experimental condition, the data have large standard deviations comparable to the psychophysical results, where only 5 of 6 adaptive runs were recorded and analyzed. The variations in model predictions are caused by the internal noise added to the internal representations of the models (see Section 4.2.4). The noise was necessary to limit detection accuracy, as otherwise the target interval energy would always have been larger than the reference interval energy, leading to unrealistically high detection performance of the models. Due to the random noise and randomized condition presentation order, the diotic model predictions differ by < 0.5 dB between the three model configurations.

The model configuration in which binaural processing was performed before modulation processing predicted a 5.5 dB threshold decrease in the dichotic stimulus conditions. This result would be in line with a linear addition of both effects as found for comodulation masking release (CMR) and BMLD (Epp and Verhey, 2009b). However, this behavior was not observed in the psychophysical results.

In the same way, the model configuration employing binaural processing after modulation processing also showed mistuning sensitivity in the dichotic data. In this case, the mistuned threshold is 3 dB smaller than the harmonic threshold.

The model configuration without modulation processing in the binaural path provided the best prediction of the psychophysical results, showing an effect of mistuning in the diotic but not in the dichotic case. The model represents a processing scheme in which the binaural processor has none or only limited access to modulation information, as proposed by Nitschmann and Verhey (2012).

In the broad-band Experiment 3, all models show higher thresholds for the harmonic conditions and lower thresholds for the mistuned conditions and thus a larger mistuning effect than in the narrow-band Experiment 2. The only exception is the dichotic prediction

of the parallel processing model (see panel f of Figure 4.4). Thus, most predictions are not in line with the psychophysical results. The deviations are likely caused by the increased bandwidth of the stimulus, leading to an increased amount of masker energy in the on-target gammatone filter employed in auditory preprocessing (see Section 4.2.4) and thus to increased thresholds. In the mistuned conditions, the 20-Hz modulation filter creates a more favorable signal-to-masker ratio for detection, leading to lower thresholds. This is not possible in the harmonic conditions, as only the F0 beat of 40-Hz occurs here, which is found in the target as well as the reference intervals. Hence, the models can not predict the vanishing mistuning effect with increased bandwidth. To achieve correct detections in the broad-band experiment, additional auditory filters for across-frequency processing would be needed. Epp and Verhey (2009a) successfully employed across-frequency processing for a model predicting the combined effects of CMR and BMLD, which could be a promising approach for the combination of mistuning and BMLD.

4.6 Conclusions

- The psychophysical data show that release from masking by mistuning and binaural disparity do not combine in a linear, additive way both in the resolved and unresolved conditions.

- Both human data and model results accentuate the need for across-frequency processing in configurations where the target tone is resolved or where the harmonic masker complex is broadband, i.e., when a large number of harmonic masker components lie outside of the passband of the on-target auditory filter.

- The single-channel model results for the unresolved narrowband conditions show that modulation processing is able to account for release from masking by mistuning.

- Regarding the combination of binaural and modulation information required for detection, the single-channel model results show that a model where the binaural processor has only limited access to modulation information was successful at predicting the data of the unresolved conditions.

Acknowledgments

This study was supported by the DFG (SFB/TRR31 “The Active Auditory System”) and by the European Union under the Advancing Binaural Cochlear Implant Technology (ABCIT) grant agreement (No. 304912). We thank the Medical Physics group and Birger Kollmeier for constant support and Astrid Klinge-Strahl, Stephan Ewert, and Georg Klump for fruitful discussions.

Chapter 5

General conclusions

The main goal of this thesis was to provide insights into the processing of binaural, periodic signals through psychophysical experiments and auditory-motivated computer models.

Chapter 2 reported the results of a psychophysical study on the influence of different segments of a periodic envelope waveform on the sensitivity to interaural time differences. The main findings were that the attack flank at the beginning of an envelope cycle and the pause time between two cycles were the most influential envelope parameters. Previous results from Bernstein and Trahiotis (2002) could be reproduced and explained in terms of the investigated envelope segments. The normalized cross-coefficient (NCC) model (e.g., Bernstein and Trahiotis, 1996, 2002) that was previously successfully employed to model ITD sensitivity could not predict parts of the obtained human data. The model predictions were improved by extending the preprocessing of the model with adaption loops (Dau et al., 1996a) that simulate neuronal stimulus adaptation. Shortly after the publication of the results presented here, another study using similar stimulus paradigms was published (Laback et al., 2011), also reporting a strong influence of pause time and rising slope in their results. The pause time results of both studies were shown to be equivalent in a comparison study (Dietz et al., 2013b). Monaghan et al. (2013) compared their results of the influence of attack time in similar stimuli to those reported in Chapter 2, finding them to be in line. Francart et al. (2012) used the extended NCC model with adaptation loops, and were able to qualitatively account for their data. Recently, I developed and conducted a psychophysical experiment that measured the influence of selected envelope segments on the extent of laterality (Dietz et al., 2015). In the study, the subject adjusted the interaural level difference of a pointer signal to match the lateralization achieved by the ITD of a target tone. Varying their durations, the influence of several envelope parameters as in Chapter 2 on lateralization was tested. The results showed a strong influence of attack and pause parameters, even outside of the physiological ITD range, which is relevant for ITD enhancement in binaural cochlear

implant processors. In an imaging study, Dietz et al. (2013a) found that the attack segment in an amplitude-modulated signal might be the region in which the auditory system is particularly sensitive to interaural timing differences, leading to reliable ITD "glimpes" even in reverberant conditions. In conclusion, the results from study 2 identify the most important localization-relevant parts of a periodically amplitude modulated signal. Binaural hearing aid processing schemes could benefit from emphasized processing of these segments. Further experiments involving hearing-impaired subjects could allow for a better understanding of individual, binaural hearing loss.

The study reported in Chapter 3 established the method of measuring detection thresholds as a means of investigating harmonicity-related effects. The obtained results were in line with mistuning studies employing different experimental paradigms (e.g., Moore et al., 1985; Hartmann et al., 1990), and a study using the same method in the free field (Klinge et al., 2011). With its results, the study prepared the ground for an experimental method that is easy to perform for subjects and computer models, allowing for a follow-up study that investigated the combination of mistuning with temporal interaural cues, reported in Chapter 4.

Chapter 4 gave insight into the combined processing of harmonicity and binaural information. The psychophysical results showed that masking releases by harmonicity and interaural phase differences do not combine in linearly additive way. The results of a single-channel detection model with three different processing strategies validated one of three hypotheses on harmonicity-IPD combination (based on McDonald and Alain, 2005; Krumbholz et al., 2009; Nitschmann and Verhey, 2012), indicating that the binaural processor has little or no access to periodicity information available to monaural processing channels. However, the single-channel model predictions for broadband stimuli were not in line with the human data, emphasizing the need of further investigation into across-frequency processing for harmonicity-IPD combination. The outcome of this study provides evidence on the processing order of binaural and periodic harmonicity cues, the role of modulation cues in harmonicity perception and the necessity of across-frequency processing for auditory models.

In summary, this thesis identified the important portions that dominate time-difference based localization of periodic signals and clarified the processing order of joint harmonicity and binaural cues in auditory scene analysis. With these results, this work constitutes a significant contribution to future processing schemes in binaural hearing devices, as well as the understanding of cue combination in auditory scene analysis, providing valuable knowledge for the creation of auditorily realistic computational ASA models.

Bibliography

Akeroyd, M. A. and Patterson, R. D. (1997). A comparison of detection and discrimination of temporal asymmetry in amplitude modulation. *The Journal of the Acoustical Society of America*, 101(1):430–439.

Akeroyd, M. A. and Summerfield, A. Q. (1999). A binaural analog of gap detection. *The Journal of the Acoustical Society of America*, 105(5):2807–2820.

American National Standards Institute (1994). *American National Standard: Acoustical Terminology*.

Bernstein, L. R. and Trahiotis, C. (1985). Lateralization of low-frequency, complex waveforms: The use of envelope-based temporal disparities. *The Journal of the Acoustical Society of America*, 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. *The Journal of the Acoustical Society of America*, 95(6):3561–3567.

Bernstein, L. R. and Trahiotis, C. (1996). On the use of the normalized correlation as an index of interaural envelope correlation. *The Journal of the Acoustical Society of America*, 100(3):1754–1763.

Bernstein, L. R. and Trahiotis, C. (2002). Enhancing sensitivity to interaural delays at high frequencies by using "transposed stimuli". *The Journal of the Acoustical Society of America*, 112(3):1026–1036.

Bernstein, L. R. and Trahiotis, C. (2008). Discrimination of interaural temporal disparities conveyed by high-frequency sinusoidally amplitude-modulated tones and high-frequency transposed tones: Effects of spectrally flanking noises. *The Journal of the Acoustical Society of America*, 124(5):3088–3094.

Bernstein, L. R. and Trahiotis, C. (2009). How sensitivity to ongoing interaural temporal disparities is affected by manipulations of temporal features of the envelopes of high-frequency stimuli. *The Journal of the Acoustical Society of America*, 125(5):3234–3242.

Bernstein, L. R. and Trahiotis, C. (2010). Accounting quantitatively for sensitivity to envelope-based interaural temporal disparities at high frequencies. *The Journal of the Acoustical Society of America*, 128(3):1224–1234.

Breebaart, J., van de Par, S., and Kohlrausch, A. (2001a). Binaural processing model based on contralateral inhibition. i. model structure. *The Journal of the Acoustical Society of America*, 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. *The Journal of the Acoustical Society of America*, 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. *The Journal of the Acoustical Society of America*, 110(2):1105–1117.

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT press.

Carhart, R., Tillman, T. W., and Johnson, K. R. (1967). Release of masking for speech through interaural time delay. *The Journal of the Acoustical Society of America*, 42(1):124–138.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, 25(5):975–979.

Christensen, H., Ma, N., Wrigley, S. N., and Barker, J. (2009). A speech fragment approach to localising multiple speakers in reverberant environments. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pages 4593–4596. IEEE.

Colburn, H. S. (1977). Theory of binaural interaction based on auditory-nerve data. ii. detection of tones in noise. *The Journal of the Acoustical Society of America*, 61(2):525–533.

Culling, J. F. and Summerfield, Q. (1995). Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay. *The Journal of the Acoustical Society of America*, 98(2):785–797.

Culling, J. F. and Summerfield, Q. (1998). Measurements of the binaural temporal window using a detection task. *The Journal of the Acoustical Society of America*, 103(6):3540–3553.

Dannenbring, G. and Bregman, A. (1978). Streaming vs. fusion of sinusoidal components of complex tones. *Perception & Psychophysics*, 24(4):369–376.

Darwin, C. and Carlyon, R. (1995). *Auditory Grouping*, pages 387 – 424. Academic Press, San Diego.

Darwin, C. and Hukin, R. (1999). Auditory objects of attention: the role of interaural time differences. *Journal of Experimental Psychology: Human perception and performance*, 25(3):617.

Darwin, C. J. (1981). Perceptual grouping of speech components differing in fundamental frequency and onset-time. *The Quarterly Journal of Experimental Psychology Section A*, 33(2):185–207.

Darwin, C. J. and Sutherland, N. S. (1984). Grouping frequency components of vowels: When is a harmonic not a harmonic? *The Quarterly Journal of Experimental Psychology Section A*, 36(2):193–208.

Dau, T., Kollmeier, B., and Kohlrausch, A. (1997). Modeling auditory processing of amplitude modulation. i. detection and masking with narrow-band carriers. *The Journal of the Acoustical Society of America*, 102(5):2892–2905.

Dau, T., Püschel, D., and Kohlrausch, A. (1996a). A quantitative model of the "effective" signal processing in the auditory system. i. model structure. *The Journal of the Acoustical Society of America*, 99(6):3615–3622. article.

Dau, T., Püschel, D., and Kohlrausch, A. (1996b). A quantitative model of the "effective" signal processing in the auditory system. ii. simulations and measurements. *The Journal of the Acoustical Society of America*, 99(6):3623–3631.

de Cheveigné, A. (1998). Cancellation model of pitch perception. *The Journal of the Acoustical Society of America*, 103(3):1261–1271.

de Cheveigné, A. (2005). *Pitch - Neural coding and perception*, chapter Pitch Perception Models, pages 169–233. Springer.

Dietz, M., Ewert, S. D., and Hohmann, V. (2009). Lateralization of stimuli with independent fine-structure and envelope-based temporal disparities. *The Journal of the Acoustical Society of America*, 125(3):1622–1635. article.

Dietz, M., Klein-Hennig, M., and Hohmann, V. (2015). The influence of pause, attack, and decay duration of the ongoing envelope on sound lateralization. *The Journal of the Acoustical Society of America*, 137(2):EL137–EL143.

Dietz, M., Marquardt, T., Salminen, N. H., and McAlpine, D. (2013a). Emphasis of spatial cues in the temporal fine structure during the rising segments of amplitude-modulated sounds. *Proceedings of the National Academy of Sciences*, 110(37):15151–15156.

Dietz, M., Wendt, T., Ewert, S. D., Laback, B., and Hohmann, V. (2013b). Comparing the effect of pause duration on threshold interaural time differences between exponential and squared-sine envelopes (l). *The Journal of the Acoustical Society of America*, 133(1):1–4.

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.

Dreyer, A. A. and Oxenham, A. J. (2008). Effects of level and background noise on interaural time difference discrimination for transposed stimuli. *The Journal of the Acoustical Society of America*, 123(1):EL1–EL7.

Durlach, N. I. (1963). Equalization and cancellation theory of binaural masking-level differences. *The Journal of the Acoustical Society of America*, 35(8):1206–1218. article.

Dye, R. H. and Hafter, E. R. (1984). The effects of intensity on the detection of interaural differences of time in high-frequency trains of clicks. *The Journal of the Acoustical Society of America*, 75(5):1593–1598.

Dye, R. H. J., Niemiec, A. J., and Stellmack, M. A. (1994). Discrimination of interaural envelope delays: The effect of randomizing component starting phase. *The Journal of the Acoustical Society of America*, 95(1):463–470.

Epp, B. and Verhey, J. (2009a). Superposition of masking releases. *Journal of Computational Neuroscience*, 26(3):393–407. J Comput Neurosci.

Epp, B. and Verhey, J. L. (2009b). Combination of masking releases for different center frequencies and masker amplitude statistics. *The Journal of the Acoustical Society of America*, 126(5):2479–2489.

Ewert, S. D. and Dau, T. (2000). Characterizing frequency selectivity for envelope fluctuations. *The Journal of the Acoustical Society of America*, 108(3):1181–1196. article.

Ewert, S. D., Dietz, M., Klein-Hennig, M., and Hohmann, V. (2010). *The Neurophysiological Bases of Auditory Perception*, book section The role of envelope wave form, adaptation, and attacks in binaural perception, page 337–346. Springer, NY.

Fletcher, N. H. (1992). *Acoustic Systems in Biology*. Oxford University Press, USA.

Francart, T., Lenssen, A., and Wouters, J. (2012). The effect of interaural differences in envelope shape on the perceived location of sounds (l). *The Journal of the Acoustical Society of America*, 132(2):611–614.

Furukawa, S. (2008). Detection of combined changes in interaural time and intensity differences: Segregated mechanisms in cue type and in operating frequency range? *The Journal of the Acoustical Society of America*, 123(3):1602–1617.

Goldstein, J. (1967). Auditory nonlinearity. *The Journal of the Acoustical Society of America*, 41(3):676–699.

Goldstein, J. L. (1973). An optimum processor theory for the central formation of the pitch of complex tones. *The Journal of the Acoustical Society of America*, 54(6):1496–1516.

Griffin, S. J., Bernstein, L. R., Ingham, N. J., and McAlpine, D. (2005). Neural sensitivity to interaural envelope delays in the inferior colliculus of the guinea pig. *J Neurophysiol*, 93(6):3463–3478.

Hafter, E., Buell, T., and Richards, V. (1988). *Auditory Function: Neurobiological Bases of Hearing*, pages 647–676. Wiley.

Hafter, E. R. and Buell, T. N. (1990). Restarting the adapted binaural system. *The Journal of the Acoustical Society of America*, 88(2):806–812.

Hafter, E. R. and Dye, R. H. (1983). Detection of interaural differences of time in trains of high-frequency clicks as a function of interclick interval and number. *The Journal of the Acoustical Society of America*, 73(2):644–651.

Hall, J. W., Grose, J. H., and Haggard, M. P. (1990). Effects of flanking band proximity, number, and modulation pattern on comodulation masking release. *The Journal of the Acoustical Society of America*, 87(1):269–283.

Hall, J. W., Haggard, M. P., and Fernandes, M. A. (1984). Detection in noise by spectro-temporal pattern analysis. *The Journal of the Acoustical Society of America*, 76(1):50–56.

Hartmann, W. M. and Doty, S. L. (1996). On the pitches of the components of a complex tone. *The Journal of the Acoustical Society of America*, 99(1):567–578.

Hartmann, W. M., McAdams, S., and Smith, B. K. (1990). Hearing a mistuned harmonic in an otherwise periodic complex tone. *The Journal of the Acoustical Society of America*, 88(4):1712–1724.

Heil, P. (2001). Representation of sound onsets in the auditory system. *Audiology and Neurotology*, 6(4):167–172.

Henning, G. B. (1974). Detectability of interaural delay in high-frequency complex waveforms. *The Journal of the Acoustical Society of America*, 55(1):84–90.

Hohmann, V. (2002). Frequency analysis and synthesis using a gammatone filterbank. *Acta Acustica united with Acustica*, 88:433–442(10).

Houtsma, A. J. M. and Smurzynski, J. (1990). Pitch identification and discrimination for complex tones with many harmonics. *The Journal of the Acoustical Society of America*, 87(1):304–310.

Hukin, R. and Darwin, C. (1995a). Comparison of the effect of onset asynchrony on auditory grouping in pitch matching and vowel identification. *Perception & Psychophysics*, 57(2):191–196.

Hukin, R. and Darwin, C. (1995b). Effects of contralateral presentation and of interaural time differences in segregating a harmonic from a vowel. *The Journal of the Acoustical Society of America*, 98(3):1380–1387.

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

Jeffress, L. A., Blodgett, H. C., Sandel, T. T., and Wood, C. L. (1956). Masking of tonal signals. *The Journal of the Acoustical Society of America*, 28(3):416–426.

Joris, P. X. and Yin, T. C. (1995). Envelope coding in the lateral superior olive. i. sensitivity to interaural time differences. *J Neurophysiol*, 73(3):1043–1062.

Kaiser, J. F. and David, E. E. (1960). Reproducing the cocktail party effect. *The Journal of the Acoustical Society of America*, 32(7):918–918.

Kepesi, M., Pernkopf, F., and Wohlmayr, M. (2007). Joint position-pitch tracking for 2-channel audio. In *Content-Based Multimedia Indexing, 2007. CBMI '07. International Workshop on*, pages 303–306.

Klein-Hennig, M., Dietz, M., Klinge-Strahl, A., Klump, G., and Hohmann, V. (2012). Effect of mistuning on the detection of a tone masked by a harmonic tone complex. *PLoS ONE*, 7(11):e48419. doi:10.1371/journal.pone.0048419.

Klinge, A., Beutelmann, R., and Klump, G. M. (2011). Effect of harmonicity on the detection of a signal in a complex masker and on spatial release from masking. *PLoS ONE*, 6(10):e26124. doi:10.1371/journal.pone.0026124.

Kohlrausch, A., Fassel, R., and Dau, T. (2000). The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers. *The Journal of the Acoustical Society of America*, 108(2):723–734.

Krumbholz, K., Magezi, D. A., Moore, R. C., and Patterson, R. D. (2009). Binaural sluggishness precludes temporal pitch processing based on envelope cues in conditions of binaural unmasking. *The Journal of the Acoustical Society of America*, 125(2):1067–1074.

Laback, B., Zimmermann, I., Majdak, P., Baumgartner, W.-D., and Pok, S.-M. (2011). Effects of envelope shape on interaural envelope delay sensitivity in acoustic and electric hearinga). *The Journal of the Acoustical Society of America*, 130(3):1515–1529.

Ladefoged, P. (1996). *Elements of Acoustic Phonetics*. University of Chicago Press.

Leakey, D. M., Sayers, B. M., and Cherry, C. (1958). Binaural fusion of low- and high-frequency sounds. *The Journal of the Acoustical Society of America*, 30(3):222–222.

Lee, J. and Green, D. M. (1994). Detection of a mistuned component in a harmonic complex. *The Journal of the Acoustical Society of America*, 96(2):716–725.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, 49(2B):467–477. article.

Licklider, J. C. R. (1948). The influence of interaural phase relations upon the masking of speech by white noise. *The Journal of the Acoustical Society of America*, 20(2):150–159.

Lindemann, W. (1986). Extension of a binaural cross-correlation model by contralateral inhibition. i. simulation of lateralization for stationary signals. *The Journal of the Acoustical Society of America*, 80(6):1608–1622.

Ma, N., Green, P., Barker, J., and Coy, A. (2007). Exploiting correlogram structure for robust speech recognition with multiple speech sources. *Speech Communication*, 49(12):874–891.

McAdams, S. (1989). Segregation of concurrent sounds. i: Effects of frequency modulation coherence. *The Journal of the Acoustical Society of America*, 86(6):2148–2159.

McDonald, K. L. and Alain, C. (2005). Contribution of harmonicity and location to auditory object formation in free field: evidence from event-related brain potentials. *The Journal of the Acoustical Society of America*, 118(3):1593–1604.

McFadden, D. and Pasanen, E. G. (1976). Lateralization at high frequencies based on interaural time differences. *The Journal of the Acoustical Society of America*, 59(3):634–639.

Meddis, R. (1986). Simulation of mechanical to neural transduction in the auditory receptor. *The Journal of the Acoustical Society of America*, 79(3):702–711.

Meddis, R., Hewitt, M. J., and Shackleton, T. M. (1990). Implementation details of a computation model of the inner hair-cell auditory-nerve synapse. *The Journal of the Acoustical Society of America*, 87(4):1813–1816.

Meddis, R. and O'Mard, L. P. (2005). A computer model of the auditory-nerve response to forward-masking stimuli. *The Journal of the Acoustical Society of America*, 117(6):3787–3798.

Metz, P. J., von Bismarck, G., and Durlach, N. I. (1968). Further results on binaural unmasking and the ec model. ii. noise bandwidth and interaural phase. *The Journal of the Acoustical Society of America*, 43(5):1085–1091.

Mitchell, O. M. M., Ross, C. A., and Yates, G. H. (1971). Signal processing for a cocktail party effect. *The Journal of the Acoustical Society of America*, 50(2B):656–660.

Monaghan, J. J. M., Krumbholz, K., and Seeber, B. U. (2013). Factors affecting the use of envelope interaural time differences in reverberationa). *The Journal of the Acoustical Society of America*, 133(4):2288–2300.

Moore, B. (1993). Frequency analysis and pitch perception. In Yost, W., Popper, A., and Fay, R., editors, *Human Psychophysics*, volume 3 of *Springer Handbook of Auditory Research*, pages 56–115. Springer New York.

Moore, B. and Glasberg, B. (1996). A revision of Zwicker's loudness model. *Acta Acustica united with Acustica*, 82:335–345(11).

Moore, B. C. J. (1973). Frequency difference limens for short-duration tones. *The Journal of the Acoustical Society of America*, 54(3):610–619.

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *The Journal of the Acoustical Society of America*, 80(2):479–483.

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (1985). Thresholds for the detection of inharmonicity in complex tones. *The Journal of the Acoustical Society of America*, 77(5):1861–1867.

Neubauer, H. and Heil, P. (2008). A physiological model for the stimulus dependence of first-spike latency of auditory-nerve fibers. *Brain Research*, 1220:208 – 223.

Nitschmann, M. and Verhey, J. L. (2012). Modulation cues influence binaural masking-level difference in masking-pattern experiments. *The Journal of the Acoustical Society of America*, 131(3):EL223–EL228.

Nuetzel, J. M. and Hafter, E. R. (1976). Lateralization of complex waveforms: Effects of fine structure, amplitude, and duration. *The Journal of the Acoustical Society of America*, 60(6):1339–1346.

Nuetzel, J. M. and Hafter, E. R. (1981). Discrimination of interaural delays in complex waveforms: Spectral effects. *The Journal of the Acoustical Society of America*, 69(4):1112–1118.

Oh, E. L. and Lutfi, R. A. (2000). Effect of masker harmonicity on informational masking. *The Journal of the Acoustical Society of America*, 108(2):706–709.

Palmer, A. and Russell, I. (1986). Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells. *Hearing Research*, 24(1):1 – 15.

Parsons, T. W. (1976). Separation of speech from interfering speech by means of harmonic selection. *The Journal of the Acoustical Society of America*, 60(4):911–918.

Patterson, R., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (1987). An efficient auditory filterbank based on the gammatone function. In *a meeting of the IOC Speech Group on Auditory Modelling at RSRE, vol. 2, no. 7. 1987*.

Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. (1992). Complex sounds and auditory images. *Auditory physiology and perception*, 83:429–446.

Plack, C. and Oxenham, A. (2005). *Pitch - Neural coding and Perception*, volume 24 of *Springer Handbook of Auditory Research*, chapter The Psychophysics of Pitch, pages 7–56. Springer.

Pressnitzer, D. and Patterson, R. (2001). Distortion products and the perceived pitch of harmonic complex tones. In *Physiological and psychophysical bases of auditory function*, pages 97–104.

Rayleigh, L. (1907). On our perception of sound direction. *Philosophical Magazine*, 13:232.

Roberts, B. and Bailey, P. J. (1996). Spectral regularity as a factor distinct from harmonic relations in auditory grouping. *Journal of Experimental Psychology: Human Perception and Performance*, 22(3):604.

Sayers, B. M. and Cherry, E. C. (1957). Mechanism of binaural fusion in the hearing of speech. *The Journal of the Acoustical Society of America*, 29(9):973–987.

Sek, A. and Moore, B. C. J. (1995). Frequency discrimination as a function of frequency, measured in several ways. *The Journal of the Acoustical Society of America*, 97(4):2479–2486.

Smith, R. L. (1979). Adaptation, saturation, and physiological masking in single auditory-nerve fibers. *The Journal of the Acoustical Society of America*, 65(1):166–178.

Smith, R. L. and Brachman, M. L. (1980). Operating range and maximum response of single auditory nerve fibers. *Brain Research*, 184(2):499 – 505.

Smoski, W. J. and Trahiotis, C. (1986). Discrimination of interaural temporal disparities by normal-hearing listeners and listeners with high-frequency sensorineural hearing loss. *The Journal of the Acoustical Society of America*, 79(5):1541–1547.

Stellmack, M. A., Viemeister, N. F., and Byrne, A. J. (2005). Discrimination of interaural phase differences in the envelopes of sinusoidally amplitude-modulated 4-khz tones as a function of modulation depth. *The Journal of the Acoustical Society of America*, 118(1):346–352.

Stern, R. M. and Shear, G. D. (1996). Lateralization and detection of low-frequency binaural stimuli: Effects of distribution of internal delay. *The Journal of the Acoustical Society of America*, 100(4):2278–2288.

Sumner, C. J., Lopez-Poveda, E. A., O'Mard, L. P., and Meddis, R. (2002). A revised model of the inner-hair cell and auditory-nerve complex. *The Journal of the Acoustical Society of America*, 111(5):2178–2188.

Thompson, E. R. and Dau, T. (2008). Binaural processing of modulated interaural level differences. *The Journal of the Acoustical Society of America*, 123(2):1017–1029. article.

van de Par, S. and Kohlrausch, A. (1997). A new approach to comparing binaural masking level differences at low and high frequencies. *The Journal of the Acoustical Society of America*, 101(3):1671–1680.

Verhey, J., Pressnitzer, D., and Winter, I. (2003). The psychophysics and physiology of comodulation masking release. *Experimental Brain Research*, 153:405–417.

von Helmholtz, H. (1863). *Die Lehre von den Tonempfindungen als Physiologische Grundlage für die Theorie der Musik*. F. Vieweg und sohn.

Weiss, T. and Rose, C. (1988). A comparison of synchronization filters in different auditory receptor organs. *Hearing Research*, 33(2):175 – 179.

Westerman, L. A. and Smith, R. L. (1984). Rapid and short-term adaptation in auditory nerve responses. *Hearing Research*, 15(3):249 – 260.

Wickesberg, R. E. and Oertel, D. (1990). Delayed, frequency-specific inhibition in the cochlear nuclei of mice: a mechanism for monaural echo suppression. *The journal of neuroscience*, 10(6):1762–1768.

Yin, T. C. and Chan, J. C. (1990). Interaural time sensitivity in medial superior olive of cat. *J Neurophysiol*, 64(2):465–488.

Young, E. D. (1988). *Auditory Function*, pages 277 – 312. Wiley, NY.

Young, L. L. and Carhart, R. (1974). Time-intensity trading functions for pure tones and a high-frequency am signal. *The Journal of the Acoustical Society of America*, 56(2):605–609.

Danksagung

Mit diesen Zeilen möchte ich mich bei all jenen Menschen bedanken, die mich bei der Anfertigung dieser Arbeit unterstützt haben.

Für die großartige Betreuung während der ganzen Zeit bedanke ich mich bei Prof. Dr. Volker Hohmann und Dr. Mathias Dietz. Sie hatten stets offene Ohren für die kleinen und großen Herausforderungen im Laufe des Projekts und waren mir mit ihrer Fachkenntnis und einer klaren, pragmatischen Herangehensweise eine große Hilfe.

Bei Prof. Dr. Dr. Birger Kollmeier bedanke ich mich für wertvolle Hinweise und Diskussionen, die Übernahme des Korreferats, das Wecken des Interesses an der Hörforschung und natürlich die Möglichkeit, in der AG Medizinische Physik zu arbeiten.

Dr. Stephan Ewert, Dr. Astrid Klinge-Strahl und Prof. Dr. Georg Klump gilt mein Dank für die gute und produktive Zusammenarbeit in den einzelnen Studien und für fruchtbare Diskussionen.

Der Arbeitsgruppe Medizinische Physik danke ich für das gute Umfeld und die Infrastruktur, die das wissenschaftliche Arbeiten braucht. Die zahlreichen Menschen, mit denen ich in Seminaren, auf Tagungen, auf dem Flur oder auch privat über fachliche und weniger fachliche Sachverhalte sprechen konnte, hatten großen Einfluss auf diese Arbeit.

Für Hilfe rund um das Labor, Büro- und Versuchsrechner, Administratives und eine angenehme Medi-Frühstücksatmosphäre danke ich Anita Gorges, Felix Grossmann, Frank Grunau, Katja Warnken und Ingrid Wusowski.

Mein besonderer Dank geht an W2 0-071, also Regina Baumgärtel, Carolin Iben, Marc René Schädler und Wiebke Schubotz. Das gemeinsame Durchleben der Höhen und Tiefen des akademischen Alltags hat für einen außerordentlichen Zusammenhalt gesorgt und innerhalb sowie außerhalb des Büros viel Freude gemacht. Ich wünsche den 71ern alles Gute auf ihren Wegen und viel Spaß dabei.

Den Versuchspersonen, die teilweise sogar die schönsten Stunden des Oldenburger Sommers in der Hörkabine verbracht haben, danke ich vielmals fürs genaue Hinhören.

Am Ende, jedoch vor Allem, danke ich meiner Familie fürs Familie-Sein.

Lebenslauf

Martin Julius Christoph Klein-Hennig

Geboren am 18.01.1984 in Oldenburg
Verheiratet, 2 Kinder

06/2014–08/2014	Wissenschaftlicher Mitarbeiter im EU-geförderten Projekt “ABCIT - Advancing Binaural Cochlear Implant Technology”
05/2011–05/2014	Wissenschaftlicher Mitarbeiter in Teilprojekt B2 des DFG Sonderforschungsbereichs TRR/31 - “Das aktive Gehör”
09/2009–05/2011	Doktorandenstipendium des Promotionskollegs “Funktion und Pathophysiologie des auditorischen Systems (HÖREN)”
06/2008–06/2009	Anfertigung der Diplomarbeit mit dem Titel “*The effect of envelope waveform on lateralization*”, betreut von Prof. Dr. Dr. Birger Kollmeier und Prof. Dr. Volker Hohmann.
10/2003–08/2009	Studium Diplom-Physik an der Carl v. Ossietzky Universität Oldenburg
09/1996–07/2003	Abitur am Gymnasium Cäcilienschule, Oldenburg

Erklärung

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt habe.

_____

Martin Klein-Hennig