PROCESSING OF HARMONICITY, ONSET, AND SPATIAL CUES IN COMPLEX ACOUSTIC ENVIRONMENTS

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

Dipl.-Biol. Astrid Klinge
geboren am 27. Oktober 1978
in Jena

Gutachter: Prof. Dr. Georg Klump

Zweitgutachter: Prof. Dr. Christine Köppl

Tag der Disputation: 24.8.2010

Für Stefan

Ja!

Abstract

The auditory system has the remarkable ability to analyze a complex auditory scene by partitioning the sound mixture that reaches both ears as a single waveform into separate auditory objects. The process of grouping frequency components belonging to one sound source and segregating them from frequency components originating from a different sound source can be accomplished by using a variety of acoustic properties (so-called “cues”) of which harmonicity, time of onset and spatial cues are among the most important. As natural sounds provide more than one cue to group or segregate frequencies, it is crucial for understanding a complex acoustic scene to investigate how combined cues are processed by the auditory system. Gaining knowledge about the mechanisms of processing more complex stimuli is valuable for the (further) development of technical solutions in audiology, for example for the application in hearing aid devices.

The experiments described in Chapter 2 were conducted to investigate the perception of a mistuned (i.e., frequency-shifted) component in an otherwise harmonic complex by Mongolian gerbils and to evaluate possible mechanisms underlying the perception of mistuning. It will be shown that the gerbil as an animal model is more sensitive than are humans in detecting a mistuned component in a harmonic complex with all harmonics starting in sine phase. This and subsequent experiments presented in Chapter 2 together with a simulation of the processing of harmonic complexes by the gerbil’s peripheral auditory filters (Chapter 4) provide evidence that temporal cues, such as phase and amplitude modulations in the filter outputs, are important for mistuning detection not only in the high but also in the low frequency region. The study described in Chapter 3 was conducted to evaluate whether differences in the

perception of frequency shifts of pure tones and of components in a complex stimulus between gerbils, humans, and various bird species can be related to differences or even constraints in the anatomy (e.g., length of cochlea, number of haircells) and physiology (e.g., width of auditory filters) of the auditory periphery. It is suggested that the superior performance of the bird species and the gerbils in detecting a mistuning in a sine phase complex may be based on an enhanced use of temporal cues due to factors limiting the exploitation of spatial patterns of excitation in the cochlea. In Chapters 4 and 5, the processing of harmonicity cues in the auditory system is evaluated in combination with onset cues (Chapter 4) or spatial cues (Chapter 5). In Chapter 4 it is examined how starting and ending a mistuned component earlier than the remaining complex affects the frequency difference limens of this component. The results indicate that gerbils exploit temporal cues for mistuning detection until the temporal overlap between mistuned harmonic and remaining complex due to an increasing onset asynchrony is too short and the gerbil switches to a pure tone frequency discrimination. Finally, in a simultaneous masking experiment in humans (Chapter 5) it will be shown that harmonicity cues affect the masked thresholds when target and tonal complex maskers are co-located whereas harmonicity cues seem to be irrelevant in determining the release from masking in a situation in which target and tonal complex masker are spatially separated.

The results of this thesis provide a basis for a further exploration of the underlying mechanism of detecting changes in the temporal waveform of complex stimuli. Especially recordings of responses of neurons in the brainstem of the gerbil may shed light on the processing of tonal complex stimuli in the auditory system.

Zusammenfassung

Das auditorische System besitzt die bemerkenswerte Fähigkeit eine komplexe akustische Szene dadurch zu analysieren, dass es ein ankommendes Tongemisch in einzelne Hörobjekte unterteilt. Diese Fähigkeit der Quellentrennung ist insofern bemerkenswert, da das Tongemisch die Ohren in Form einer einzelnen Schallwelle erreicht. Für die Gruppierung von Frequenzkomponenten einer Schallquelle und der Trennung dieser von Frequenzkomponenten einer anderen Schallquelle kann das auditorische System eine Reihe von akustischen Eigenschaften nutzen. Einige der wichtigsten Eigenschaften in einem komplexen akustischen Reiz sind Harmonizität, Stimulusbeginn und Raumrichtungsinformationen. Da jedoch in natürlichen akustischen Reizen mehr als nur eine Eigenschaft für die Gruppierung oder Trennung von Frequenzkomponenten vorhanden ist, ist es wichtig zu untersuchen, wie verschiedene Merkmalskombinationen durch das Hörsystem verarbeitet werden. Ein besseres Verständnis der Mechanismen für die Verarbeitung komplexerer akustischer Reize im Hörsystem ist nötig zur Entwicklung sowie Weiterentwicklung technischer Lösungen in der Audiologie, zum Beispiel für die Anwendung in technischen Hörhilfen.

In den im Kapitel 2 beschriebenen Experimenten wurde untersucht, wie gut Mongolische Rennmäuse Frequenzverschiebungen in einer Komponente eines harmonischen Tonkomplexes wahrnehmen können und welche Verarbeitungsmechanismen der Detektionsleistung zu Grunde liegen. Es wird gezeigt, dass die Rennmaus als gewähltes Tiermodell Frequenzverschiebungen in einer Komponente eines Tonkomplexes, in der alle Komponenten mit Sinusphase beginnen, besser detektieren kann als der Mensch. Sowohl dieses Experiment und daran anschließende Experimente aus Kapitel 2 als auch eine Simulation der Filterantwort auf die Präsentation von Tonkomplexen im peripheren

Hörapparat der Rennmaus (Kapitel 4) geben Hinweise darauf, dass die Detektion der Frequenzverschiebungen im hohen als auch im niedrigen Frequenzbereich auf zeitlichen Verarbeitungsmustern beruht. Diese können Phasen- als auch Amplitudenmodulationen im Zeitmuster der Filterantwort sein. In Kapitel 3 wird die Frage erörtert, ob Unterschiede, die man zwischen verschiedenen Spezies (Rennmäusen, Menschen und diversen Vogelarten) sowohl bei der Wahrnehmung von Frequenzunterschieden zwischen zwei Reintönen als auch bei der Detektion von Frequenzverschiebungen in einer Komponente eines Tonkomplexes findet darauf zurückzuführen sind, dass sich anatomische Gegebenheiten oder physiologische Eigenschaften des peripheren Hörsystems zwischen den Arten unterscheiden oder sogar gewissen Zwängen unterworfen sind. Die Untersuchungen lassen vermuten, dass die bei der Detektion von Frequenzverschiebungen in Komponenten eines Tonkomplexes dem Menschen überlegenen Vögel und Rennmäuse auf einer verbesserten Verarbeitungsleistung von zeitlichen Mustern basiert. Ein Grund hierfür könnten anatomische oder physiologische Faktoren sein, die die Ausnutzung räumlicher Erregungsmuster auf der Basilarmembran einschränken. Kapitel 4 und 5 beschäftigen sich mit der interessanten Frage, wie Merkmalskombinationen im Hörsystem verarbeitet werden. In Kapitel 4 wird bei der Rennmaus untersucht, wie es sich auf die Frequenzdetektionsschwelle auswirkt, wenn eine frequenzverschobene Komponente zeitlich früher beginnt und endet als die restlichen Komponenten eines Tonkomplexes. Die Ergebnisse dieser Studie deuten darauf hin, dass auch hier die Detektion der Frequenzverschiebung auf die Wahrnehmung von Änderungen im zeitlichen Muster in der Filterantwort basiert, was jedoch nur bis zu einem gewissen Grad an Asynchronität möglich ist. Ist durch einen deutlich verschobenen Beginn die zeitliche Überlappung zwischen frequenzverschobener Komponente und Tonkomplex zu gering, dann lassen die Ergebnisse vermuten, dass die Rennmaus zur Reinton-Frequenzdiskrimination wechselt, um die Frequenzverschiebung in der Komponente zu detektieren. Im letzten Kapitel (Kapitel 5) wird der Frage nachgegangen, ob es eine Interaktion zwischen der Verarbeitung von Harmonizität und der Verarbeitung von Raumrichtungsinformationen bei der Detektion von Zieltönen in einem Maskierer gibt. Es wird gezeigt, dass Harmonizität die Maskierungsschwelle beeinflusst, wenn Zielton und Maskierer (ein Tonkomplex) aus derselben Raumrichtung kommen. Harmonizität als akustische Eigenschaft scheint jedoch keine Rolle bei der Reduzierung der Maskierungswirkung zu spielen, wenn Zielton und Maskierer (Tonkomplex) räumlich getrennt sind.

Die Ergebnisse der hier vorgelegten Dissertation bieten eine Basis für weitergehende Untersuchungen am Tiermodell der Mongolischen Rennmaus. Interessant dabei wären weitere Experimente, die zur Aufklärung der zu Grunde liegenden Mechanismen bei der Detektion von feinen Änderungen im zeitlichen Muster komplexer akustischer Reize (wie z.B. harmonischer Tonkomplexe) beitragen könnten. Wichtig wären im Folgenden auch elektrophysiologische Messungen im Hirnstamm der Rennmaus, um ein Verständnis für die neuronale Verarbeitung komplexer akustischer Reize im Hörsystem zu bekommen.

Contents

1 Introduction 1

1.1 General Introduction 2
1.2 Harmonicity as a grouping cue 5
1.3 Perceptual interaction of onset cues and harmonicity 8
1.4 Perceptual interaction of spatial cues and harmonicity 10
1.5 The Mongolian gerbil as a suitable animal model for studying the effects of grouping cues on the processing of complex stimuli 12
1.6 Aim and scope of this thesis 13
1.7 References 17

2 Frequency difference limens of pure tones and harmonics within complex stimuli in Mongolian gerbils and humans 25

2.1 Abstract 26
2.2 Introduction 27
2.3 Materials and methods 29
  2.3.1 Animal subjects 29
  2.3.2 Apparatus and stimulus generation 29
  2.3.3 Procedure 33
  2.3.4 Testing of human subjects 34

2.3.5 Data analysis 34
2.3.6 Prior testing of the hearing sensitivity of gerbils 35
2.4 Results 37
2.4.1 Experiment 1: Pure tone frequency discrimination 37
2.4.2 Experiment 2: Frequency difference limens of mistuned components in a sine phase harmonic complex 38
2.4.3 Experiment 3: Frequency difference limens of mistuned components in a random phase harmonic complex 39
2.4.4 Comparison of all three experiments 41
2.4.5 Comparison with human data 42
2.5 Discussion 44
2.5.1 Pure tone frequency difference limens 44
2.5.2 Frequency difference limens of mistuned components in a harmonic complex 48
2.5.3 Comparison with other species 54
2.6 Acknowledgements 56
2.7 References 57

3 A Comparative View on the Perception of Mistuning: Constraints of the Auditory Periphery 63
3.1 Abstract 64
3.2 Introduction 65
3.3 Detecting Frequency Shifts of Pure Tones 66
3.4 Detecting a Mistuned Component in Harmonic Complexes 71
3.4.1 Mistuning Detection in Sine Phase Harmonic Complexes 72

3.4.2 Mistuning Detection in Random Phase Harmonic Complexes . . 73
3.4.3 Neural Basis of Mistuning Detection 74
3.5 References 78

4 Mistuning detection and onset asynchrony in harmonic complexes in Mongolian gerbils 83

4.1 Abstract 84
4.2 Introduction 85
4.3 Materials and methods 88
4.3.1 Animal subjects 88
4.3.2 Apparatus and stimulus generation 89
4.3.3 Procedure 91
4.3.4 Data analysis 92
4.4 Results 93
4.4.1 Detection of mistuning in harmonic complexes with different fundamentals 93
4.4.2 FDLs of mistuned components with an onset asynchrony 96
4.5 Discussion 100
4.5.1 Detection of mistuning in harmonic complexes with different fundamentals 100
4.5.2 FDLs of mistuned components with an onset asynchrony 108
4.6 Conclusion 113
4.7 Acknowledgments 114
4.8 References 114

5 Effect of harmonicity on the detection of a signal in a complex masker and on spatial release from masking 121

5.1 Abstract . . . 122

5.2 Introduction . . . 123

5.3 Materials and methods . . . 125

  5.3.1 Subjects . . . 125

  5.3.2 Apparatus and stimulus generation . . . 125

  5.3.3 Procedure . . . 129

  5.3.4 Data analysis . . . 130

5.4 Results . . . 130

  5.4.1 Effect of masker type on the masked thresholds . . . 130

  5.4.2 Effect of masker type on the spatial release from masking . . . 134

5.5 Discussion . . . 136

  5.5.1 Influence of masker type on the detection threshold . . . 136

  5.5.2 Influence of masker type on spatial release from masking . . . 141

5.6 Acknowledgements . . . 143

5.7 References . . . 143

1 Introduction

1.1 General Introduction

The natural environment is acoustically complex. At every point in time animals and humans are surrounded by a number of sound sources that often overlap in time and in frequency. In this constantly varying mixture of sound, animals and humans have to communicate and attend to behaviorally important sounds that might be masked by competing sounds. Thus, decomposing the acoustic information entering the ears and grouping them into packages of acoustic evidence is crucial to permit or at least facilitate sound source identification. This process which has been referred to as 'auditory scene analysis' by Bregman (1990) is made even more complex inasmuch as the sound mixture that reaches each ear is a single pattern formed by pressure changes over time. The auditory system is able to partition the incoming sound mixture into separate groups of sound by using primitive grouping mechanisms. They are based on simple stimulus properties, function largely automatic (pre-attentive) without any specific prior knowledge and are probably innate or acquired early. Bregman (1990) referred to the processing of the acoustic information based on these so-called bottom-up mechanisms as 'primitive auditory scene analysis'. The primitive grouping mechanisms can be divided into mechanisms that group frequency components along the time axis (sequential grouping) and mechanisms that group frequency components along the frequency axis (simultaneous grouping).

The main objective of this thesis was to investigate how different acoustic properties (so-called 'cues') are used for processing sounds that occur simultaneously. In addition, it was of special interest to examine the interaction of several cues that are naturally present in more complex stimuli. Among the most important cues for grouping of concurrent sounds are the harmonic relation between frequency components of a complex sound (Moore *et al.*, 1984, 1985; Hartmann *et al.*, 1990), the synchrony of the onset of signal components (Darwin and Ciocca, 1992; Darwin and Hukin, 1998; Gockel *et al.*, 2005; Bleeck *et al.*, 2008), and cues related to the spatial location (e.g. Hirsh, 1948; Jeffress *et al.*, 1956; but see also Hukin and Darwin, 1995).

As many natural sounds are harmonic complexes, harmonicity cues are an important focus of auditory research (see, for example, reviews by Darwin, 1997; Deutsch, 1999; Cheveigné, 2005; Micheyl and Oxenham, 2010). For example, harmonic complexes occur in human speech and in animal vocalizations, and they are produced by musical instruments. They are composed of frequency components that are integer multiples of a fundamental frequency F0. The auditory system tends to group these simultaneously occurring and harmonically related frequency components and assigns them to a common pitch that corresponds to the frequency of the F0. Thus, harmonicity provides strong cues aiding the grouping of sounds from a single source to a perceived auditory object and the segregation of sounds arising from different sources.

Another general regularity of natural sounds composed of more than one component is the tendency of components that belong to the same sound source to start and end synchronously. Thus, frequencies with an asynchronous onset and offset tend to be segregated from other components. The importance of onset as a cue can be seen, for example, in experiments in which the level of a continuous sound is suddenly increased (Warren *et al.*, 1972). The auditory system interprets these sudden changes in level as if a new stimulus is added to a continuous one ('old-plus-new' heuristic, Bregman, 1990).

The ability to localize a sound source is crucial for humans and animals alike. Analyzing spatial cues present in the incoming sound facilitates the grouping of frequency components belonging to one sound source and the separation from simultaneously presented frequency components belonging to a different sound source. This is especially advantageous in an acoustically complex situation in which behaviorally relevant sounds are masked by competing sounds. The two main binaural cues for localizing a sound source in the horizontal plane are differences in the intensity or in the time of arrival of the sound at the two ears. Interaural time differences (ITDs) occur when sound reaches the ear that is closer to the source earlier than the ear that is further away from the source. ITDs are the predominant cue for lower frequencies due to the limited phase-locking ability of neurons. Interaural level differences (ILDs) occur due to shadowing effects of the head and the torso of the listener and are the predominant cue for higher frequencies (e.g., above approximately 2 kHz in humans). In complex listening situations in which several sounds come from different directions, the auditory system may analyze the perceived location of each sound source or use monaural cues to detect a behaviorally relevant sound in a masking background (such as differences in the level between the targeted and the masking sound).

Natural sound sources generally provide more than one cue to the auditory system to group or segregate frequency components. Thus, besides the importance to investigate processing mechanisms of single cues, it is crucial to analyze how a combination of different cues, such as onset asynchrony and harmonicity, or spatial location and harmonicity, is processed by the auditory system. For example, frequency components which can all be assigned to a common F0 will most likely be grouped by the auditory system. The grouping into a single perceived auditory object is facilitated if all components start at roughly the same time or if the spatial information across all components is identical. An interesting approach that might contribute to the understanding of underlying processing mechanisms if more than one cue is present in a stimulus is to examine how the auditory system copes with signals providing conflicting evidence. For example, an asynchronous onset or a spatial separation of components in a harmonic complex should lead to their segregation while harmonicity is a strong cue to group the components. The response of the auditory system to the signals providing conflicting cues can then be compared to responses to signals providing non-conflicting evidence (a single cue or combined cues that are non-conflicting). Possible differences in the responses might give insight into the processing of combined cues in the auditory system.

The following sections will provide a short overview over the effect of harmonicity on the grouping of frequencies that belong to one sound source (Section 1.2) and the effect of combining two cues, like harmonicity and onset cues (Section 1.3) or harmonicity and spatial cues (Section 1.4), respectively, on the processing of these cues in the auditory system. Section 1.5 discusses if the Mongolian gerbil is a suitable animal model to investigate the processing of grouping cues in the auditory system.

1.2 Harmonicity as a grouping cue

Harmonicity cues can be used to segregate two concurrent harmonic complexes by means of differences between the F0s of the two sounds (e.g., Scheffers, 1984; Assmann and Summerfield, 1990; McKeown and Patterson, 1995) or can be used to detect a frequency shift (mistuning) in one of the components of a harmonic complex stimulus (e.g., Moore *et al.*, 1985; Hartmann *et al.*, 1990). The sensitivity of the auditory system for harmonicity cues can be investigated using the “mistuned” harmonic paradigm (e.g., Moore *et al.*, 1984, 1985; Lin and Hartmann, 1998). In such an experiment, a harmonic complex serves as a reference stimulus to an inharmonic complex stimulus in which one of the components is mistuned. Moore *et al.* (1985) observed that for lower harmonics human subjects heard the mistuned component as a separate entity whereas for higher harmonics the mistuning produced a “beating” that could be used by the listener to identify the complex containing the mistuned component. These results indicated that in a harmonic complex with a mistuned component not only changes in the spectral domain but also changes in the temporal structure could have been used to detect inharmonicity. A subsequent study (Moore *et al.*, 1986) revealed that in the higher harmonics much larger frequency shifts than those for detecting mistuning were necessary to actually segregate the frequency-shifted harmonic from the remaining complex.

The mechanisms that underlie the detection of a mistuned component in a harmonic complex are still not fully understood. A mechanism put forward by Moore *et al.* (1985) to explain the low detection thresholds for the higher harmonics was the detection of beats or a “roughness”. They proposed that this percept is due to periodic waveform fluctuations at the output of an auditory filter caused by the changing relative phase between the mistuned component and adjacent harmonics. Several psychophysical and physiological studies support this hypothesis by showing that the auditory system is very sensitive in detecting changes in the relative phase of a single component in a harmonic complex (e.g., Horst *et al.*, 1986, 1990; Moore and Glasberg, 1989; Simmons *et al.*, 1993). LeGoff and Kohlrausch (2005) suggested that even for a resolved harmonic like the F0 component, changes in the temporal envelope at the output of an auditory filter might be the cue that is used for detecting the mistuning under certain circumstances.

For two bird species, the zebra finch and the budgerigar, a similar experiment using the “mistuned harmonic” paradigm has been carried out by Lohr and Dooling (1998). They determined the frequency difference limens (FDLs) for mistuned harmonics in harmonic complexes that resembled natural contact calls of zebra finches. They observed much smaller detection thresholds for both bird species compared to the FDLs in humans for both low and high harmonics. Lohr and Dooling (1998) raised the question whether the superior thresholds may reflect an enhanced ability of birds to process the same cues as humans or whether the bird auditory system has a superior temporal resolving power compared to humans. In a later study, Dooling *et al.* (2002) showed that budgerigars, zebra finches and canaries were better than humans in discriminating between complexes with different temporal fine structures. For example, budgerigars could discriminate cosine phase harmonic complexes from random phase complexes (a set of seven different phase randomizations were used in the experiment) at much shorter harmonic periods (1-2 ms) than could humans (3-4 ms). Compound action potential (CAP) measurements in the VIIIth nerve (a measure of the synchronization of neural firing to a stimulus) in response to positive and negative Schroeder-phase harmonic complexes showed differences between the bird species and the Mongolian gerbil that paralleled the behavioral differences between the birds and the humans. Dooling *et al.* (2002) found a significant difference between the CAPs up to a F0 of 800 Hz in the birds but no systematic differences in gerbils. They suggest that birds have an enhanced temporal resolution in the processing of harmonic complex stimuli that is not found in humans and other mammals. However, this study could not reveal if gerbils are truly not able to discriminate between harmonic complexes with different temporal fine structures in behavioral experiments.

The performance in detecting a mistuned component in an otherwise harmonic complex has so far not been investigated in any other mammalian species. That other mammals besides humans are able to detect mistuning has only been shown by Sinex and colleagues (Sinex *et al.*, 2002, 2003) in neurophysiological studies in the chinchilla. They recorded neuronal responses from single units at various levels of the auditory pathway. Neurons located in the inferior colliculus, for example, responded with a distinctively modulated temporal pattern when presented with mistuned harmonic complexes (Sinex *et al.*, 2002). However, the question is left open if the performance in mistuning detection in the gerbil as another mammalian species is more similar to the performance seen in humans or similar to the performance seen in the two bird species.

1.3 Perceptual interaction of onset cues and harmonicity

Sounds originating from the same sound source tend to start and stop at approximately the same time. The auditory system is able to use these onset cues to group frequencies belonging to the same sound source and segregate them from frequencies originating from a different sound source. Thus, a frequency component that starts and stops asynchronously to other components is judged as far more likely to belong to a different sound source. An interesting question hereby is which amount of onset asynchrony is necessary to segregate two sounds. As mentioned in Section 1.1, natural sounds generally provide more than one cue to the auditory system and these cues might be contradictory in their information for grouping or segregating components in a complex stimulus. For example, a frequency component is most likely to be grouped to a harmonic complex if it is harmonically related to the other components of the complex. However, if that frequency component starts considerably before the onset of the remaining complex, the auditory system tends to segregate the component from the complex. How the auditory system copes with this conflicting evidence and how much onset asynchrony is needed to segregate a component from a harmonic complex or vowel has been only investigated in humans so far using two different approaches: (1) the pitch matching paradigm and (2) the paradigm of identifying vowels differing in their first formant frequency (e.g., Darwin, 1984; Darwin and Ciocca, 1992; Hukin and Darwin, 1995; Darwin and Hukin, 1998).

In pitch-matching studies, the pitch of a harmonic complex is determined by matching the frequency of a pure tone to the pitch of the complex. If a frequency-shift is applied to a harmonic of this complex, the pitch of the overall complex changes. The pitch-shift can be gradually reversed if this mistuned component is perceptually removed, for example, by an additional onset asynchrony applied onto the mistuned component. In a study by Darwin and Ciocca (1992), the contribution of a mistuned component to the pitch of a harmonic complex is reduced if it starts 160 ms before (but ends together with) the complex. A component with an onset asynchrony of 300 ms does not contribute to the pitch of the complex.

Vowel identification studies, on the other hand, take advantage of a phoneme-boundary shift (i.e., a shift from perceiving the vowel /ɪ/ to perceiving the vowel /ɛ/) that occurs if one of the harmonic components in the vowel is physically or perceptually removed. Darwin (1984) showed that a harmonic that starts at least 32 ms before the onset of a vowel makes less of a contribution to vowel quality than one that starts at the same time as the vowel.

As a possible mechanism underlying the segregation effects with increasing onset asynchrony, Darwin and Ciocca (1992) and Ciocca and Darwin (1993) suggested perceptual grouping in higher auditory centers rather than peripheral adaptation processes. A component with a sufficiently large onset asynchrony would be grouped separately and thus, given less weight in determining the pitch of the overall complex. Referring to results of a vowel identification task from Darwin (1984), Bleeck *et al.*

(2008) proposed that adaptation processes in neurons, for example at the level of the ventral cochlear nucleus, are sufficient to explain the observed reduced contribution of the asynchronous harmonic to the quality of a vowel.

The processing of combined harmonicity and onset cues has not been examined in other animal species so far. Thus, an onset asynchrony applied to a mistuned component of a harmonic complex might affect the detection of mistuning differently than it has in the human studies. For example, an increasing onset asynchrony (if asynchrony is achieved by time-shifting on- and offset of the mistuned harmonic) also leads to a reduced temporal overlap between the mistuned harmonic and the remaining complex. Therefore, a possible mechanism of interaction between onset asynchrony and mistuning detection could consist of a reduced duration of using harmonicity cues for mistuning detection.

1.4 Perceptual interaction of spatial cues and harmonicity

The ability to process binaural cues is advantageous in an acoustically complex situation in which behaviorally relevant sounds are masked by competing sounds. One form of masking is the energetic masking that arises due to an overlap of acoustic energy of the target and the masker in an auditory filter. The detectability of a target in a masker is generally improved by spatially separating the target from the masker. While the processing of binaural cues has often been examined with noise or speech maskers, there are only a few studies focussing on the mechanisms underlying the perceptual interaction of binaural and harmonicity cues (e.g., Buell and Hafter, 1991; Hill and

Darwin, 1996; Darwin and Hukin, 1998; McDonald and Alain, 2005). However, the fact that many natural sounds are harmonic complexes makes it especially interesting to investigate the processing of the combined binaural and harmonicity cues in the auditory system.

The above mentioned human studies showed that mistuning a component in a harmonic complex (a vowel in the case of Darwin and Hukin, 1998) changes either the perceived lateralization of the mistuned component (Buell and Hafter, 1991; Hill and Darwin, 1996), the vowel's identity (Darwin and Hukin, 1998), or enhances the segregation of the mistuned component from the complex (McDonald and Alain, 2005). However, Darwin and Hukin (1998) and McDonald and Alain (2005) showed different results for processing binaural cues in combination with a harmonic component in a complex than Buell and Hafter (1991). Based on the results of their study, Buell and Hafter (1991) suggested that harmonically related components in a complex are combined into a single auditory object, regardless of their interaural coherence. Lateralization of a harmonic component in a vowel in the study of Darwin and Hukin (1998) revealed a small effect towards segregation of the harmonic from the vowel. McDonald and Alain (2005) showed in a free-field study that the probability of perceiving two separate sounds increased if the harmonic third component of a 200 Hz complex was presented from a different location as the complex. However, the results of the latter two studies also revealed that the segregation of the harmonic from the complex on the basis of binaural cues seems to be independent from the harmonic relation between the component and the complex.

The above mentioned studies leave open the question how the auditory system processes harmonicity cues and interaural disparities in higher frequency regions in which harmonic components are unresolved (i.e., more than one component falling into an auditory filter) and may interact in an auditory filter. Thus, additional experiments with a systematic investigation of harmonicity cues and interaural disparities using different harmonic and inharmonic maskers in different frequency regions may contribute to the understand of the interaction of the processing mechanisms if more than one cue is present in a stimulus.

1.5 The Mongolian gerbil as a suitable animal model for studying the effects of grouping cues on the processing of complex stimuli

The Mongolian gerbil (*Meriones unguiculatus*) is a well-established animal model that has been used in many behavioral and neurophysiological studies (e.g., Ryan, 1976; Schmiedt and Zwislocki, 1980; Schmiedt *et al.*, 1980; Plassmann *et al.*, 1987; Frisina *et al.*, 1990a,b; Sinnott *et al.*, 1992; Kittel *et al.*, 2002; Maier and Klump, 2006; Gleich *et al.*, 2007). These studies showed that the auditory system of the gerbil is comparable to that of humans for various parameters. Compared to other rodent species, the hearing sensitivity of gerbils is relatively well developed in the low-frequency range and similar to that of humans (Ryan, 1976). The estimated auditory filter bandwidths in the Mongolian gerbil are comparable in width to the critical bandwidths estimated in humans (Kittel *et al.*, 2002). Furthermore, Wagner *et al.* (2003) showed that the temporal processing, measured with gap detection thresholds, is similar to that of humans and other mammals. On the contrary, Sinnott *et al.* (1992) found extraordinarily high pure tone frequency discrimination thresholds in gerbils compared to those in humans and other tested animal species. This might lead to the assumption that the gerbil also has a poor capacity to detect a frequency shift in a component of a harmonic complex. However, comparing psychoacoustical studies in other animals reveals that the performance in pure tone frequency discrimination does not necessarily reflect the performance in detecting frequency shifts in components of harmonic complex stimuli or in discriminating between different speech stimuli (e.g., budgerigar: Dent *et al.*, 2000 and Lohr and Dooling, 1998; monkey: Sinnott *et al.*, 1992 and Sinnott and Brown, 1997; starling: Langemann and Klump, 1992 and Klump and Groß, submitted). For the gerbil, Sinnott and Mosteller (2001) showed that the capacity of discriminating speech stimuli was much better than would have been expected on the basis of the gerbil's pure tone frequency discrimination thresholds. Taking into account the findings of the above mentioned studies, it is suggested that the gerbil is a suitable animal model to investigate the effect of grouping cues on the processing of complex stimuli in the auditory system of mammals.

A further advantage of using an animal model for studying auditory perception is the possibility to combine psychophysical with invasive approaches like neural recordings from different regions of the brain.

1.6 Aim and scope of this thesis

The aim of this thesis is to contribute to the understanding of how simultaneous grouping mechanisms are used by the auditory system of mammals to analyze complex auditory scenes. More specifically, it will be investigated how the auditory system processes not only harmonicity cues in isolation but beyond that in conjunction with additional onset and spatial cues when detecting a signal in a tonal complex stimulus. A further objective was to examine if variations in the data between different species can be related to anatomical and physiological constraints in the auditory periphery.

The aim of the experiments conducted in Chapter 2 is to evaluate the sensitivity of the gerbil auditory system for harmonicity cues in harmonic complexes in which one of the components is mistuned. The performance in detecting a mistuned component in a harmonic complex has been investigated in humans and in several bird species but not in any other mammalian species so far (see Section 1.2). As discussed in Section 1.5, the Mongolian gerbil seems to be a suitable animal model for studying the processing of harmonicity cues in harmonic complex stimuli. Frequency difference limens (FDLs) were determined for mistuned components in a harmonic complex and compared to FDLs for pure tones in the gerbil as well as to FDLs of mistuned components in a harmonic complex in humans. Generally, the thresholds for detecting mistuning in a harmonic complex should be much better than for pure tone frequency discrimination as has been found in bird species (see Section 1.2 and 1.5). The hypothesis is investigated that the detection of a mistuned harmonic in a complex by the gerbil is based on changes in the temporal waveform occurring with mistuning a component of the complex. This should result in a performance that is best if the temporal structure in a harmonic complex is held constant by starting all components in sine phase. Performance should deteriorate if changes in the temporal waveform are made less useful as a cue by constantly varying the phase relations between the harmonic components in the random phase harmonic complex. The comparison of the thresholds for detecting a frequency shift in the three conditions will provide insight into the mechanisms of processing harmonic complex stimuli.

The study presented in Chapter 3 discusses in detail differences in thresholds for pure tone discrimination and mistuning detection when comparing different animal species and whether this can be related to anatomical and physiological differences or even constraints in the periphery of the auditory system. As discussed in Sections 1.2 and 1.5, humans are better than all animals tested so far in pure tone frequency discrimination whereas a study by Lohr and Dooling (1998) showed that birds are better in mistuning detection than are humans. Thus, the performance in detecting frequency shifts in components of harmonic complexes and in pure tones will be compared across selected animal species (gerbil, human, starling, and budgerigar) for which behavioral data for both types of experiments were available. The psychoacoustical data is then compared to physiological measures of the cochlea (exemplified by the cochlear map function) in the animal species. The hypothesis under investigation is that different processing mechanisms are underlying the differences found between pure tone FDLs and FDLs for detecting a mistuned component in a complex in the animal species and that the exploitation of different mechanisms is a result of constraints in the auditory periphery.

The rationales of the behavioral experiments presented in Chapter 4 are twofold. In the first experiment it is examined whether changing the fundamental frequency of the harmonic complex results in changes in the threshold for detecting the mistuned harmonic in the complex in the gerbil. The results of this experiment are compared to those of the experiment presented in Chapter 2 using the sine phase harmonic complex. A simulation of the output signals of the estimated auditory filters of the gerbil (Kittel *et al.*, 2002) aimed at affirming the hypothesis that the gerbil exploits temporal cues for mistuning detection. The second experiment examines how onset cues in combination with harmonicity cues are processed in the gerbil auditory system. As discussed in Section 1.3, several human studies investigated onset asynchrony in a component of a harmonic complex as a cue exploited by the auditory system to segregate the asynchronous component from the remaining complex. The present experiment aimed at examining the effect of onset cues on the detection of a mistuned component in a harmonic complex. The hypothesis under investigation is that the onset asynchrony of the mistuned component determines the duration of the temporal overlap between mistuned component and remaining complex which influences the ability to detect changes in the temporal waveform that occur with mistuning a component of a complex.

The aim of the experiment in Chapter 5 is to evaluate how a combination of harmonicity and spatial cues affects the processing of tonal complex stimuli in the human auditory system. In a simultaneous masking experiment, the effect of harmonicity cues on the detection of a pure tone target in different harmonic and inharmonic complex maskers is determined. Furthermore, by spatially separating the target from the masking signal it is examined if the processing of harmonicity cues interacts with the processing of spatial cues in reducing the masking effect. Several studies that were discussed in Section 1.4 showed that mistuning a component provides an additional cue aiding the segregation of the component from the complex. However, the results were different with respect to the interaction between harmonicity cues and spatial cues. Based on the results of these studies it is hypothesized that cues provided by mistuning the target should have an additional effect on the detectability of the spatially separated target in the masker. It is investigated if harmonicity is a strong grouping cue interfering with spatial cues when detecting the target in a harmonic complex masker. A systematic investigation using different harmonic and inharmonic maskers and also high frequency targets and maskers will provide a more detailed view of the interaction of harmonicity cues and spatial cues in the auditory system.

1.7 References

Assmann, P. F. and Summerfield, Q. (**1990**). “Modeling the perception of concurrent vowels: vowels with different fundamental frequencies”, The Journal of the Acoustical Society of America **88**, 680–697.

Bleeck, S., Ingham, N. J., Verhey, J. L., and Winter, I. M. (**2008**). “Rebound depolarization in single units of the ventral cochlear nucleus: a contribution to grouping by common onset?”, Neuroscience **154**, 139–146.

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

Buell, T. N. and Hafter, E. R. (**1991**). “Combination of binaural information across frequency bands”, The Journal of the Acoustical Society of America **90**, 1894–1900.

Cheveigné, A. (**2005**). “Pitch perception models”, in *Pitch: Neural coding and perception*, edited by C. J. Plack, A. J. Oxenham, R. Fay, and A. Popper, 169–233 (Springer, New York).

Ciocca, V. and Darwin, C. J. (**1993**). “Effects of onset asynchrony on pitch perception: adaptation or grouping?”, The Journal of the Acoustical Society of America **93**, 2870–2878.

Darwin, C. J. (**1984**). “Perceiving vowels in the presence of another sound: constraints on formant perception”, The Journal of the Acoustical Society of America **76**, 1636–1647.

Darwin, C. J. (**1997**). “Auditory grouping”, Trends in Cognitive Sciences **1**, 327–333.

Darwin, C. J. and Ciocca, V. (**1992**). “Grouping in pitch perception: effects of onset asynchrony and ear of presentation of a mistuned component”, The Journal of the Acoustical Society of America **91**, 3381–3390.

Darwin, C. J. and Hukin, R. W. (**1998**). “Perceptual segregation of a harmonic from a vowel by interaural time difference in conjunction with mistuning and onset asynchrony”, The Journal of the Acoustical Society of America **103**, 1080–1084.

Dent, M. L., Dooling, R. J., and Pierce, A. S. (**2000**). “Frequency discrimination in budgerigars (*Melopsittacus undulatus*): effects of tone duration and tonal context”, The Journal of the Acoustical Society of America **107**, 2657–2664.

Deutsch, D. (**1999**). *Psychology of Music, Second Edition* (Academic Press, San Diego).

Dooling, R. J., Leek, M. R., Gleich, O., and Dent, M. L. (**2002**). “Auditory temporal resolution in birds: discrimination of harmonic complexes”, The Journal of the Acoustical Society of America **112**, 748–759.

Frisina, R. D., Smith, R. L., and Chamberlain, S. C. (**1990**a). “Encoding of amplitude modulation in the gerbil cochlear nucleus: I. A hierarchy of enhancement”, Hearing Research **44**, 99–122.

Frisina, R. D., Smith, R. L., and Chamberlain, S. C. (**1990**b). “Encoding of amplitude modulation in the gerbil cochlear nucleus: II. Possible neural mechanisms”, Hearing Research **44**, 123–141.

Gleich, O., Kittel, M. C., Klump, G. M., and Strutz, J. (**2007**). “Temporal integration in the gerbil: the effects of age, hearing loss and temporally unmodulated and modulated speech-like masker noises”, Hearing Research **224**, 101–114.

Gockel, H., Plack, C. J., and Carlyon, R. P. (**2005**). “Reduced contribution of a nonsimultaneous mistuned harmonic to residue pitch”, The Journal of the Acoustical Society of America **118**, 3783–3793.

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**, 1712–1724.

Hill, N. I. and Darwin, C. J. (**1996**). “Lateralization of a perturbed harmonic: effects of onset asynchrony and mistuning”, The Journal of the Acoustical Society of America **100**, 2352–2364.

Hirsh, I. J. (**1948**). “The influence of interaural phase on interaural summation and inhibition”, The Journal of the Acoustical Society of America **20**, 536–544.

Horst, J. W., Javel, E., and Farley, G. R. (**1986**). “Coding of spectral fine structure in the auditory nerve. I. Fourier analysis of period and interspike interval histograms”, The Journal of the Acoustical Society of America **79**, 398–416.

Horst, J. W., Javel, E., and Farley, G. R. (**1990**). “Coding of spectral fine structure in the auditory nerve. II. Level-dependent nonlinear responses”, The Journal of the Acoustical Society of America **88**, 2656–2681.

Hukin, R. W. and Darwin, C. J. (**1995**). “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**, 1380–1387.

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

Kittel, M., Wagner, E., and Klump, G. M. (**2002**). “An estimate of the auditory-filter bandwidth in the mongolian gerbil”, Hearing Research **164**, 69–76.

Klump, G. M. and Groß, S. (**submitted**). “Detection of frequency shifts and mistuning in complex tones in the european starling”, Journal of Comparative Physiology A .

Langemann, U. and Klump, G. M. (**1992**). “Frequency discrimination in the european starling (*Sturnus vulgaris*): a comparison of different measures”, Hearing Research **63**, 43–51.

LeGoff, N. and Kohlrausch, A. (**2005**). “The use of temporal cues for frequency discrimination of the fundamental component in a complex tone”, Proceedings of the Forum Acusticum, Budapest, 1503–1507.

Lin, J. Y. and Hartmann, W. M. (**1998**). “The pitch of a mistuned harmonic: evidence for a template model”, The Journal of the Acoustical Society of America **103**, 2608–2617.

Lohr, B. and Dooling, R. J. (**1998**). “Detection of changes in timbre and harmonicity in complex sounds by zebra finches (*Taeniopygia guttata*) and budgerigars (*Melopsittacus undulatus*)”, Journal of Comparative Psychology (Washington, D.C.: 1983) **112**, 36–47.

Maier, J. K. and Klump, G. M. (**2006**). “Resolution in azimuth sound localization in the mongolian gerbil (*Meriones unguiculatus*)”, The Journal of the Acoustical Society of America **119**, 1029–1036.

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**, 1593–1604.

McKeown, J. D. and Patterson, R. D. (**1995**). “The time course of auditory segregation: concurrent vowels that vary in duration”, The Journal of the Acoustical Society of America **98**, 1866–1877.

Micheyl, C. and Oxenham, A. J. (**2010**). “Pitch, harmonicity and concurrent sound segregation: psychoacoustical and neurophysiological findings”, Hearing Research **266**, 36–51.

Moore, B. C. and Glasberg, B. R. (**1989**). “Difference limens for phase in normal and hearing-impaired subjects”, The Journal of the Acoustical Society of America **86**, 1351–1365.

Moore, B. C., 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**, 479–483.

Moore, B. C., Glasberg, B. R., and Shailer, M. J. (**1984**). “Frequency and intensity difference limens for harmonics within complex tones”, The Journal of the Acoustical Society of America **75**, 550–561.

Moore, B. C., 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**, 1861–1867.

Plassmann, W., Peetz, W., and Schmidt, M. (**1987**). "The cochlea in gerbilline rodents", Brain, Behavior and Evolution **30**, 82–101.

Ryan, A. (**1976**). "Hearing sensitivity of the mongolian gerbil (*Meriones unguiculatis*)", The Journal of the Acoustical Society of America **59**, 1222–1226.

Scheffers, M. T. (**1984**). "Discrimination of fundamental frequency of synthesized vowel sounds in a noise background", The Journal of the Acoustical Society of America **76**, 428–434.

Schmiedt, R. A. and Zwislocki, J. J. (**1980**). "Effects of hair cell lesions on responses of cochlear nerve fibers. II. Single- and two-tone intensity functions in relation to tuning curves", Journal of Neurophysiology **43**, 1390–1405.

Schmiedt, R. A., Zwislocki, J. J., and Hamernik, R. P. (**1980**). "Effects of hair cell lesions on responses of cochlear nerve fibers. I. Lesions, tuning curves, two-tone inhibition, and responses to trapezoidal-wave patterns", Journal of Neurophysiology **43**, 1367–1389.

Simmons, A. M., Reese, G., and Ferragamo, M. (**1993**). "Periodicity extraction in the anuran auditory nerve. II: Phase and temporal fine structure", The Journal of the Acoustical Society of America **93**, 3374–3389.

Sinex, D. G., Guzik, H., Li, H., and Sabes, J. H. (**2003**). "Responses of auditory nerve fibers to harmonic and mistuned complex tones", Hearing Research **182**, 130–139.

Sinex, D. G., Sabes, J. H., and Li, H. (**2002**). "Responses of inferior colliculus neurons to harmonic and mistuned complex tones", Hearing Research **168**, 150–162.

Sinnott, J. M. and Brown, C. H. (**1997**). “Perception of the american english liquid /ra-la/ contrast by humans and monkeys”, The Journal of the Acoustical Society of America **102**, 588–602.

Sinnott, J. M., Brown, C. H., and Brown, F. E. (**1992**). “Frequency and intensity discrimination in mongolian gerbils, african monkeys and humans”, Hearing Research **59**, 205–212.

Sinnott, J. M. and Mosteller, K. W. (**2001**). “A comparative assessment of speech sound discrimination in the mongolian gerbil”, The Journal of the Acoustical Society of America **110**, 1729–1732.

Wagner, E., Klump, G. M., and Hamann, I. (**2003**). “Gap detection in mongolian gerbils (*Meriones unguiculatus*)”, Hearing Research **176**, 11–16.

Warren, R. M., Obusek, C. J., and Ackroff, J. M. (**1972**). “Auditory induction: perceptual synthesis of absent sounds”, Science (New York, N.Y.) **176**, 1149–1151.

In the context of this dissertation, the following peer-reviewed publications have been published or submitted:

1. Astrid Klinge and Georg M. Klump "Frequency difference limens of pure tones and harmonics within complex stimuli in Mongolian gerbils and humans" in *Journal of the Acoustical Society of America*, **125** (1), pp. 304−314 (2009).

2. Astrid Klinge, Naoya Itatani and Georg M. Klump "A Comparative View on the Perception of Mistuning: Constraints of the Auditory Periphery," in *The Neurophysiological Bases of Auditory Perception*, edited by E. A. Lopez-Poveda, A. R. Palmer, and R. Meddis (Springer, New York), pp. 465−475 (2010).

3. Astrid Klinge and Georg M. Klump "Mistuning detection and onset asynchrony in harmonic complexes in Mongolian gerbils" in *Journal of the Acoustical Society of America*, **128** (1), pp. 280−290 (2010).

4. Astrid Klinge, Rainer Beutelmann and Georg M. Klump "Effect of harmonicity on the detection of a signal in a complex masker and on spatial release from masking" in *Journal of the Association for Research in Otolaryngology*, *(submitted 25.06.2010).*

2 Frequency difference limens of pure tones and harmonics within complex stimuli in Mongolian gerbils and humans[1]

---

[1] This chapter has been published in the present form in JASA 125(1):304−314 (2009).

2.1 Abstract

Frequency difference limens (FDLs) for pure tones between 200 and 6400 Hz and for the first, the second, or the eighth harmonic of an 800 Hz complex in four Mongolian gerbils (*Meriones unguiculatus*) were determined using a Go/NoGo-procedure. The 12 harmonics of the complex started either in sine phase or at a random phase. Gerbils showed very high pure tone FDLs ranging from 17.1% Weber fraction (200 Hz) to 6.7% (6400 Hz). They performed much better in detecting mistuning of a harmonic in the complex in the sine phase condition with FDLs decreasing from 0.07% for the first harmonic to 0.02% for the eighth harmonic. FDLs were about one order of magnitude higher when temporal cues were degraded by randomizing the starting phase of every component in the harmonic complex for every stimulus. These results are strikingly different from those obtained in four human subjects who needed about four times higher frequency shifts than gerbils for detecting a mistuned component in a sine phase complex and showed similar detection of mistuning in the random phase condition. The results are discussed in relation to possible processing mechanisms for pure tone frequency discrimination and for detecting mistuning in harmonic complex stimuli.

2.2 Introduction

The natural environment is acoustically complex. The auditory system receives acoustic information from many different sound sources that often overlap in time and in frequency. The auditory system has the capability to segregate these sound sources and to group together sounds that belong to one source. Many of the natural sounds are not simple pure tones but are harmonic complexes composed of tonal components with a frequency that is an integer multiple of a fundamental frequency F0. Such harmonic complexes are produced by the human voice, musical instruments, and animal calls, but can even be generated incidentally by mechanical action of water or wind. The harmonic complex is generally perceived as one sound with a pitch corresponding to the fundamental frequency. Thus, harmonicity is a strong cue to group together frequencies aiding the segregation of sounds from different sources and the grouping of sounds from a single source to a perceived auditory object.

If a component of a harmonic complex is shifted in frequency, the percept of the harmonic complex changes for small frequency shifts until the mistuned component falls out of the harmonic series producing a separate auditory object at a large frequency shift. The "mistuned" harmonic paradigm is one way of studying the influence of harmonicity on the segregation of concurrent auditory objects (e.g., Moore *et al.* 1984, 1985, 1986; Lin and Hartmann, 1998). In such an experiment, a harmonic complex serves as a reference stimulus and in the inharmonic test stimulus one of the components of the harmonic complex is shifted in frequency. Moore *et al.* (1985) reported that the cues for detecting the inharmonicity differed for the lower and higher harmonics. Mistuned harmonics of low frequencies that are resolved by the auditory system "stand out" as a separate pure tone whereas harmonics with high unresolved frequencies are heard as a “beating” of the stimulus. These results indicate that in a harmonic complex with a mistuned component not only changes in the spectral domain but also changes in the temporal structure can be used to detect inharmonicity.

In this paper, we evaluate the perception of mistuned components in an otherwise harmonic complex in the Mongolian gerbil and compare these results to human data. Mongolian gerbils are a widely used animal model for studying auditory processing mechanisms in mammals since their hearing sensitivity in the low-frequency region is similar to that of humans. We measured the gerbil’s ability to detect a frequency shift in three behavioral experiments. In the first experiment, we obtained frequency difference limens (FDLs) for pure tones at different frequencies that were used as a reference for the two subsequent experiments. In the second and third experiment, FDLs for mistuned components of a harmonic complex stimulus were measured. The harmonic complex in the second experiment consisted of components that all started in sine phase. This resulted in a periodic temporal pattern that may provide a reference for detecting a change caused by the mistuning of a harmonic. If gerbils are especially sensitive to such temporal cues, we expect them to perform much better in detecting a frequency shift leading to the inharmonicity in a complex stimulus than in frequency discrimination of pure tones as has been found in birds (e.g., Lohr and Dooling, 1998). By randomizing the phase of every component in the harmonic complex in every presented stimulus during the third experiment, we provided a reference with a constant spectral composition but a varying temporal structure. We thereby made a change in the temporal pattern less useful as a cue. The comparison of the thresholds for detecting a frequency shift in the three conditions will provide insight into the mechanisms of processing of harmonic complexes that affects the formation of auditory objects.

2.3 Materials and methods

2.3.1 Animal subjects

Four adult Mongolian gerbils (*Meriones unguiculatus*) – three females, one male – were trained to serve as subjects in the experiments. For one of the females, we only obtained hearing thresholds and FDLs of mistuned harmonics in a complex with all harmonics starting in sine phase. During the time of testing, the animals were between 15 and 27 months old. All four gerbils had normal hearing within the frequency range of 100 Hz - 8 kHz (see below for further explanation). Animals were housed in individual cages (Type III, 42 × 26 × 15 cm^3) and had unlimited access to water. The daily amount of food was restricted to maintain a weight of approximately 80% of their ad libitum weight.

The care and treatment of the gerbils were in accordance with the procedures of animal experimentation approved by the Government of Lower Saxony, Germany. All procedures were performed in compliance with the NIH Guide on Methods and Welfare Considerations in Behavioural Research with Animals (NIH Publication No. 02-5083).

2.3.2 Apparatus and stimulus generation

Experiments were conducted in a single-walled sound attenuating booth (Industrial Acoustics, Type IAC 401-A). The interior of the chamber was lined with a 15 cm thick layer of sound absorbing foam with a pyramidal surface (Illbruck Illtec Pyramide 100/50, Illbruck Illtec PLANO Type 50/0). The reverberation time T_{60} of a broadband white noise measured through a microphone placed at the location where the head of the gerbil would normally be during the experiment was 12 ms and thus can still be referred to as anechoic.

The experimental cage was mounted on a rack in the middle of the chamber. The rack was built out of 5 mm thin metal bars that produced no substantial reflections. The cage was doughnut shaped and consisted of wire mesh (mesh size of 1 cm) on all sides. The floor of the cage was covered with fine wire mesh (mesh size of 1 mm) to provide for an even surface. Inside the cage, there was an elevated platform where the gerbils had to sit and wait for the test stimulus being played. Two light barriers were installed, one monitoring the pedestal and one the access to it. By forcing the gerbil to interrupt the light barriers in a specific sequence, we ensured that test stimuli were only presented if the gerbil was facing the loudspeaker. The loudspeaker (Canton Plus XS, frequency range: 150 Hz - 21 kHz) was mounted about 30 cm in front of the elevated platform at 0° azimuth and 0° elevation in relation to the head of the gerbil. For correct responses, food rewards (20 mg round pellets, Bioserve Dustless Precision Pellets Formula F0163-J50) were dispensed from a custom-built feeder. It was connected via a flexible tube to a food tray mounted close to the platform. A red feeder light emitting diode (LED) served as an additional reinforcer. Custom software controlled feeder and feeder light, registered switches of the light barriers, and generated and delivered the stimuli. Experiments took place with no visible light in the chamber. The sessions could be monitored by a charge coupled device camera and infrared LEDs.

All stimuli were generated using a Linux workstation with an AMD processor and an RME sound card (Hammerfall DSP Multiface II). The analog signal from the sound card (sample frequency of 44.1 kHz) was delivered to a manual attenuator (Kenwood type RA-920A, presetting of overall signal level) and from there to a programmable attenuator (PA5 from Tucker Davis Technologies System 3, for computer controlled variations in signal level). The signal was then passed through an acoustic isolation transformer (1:1) to an amplifier (Rotel High Current 8 Channel Power Amplifier RMB-1048) and from there to the free-field loudspeaker. Stimuli were monitored on an oscilloscope (Conrad Voltcraft Type 630-2).

The setup was calibrated before every testing day using a spectrum analyzer (Stanford Research SR780) in conjunction with a G.R.A.S. Type 21SP Acoustic Transducer (microphone Type 40AF plus preamplifier) which was permanently mounted next to the pedestal.

We conducted three experiments with three different stimuli. In the first experiment, we obtained FDLs for pure tones. Test frequencies were 200, 400, 800, 1200, 1600, 3200, 6000, and 6400 Hz. Tone duration was 400 ms including a 25 ms Hanning ramp at stimulus onset and offset. The overall sound-pressure level (SPL) of 60 dB was roved by ±3 dB to eliminate any possible level cue for the gerbils.

In the second and third experiment, we obtained FDLs for mistuned components in a harmonic complex. The harmonic complex consisted of the first 12 harmonics with a fundamental frequency F0 of 800 Hz and a SPL of 60 dB per component (with levels adjusted to the frequency response of the loudspeaker). The SPL of the overall complex was varied randomly by ±3 dB. As in the first experiment, the duration of the stimulus was 400 ms including a 25 ms Hanning ramp at onset and offset. In the reference stimuli, all components were harmonically related whereas in the test stimuli one of the harmonics was mistuned. For the second experiment, all 12 components of the complex started in sine phase [0°, Fig. 2.1(a)]. This resulted in reference stimuli that had exactly the same peaky waveform from one presentation to the next and the waveform was only different if a test stimulus with a mistuned harmonic was presented

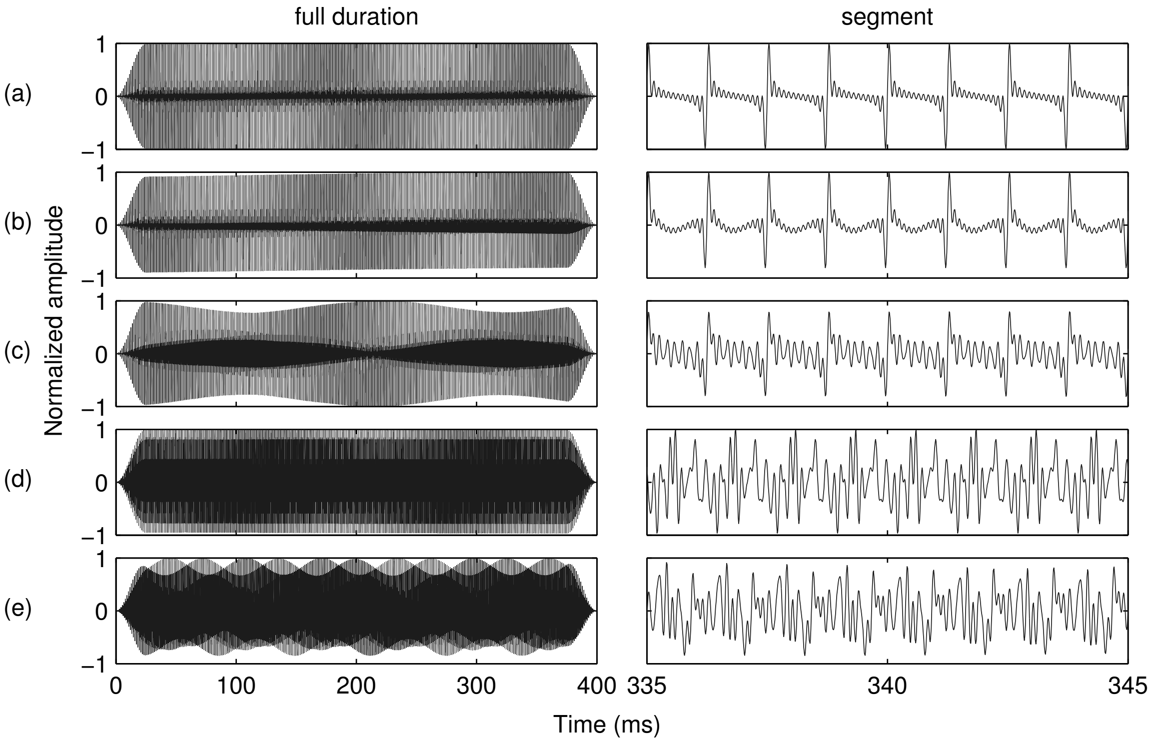

Figure 2.1: Complex stimulus consisting of 12 harmonics and a fundamental frequency of 800 Hz. Left column shows complete stimulus length of 400 ms, right column a close up to compare the changes in the fine structure. [(a)-(c)] harmonic complex stimulus with all harmonics starting at sine phase with (a) no mistuning, (b) first harmonic mistuned by 0.07% (0.6 Hz threshold value), and (c) eighth harmonic mistuned by 0.07% (4.7 Hz); [(d)-(e)] harmonic complex stimulus with all harmonics starting at random phase with (d) no mistuning and (e) first harmonic mistuned by 2% (15.8 Hz threshold value). Note the differences in fine structure of every stimulus.

[Figs. 2.1(b) and 2.1(c)]. For the third experiment, all components started at a random phase which resulted in reference stimuli [Fig. 2.1(d)] and test stimuli [Fig. 2.1(e)] with a different temporal pattern for every stimulus presentation. We determined FDLs for frequency shifts of the first (800 Hz), the second (1600 Hz), and the eighth (6400 Hz) harmonic.

2.3.3 Procedure

Animals were trained using operant conditioning with positive reinforcement (food pellets) in a Go/NoGo paradigm. During the session, we continuously repeated a reference stimulus every 1.3 s. The stimulus was either a pure tone (experiment 1) or a harmonic complex (experiments 2 and 3). To initialize a trial, the gerbil had to jump onto the pedestal and interrupt the light barrier. After a randomly chosen waiting interval between 1 and 7 s, the test stimulus was played instead of the reference stimulus. The test stimulus could either be the same as the reference stimulus during a catch trial or be different from the reference stimulus. For the first experiment, the test stimulus was a pure tone with a higher frequency than the reference stimulus. In the second and third experiment, the test stimulus was an inharmonic complex with one of the harmonics shifted upwards in frequency.

A correct response after a test stimulus ("Hit") was rewarded by switching on the feeder LED for 3 s and delivering a food pellet with a probability of 80%. Sessions consisted of a block of ten warm-up trials, followed by ten blocks of ten trials each. Within each block, three catch trials and seven different test trials were presented in a randomized order. Frequency increments in test trials consisted of fixed steps of a preselected Weber fraction. The step size was adjusted for the FDL measurements at the different harmonics. To provide at least some salient frequency differences to the animals, the step size for the two largest frequency increments that were not included in determining thresholds was double that of the step size used for determining thresholds. For example, a possible step size of 0.00005 Weber fraction around the threshold value (1.1 Hz in the 6400 Hz component) in the sine phase condition would result in frequency shifts of 0.32, 0.64, 0.96, 1.28, 1.60, 2.24, and 2.88 Hz for the seven test stimuli. In the block of warm-up trials, only the largest frequency increment was presented. The performance during these trials was not included in the data analysis.

2.3.4 Testing of human subjects

We obtained FDLs from four human subjects listening to the same harmonic complex stimuli that were generated by similar hardware as in the experiments with the gerbils. Subjects included the first author and three students, who took part in a practical course. All four subjects had normal hearing. Informed written consent was obtained from each subject. Except for the first author, the human subjects had no prior experience in psychoacoustic tasks. The students got course credits for completing the experiments. The experiments were conducted in a single-walled sound attenuating chamber (IAC Mini 250). The stimuli were presented via headphones (Sennheiser HDA 200). The headphones were calibrated using an artificial ear (Brüel&Kjær Type 4153) mounted on a microphone (Brüel&Kjær Type 4134) and a measuring amplifier (Brüel&Kjær Type 2610). All harmonics were presented at 60 dB SPL. The procedure being used in the gerbil experiments was adapted for humans, replacing the light barriers with a two-button response box. The human subjects had to press one of the buttons to initiate a trial and the other button to respond to a difference between reference and test stimulus. Feedback was given at correct responses by switching on a LED on the response box for 2 s.

2.3.5 Data analysis

A session was accepted as being valid if the subject responded correctly to at least 80% of the two test stimuli with the largest frequency difference and their false alarm rate did not exceed 20%. A psychometric function was constructed relating d' to the frequency shift in the test trials. The threshold was determined by linearly interpolating between adjacent points of the psychometric function as the amount of frequency shift resulting in a d' of 1.8 (Green and Swets, 1966). Two consecutive valid sessions in which thresholds differed no more than 1% Weber fraction for experiment 1 and experiment 3 and 0.1% Weber fraction for experiment 2 were combined to calculate the final FDL.

To exclude any training effects, we randomized the presentation order of every threshold for every animal within one experiment and repeated the first threshold after finishing each experiment. If the repeated threshold did not differ more than 3% (experiment 1), 1% (experiment 3), or 0.1% (experiment 2), respectively, from the threshold obtained the first time, then we assumed that no training effect affected the results. If it differed by more than that, the second threshold had to be repeated, then the third and so on until the repetition matched the original results. Data were always taken from the last measurement. The order in which the experiments were conducted was randomized and different for each animal. For a statistical analysis we used the statistical software package STATISTICA (StatSoft®, Version 7.0). Throughout the text, average results are reported as mean ± two times the standard error (mean± 2SE). All p-values are two tailed unless stated otherwise.

2.3.6 Prior testing of the hearing sensitivity of gerbils

Prior to the frequency discrimination experiments, we tested the hearing sensitivity of all four gerbils at the frequencies of 100, 500, 1000, 2000, 4000, and 8000 Hz. Testing took place in the same chamber with the same procedure as was used for the frequency discrimination experiments. No stimulus was played except in test trials in which a single 800 ms pure tone was presented as the test signal. During a catch trial, no signal was presented. The method of constant stimuli was used with a step size in signal level of 3 dB. A session was accepted if at least 80% of responses to the two stimuli with the highest level were "Hits" and no more than 20% false alarms were reported. The hearing threshold was defined as the SPL at a d' of 1.8 (Green and Swets, 1966). Two consecutive valid sessions in which thresholds differed no more than 3 dB were combined to calculate the final hearing threshold based on 20 responses from 20 trials at each level and 60 catch trials. The presentation order of all tested frequencies was randomized for every animal and the first threshold was repeated after finishing the test series of the experiment to exclude any training effects. If the repeated threshold differed more than 3 dB from the threshold obtained the first time, then the next threshold of the series was repeated. The range of tested frequencies with the best hearing was between 1 kHz (-1.0 ± 4.1 dB SPL, mean ± standard deviation) and 8 kHz (-8.9 ± 1.5 dB SPL) with an average of -5.8 dB SPL. At the frequencies of 2 kHz and 4 kHz, we determined hearing sensitivity thresholds of -4.7 ± 2.7 dB SPL and -8.4 ± 1.3 dB SPL, respectively. Below 1 kHz, the hearing sensitivity declined and thresholds increased up to 16.2 ± 3.7 dB SPL at a frequency of 500 Hz and 38.0 ± 1.3 dB SPL at 100 Hz. In the low-frequency range up to 1 kHz, the gerbils' hearing did not differ much from the hearing sensitivity thresholds obtained by Ryan (1976). At frequencies from 1 kHz to 8 kHz we observed between 7 and 14 dB SPL lower hearing thresholds than Ryan (1976). All four gerbils could be considered as having a normal hearing.

2.4 Results

2.4.1 Experiment 1: Pure tone frequency discrimination

Figure 2.2 shows the mean pure tone FDLs of the gerbils in percent Weber fraction (solid line with open squares) in relation to the reference frequency. The corresponding FDLs in Hertz are listed in Table 2.1. The relative frequency shift necessary for the detection decreased from a Weber fraction of 20.8% ± 10.4% (mean ± 2SE) at 200 Hz to 6.7% ± 1.6% at 6400 Hz. This corresponds to an absolute frequency difference of 41.7 ± 20.9 Hz at a reference frequency of 200 Hz. The absolute frequency shift necessary for detection increased gradually with increasing frequency of the reference signal up to 428.2 ± 100.2 Hz at 6400 Hz. A one-way repeated measure ANOVA revealed a significant effect of the reference frequency on the relative frequency shift at detection threshold ($p < 0.05; \eta^2 = 0.56$) as well as on the absolute frequency shift at detection threshold ($p < 0.05; \eta^2 = 0.95$). The *post hoc* pairwise comparisons revealed a significant difference between the mean thresholds for the absolute frequency discrimination determined for 800 and 6400 Hz reference tones (Tukey test, $p < 0.01$).

Table 2.1: FDLs of pure tones. Individual data of three gerbils and their mean with 2SE in Hz

Frequency (Hz)	Lv	Sa	Th	Mean ± 2SE
200	45	22	58	42 ± 21
400	67	25	56	49 ± 25
800	84	77	72	78 ± 7
1200	84	112	86	94 ± 18
1600	87	147	153	129 ± 42
3200	158	196	295	216 ± 81
6000	377	327	345	350 ± 29
6400	383	528	374	428 ± 100

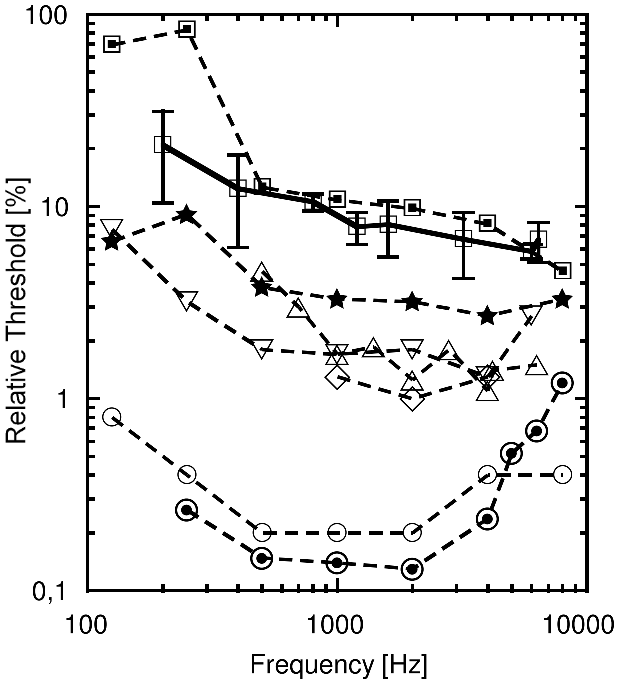

Figure 2.2: Relative thresholds of pure tone frequency discrimination of the gerbils of the present study (open squares, continuous line, error bars = 2SEs), in comparison to literature data: gerbil (filled squares), African monkey (*Cercopithecus mitis*, *Cercocebus albigena*, stars), and human (open circles): Sinnott *et al.* (1992); human (filled circles): Moore (1973); budgerigar (open diamond): Dent *et al.* (2000); starling (open standing triangles): Klump *et al.* (2000); and pigeon (open hanging triangles): Sinnott *et al.* (1980)).

None of the other *post hoc* comparisons revealed significant differences.

2.4.2 Experiment 2: Frequency difference limens of mistuned components in a sine phase harmonic complex

In this experiment, we presented a harmonic complex to the gerbils in which every component started in sine phase [0°, Fig. 2.1(a)]. Applying a frequency shift on one of the harmonics produced a change in the fine structure of the waveform over the duration of the stimulus and a slight amplitude modulation occurred [Figs. 2.1(b) and

2.1(c)]. This stimulus condition resulted in the lowest FDLs of all three experiments. Figure 2.3 shows the amount of mistuning at threshold expressed as the Weber fraction relative to the frequency of the respective harmonic. The corresponding absolute threshold values in hertz are shown in Table 2.2. The Weber fraction necessary for detecting the mistuning decreased from $0.073\% \pm 0.027\%$ (mean ± 2SE) at the 800 Hz harmonic to $0.017\% \pm 0.001\%$ at the 6400 Hz harmonic. Harmonic number significantly affected the Weber fraction necessary for detection of the mistuning (one-way repeated measure ANOVA, $p < 0.05, \eta^2 = 0.81$). Pairwise *post hoc* tests revealed a significant decrease of the threshold Weber fraction between the data for the 800 and 6400 Hz harmonic (Tukey test, $p < 0.01$) and a trend for the decrease comparing the 1600 and 6400 Hz harmonic (Tukey test, $p = 0.08$). The corresponding absolute frequency shift significantly increased from 0.59 Hz ± 0.21 Hz at the 800 Hz harmonic to 1.09 Hz ± 0.08 Hz at the 6400 Hz harmonic [one-way repeated measure analysis of variance (ANOVA), $p < 0.05, \eta^2 = 0.69$, a *post hoc* Tukey test revealed a significant difference between the data for the 800 Hz and the 6400 Hz harmonic, $p < 0.05$].

2.4.3 Experiment 3: Frequency difference limens of mistuned components in a random phase harmonic complex

During the third experiment, we randomized the starting phase of every component of the 800 Hz complex for every stimulus we presented throughout the experiment in order to degrade temporal cues. Figure 2.3 shows the relative frequency shifts necessary for the detection of a mistuning. The corresponding absolute frequency shifts at threshold are listed in Table 2.2. A one-way repeated measure ANOVA revealed a significant difference in the relative frequency shifts at the detection threshold from

2.0% ± 0.8% (mean ± 2SE) at the 800 Hz harmonic to 0.22% ± 0.16% at the 6400 Hz harmonic ($p < 0.05; \eta^2 = 0.79$). However, pairwise *post hoc* comparisons only revealed a significant decrease in the relative frequency shift necessary for the detection of a mistuning from the 800 Hz harmonic to the 6400 Hz harmonic (Tukey test, $p < 0.05$).

Table 2.2: FDLs for three harmonics of an 800 Hz complex with all components starting at sine phase and for three harmonics of an 800 Hz complex with all components starting at random phase. Individual data of three and four gerbils, respectively, and of four human subjects and their mean with 2SE in Hz.

GERBILS						
Sine phase condition						
Harmonic number	Frequency of harmonic (Hz)	Lv	Sa	Th	Ls	mean ± 2SE
1	800	0.51	0.60	0.36	0.87	0.59 ± 0.21
2	1600	0.44	0.97	0.60	0.99	0.75 ± 0.27
8	6400	1.14	1.03	1.18	1.01	1.09 ± 0.08
Random phase condition						
Harmonic number	Frequency of harmonic (Hz)	Lv	Sa	Th	Ls	mean ± 2SE
1	800	19.9	9.2	18.4		15.8 ± 6.7
2	1600	33.4	16.6	11.4		20.5 ± 13.3
8	6400	5.8	19.8	12.5		12.7 ± 8.1
HUMANS						
Sine phase condition						
Harmonic number	Frequency of harmonic (Hz)	AK	AV	NG	AT	mean ± 2SE
1	800	1.9	3.1	2.2	1.5	2.2 ± 0.7
2	1600	7.6	6.5	1.5	8.2	6.0 ± 3.0
8	6400	2.9	10.5	5.5	1.3	5.1 ± 4.0
Random phase condition						
Harmonic number	Frequency of harmonic (Hz)	AK	AV	NG	AT	mean ± 2SE
1	800	1.9	3.5	2.5	5.4	3.3 ± 1.5
2	1600	9.8	4.7	8.9	10.5	8.3 ± 2.6
8	6400	2.4	15.1	7.1	5.5	8.8 ± 5.4

All other pairwise comparisons were not significant. A one-way repeated measure ANOVA revealed no significant effect of the harmonic on the absolute frequency shift in Hertz.

2.4.4 Comparison of all three experiments

Comparing the results of all three experiments with each other (Fig. 2.3), we observed a highly significant difference between the mean relative frequency shift at the detection threshold among the three different experimental conditions. A two-way repeated measure ANOVA with experimental conditions (stimulus type) and reference frequency as factors including the data from the three gerbils that completed all experiments resulted in values of $p < 0.001$ and $p < 0.05$, respectively. The stimulus type had a

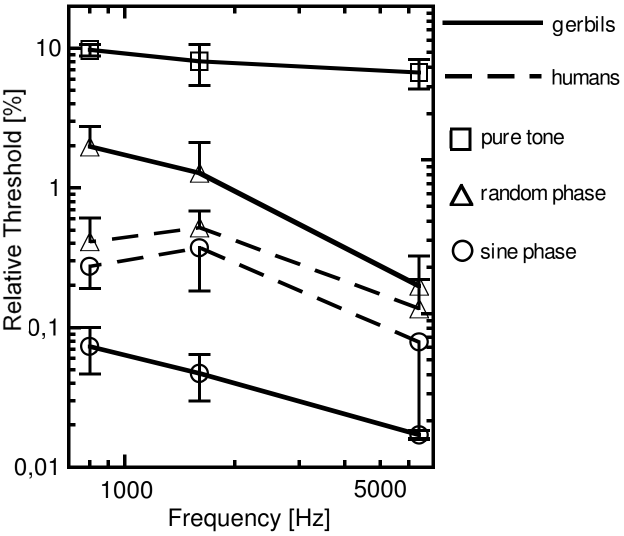

Figure 2.3: Mean thresholds for gerbils and human subjects tested on the frequency discrimination ability of pure tones (only gerbils, squares) and on the frequency discrimination ability of the mistuned 800, 1600, and the 6400 Hz harmonic within a complex with all harmonics starting at random phase (triangles) and within a complex with all harmonics starting at sine phase (circles). Error bars represent 2 SF.

strong effect on the discrimination threshold (effect size $\eta^2 = 0.99$). The thresholds for the relative frequency shifts were the lowest for harmonics in sine phase stimuli, about one order of magnitude higher for harmonics in random phase stimuli, and about another order of magnitude higher for pure tones (all pairwise planned comparisons of the experimental condition within a reference frequency revealed significant differences, all $p < 0.05$). The second main effect reference frequency also showed a high effect size η^2 of 0.78. There was no significant interaction between the factors experimental condition and reference frequency (two-way repeated measure ANOVA, $p = 0.26$).

The performance of the gerbils during the sessions was different for the second and third experiments. Corresponding to the large difference in FDL, it was more difficult to obtain data in the random phase condition, than in the sine phase condition. In the sine phase condition, the gerbils needed on an average five sessions to obtain a threshold which was significantly less than in the random phase condition with an average of 14 sessions to obtain a threshold (paired t-test, $p < 0.05$). When the gerbil switched from the sine phase condition to the random phase condition, their false alarm rate was significantly increased (from 98 to 160 false alarms, averaged over five sessions of two gerbils, paired t-test, $p < 0.05$), whereas a slight decrease was observed when switching from random phase to sine phase (from 104 to 101 false alarms, averaged over five sessions of one gerbil, paired t-test, $p = 0.92$).

2.4.5 Comparison with human data

Human FDLs for frequency shifts in the harmonic complex were obtained by using the same procedure and stimuli as for the gerbils to be able to compare FDLs directly (Table 2.2, Fig. 2.3). A two-way repeated measure ANOVA with experimental condition (sine phase and random phase) and reference frequency (800, 1600, and 6400 Hz) as factors revealed no significant difference between the Weber fraction of the shift at threshold for sine phase and random phase stimuli ($p = 0.129$). However, the reference frequency as a main effect had a significant effect on the Weber fraction at threshold ($p < 0.01, \eta^2 = 0.79$). *Post hoc* tests revealed significant differences for the comparison of the data obtained for reference frequencies of 6400 and 1600 Hz (Tukey test, $p < 0.01$) and for reference frequencies of 6400 and 800 Hz (Tukey test, $p < 0.05$). No significant interactions between reference frequency and experimental condition were found. Neither experimental condition nor reference frequency had a significant effect on the absolute frequency differences necessary for the detection of the mistuning (two-way repeated measure ANOVA).

For sine phase stimuli, the Weber fraction of the shift at threshold was smaller in gerbils than in humans, i.e., gerbils were more sensitive for detecting the mistuning. This difference was significant for the 800 Hz harmonic (t-test, $p < 0.005$) and the 1600 Hz harmonic (t-test, $p < 0.02$), and there was a trend in the same direction for the 6400 Hz harmonic (t-test, $p = 0.096$). For random phase stimuli, the Weber fraction of the shift at threshold was larger in gerbils than in humans. This difference was significant for the 800 Hz harmonic (t-test, $p < 0.005$) and there was a trend for the 1600 Hz harmonic (t-test, $p = 0.091$).

2.5 Discussion

2.5.1 Pure tone frequency difference limens

Compared to the discrimination ability of other animal species (Fig. 2.2), the gerbil exhibits a poor ability to discriminate the frequency of pure tones (Fig. 2.2, see also Sinnott *et al.*, 1992), especially in the low-frequency region. Human data are shown from Moore (1973) and Sinnott *et al.* (1992) for comparison. The highest frequency discrimination acuity in humans being as low as 0.2% shift in frequency is found between 500 Hz and 2 kHz. Bird species vary in their ability to discriminate between pure tones of different frequencies. Pigeons, for example, exhibit fairly high FDLs at frequencies higher than 1 kHz that are, however, within the range of nonhuman mammals (Sinnott *et al.*, 1980). Klump *et al.* (2000) reported pure tone FDLs in European starlings from 4.6% for a frequency of 500 Hz to 1.5% for 6300 Hz. Studying budgerigars, Dent *et al.* (2000) found FDLs of around 1.3% at frequencies of 1, 2, and 4 kHz. The SPLs used for the experiments in the different studies were all around 60-65 dB SPL which compares well to the sensation level used in the present study.

Pure tone frequency discrimination can be achieved based on two different mechanisms. The first mechanism applies to the low-frequency region where phase locking of auditory nerve (AN) fibers to the stimulus waveform occurs. For these frequencies, the distribution and change of interspike intervals in the AN fibers can be exploited to estimate the change in frequency (e.g., Wever, 1949). The second mechanism mainly applies to the high frequency region where phase locking is no longer possible. It evaluates the place of excitation in the cochlea for frequency discrimination. Ideas for a place mechanism in which frequencies are represented tonotopically along the basilar membrane reach as far back as to the 17th century (Du Verney, 1683, c.f. Cheveigné, 2005). Later on, von Békésy (1960) and Zwicker (1970) proposed a place mechanism in which a frequency is represented as a certain excitation pattern along the basilar membrane. The detection of a change in frequency is based on changes in the excitation pattern.

Comparing a wide range of species, Fay (1992) favored a place mechanism in most animal species. He stated, based on analyses of cochlear place-frequency functions by Greenwood (1961, 1990), that perceptual units in the frequency domain such as the pure tone FDL, the critical bandwidth (CB), critical masking ratio bandwidth, and the psychophysical tuning curve can all be related to the spatial frequency map on the basilar membrane. In the Mongolian gerbil, a physiologically determined cochlear map (Müller, 1996) fitted by the cochlear place-frequency function (Greenwood, 1961) can be used to calculate the relation between the critical ratio bandwidth or the CB and the FDL (Kittel *et al.*, 2002). Figure 2.4 shows an estimate of the CB as a function of frequency (gray line) in comparison to the FDLs determined in this study. CBs were calculated from a cochlear placefrequency function that was fitted to the behavioral estimates of the CB (Kittel *et al.*, 2002). They range from 56 Hz at a center frequency of 200 Hz to 638 Hz at 6400 Hz. The function of the observed FDLs (black dotted line on top of the gray line) runs parallel to the CB function (i.e., it is offset by a scaling factor of 1.5) over the whole range of tested frequencies (Fig. 2.4, axes were adjusted accordingly). This parallel run of the CB function and the pure tone FDL function in the gerbil suggests that at least over the tested frequency range of 200-6400 Hz, a pure place mechanism might explain the observed FDLs.

What remains unexplained is the small scaling factor relating the FDL to the CB of the gerbil. Fay (1992) showed that the best FDLs in humans are related to the CB by a factor of 50. He proposed that this scaling factor applies to most mammals

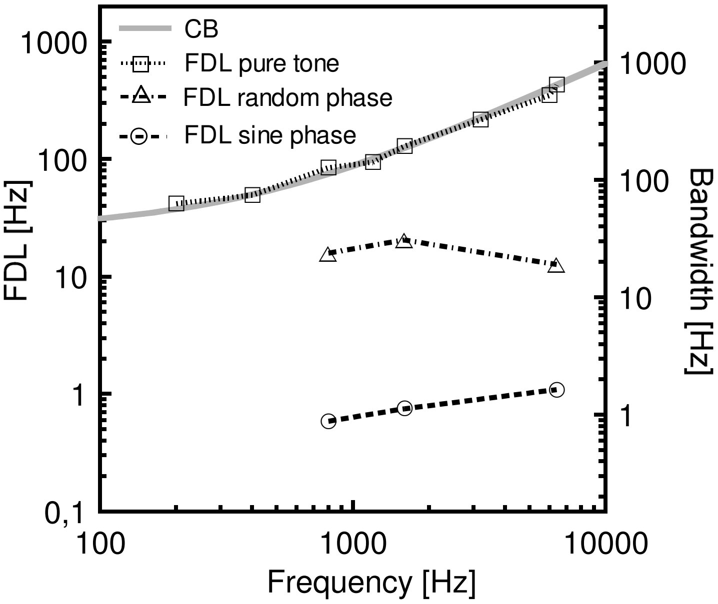

Figure 2.4: Pure tone FDLs (black dotted line, squares), random phase FDLs (black dash-dotted line, triangles), and sine phase FDLs (black dashed line, circles) of gerbils are plotted in comparison to the cochlear place-frequency function (gray solid line). The CB function is derived from the general bandwidth function from Greenwood (1961) with parameters obtained from physiological data of gerbils (Müller, 1996).

and birds at least for frequencies above 1 kHz. However, Sinnott *et al.* (1992) argued against this scaling factor by presenting data from African monkeys (*Cercopithecus mitis*, *Cercocebus albigena*) that show much larger pure tone FDLs than humans in spite of having a CB that is similar to the human CB. The FDLs observed in monkeys can only be matched to the CB function by a scaling factor of 3 (see Sinnott *et al.*, 1992) which is more similar to the scaling factor of 1.5 found for the gerbil in the present study. Studying humans, Sek and Moore (1995) related different measures of frequency discrimination to the auditory filter bandwidth. They observed that the scaling factor is reduced considerably at frequencies for which humans have to rely on a place mechanism for frequency analysis. It deteriorated from about 18 at 4 kHz to about 5 at 8 kHz. This is similar to the range of scaling factors from about 7 to 17 that was found by Buus *et al.* (1995) fitting the FDL data from European starlings to the starling's CB function. These scaling factors, however, are much larger than the value found for the gerbil.

What may determine the value of the scaling factor and the value of the FDL that can vary considerably between species? Fay (1992) proposed that the largest absolute frequency discrimination acuity will be found in mammals with a poor high frequency hearing (thus perceiving a small frequency range) and a long cochlea. Frequency discrimination ability will be poor when the cochlea is short but the frequency range of sensitive hearing is wide. The gerbil cochlea is only 11 mm long (Müller, 1996) which is short in comparison to 34 mm reported for the human cochlea (Miller, 2007). Corresponding to the difference in length, the number of hair cells along the basilar membrane is reduced from about 3500 inner hair cells (IHCs) in the human cochlea (Wright *et al.*, 1987) to about 1400 IHCs in the gerbil cochlea (Plassmann *et al.*, 1987). Moreover, the hearing range of the gerbil extending from approximately 0.1-60 kHz (Ryan, 1976) is much larger than the human hearing range extending from about 0.01 to 16 kHz (ISO 389-7, 1996). Furthermore, the gerbil's larger frequency range is mapped onto more critical bands than the frequency range of humans (estimated on the basis of data from Kittel *et al.*, 2002 for gerbils, and Greenwood, 1961 and Moore and Glasberg, 1983 for humans). The distance on the cochlear map covered by a critical band in the gerbil is about 0.21 mm which is small compared to the distance of 1.15 mm estimated for humans (Fay, 1992). Thus, a specific shift in frequency relative to the reference will result in a smaller change of the spatial pattern of excitation in the cochlea involving a much smaller number of hair cells in gerbils compared to humans (a difference by a factor of 5-8). The difference between human and gerbil FDLs may thus be related to the higher number of hair cells in humans and encoding of frequencies by a larger number of AN fibers. This might result in a more fine-scaled frequency representation in the cochlea.

2.5.2 Frequency difference limens of mistuned components in a harmonic complex

The results of the “mistuned harmonic” experiments appeared in stark contrast to the pure tone FDLs. While the performance of the gerbils in pure tone frequency discrimination was unusually poor and about 20 times worse than the performance of humans, gerbils were about four times better than humans in detecting a mistuned harmonic in the sine phase condition. When gerbils detected a mistuned harmonic in a harmonic complex with all components starting at a random phase, FDLs were about one order of magnitude worse than FDLs obtained in the sine phase condition. This makes gerbils about two to four times less sensitive in reporting the mistuning compared to humans in the random phase condition.

Models for pitch perception and frequency discrimination

Two classes of models have been applied to explain pitch perception and discrimination of complex tone stimuli: (1) spectral pattern recognition models and (2) autocorrelation models. The spectral pattern recognition models (early models by Goldstein, 1973, Wightman, 1973, and Terhardt, 1974) assume that the frequency of every harmonic is determined either by place or by temporal mechanisms and then the best fitting fundamental frequency is estimated. An internal central processor or pattern recognizer decides if the frequency of a component of the complex is belonging to the harmonic series or not and has therefore been called a “harmonic sieve” by Duifhuis *et al.* (1982). One constraint of this type of model is, however, that the harmonics of the complex have to be resolved. Furthermore, the phase of components of complex stimuli is not taken into account as the pitch is extracted from the peripheral excitation pattern. A spectral pattern recognition model thus implies that the frequency selectivity of the basilar membrane has to be sufficiently good in order to reach low FDLs. To evaluate the involvement of place mechanisms in the frequency discrimination ability of mistuned harmonics in a harmonic complex, we plotted the FDLs of the random phase condition and of the sine phase condition of our study in relation to the placefrequency map in Fig. 2.4. If a place model would explain the FDLs for mistuned harmonics in complex stimuli at both condition, then the FDLs would be parallel to the cochlear place-frequency function determined by Greenwood (1961). The FDLs of mistuned harmonics at the random phase and the sine phase condition cannot be fitted to the cochlear map function by the same scaling factor at least over the tested frequency range which makes place mechanisms more unlikely to operate over the whole frequency range tested. Furthermore, we would expect much higher FDLs for mistuned harmonics in a harmonic complex in gerbils than in humans due to the gerbil’s poor pure tone FDLs, if we assume that the pattern recognition model applies. Our results do not support the prediction based on the spectral pattern recognition model.

Instead, the FDLs of mistuned harmonics in a complex with all harmonics starting at sine phase are two orders of magnitude smaller than pure tone FDLs. This applies not only to the 6400 Hz harmonic that is unresolved in the gerbil auditory system but also for the 800 and the 1600 Hz harmonic that are resolved. Detecting an inharmonicity in the sine phase complex stimuli by the Mongolian gerbil therefore seems not to involve the analysis of the spectral pattern of a complex stimulus or the evaluation of temporal patterns within a single auditory filter but rather points to processing of temporal patterns beyond the stage of the basilar membrane.

Most of the temporal models for the processing of harmonic tone complexes rely on autocorrelation which is computed on the basis of the temporal envelope or temporal fine structure in all frequency channels. Such models were first introduced by Licklider (1951) and further developed by Meddis and Hewitt (1991a, 1991b), Meddis and O'Mard (1997), Patterson *et al.* (1995), and Denham (2005) using a summary autocorrelation function over all channels. The autocorrelation function suggested by Licklider (1951) is calculated on the basis of spike occurrences from a single neuron or a group of neurons and depends on the waveform envelope of unresolved harmonics to extract the pitch of a complex. A summary autocorrelation function combines the periodicity estimates derived from all frequency channels and therefore they do not depend only on unresolved components of a harmonic complex.

Temporal cues available for detecting a mistuned harmonic in a complex

Temporal cues that might be used for detecting mistuning of a component in a complex with all harmonics in sine phase could be a beating in the mistuned harmonic complex or a gradual phase shift over the duration of the stimulus resulting in a change of the fine structure of the ongoing stimulus. A beating is produced when the frequency-shifted harmonic in the mistuned harmonic complex is unresolved and interacts with the adjacent harmonics in the complex. Human subjects in the study by Moore *et al.* (1985) identified the mistuning of unresolved harmonics by hearing a kind of "beat." This detection based on the amplitude modulation suggests that for humans temporal cues may play a greater role than spectral cues for detecting an unresolved mistuned harmonic in a complex stimulus. Temporal mechanisms for detecting a mistuned component in an otherwise harmonic complex in humans have also been suggested by Le Goff and Kohlrausch (2005). Their simulations showed that the excitation patterns for flat-spectrum and sloped-spectrum complex stimuli are less affected by mistuning the fundamental frequency component than the excitation pattern of the frequency-shifted pure tone. In a simulation of the envelope pattern of the output signal of the auditory filters, however, Le Goff and Kohlrausch (2005) showed that mistuning the fundamental frequency component in a sloped spectrum changed the temporal envelope of the signals in the auditory filters considerably. This points to the use of temporal mechanisms by humans even for resolved mistuned harmonics under certain circumstances.

Some authors proposed the evaluation of distortion products producing beating as a possible explanation for the good performance in detecting a mistuned harmonic in a complex (Moore *et al.*, 1985; Pressnitzer and Patterson, 2001). Distortion products emerge at the point on the basilar membrane where the traveling waves of at least two nearby tones (with frequencies $f1$ and $f2$) overlap. Generally, the most prominent distortion product is the cubic $2f1 - f2$ distortion which may under certain circumstances interact considerably with components of the presented stimulus on the basilar membrane. We estimated the possible interaction of the mistuned harmonic and the cubic distortion product which would develop at the place where the shifted harmonic has been. Taking into account measurements of distortion product otoacoustic emissions and neuronal responses by Faulstich and Kössl (1999) and Mom *et al.* (2001), we concluded that the SPL of the distortion product would be at least 20 dB lower than the level of the interacting mistuned harmonic. The estimated modulation depth due to the beating between the distortion product and the mistuned harmonic was around 1.1 dB. Since the gerbil's intensity discrimination threshold is at least 3 dB (Sinnott *et al.*, 1992), we conclude that the depth of modulation resulting from the beating of a distortion product with the mistuned component would not allow the gerbils to detect the beating and to use it as a detection cue.

A more likely temporal cue available to the gerbils for detecting the mistuning is an ongoing change in the fine structure over the duration of the mistuned harmonic complex. The change results from the gradual phase shift of the mistuned component compared to the other components. The peaky waveform of the sine phase stimulus may aid the gerbil in using this cue. In the random phase condition, however, only some of the randomly produced waveforms have a peaky temporal structure. Sensitivity to the phase or to changes of the phase in harmonic complex stimuli has been proposed in several physiological and psychophysical studies (e.g., Bilsen, 1973; Horst *et al.*, 1986, 1990; Moore and Glasberg, 1989; Simmons *et al.*, 1993). Phase relations might be detected on the basilar membrane when unresolved harmonics fall within one auditory filter. Phase relations for resolved harmonics might be preserved and detected in more central stages of the auditory system where they can be analyzed across channels. Hartmann (1988) showed that reversing the phase of the resolved fourth harmonic in comparison to the other components of a 200 Hz complex influenced the detection of a mistuning in relation to the duration of the stimulus. Since the mistuned component with the phase shift of 180° and the adjacent harmonics fall in different auditory filters, a processing of the phase relationship between the mistuned harmonic and the adjacent components is less likely to be analyzed on the basilar membrane. Thus, the phase effects observed by Hartmann (1988) suggest that the processing of these temporal cues in humans is not limited to cochlear within-channel mechanisms.

A further improvement for detecting a mistuned harmonic in the sine phase compared to the random phase stimulus could be explained by the stimulus statistics. In the sine phase condition, a constant temporal pattern is provided with every background stimulus that only changes during the presentation of a test stimulus. In the random phase stimulus condition, we continuously varied the temporal fine structure and phase relations by randomizing the starting phase of every harmonic in every presented stimulus. Changing the temporal fine structure with every presented stimulus removed the possibility to compare temporal patterns of the harmonic complex sequentially. This high level of uncertainty in the stimuli might have resulted in the observed decline of FDLs in the random phase condition compared to the sine phase condition. The observation that gerbils needed more sessions to obtain a threshold in the random phase condition than in the sine phase condition and that they showed an increased false alarm rate when switching from the sine phase to the random phase condition also suggests that available cues in the random phase condition may have been less salient.

The amplitude modulation of the stimulus envelope resulting from mistuning could provide another possible cue for detecting a mistuned harmonic in a complex in the random phase condition. The depth of the amplitude modulation, however, has to be sufficiently large to be perceived by the gerbil. In the sine phase condition, the FDLs are too small to produce an exploitable modulation. Also in the random phase condition, the modulation spectrum of the mistuned complex stimulus at threshold value indicated that the modulation depth is not large enough to be detectable by the gerbil.

2.5.3 Comparison with other species

Figure 2.5 gives an overview of FDLs of mistuned components in harmonic complexes in several animal species and humans. Zebra finches and budgerigars exhibit very low FDLs for detecting mistuning of components in harmonic complexes in the sine phase condition (Lohr and Dooling, 1998) that were similar to the FDLs of gerbils determined in the present study. An increased FDL in the random phase condition compared to the sine phase condition could be shown for zebra finches in the study by Lohr and Dooling (1998). The European starling also exhibits discrimination thresholds that are similar to those found in gerbils (Klump *et al.*, 2008). The FDLs in the bird species were significantly lower compared to those of humans when all components of the complex stimulus started in sine phase. The studies show that gerbils are not the only species that have such low FDLs for detecting a frequency shift in harmonic complexes in the sine phase condition. The difference in the salience of mistuning in the sine phase and the random phase conditions in gerbils and birds suggests that temporal fine structure of the stimuli may be of particular importance. Sinnott and Mosteller (2001) found gerbils to be equally sensitive in detecting frequency changes in speech formants compared to detecting a frequency shift of a harmonic in the random phase experiment as determined in the present study. Threshold values for detecting frequency changes in speech formants were similarly decreased compared to pure tone FDLs in gerbils as is described here for detecting a mistuning of a harmonic in a random phase complex. Furthermore, the ratio between the gerbil and human formant FDLs in the speech discrimination task and between the gerbil and human FDL in the random phase condition was found to be similar. These comparisons could be interpreted as a hint that humans and gerbils may employ similar mechanisms in processing frequency

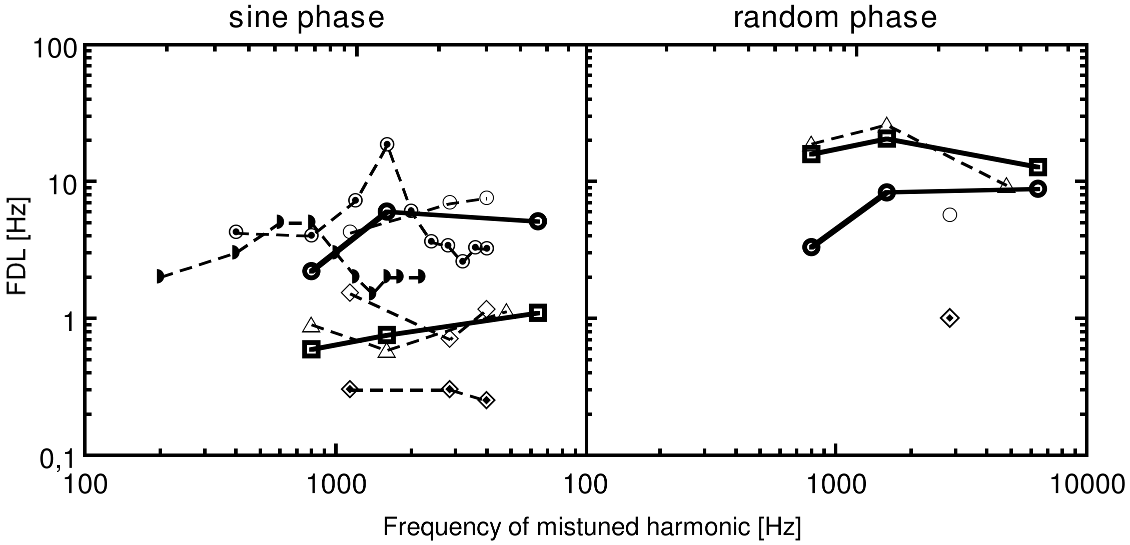

Figure 2.5: Absolute thresholds in Hz of mistuned components of sine phase (left panel) and random phase (right panel) complex stimuli of gerbils (thick square, continuous line) and humans (thick circle, continuous line) determined in the present study in comparison to literature data. Zebra finch (filled diamond), budgerigar (open diamond) and human (open circle): Lohr and Dooling (1992); starling (open triangle): Klump *et al.* (2008); human (filled circle): Moore *et al.* (1985); human (half circle): Lee and Green (1994).

changes in both experiments. However, it may well be that both species use different mechanisms resulting in a similar sensitivity for detecting the frequency shift.

Human subjects were not as sensitive as gerbils or zebra finches (Lohr and Dooling, 1998) to the loss of distinctive temporal features of the stimulus during the presentation of random phase harmonic complexes. In humans, the FDLs for mistuned harmonics in harmonic complexes with all components starting at random phase were never significantly different to the FDLs for mistuned harmonics in the sine phase condition (see present study or Lohr and Dooling, 1998). These results are supported by studies from Dooling *et al.* (2002) and Patterson (1987) who showed that humans cannot discriminate cosine phase complex stimuli from random phase stimuli at a fundamental frequency of 800 Hz. This points to a reduced importance of temporal fine structure as a cue in humans compared to gerbils and birds.

Which constraints may bias gerbils and birds to rely more and humans to rely less on temporal cues when detecting mistuning of components in a harmonic complex? The length of the cochlea may hint at a possible explanation. The Mongolian gerbil has a short cochlea compared to the human cochlea with much fewer hair cells covering a CB. Thus, the same change in frequency in a gerbil and a human cochlea results in a much smaller change of the spatial pattern of excitation in the gerbil than in humans. In order to compensate for this, the gerbil and possibly the birds (also having a short cochlea) might have switched to exploiting temporal cues in complex stimuli to detect small amounts of mistuning. However, by proposing the use of temporal fine structure as a cue the question remains why gerbils seem not to exploit any temporal cues for detecting frequency differences in pure tones (at least in the frequency range below 2 kHz for which this is suggested in humans, e.g., Moore, 1973). Since humans have a much longer cochlea than birds or the gerbil, the evolutionary pressure to use changes in the temporal fine structure as a cue in the processing of frequency shifts in harmonic complexes may have been much lower than in the other species discussed here.

2.6 Acknowledgements

We thank Susanne Groß and Anne Vellage for helping with the experiments, Antje Tjarks, Nantje Gelszat, and Anne Vellage for being motivated subjects in the human experiments, Holger Dierker for technical support, Stephan Ewert for assisting in the reverberation measurement, and Stefan Strahl for helping out in computational problems. This study was supported by the DFG within the Sonderforschungsbereich

SFB TR31 “The Active Auditory System” and the International Graduate School GRK 591 “Neurosensory science, systems and applications”.

2.7 References

Bilsen, F. A. (**1973**). “On the influence of the number and phase of harmonics on the perceptibility of the pitch of complex signals,” Acustica **28**, 60-65.

Buus, S., Klump, G. M., Gleich, O., and Langemann, U. (**1995**). “An excitation-pattern model for the starling (*Sturnus vulgaris*),” J. Acoust. Soc. Am. **98**, 112-124.

Cheveigné, A. (**2005**). “Pitch perception models,” in *Pitch: Neural Coding and Perception*, edited by C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper, (Springer, New York), pp. 169-233.

Denham, S. L. (**2005**). “Pitch detection of dynamic iterated rippled noise by humans and a modified auditory model,” BioSystems **79**, 199-206.

Dent, M. L., Dooling, R. J., and Pierce, A. L. (**2000**). “Frequency discrimination in budgerigars (*Melopsittacus undulatus*): Effects of tone duration and tonal context,” J. Acoust. Soc. Am. **107**, 2657-2664.

Dooling, R. J., Leek, M. R., Gleich, O., and Dent, M. L. (**2002**). “Auditory temporal resolution in birds: Discrimination of harmonic complexes,” J. Acoust. Soc. Am. **112**, 748-759.

Duifhuis, H., Willems, L. F., and Sluyter, R. J. (**1982**). “Measurement of pitch in speech: An implementation of Goldstein’s theory of pitch perception,” J. Acoust. Soc. Am. **71**, 1568-1580.

Faulstich, M., and Kössl, M. (**1999**). “Neuronal response to cochlear distortion products in the anteroventral cochlear nucleus of the gerbil,” J. Acoust. Soc. Am. **105**, 491-502.

Fay, R. R. (**1992**). “Structure and function in sound discrimination among vertebrates,” in *The Evolutionary Biology of Hearing*, edited by D. Webster, R. R. Fay, and A. Popper (Springer, New York), pp. 229-263.

Goldstein, J. L. (**1973**). “An optimum processor theory for the central formation of the pitch of complex tones,” J. Acoust. Soc. Am. **54**, 1496- 1516.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Greenwood, D. D. (**1961**). “Critical bandwidth and the frequency coordinates of the basilar membrane,” J. Acoust. Soc. Am. **33**, 1344-1356.

Greenwood, D. D. (**1990**). “A cochlear frequency-position function for several species–29 years later,” J. Acoust. Soc. Am. **87**, 2592-2605.

Hartmann, W. M. (**1988**). “Pitch perception and the segregation and integration of auditory entities,” in *Auditory Function*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 623-645.

Horst, J. W., Javel, E., and Farley, G. R. (**1986**). “Coding of spectral fine structure in the auditory nerve. I. Fourier analysis of period and interspike interval histograms,” J. Acoust. Soc. Am. **79**, 398-416.

Horst, J. W., Javel, E., and Farley, G. R. (**1990**). “Coding of spectral fine structure in the auditory nerve. II. Level-dependent nonlinear responses,” J. Acoust. Soc. Am. **88**, 2656-2681.

ISO 389-7, (**1996**). “Acoustics–Reference zero for the calibration of audiometric equipment–Part 7: Reference threshold of hearing under free-field and diffuse-field listening conditions."

Kittel, M., Wagner, E., and Klump, G. M. (**2002**). "An estimate of the auditory-filter bandwidth in the Mongolian gerbil," Hear. Res. **164**, 69-76.

Klump, G. M., Groß, S., and Itatani, N. (**2008**). "Processing of a mistuned harmonic by the European starling auditory system," 31st MidWinter Research Meeting [J. Assoc. Res. Otolaryngol. **31**, 139].

Klump, G. M., Langemann, U., and Gleich, O. (**2000**). "The European starling as a model for understanding perceptual mechanisms," in *Auditory Worlds: Sensory Analysis and Perception in Animals and Man*, edited by G. A. Manley, H. Fastl, M. Kössl, H. Oeckinghaus, and G. M. Klump (Wiley-VCH, Weinheim), pp. 193-211.

Le Goff, N., and Kohlrausch, A. (**2005**). "The use of temporal cues for frequency discrimination of the fundamental component in a complex tone," Proceedings of the Forum Acusticum, Budapest, pp. 1503-1507.

Lee, J., and Green, D. M. (**1994**). "Detection of a mistuned component in a harmonic complex," J. Acoust. Soc. Am. **96**, 716-725.

Licklider, J. C. (**1951**). "A duplex theory of pitch perception," Experientia **7**, 128-134.

Lin, J.-Y., and Hartmann, W. M. (**1998**). "The pitch of a mistuned harmonic: Evidence for a template model," J. Acoust. Soc. Am. **103**, 2608-2617.

Lohr, B., and Dooling, R. J. (**1998**). "Detection of changes in timbre and harmonicity in complex sounds by zebra finches (*Taeniopygia guttata*) and budgerigars (*Melopsittacus undulatus*)," J. Comp. Psychol. **112**, 36-47.

Meddis, R., and Hewitt, M. J. (**1991a**). "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification," J. Acoust. Soc. Am. **89**, 2866-2882.

Meddis, R., and Hewitt, M. J. (**1991b**). “Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: Phase sensitivity,” J. Acoust. Soc. Am. **89**, 2883-2894.

Meddis, R., and O’Mard, L. (**1997**). “A unitary model of pitch perception,” J. Acoust. Soc. Am. **102**, 1811-1820.

Miller, J. D. (**2007**). “Sex differences in the length of the organ of Corti in humans,” J. Acoust. Soc. Am. **121**, EL151-EL155.

Mom, T., Bonfils, P., Gilain, L., and Avan, P. (**2001**). “Origin of cubic difference tones generated by high-intensity stimuli: Effect of ischemia and auditory fatigue on the gerbil cochlea,” J. Acoust. Soc. Am. **110**, 1477-1488.

Moore, B. C. (**1973**). “Frequency difference limens for short-duration tones,” J. Acoust. Soc. Am. **54**, 610-619.

Moore, B. C., and Glasberg, B. R. (**1983**). “Suggested formulae for calculating auditory-filter bandwidths and excitation patterns,” J. Acoust. Soc. Am. **74**, 750-753.

Moore, B. C., and Glasberg, B. R. (**1989**). “Difference limens for phase in normal and hearing-impaired subjects,” J. Acoust. Soc. Am. **86**, 1351- 1365.

Moore, B. C., Glasberg, B. R., and Peters, R. W. (**1984**). “Frequency and intensity difference limens for harmonics within complex tones,” J. Acoust. Soc. Am. **75**, 550-561.

Moore, B. C., Peters, R. W., and B. R. Glasberg (**1985**): “Thresholds for the detection of inharmonicity in complex tones,” J. Acoust. Soc. Am. **77**, 1861-1867.

Moore, B. C., Glasberg, B. R., and Peters, R. W. (**1986**). “Thresholds for hearing mistuned partials as separate tones in harmonic complexes,” J. Acoust. Soc. Am. **80**, 479-483.

Müller, M. (**1996**). "The cochlear place-frequency map of the adult and developing Mongolian gerbil," Hear. Res. **94**, 148-156.

Patterson, R. D. (**1987**). "A pulse ribbon model of monaural phase perception," J. Acoust. Soc. Am. **82**, 1560-1586.

Patterson, R. D., Allerhand, M. H., and Giguere, C. (**1995**). "Time domain modelling of peripheral auditory processing: A modular architecture and a software platform," J. Acoust. Soc. Am. **98**, 1890-1894.

Plassmann, W., Peetz, W., and Schmidt, M. (**1987**). "The cochlea in Gerbilline rodents," Brain Behav. Evol. **30**, 82-101.

Pressnitzer, D. and Patterson, R. D. (**2001**). "Distortion products and the perceived pitch of harmonic complex tones," in *Physiological and Psychophysical Bases of Auditory Function*, edited by D. J. Breebart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonoven (Shaker, Maastricht), pp. 97-104.

Ryan, A. (**1976**). "Hearing sensitivity of the Mongolian gerbil, *Meriones unguiculatis*," J. Acoust. Soc. Am. **59**, 1222-1226.

Sek, A., and Moore, B. C. J. (**1995**). "Frequency discrimination as a function of frequency, measured in several ways," J. Acoust. Soc. Am. **97**, 2479-2486.

Simmons, A. M., Reese, G., and Ferragamo, M. (**1993**). "Periodicity extraction in the anuran auditory nerve II. Phase and temporal fine structure," J. Acoust. Soc. Am. **93**, 3374-3389.

Sinnott, J. M., and Mosteller, K. W. (**2001**). "Comparative assessment of speech sound discrimination in the Mongolian gerbil," J. Acoust. Soc. Am. **110**, 1729-1732.

Sinnott, J. M., Brown, C. H., and Brown, F. E. (**1992**). "Frequency and intensity discrimination in Mongolian gerbils, African monkeys and humans," Hear. Res. **59**, 205-212.

Sinnott, J. M., Sachs, M. B., and Hienz, R. D. (**1980**). “Aspects of frequency discrimination in passerine birds and pigeons,” J. Comp. Physiol. Psychol. **94**, 401-415

Terhardt, E. (**1974**). “Pitch, consonance and harmony,” J. Acoust. Soc. Am. **55**, 1061-1069.

Von Békésy, G. (**1960**). *Experiments in Hearing* (McGraw-Hill, New York).

Wever, E. G. (**1949**). *Theory of Hearing* (Wiley, New York).

Wightman, F. L. (**1973**). “The pattern-transformation model of pitch,” J. Acoust. Soc. Am.**54**, 407-416.

Wright, A., Davis, A., Bredberg, G., Ulehlova, L., and Spencer, H. (**1987**). “Hair cell distributions in the normal human cochlea,” Acta OtoLaryngol., Suppl. **444**, 1-48.

Zwicker, E. (**1970**). “Masking and psychological excitation as consequences of the ear’s frequency analysis,” in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden).

3 A Comparative View on the Perception of Mistuning: Constraints of the Auditory Periphery[1]

---

[1] This chapter has been published in the present form in *The Neurophysiological Bases of Auditory Perception*, edited by E. A. Lopez-Poveda, A. R. Palmer, and R. Meddis (Springer, New York), pp. 465–475 (2010).

3.1 Abstract

Harmonicity serves to group together frequencies from a single source to a perceived auditory object. The peripheral auditory system may exploit either spectral cues resulting in a specific spatial pattern of excitation or temporal cues that are due to the interaction of frequency components in the complex to evaluate whether a component does not belong to a harmonic complex, i.e., is mistuned. Which cues are useful for mistuning detection may depend on the anatomical and physiological constraints in the peripheral auditory system. Here we compare the perception of frequency shifts in harmonic complexes (i.e., mistuning) and in pure tones across species. Mongolian gerbils and birds are superior to humans in detecting small amounts of mistuning in sine phase harmonic complexes. This difference is reduced in the detection of mistuning in random phase harmonic complexes (but not in harmonic complexes with "frozen random phase"). Humans are superior to birds and gerbils in detecting pure-tone frequency shifts. The results suggest that species with a short cochlea and only few hair cells per critical band tend to rely more on temporal fine structure in the analysis of mistuning of components in harmonic complexes whereas excitation patterns may play a larger role in humans with a much longer cochlea. For the analysis of pure-tone frequency shifts, excitation patterns appear to play a more prominent role in all species. Exemplary neurophysiological data obtained in starlings support this view.

3.2 Introduction

Harmonicity is one of the most important grouping cues in the analysis of sound. To determine the effect of harmonicity on the grouping or the segregation of frequencies in a complex stimulus, the "mistuned harmonic" paradigm (Moore *et al.*, 1984) can be used. Shifting the frequency of one component in a harmonic complex (i.e., "mistuning" the component) changes the percept of the whole complex for small frequency shifts and results in hearing the mistuned component as a separate object for large frequency shifts (Moore *et al.*, 1985, 1986). Frequency difference limens (FDLs) for detecting the mistuning of a component in a harmonic complex vary considerably between humans and animal species such as gerbils, starlings, zebra finches, and budgerigars (Klinge and Klump, 2009, 2010; Klump and Groß, submitted; Lohr and Dooling, 1998). This suggests that different mechanisms may be employed by the different species that will be discussed here.

There is an ongoing discussion whether temporal or spectral pattern recognition mechanisms play a greater role in detecting such mistuning. According to the spectral pattern recognition hypothesis (e.g. Goldstein, 1973), the frequency of each component of a complex is determined and compared to a template pattern of harmonic frequencies to decide whether the frequency of the component belongs to the harmonic series or not. Models based on this concept (e.g. Duifhuis *et al.*, 1982) imply that at least some frequency components of the harmonic complex are resolved and can be separately analyzed. Another class of models that was introduced by Licklider (1951) to account mainly for the processing of unresolved harmonics relies on exploiting temporal patterns created by the interacting components on the basilar membrane (e.g. de Cheveigné, 1998; Meddis and Hewitt, 1991a,b). While in some studies separate mechanisms have been proposed for low and high frequencies, i.e., resolved and unresolved harmonics, respectively (e.g. Houtsma and Smurzynski, 1990), others have suggested that one common mechanism based on temporal processing may be sufficient to explain the human psychophysical data (e.g. Meddis and O'Mard, 1997; Gockel *et al.*, 2004). Here we will discuss whether one mechanism may account for the variation in the animal data or whether there is evidence for different mechanisms. Furthermore, we will investigate the possible relation between differences in the frequency representation by the peripheral auditory system (e.g., as exemplified by the cochlear-map functions) of humans, gerbils and two bird species (starlings and budgerigars) and the perception of frequency shifts in components of a harmonic complex and of pure tones.

3.3 Detecting Frequency Shifts of Pure Tones

Before discussing the detection of mistuned harmonics, we will turn to the FDL of pure tones as a baseline measure of frequency discrimination. Two mechanisms have been suggested to represent pure-tone frequencies in the auditory periphery. For low frequencies at which phase locking of auditory nerve (AN) fibres to the period of the pure-tone frequency is possible, frequency can be encoded in the temporal spiking pattern of the neurons. At frequencies higher than about 2-4 kHz, the ability of neurons in the AN to phase lock to the period of the pure-tone frequency degrades (Köppl, 1997). Thus, a place mechanism on the cochlear frequency map is assumed to operate and frequency shifts are detected by the change in the distribution of neural activity in the peripheral auditory system across the AN fibres tuned to different frequencies (i.e., a rate-place code is used).

Let us first compare the pure-tone FDLs across species and then evaluate their relation to the species' cochlear frequency map. Gerbils show the poorest pure-tone frequency discrimination ability (Fig. 3.1a) if compared to starlings and humans. The gerbil's FDLs range from 21% (Weber fraction) at 200 Hz to 7% at 6400 Hz (Klinge and Klump, 2009) whereas the pure-tone FDLs of starlings range from 9% at 200 Hz to 1.5% at 4800 Hz (Buus *et al.*, 1995; Klump and Groß, submitted). The budgerigar's performance in pure-tone frequency discrimination lies between that of starlings and humans decreasing from 2.8% at 500 Hz to about 0.7% for frequencies from 1000 to 2000 Hz and then increasing to 1.3% at 5700 Hz (Dooling and Saunders, 1975). In contrast to most animal data, humans show low FDLs ranging from 0.2% within the frequency range of 500-2000 Hz and increasing to 1.2% at about 8000 Hz (e.g. Moore, 1973). How can these differences between pure-tone FDLs be related to the change in the peripheral excitation pattern and to the ability to represent frequencies by the temporal firing patterns of the AN fibres? For humans it has been suggested that in the frequency region in which phase locking of AN fibres is still possible, temporal mechanisms are exploited to reach the very low FDLs (e.g. Sek and Moore, 1995). A similar performance would be expected for the starling and possibly also for the gerbil (assuming that it has a similar phase locking ability as the Guinea pig, see Köppl, 1997) if they also would rely on temporal mechanisms. However, the much higher FDL in these species compared to the human FDL suggests that no such temporal mechanism is exploited at low pure-tone frequencies.

Based on the cochlear place-frequency functions derived by Greenwood (1990), Fay (1992) proposed that pure-tone FDLs, the critical bandwidth, the critical masking ratio bandwidth and the width of psychophysical tuning curves can all be related to the spatial frequency map on the basilar membrane. Thus, if the pure-tone FDLs

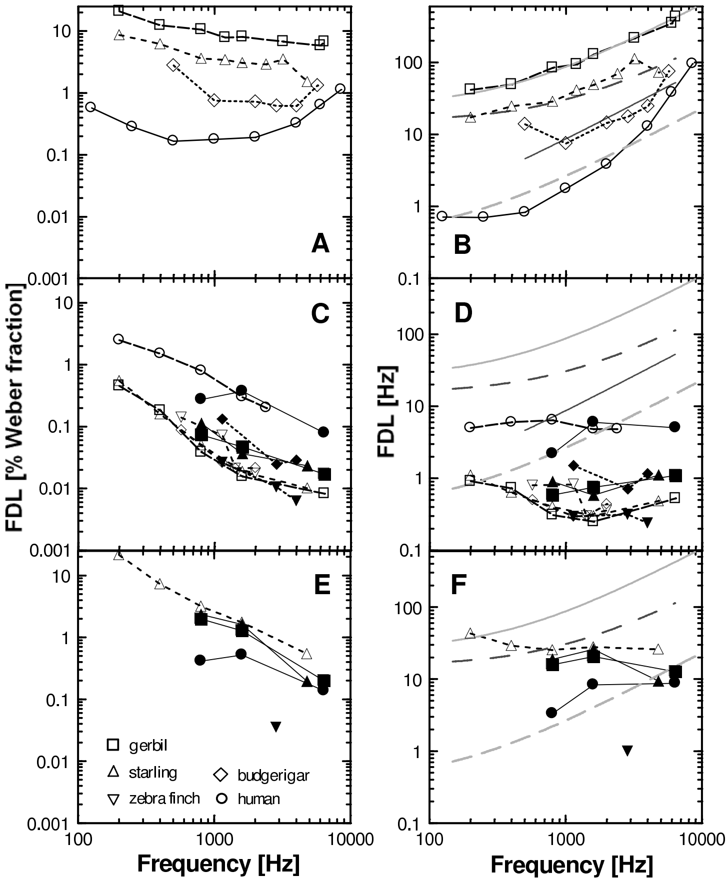

Figure 3.1: Mean FDLs of pure tones (**a**, **b**), of mistuned harmonics in sine phase complexes (**c**, **d**), and of mistuned harmonics in random phase complexes (**e**, **f**) for different species (indicated by the symbol shape, see legend). In the panels on the left side the FDLs are displayed in % Weber fraction ($\Delta f/f$). On the right side the FDLs are shown in Hz in comparison to the prediction for pure-tone FDLs generated using cochlear-map functions of the different animal species (gerbil light grey solid line, budgerigar black solid line, starling black dashed line) and humans (light grey dashed line) based on the Greenwood (1990) functions. In (**c**)-(**f**) data obtained for complexes using low fundamentals (e.g., 200-285 Hz) are shown with open symbols whereas data for complexes with high fundamentals (e.g., 570-800 Hz) are shown with filled symbols. Gerbil data are from Klinge and Klump (2009), starling data from Klump and Groß (submitted), zebra finch and budgerigar data from Lohr and Dooling (1998), and human data from Moore (1973, pure-tone FDLs) and Klinge and Klump (2009, mistuned harmonic FDLs).

increase with increasing frequency proportional to the cochlear-map functions derived on the basis of critical bandwidths, critical-ratio bandwidths or physiological frequency maps, then a place mechanism would explain the FDLs. Figure 3.1bshows the FDLs of humans, gerbils, starlings, and budgerigars as a function of frequency together with a function predicting the FDL (based on the Greenwood function) assuming that it represents a constant spatial shift on the cochlear-map function of these species. For gerbils and starlings, the functions of the observed FDLs are parallel and differ (i.e., are smaller) by a scaling factor of 1.5 for gerbils and a scaling factor between 5 and 7 for starlings from the critical-bandwidth functions calculated on the basis of the species' physiological data (Buus *et al.*, 1995; Kittel *et al.*, 2002; Klump and Groß, submitted). For budgerigars, only the FDLs for frequencies higher than 1 kHz fit the cochlear-map function estimated on the basis of the critical-ratio bandwidth and the average scaling factor is about 30 (however, Dooling and Saunders (1975) point out that the shapes of the FDL and CR functions deviate). The good fit of these functions to the pure-tone FDLs in the gerbil, the starling and for part of the frequency range in the budgerigar suggest that a pure place mechanism might explain the observed FDLs over most of the frequency range in these species. In humans, the FDLs do not follow the function predicted by the CBs over the total frequency range tested. An average scaling factor of 70 can be calculated on the basis of the data provided by Moore (1973) that is reduced to below 20 at 8 kHz. This high value may indicate that temporal mechanisms in addition to the proposed place mechanism are exploited in the human auditory system (e.g. Sek and Moore, 1995).

It remains to be discussed why humans show so much better pure-tone FDLs at high frequencies at which they have to rely on place mechanisms similar to the other species described here. Fay (1992) proposed that these differences might arise from differences in the anatomy of the auditory periphery. The best frequency discrimination ability should be found in animals with a long cochlea and a narrow frequency range of hearing sensitivity. The lowest frequency discrimination acuity should be found in species with a short cochlea and a hearing sensitivity extending to high frequencies. The gerbil cochlea is only 11 mm long (Müller, 1996) which is short in comparison to 34 mm reported for the human cochlea (e.g. Wright *et al.*, 1987). The bird species considered here have a much shorter basilar papilla than the cochlea of the gerbil (2.9 mm for starlings, Buus *et al.* (1995); 2.5 mm for budgerigars, Manley *et al.* (1993)). In mammals, the number of hair cells corresponds to the difference in the length of the cochlea, and thus is reduced from about 3,500 inner hair cells (IHCs) in the human cochlea (Wright *et al.*, 1987) to about 1,400IHCs in the gerbil cochlea (Plassmann *et al.*, 1987). Birds have several rows of hair cells across the basilar papilla that function like mammalian IHCs and may provide a large afferent input in a critical band (total number of hair cells on the starling and budgerigar basilar papilla is 5,800 and 5,400, respectively, see Gleich and Manley (1988); Manley *et al.* (1993). If we now compare the hearing range between the different species, we find that gerbils have the largest hearing range extending from approximately 0.1 to 60 kHz (Ryan, 1976). The human hearing range extends from about 0.01 to 16 kHz (ISO 389-7, 1996) whereas the hearing range of budgerigars and starlings extend from about 0.1 to 6.1 and 6.3 kHz, respectively (Dooling and Saunders, 1975; Buus *et al.*, 1995). Comparing the bird species to the mammal species, we can estimate that one critical band covers the shortest distance in the cochlea of the birds (0.1 mm for the starling and 0.13 mm for the budgerigar, using parameters for the Greenwood function suggested by Buus *et al.* (1995); Fay (1992)). The distance on the cochlear map covered by each critical band is about 0.2 mm in the gerbil (Kittel *et al.*, 2002) and about 1.15 mm in humans (Fay, 1992). Given these parameters of the species, we find that in the gerbil the number of hair cells in a critical band is the lowest from all four species, which suggests that due to the reduced afferent input only large changes in the excitation pattern may be detectable by the gerbil. Viewed on the basis of distance in the cochlea corresponding to the FDL, the gerbil needs the largest spatial shift to detect a frequency shift (0.13 mm), whereas in budgerigars, starlings, and humans the spatial shift corresponding to the pure-tone FDL is about an order of magnitude smaller than in the gerbil.

3.4 Detecting a Mistuned Component in Harmonic Complexes

The fundamental frequency, the number of frequency components in a harmonic complex, their resolvability, their phase relationship, and the shape of the resulting temporal waveform in different auditory filters may affect the FDL for mistuned harmonics in an otherwise harmonic complex. Some of these parameters are important for mechanisms of mistuning detection relying on the spatial pattern of excitation in the cochlea or on the temporal processing. By varying the fundamental frequency of a harmonic complex, the spacing between two adjacent harmonics and thus their resolvability and the period of the temporal waveform of the harmonic complex can be changed. Given the limited ability of the auditory system to follow fast temporal modulations of the stimulus (e.g., see review by Joris *et al.*, 2004), temporal patterns corresponding to the 5-ms period in a 200 Hz complex should be much better represented by the temporal discharge of neurons in the auditory system than the patterns corresponding to the 1.25 ms period in an 800 Hz complex (see also Hartmann *et al.*, 1990). The phase relationship between the frequency components of the complex will mainly be important for the temporal mechanisms. A starting phase of 0° (sine phase) for all harmonics in a complex results in a waveform with a non-varying and peaky temporal fine structure that will only change if mistuned harmonics are present. A randomized starting phase, however, degrades the distinctiveness of the temporal waveform and, if the phase is randomized for every presented stimulus, the constantly varying temporal fine structure makes it difficult to detect the change that is due to the mistuning. Thus, it is a hallmark of temporal mechanisms of mistuning detection that they are sensitive to the phase relation of the components of the harmonic complex.

3.4.1 Mistuning Detection in Sine Phase Harmonic Complexes

Contrary to the pattern observed for the pure-tone FDL, the ability to detect a frequency shift of a mistuned component in a sine phase harmonic complex was much better in gerbils and the three bird species than in humans. The animal species' FDLs for this stimulus condition were the lowest of all experiments described here (Fig. 3.1c in % Weber fraction and Fig. 3.1d in Hz). Humans showed about one order of magnitude higher thresholds for detecting a mistuned harmonic in a sine phase complex than the animal species. At a specific frequency, the FDLs for the harmonic complexes varied with the fundamental frequency in the animal species but not so much in the humans (e.g. Klinge and Klump, 2010; Moore *et al.*, 1985). In the gerbil and the three bird species, the FDLs for a mistuned component in a complex with a high fundamental (800 Hz for gerbil and starling, 570 Hz for budgerigar and zebra finch) were higher than the FDLs for a mistuned component in a complex with a low fundamental (200 Hz for gerbil and starling, 285 Hz for budgerigar and zebra finch). If we compare the prediction from the cochlear-map function (that fitted the pure-tone FDLs quite well) with the FDLs for mistuned harmonics in a sine phase complex, it becomes evident that mechanisms relying on the spatial pattern of excitation in the cochlea are unlikely to explain the data.

3.4.2 Mistuning Detection in Random Phase Harmonic Complexes

Also in the random phase condition, in which the spectral composition of the signal equals that of a sine phase complex but has an altered temporal waveform, the FDLs of the animal species are much smaller than their pure-tone FDLs (Fig. 3.1e, f). In starlings (Klump and Groß, submitted) and gerbils (Klinge and Klump, 2009), the FDLs for detecting a mistuned harmonic in an 800 Hz complex with all harmonics in random phase are similar and range from between approximately 2% for the first harmonic (800 Hz) to 0.2% for the sixth (4,800 Hz) or eighth (6,400 Hz) harmonic, respectively. FDLs for the 200 Hz complex obtained in starlings (Klump and Groß, submitted) resulted in similar values as for the 800 Hz complex. As for pure-tone frequency discrimination, FDLs for mistuned harmonics in random phase complexes with a fundamental of 800 Hz are lower in humans (FDLs ranging from 0.5 to 0.1%) than in gerbils or starlings. However, in a study using the same stimuli and procedures these differences were only significant for the first harmonic (800 Hz Klinge and Klump, 2009). The human's FDLs in the random phase condition were not significantly different from those for detecting a mistuned harmonic in a sine phase complex. The zebra finch (diamond in Fig.3.1e, f) has the lowest FDL of all species tested in the random phase condition (Lohr and Dooling, 1998) which might be due to differences in the variability of the waveforms depending on how the stimulus is generated. To obtain FDLs for mistuned harmonics in a random phase complex with a fundamental frequency of 800 Hz, the humans, starlings, and gerbils in the study of Klinge and Klump (2009) were provided with a stimulus that had a randomized starting phase of every component for every presented stimulus. For the 200 Hz complex, the starlings were presented with a large set of at least 10 different random phase renditions of every test stimulus and of 30 random phase renditions of the reference stimulus that were newly produced before each session. Lohr and Dooling (1998), however, used only five different "frozen" random phase test stimuli and ten different "frozen" random phase reference stimuli and presented them to a subject throughout the experiment. The low number of different stimuli might have allowed the zebra finches to memorize the temporal waveforms of the stimuli and thus made detecting a change in the waveform due to a mistuned harmonic easier for the birds. In their recent study on starlings, Klump and Groß (submitted) found that the FDL decreased nearly to the FDL obtained for components in a sine phase harmonic complex if the birds were presented with a single "frozen" random phase stimulus as the reference for multiple sessions. Taken together these data suggest that also in the random phase condition the detection of mistuned harmonics may rely on temporal mechanisms.

3.4.3 Neural Basis of Mistuning Detection

The results of the different experiments described above showed that humans seem to process the information in a mistuned harmonic complex differently in their auditory system than the animal species. The gerbils and the birds showed much higher pure-tone FDLs than humans but were superior to humans when detecting a mistuned harmonic in a sine phase complex. In the animal species, the FDLs for the random

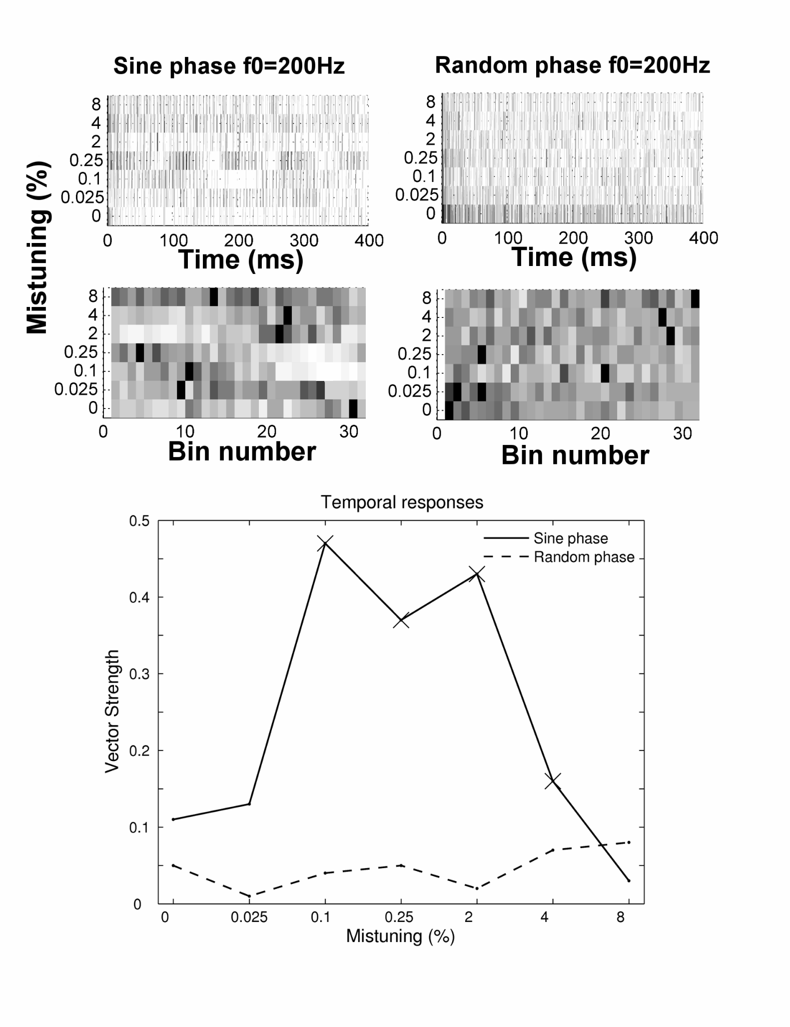

Figure 3.2: Multiunit response observed in the starling forebrain when stimulated with a harmonic complex (fundamental 200 Hz, 48 harmonics, component at 4,800 Hz mistuned, level of components 60dB SPL; excitatory frequency range of tuning curve at 60dB SPL ranged from about 3,500 to 5,200 Hz). Normalized peri-stimulus time histograms (top graphs) and beat frequency period histograms (middle graphs, the first 30ms of the response were excluded) of neuronal activity elicited by a mistuned harmonic are shown; the numbers on the ordinate of the graphs indicate the amount of mistuning (Weber fraction $\Delta f/f$ in %). The bottom graph shows the vector strength computed from the period histograms demonstrating a significant locking (marked with X) of the response to the beating period in the sine phase condition.

phase complex were about one order of magnitude higher than the FDLs for the sine phase complex. In contrast, the FDLs in humans did not differ significantly in the random phase and the sine phase complex. This raises the question whether the processing mechanisms may differ between humans and the animal species that were tested. Observing substantially lower FDLs for zebra finches and budgerigars in comparison to humans, Lohr and Dooling (1998) suggested either temporal processing mechanisms in the birds and spectral mechanisms in the humans to detect a mistuned harmonic in sine phase or an enhanced ability of birds to discriminate the same cues as humans. Klinge and Klump (2009, 2010) and Klump and Groß (submitted) suggest that temporal mechanisms might also be employed by the gerbil and the starlings to detect a mistuned component in a harmonic complex for which the evidence was presented above.

We have argued above that the animal species having a short cochlea cannot rely on the spatial pattern of excitation on the basilar papilla but have to rely on the analysis of temporal discharge patterns of AN fibres to achieve the low FDLs for mistuned harmonics. Three temporal cues are available in the auditory periphery allowing the discrimination of a complex with a mistuned harmonic from a fully harmonic complex (e.g. Moore *et al.*, 1985; Klinge and Klump, 2010). The first cue that is available if more than one component of the complex drives the neural response is the slow amplitude modulation that is due to the beats created by the interaction of a mistuned component with a neighbouring harmonic frequency. We have observed a temporal neural response pattern in the starling auditory forebrain that indicates a change in the discharge rate with the period of the beating (Fig. 3.2). This periodic rate change could be used for mistuning detection. The lower limit for the FDL of mistuning detection will be set by the neurons' integration time window and the duration of the complex used for stimulation. For small amounts of mistuning, the change in the amplitude created by the beating of neighbouring frequencies throughout the stimulus duration may be too small to be usable for mistuning detection.

The second available cue lies in the temporal fine structure of the waveform that results from the phase relationship of the components and which differs, for example, between the mistuned and the harmonic condition of sine phase complexes. As in the case of beating, more than one component of the complex must be able to drive the neural response to make this cue effective, i.e., at least two frequency components must lie in the frequency range to which the auditory neurons are tuned. In the sine phase condition, for example, the fine structure of the waveform changes throughout the mistuned stimulus while it is constant in the fully harmonic stimulus. Previous studies have demonstrated a sensitivity of auditory neurons in the brainstem to the shape of the periodic amplitude change of the waveform (i.e., ramped and damped sounds, see Pressnitzer *et al.*, 2000). These neurons should also be sensitive to the change in the waveform occurring with mistuning in a sine phase complex. If the phase of the components of a complex is random and constantly changes from one stimulus to the next, this cue is not suited for mistuning detection.

The third cue that can be used for mistuning detection is provided by the phase of individual resolved components of the stimulus represented in different frequency channels. If the phase information is transmitted by the auditory neurons (e.g., in AN fibres this should be possible up to 2-4 kHz, see Köppl, 1997), wide-band units functioning as coincidence detectors above the level of the auditory periphery would be able to detect a gradual phase shift occurring throughout the stimulus that is due to the mistuning. This gradual phase shift is found only in those auditory filters (and the respective neurons) that contain the mistuned harmonic and could be compared to the output driven by filters containing only in-tune harmonics of the complex. This cue and the second cue complement each other in the frequency range of hearing. In the low frequency range in which the harmonics are resolved in the response of the peripheral auditory system the direct evaluation of the phase is possible, whereas in the high frequency range with unresolved harmonics the direct interaction of components in the cochlea allows for an indirect evaluation of the phase relation between components.

3.5 References

Buus, S., Klump, G. M., Gleich, O., and Langemann, U. (**1995**). "An excitation-pattern model for the starling (*Sturnus vulgaris*)", The Journal of the Acoustical Society of America **98**, 112–124.

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

Dooling, R. J. and Saunders, J. C. (**1975**). "Hearing in the parakeet (*Melopsittacus undulatus*): absolute thresholds, critical ratios, frequency difference limens, and vocalizations", Journal of Comparative and Physiological Psychology **88**, 1–20.

Duifhuis, H., Willems, L. F., and Sluyter, R. J. (**1982**). "Measurement of pitch in speech: an implementation of goldstein's theory of pitch perception", The Journal of the Acoustical Society of America **71**, 1568–1580.

Fay (**1992**). "Structure and function in sound discrimination among vertebrate", in *The evolutionary biology of hearing*, edited by D. Webster, R. Fay, and A. Popper, 229–263 (Springer, New York).

Gleich, O. and Manley, G. A. (**1988**). “Quantitative morphological analysis of the sensory epithelium of the starling and pigeon basilar papilla”, Hearing Research **34**, 69–85.

Gockel, H., Carlyon, R. P., and Plack, C. J. (**2004**). “Across-frequency interference effects in fundamental frequency discrimination: questioning evidence for two pitch mechanisms”, The Journal of the Acoustical Society of America **116**, 1092–1104.

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**, 1496–1516.

Greenwood, D. D. (**1990**). “A cochlear frequency-position function for several species–29 years later”, The Journal of the Acoustical Society of America **87**, 2592–2605.

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**, 1712–1724.

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**, 304–310.

ISO 389-7 (**1996**). “ISO 389-7 acoustics – reference zero for the calibration of audiometric equipment – part 7: Reference threshold of hearing under free-field and diffuse-field listening conditions”, .

Joris, P. X., Schreiner, C. E., and Rees, A. (**2004**). “Neural processing of amplitude-modulated sounds”, Physiological Reviews **84**, 541–577.

Kittel, M., Wagner, E., and Klump, G. M. (**2002**). “An estimate of the auditory-filter bandwidth in the mongolian gerbil”, Hearing Research **164**, 69–76.

Klinge, A. and Klump, G. M. (**2009**). “Frequency difference limens of pure tones and harmonics within complex stimuli in mongolian gerbils and humans”, The Journal of the Acoustical Society of America **125**, 304–314.

Klinge, A. and Klump, G. M. (**2010**). “Mistuning detection and onset asynchrony in harmonic complexes in mongolian gerbils”, The Journal of the Acoustical Society of America **128**, 280–290.

Klump, G. M. and Groß, S. (**submitted**). “Detection of frequency shifts and mistuning in complex tones in the european starling”, Journal of Comparative Physiology A .

Köppl, C. (**1997**). “Phase locking to high frequencies in the auditory nerve and cochlear nucleus magnocellularis of the barn owl, tyto alba”, The Journal of Neuroscience: The Official Journal of the Society for Neuroscience **17**, 3312–3321.

Licklider, J. C. R. (**1951**). “A duplex theory of pitch perception”, Experientia **7**, 128–134.

Lohr, B. and Dooling, R. J. (**1998**). “Detection of changes in timbre and harmonicity in complex sounds by zebra finches (*Taeniopygia guttata*) and budgerigars (*Melopsittacus undulatus*)”, Journal of Comparative Psychology (Washington, D.C.: 1983) **112**, 36–47.

Manley, G. A., Schwabedissen, G., and Gleich, O. (**1993**). “Morphology of the basilar papilla of the budgerigar (*Melopsittacus undulatus*)”, Journal of Morphology **218**, 153–165.

Meddis, R. and Hewitt, M. J. (**1991**a). “Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification”, The Journal of the Acoustical Society of America **89**, 2866–2882.

Meddis, R. and Hewitt, M. J. (**1991**b). “Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: Phase sensitivity”, The Journal of the Acoustical Society of America **89**, 2883–2894.

Meddis, R. and O’Mard, L. (**1997**). “A unitary model of pitch perception”, The Journal of the Acoustical Society of America **102**, 1811–1820.

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

Moore, B. C., 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**, 479–483.

Moore, B. C., Glasberg, B. R., and Shailer, M. J. (**1984**). “Frequency and intensity difference limens for harmonics within complex tones”, The Journal of the Acoustical Society of America **75**, 550–561.

Moore, B. C., 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**, 1861–1867.

Müller, M. (**1996**). “The cochlear place-frequency map of the adult and developing mongolian gerbil”, Hearing Research **94**, 148–156.

Plassmann, W., Peetz, W., and Schmidt, M. (**1987**). “The cochlea in gerbilline rodents”, Brain, Behavior and Evolution **30**, 82–101.

Pressnitzer, D., Winter, I. M., and Patterson, R. D. (**2000**). “The responses of single units in the ventral cochlear nucleus of the guinea pig to damped and ramped sinusoids”, Hearing Research **149**, 155–166.

Ryan, A. (**1976**). “Hearing sensitivity of the mongolian gerbil (*Meriones unguiculatis*)”, The Journal of the Acoustical Society of America **59**, 1222–1226.

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

Wright, A., Davis, A., Bredberg, G., Ulehlova, L., and Spencer, H. (**1987**). “Hair cell distributions in the normal human cochlea”, Acta Oto-Laryngologica. Supplementum **444**, 1–48.

4 Mistuning detection and onset asynchrony in harmonic complexes in Mongolian gerbils[1]

---

[1] This chapter has been published in the present form in JASA 128(1):280−290 (2010).

4.1 Abstract

By applying a Go/NoGo paradigm, thresholds for detecting mistuning of components of a 200 Hz complex were determined in the Mongolian gerbil and compared with thresholds obtained in a previous study with an 800 Hz complex. Frequency difference limens (FDLs) for detecting mistuning decreased with increasing harmonic frequency and harmonic number (0.5% to 0.01% Weber fraction). It was furthermore examined how starting and ending the mistuned component earlier than the remaining complex affects the FDL (duration of all components 400 ms, time shift 30 to 500 ms). Large FDLs that are similar to pure tone FDLs (between 21% and 6.7%) were found for onset asynchronies of 300 ms and more, indicating separate processing of the mistuned component. Small FDLs that are similar to FDLs of the synchronous condition were found if the temporal overlap between the mistuned component and the remaining complex was 100 ms or more. These experimental data in combination with a simulation of processing of the harmonic complexes by the gerbil's peripheral auditory filters led to the conclusion that the phase and amplitude modulations in the filter outputs can provide cues that allow gerbils a sensitive detection of mistuning across a wide range of frequencies.

4.2 Introduction

How the hearing system segregates overlapping sounds from different sources and which cues are used to group components of a sound from one source are central questions in auditory research (Bregman, 1990; Darwin, 1997; Shinn-Cunningham, 2008). Among the most important cues for grouping of concurrent sounds are the harmonicity of frequency components of a complex sound (Moore *et al.*, 1984, 1985; Hartmann *et al.*, 1990), the synchrony of the onset of signal components (Darwin and Ciocca, 1992; Darwin and Hukin, 1998; Gockel *et al.*, 2005; Bleeck *et al.*, 2008), and correlated amplitude modulations of the signal envelope in different frequency bands (e.g., Hall and Grose, 1991). Grouping by harmonicity has been investigated using the "mistuned harmonic" paradigm (e.g., Moore *et al.*, 1985). However, the mechanisms that underlie the detection of a mistuned component in a harmonic complex are still not fully understood and are a topic of intense discussion (Moore *et al.*, 1985; Houtsma and Smurzynski, 1990; Carlyon and Shackleton, 1994; de Cheveigné, 1998; Lin and Hartmann, 1998; Gockel *et al.*, 2004; Oxenham *et al.*, 2004). Most of the proposed mechanisms address pitch perception problems rather than explain mistuning detection per se. Thus, we discussed in a previous study (Klinge and Klump, 2009) if two of the major classes of pitch perception models, the spectral pattern recognition models (e.g., Goldstein, 1973; Wightman, 1973; Terhardt, 1974) and the temporal models based on neural autocorrelation processes (e.g., Licklider, 1951; Meddis and Hewitt, 1991a, 1991b; Patterson *et al.*, 1995), might provide an explanation for mistuning detection.

As a well-established animal model, the Mongolian gerbil (*Meriones unguiculatus*) has been used in Klinge and Klump (2009) as well as in the present experiments to study auditory processing mechanisms in mammals since its hearing sensitivity in the low-frequency range is similar to that of humans. Such an animal model gives the opportunity to combine psychophysical approaches that can be compared to human psychophysical experiments with invasive approaches like neural recordings from different regions of the brain. So far, only Klinge and Klump (2009) examined the effects of grouping cues on the processing of complex sounds in gerbils. The results of the behavioral study showed that Mongolian gerbils are highly sensitive at detecting a mistuned component in an 800 Hz harmonic complex with all harmonics starting in sine phase. They even showed a better performance than humans in this task. The extraordinarily high sensitivity for mistuning detection in a harmonic complex stands in contrast to the gerbils' poor performance in pure tone frequency discrimination. Furthermore, this study showed that changing the temporal but not the spectral pattern of the waveform (e.g., in a complex with all harmonics in random phase) resulted in an increase in FDLs for mistuned components of the harmonic complex by about an order of magnitude. Based on these data we proposed that temporal processing mechanisms rather than spectral pattern recognition mechanisms are employed by the gerbil to detect a mistuned component in a harmonic complex for both resolved and unresolved harmonics (Klinge and Klump, 2009).

The present study focused on how changes in two sound characteristics, the fundamental frequency F0 of a harmonic complex and the synchrony between a mistuned harmonic and the remaining complex, affect the way complex sounds are processed in the auditory system (e.g., Moore *et al.* 1985; Roberts and Moore, 1990). In the first experiment, frequency difference limens (FDLs) of mistuned harmonics in a 200 Hz complex with all harmonics starting synchronously were determined. This lower F0 compared to the 800 Hz fundamental of the previous study (Klinge and Klump, 2009) results in a longer period duration which might be beneficial for detecting changes in the temporal structure of the waveform of the harmonic complex. Furthermore, a lower fundamental increases the number of harmonics within one auditory filter which results in larger interactions between temporal waveforms of the harmonics at the filter output. If Mongolian gerbils actually rely on the processing of temporal cues such as fast envelope fluctuations or phase shifts, we would expect lower FDLs for mistuning detection in a 200 Hz complex than in an 800 Hz complex when matching the same frequencies.

The second experiment examined how onset asynchrony applied to a mistuned component of a 200 Hz harmonic complex [Fig. 4.1(A)] affects the detection of the mistuning in gerbils. Natural sound sources generally provide more than one cue to the auditory system to group or segregate frequency components. Past research has strived to understand how a combination of different cues, e.g., onset asynchrony and harmonicity, is processed to analyze natural acoustic scenes (e.g., Darwin and Ciocca, 1992; Darwin *et al.*, 1994; Darwin and Hukin, 1998; Ciocca and Darwin, 1999; Gockel *et al.*, 2005; Bleeck *et al.*, 2008). Frequency components with a common onset are more likely to originate from the same sound source than frequency components that have a different onset. It is an interesting question regarding the interaction of cues how the auditory system copes with signals providing conflicting evidence. For example, asynchronous onset of components in a harmonic complex should lead to their segregation while harmonicity is a strong cue to group the components. An increasing onset asynchrony when time-shifting the mistuned harmonic also leads to a reduced temporal overlap between the mistuned harmonic and the remaining complex. If temporal cues that are available in the mistuned harmonic complex are important for the gerbil to detect the mistuning, the duration of the overlap between mistuned harmonic and remaining complex might be crucial for a good performance. The use of temporal processing mechanisms underlying the detection of mistuning is discussed on the basis of simulated output signals of the estimated auditory filters of the gerbil (Kittel *et al.*, 2002).

4.3 Materials and methods

4.3.1 Animal subjects

Five adult Mongolian gerbils – 3 females, 2 males – served as subjects in the experiments. During the time of testing, the animals were between 20 and 36 months old. Except one subject (Lv), different animals than in Klinge and Klump (2009) served as subjects in the present study. Data acquisition was terminated when the subjects were older than 36 months (Lv, Mo) to minimize effects of possible hearing loss occurring in old gerbils (Hamann *et al.*, 2004). One female (Re) did not complete all experiments (for details see results in Table 4.1). Prior to the frequency discrimination experiments, the hearing sensitivity of the gerbils at low (200 Hz), middle (1000 Hz, 1200 Hz, respectively), and high frequencies (6400 Hz, 8000 Hz, respectively) was tested. Hearing thresholds for Mg, Fa, Re, and Mo ranged between -3.9 and -7 dB SPL at 6400 Hz, between -2.5 and -8.2 dB SPL at 1200 Hz, and between 30.6 and 28.6 dB SPL at 200 Hz. Hearing thresholds for Lv were -7.2 dB SPL at 8000 Hz, 1.2 dB SPL at 1000 Hz, and 28.4 dB SPL at 200 Hz. Similar hearing thresholds have been found in gerbils by Ryan (1976) and Klinge and Klump (2009). Thus, all five gerbils could be considered as having normal hearing.

The care and treatment of the gerbils were in accordance with the procedures of animal experimentation approved by the Government of Lower Saxony, Germany. All procedures were performed in compliance with the NIH Guide on Methods and Welfare Considerations in Behavioral Research with Animals (NIH Publication No. 02-5083).

4.3.2 Apparatus and stimulus generation

The apparatus has been described in detail by Klinge and Klump (2009). Briefly, experiments were conducted in a doughnut-shaped experimental cage mounted in a single-walled, echo-reduced, sound-attenuating booth (Industrial Acoustics, Type IAC 401-A, $T_{60} = 12$ ms). The gerbil was trained to sit and wait on an elevated platform facing the loudspeaker (Canton Plus XS, frequency range: 150 Hz-21 kHz, mounted about 30 cm in front of the platform at 0° azimuth and 0° elevation) until the test stimulus was played. Correct responses to the test stimulus were rewarded with food (20 mg Bioserve Dustless Precision Pellets Formula F0163-J50). A red feeder light emitting diode (LED) served as an additional reinforcer. Experiments took place with no visible light in the chamber. The sessions could be monitored by a charge coupled device (CCD) camera and infrared LED illumination.

All stimuli were generated using a Linux workstation with an AMD processor and an RME sound card (Hammerfall DSP Multiface II). The analog signal from the sound card (sample frequency 44.1 kHz) was attenuated (Kenwood type RA-920A and Tucker Davis Technologies System 3 PA5), passed through an acoustic isolation transformer (1:1), and then amplified (Rotel RMB-1048). The stimuli were pre sented via a free-field loudspeaker.

The stimuli were harmonic complexes comprised of the first 48 harmonics of a 200 Hz F0 added in sine phase. The sound pressure level was set at 60 dB SPL per component with levels adjusted to the frequency response of the loudspeaker. The level of the overall complex was randomly varied by ±3 dB to eliminate any possible level cue for the gerbils. The duration of the complex stimulus was 400 ms including 25-ms raised cosine ramps at stimulus onset and offset. In a control experiment, the duration of the complex was shortened to 100 ms including the 25-ms cosine ramps. The FDLs of mistuned harmonics in the 200 Hz complex were measured for the first (200 Hz), second (400 Hz), fourth (800 Hz), eighth (1600 Hz), and 32nd (6400 Hz) harmonic of the complex with all harmonics starting and ending synchronously. In a second experiment, the mistuned harmonic was presented asynchronously (i.e., starting and ending earlier) to the remainder of the complex [Fig. 4.1(A)]. FDLs were obtained in two experimental series for the mistuned second (resolved) harmonic with onset asynchronies of 30, 70, 100, 200, 300, or 400 ms and for the mistuned 32nd (unresolved) harmonic with onset asynchronies of 100, 200, 300, 400, or 500 ms. Initial experiments revealed that thresholds obtained for the mistuned second harmonic with onset asynchronies of 30 and 70 ms were approximately the same as in the synchronous condition. Thus, these conditions were omitted from the threshold measurements for the 32nd harmonic. Similarly, for the mistuned 32nd harmonic it has been shown that the threshold of the 500 ms onset asynchrony condition was similar to the result of the 400 ms asynchrony condition. Therefore, thresholds were not obtained for an onset asynchrony of 500 ms in the second harmonic. After finishing the main experiments, two control experiments were conducted. In the first control, FDLs were determined for the second and 32nd harmonic of a complex with an overall duration of 100 ms and all harmonics starting synchronously [Fig. 4.1(B)]. During the second control experiment, the mistuned harmonic was prolonged [it started earlier but stopped with the other harmonics of the complex, see Fig. 4.1(C)]. Here, FDLs were obtained for the second and 32nd mistuned harmonic with onset asynchronies of 100 ms and 300 ms.

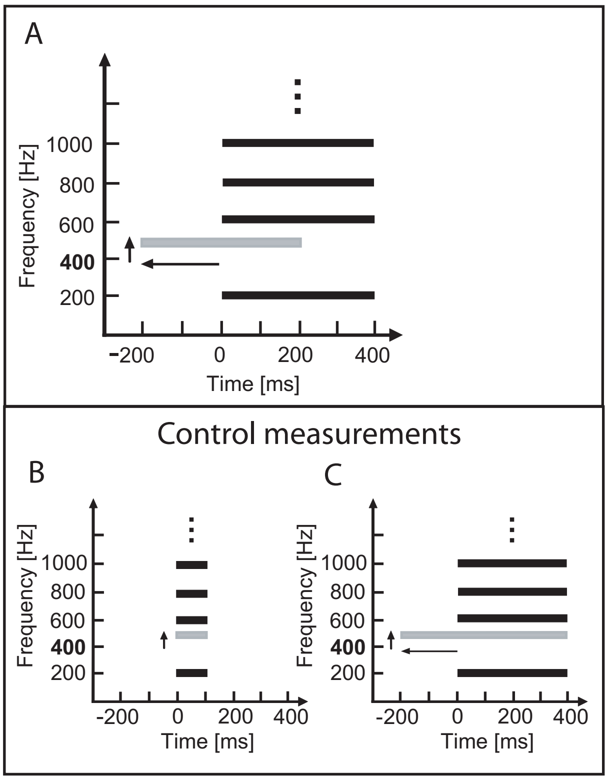

Figure 4.1: Examples of the stimuli used in the main experiment (A) and the control measurements [(B) and (C)]. The 200 Hz complex consisted of the first 48 harmonics (200 to 9600 Hz). One of the harmonics was mistuned by shifting the frequency upwards. In the main experiment (A), an onset asynchrony between 0 and 500 ms was introduced by time-shifting the mistuned harmonic. In the first control measurement (B), the duration of the whole complex was shortened from 400 ms to 100 ms, and the mistuned harmonic was synchronous to the remaining complex. In the second control measurement (C), onset asynchrony was introduced by prolonging the mistuned harmonic.

4.3.3 Procedure

Animals were trained using a Go/NoGo operant paradigm with positive reinforcement (food rewards). During the session, a reference stimulus was continuously repeated every 1.3 s. The reference stimulus was a complex in which none of the harmonics were mistuned. The gerbil initialized a trial by jumping onto the pedestal. After a random waiting interval between 1 and 7 s, the test stimulus was presented instead of the reference stimulus. The test stimulus could either be a complex with one of the harmonics shifted upwards in frequency (test trial, Go-stimulus) or be the same as the reference stimulus (catch trial, NoGo-stimulus). The gerbil indicated the detection of mistuning by jumping off the platform (“Hit”). If it jumped off the platform during a NoGo-stimulus, the trial was counted as a “False Alarm”. Hit rates and false alarm rates were used to calculate the sensitivity measure d'.

Sessions consisted of blocks of ten trials each and had a total of 110 trials. Within each block three catch trials and seven test trials were presented in a randomized order. Frequency increments in test trials consisted of fixed steps of a pre-selected Weber fraction. The step size was adjusted for the FDL measurements and was doubled for the two largest frequency increments that were not included in determining thresholds. These adjustments provided both sub-threshold stimuli and stimuli well above threshold. The first block in each session was a block of warm-up trials in which only the largest frequency increment was presented. The warm-up trials were not included in the data analysis (for details see Klinge and Klump, 2009).

4.3.4 Data analysis

A session was accepted as being valid if the gerbils responded correctly to at least 80% of the two test stimuli with the largest frequency increments and if the false alarm rate did not exceed 20%. A psychometric function was constructed relating d' to the frequency shift and linearly interpolating between adjacent points. The threshold was determined as the amount of frequency shift resulting in a d' of 1.8. Two consecutive valid sessions in which thresholds differed by no more than 0.1% Weber fraction (synchronous conditions) or 1% Weber fraction (asynchronous conditions) were combined to calculate the final FDL. After finishing an experimental series the first FDL measurement of this series was repeated. If the repeated threshold differed by less than 0.1% (synchronous conditions) or 1% (asynchronous conditions) from the first FDL measurement, it was assumed that no training effect affected the results and the experimental series was terminated. If it differed by more, the following FDL measurements from the experimental series were repeated until the repeated threshold reached the criterion of similarity to the first measured threshold. From all repeated thresholds only the data of the repetition were taken into account for further data analysis.

The presentation order of the experiments (synchronous and asynchronous) and of all FDL measurements within the two experiments was randomized for every animal to exclude training effects. Data were were analyzed using the statistical software package SPSS 15.0.

4.4 Results

4.4.1 Detection of mistuning in harmonic complexes with different fundamentals

In the first experiment, FDLs were determined for three resolved (1st, 2nd, and 4th) and two unresolved (8th and 32nd) harmonics in a 200 Hz complex (Table 4.1) with all harmonics starting synchronously. The gerbils' FDLs for detecting a mistuned harmonic were generally small, decreasing from a Weber fraction of 0.46% ± 0.28% (mean ± 2SE) at the first harmonic (200 Hz) to 0.008% ± 0.003% at the 32nd harmonic

(6400 Hz). An univariate general linear model (GLM) ANOVA revealed a significant effect of the frequency of the harmonic on the relative frequency shift at detection threshold [$F(4, 15) = 12.43, p < 0.001$]. The results from the 200 Hz complex were compared with those obtained in a previous study using an F0 of 800 Hz (Klinge and Klump, 2009). In both studies, the harmonic complexes included all harmonics up to 9600 Hz (i.e., 48 harmonics for the 200 Hz complex and 12 harmonics for the 800 Hz complex). Similar to the data of the 200 Hz complex, the FDLs of the first, second (both were resolved), and eighth harmonic (unresolved) of the 800 Hz complex decreased significantly from 0.073% ± 0.027% Weber fraction at the 800 Hz harmonic to 0.017% ± 0.001% at the 6400 Hz harmonic (Klinge and Klump, 2009). In a first analysis, the FDLs of the matched harmonic numbers [1st, 2nd, 8th, Fig. 4.2(A)] of both complexes were compared using a general linear mixed model (GLMM) ANOVA. The main effects of the fundamental of the complex and the harmonic number on the FDLs were significant [$F(1, 17) = 19.35, p < 0.001; F(2, 17) = 13.18, p < 0.001$] and so was the interaction between the two main effects [$F(2, 17) = 7.96, p < 0.005$]. The

Table 4.1: FDLs of mistuned harmonics (1st, 2nd, 4th, 8th, 32nd) of a 200 Hz complex with all harmonics starting synchronously. Results are displayed as individual FDLs (Hz) for all tested gerbils plus their mean and two times the standard error (2SE).

Frequency (Hz)	Thresholds (Hz)					
	Fa	Mg	Mo	Re	Lv	Mean ± 2SE
200	1.44	0.82	...	...	0.47	0.91 ± 0.57
400	0.39	0.48	1.14	...	0.89	0.73 ± 0.35
800	0.34	0.37	0.27	0.32	0.25	0.31 ± 0.04
1600	0.37	0.13	0.34	...	0.16	0.25 ± 0.12
6400	0.60	0.58	0.67	...	0.26	0.48 ± 0.18

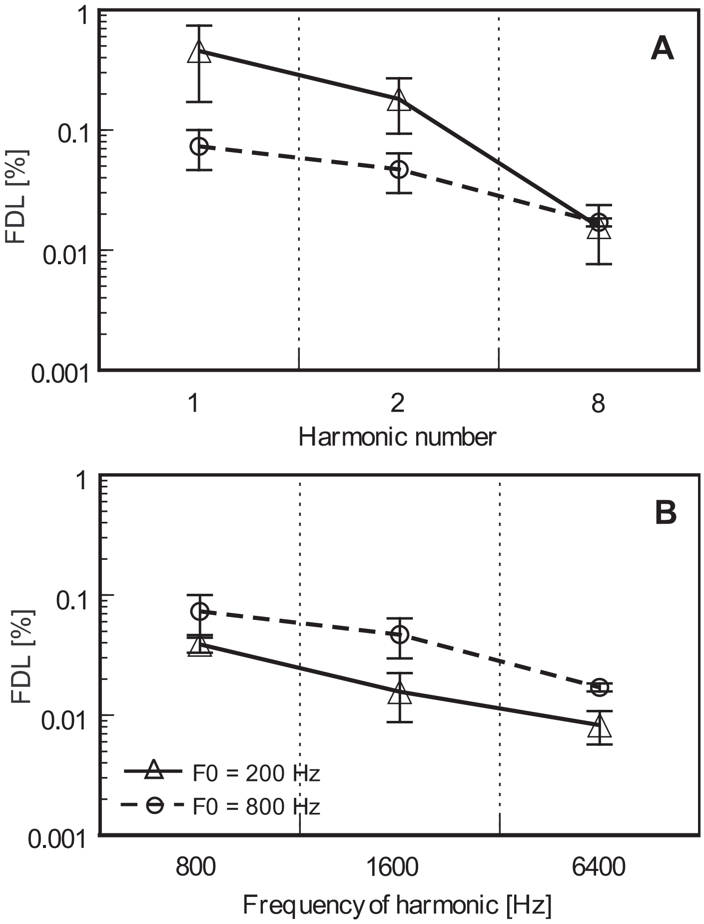

Figure 4.2: FDLs of three harmonics of the 200 Hz complex (triangles, solid line) compared to FDLs of three harmonics of the 800 Hz complex (circles, dashed line, data from Klinge and Klump, 2009) matched by harmonic number (1st, 2nd, 8th harmonic, panel A) or matched by the frequency of the harmonic (800 Hz, 1600 Hz, 6400 Hz, panel B). Data is plotted in % Weber fraction. Error bars equal to two times the standard error.

FDLs for the first and second harmonic were larger for the 200 Hz complex than for the 800 Hz complex whereas they were similar at the eighth harmonic. In a second analysis, the FDLs of the matched frequencies of the mistuned harmonics of both complexes [800, 1600, or 6400 Hz, Fig. 4.2(B)] were compared. A GLMM ANOVA revealed significant main effects of the fundamental [$F(1, 15.24) = 17.12, p < 0.05$] and of the frequency of the mistuned harmonic [$F(2, 13.87) = 30.84, p < 0.001$] on the FDLs. There was no significant interaction between both main effects [$F(2, 13.87) = 3.17, p = 0.074$]. For each given frequency, a smaller FDL was observed for detecting the mistuning in the 200 Hz complex than in the 800 Hz complex.

4.4.2 FDLs of mistuned components with an onset asynchrony

In the second experiment, FDLs were measured for a resolved (2nd) and an unresolved (32nd) harmonic of the 200 Hz complex in which the mistuned harmonic was asynchronous to the remaining complex. The mistuned harmonic started and stopped

Table 4.2: FDLs of mistuned harmonics (1st, 2nd, 4nd, 8nd, 32nd) of a 200 Hz complex with all harmonics starting synchronously. Results are displayed as individual FDLs (Hz) for all tested gerbils plus their mean and two times the standard error (2SE).

	Thresholds (Hz)					
Onset asynchrony (ms)	Fa	Mg	Mo	Re	Lv	Mean ± 2SE
	2nd harmonic (400 Hz component)					
30	0.53	0.56	0.92	0.62	…	0.66 ± 0.18
70	0.67	0.61	1.22	0.93	…	0.86 ± 0.28
100	0.75	0.5	0.94	0.93	…	0.78 ± 0.21
200	0.67	0.79	1.16	0.69	…	0.83 ± 0.23
300	17.03	18.25	21.63	13.9	…	17.70 ± 3.20
400	15.76	27.59	14.67	…	…	19.34 ± 8.27
	32nd harmonic (6400 Hz component)					
100	0.36	1.31	0.42	…	…	0.70 ± 0.61
200	4.78	1.39	0.35	…	1.39	1.98 ± 1.93
300	318.33	1.22	1.57	…	…	107.04 ± 211.29
400	319.25	201.94	240.94	…	239.22	250.34 ± 49.34
500	497.78	246.97	269.21	…	247.47	315.36 ± 122.06

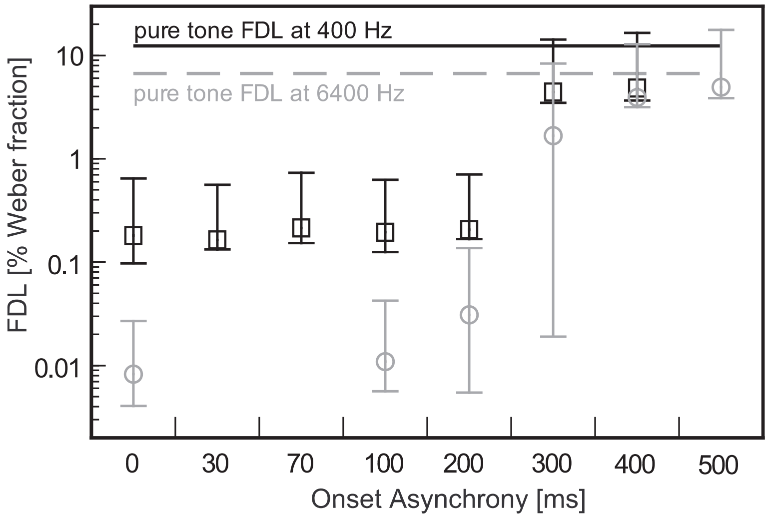

Figure 4.3: Mean FDLs (% Weber fraction) at different onset asynchronies for the 2nd harmonic mistuned (black squares) and for the 32nd harmonic mistuned (gray circles). The two lines represent the mean threshold of pure tone FDLs that were obtained in our previous study (Klinge and Klump, 2009). Error bars represent the range of the threshold values of the four tested gerbils.

between 30 ms and 500 ms earlier than the remaining harmonics of the complex, keeping the duration of both mistuned component and tone complex the same [see Fig. 4.1(A)]. Figure 4.3 shows mean FDLs for detecting the mistuning of the 2nd and 32nd harmonic as a function of onset asynchrony and Table 4.2 depicts the individual and mean thresholds in Hz. The two horizontal lines in Fig. 4.3 indicate FDLs for pure tones at the frequency of the shifted harmonic (400 Hz and 6400 Hz, respectively; data from Klinge and Klump, 2009). The observed FDLs were compared using a GLMM ANOVA. The main effects of the harmonic number and of the amount of onset asynchrony on the FDL were significant (harmonic number: $F(1, 34.06) = 5.87, p < 0.05$, onset asynchrony: $F(7, 32.61) = 24.31, p < 0.001$?, but there was no significant interaction $[F(4, 32.8) = 2.08, p = 0.11]$. To assess which amount of onset asynchrony is needed to significantly increase the FDL *post hoc* tests (t-tests with Bonferroni correction) were conducted. These pair-wise comparisons revealed no significant differences between FDLs of a mistuned harmonic with onset asynchronies of 30, 70, 100 and 200 ms and FDLs of a mistuned harmonic of a complex with all harmonics starting synchronously. A significant increase in the FDL of the mistuned harmonic in the harmonic complex was found for an onset asynchrony of 300 ms ($p < 0.001$). At this asynchrony, FDLs for the 2nd harmonic (400 Hz component) were similar to the 400 Hz pure tone FDL in all gerbils. However, for the 32nd harmonic (6400 Hz) two out of three gerbils showed FDLs that were similar to FDLs found for the synchronous presentation of harmonics while the FDL for one gerbil was similar to the pure tone FDL (Table 4.2). At onset asynchronies of 400 ms and 500 ms, there was no temporal overlap between the mistuned harmonic and the remaining complex. A student's t-test revealed no significant difference between the FDLs of the mistuned component for these onset asynchronies and the pure tone FDLs at 400 Hz and 6400 Hz, respectively.

To assess the effect of duration on the detection of the mistuned harmonic in a complex, two control measurements with two gerbils each were conducted. Since the 300 ms onset asynchrony condition was a transition point in the results, the first control stimulus was a 200 Hz complex whose duration was shortened to 100 ms and with the mistuned component synchronous to the remaining components [Fig. 4.1(B)]. This matched the duration of the control with the duration of the overlap in the 300 ms onset asynchrony condition. The results showed smaller FDLs for the control stimulus than for the stimulus in the 300 ms onset asynchrony condition [Fig. 4.4(A)], indicating that other effects may play a role which will be discussed in Sec. 4.5.2. The FDLs obtained in the first control experiment (100 ms duration) were approximately four times higher than the FDLs determined for the 400 ms long synchronous stimuli.

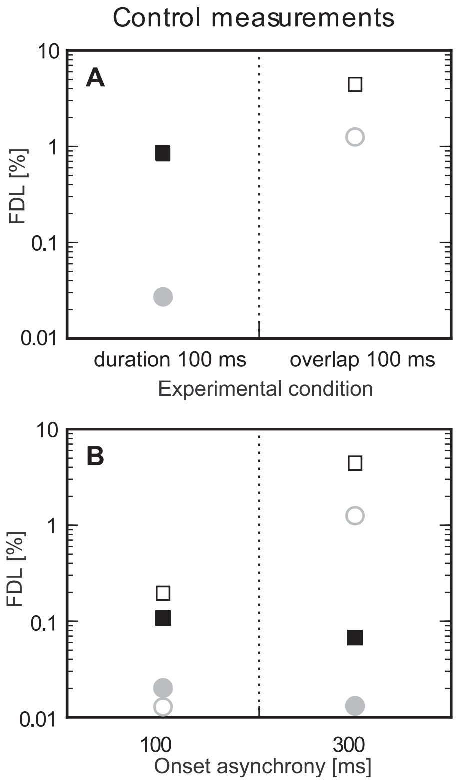

Figure 4.4: Mean FDLs (% Weber fraction) for the two control measurements (filled symbols) compared to the corresponding data of the main experiment (open symbols). Data for the second harmonic are depicted in black squares and data for the 32nd harmonic are depicted in light gray circles. In the first control measurement (A), the duration of the whole complex was shortened to 100 ms [left panel, filled symbols, see also Fig. 4.1(B)] and compared to the 300 ms onset asynchrony condition of the main experiment in which the temporal overlap was 100 ms (right, open symbols). In the second control measurement (B), the onset asynchrony was introduced by prolonging the mistuned harmonic [see Fig. 4.1(C)]. The results for the 100 ms and for the 300 ms onset asynchrony condition of the control measurements (filled symbols) are compared with the 100 ms and the 300 ms onset asynchrony condition of the main experiment (open symbols).

In the second control experiment, onset asynchrony was obtained by prolonging the mistuned component in its duration (onset asynchronies of 100 and 300 ms were tested) rather than only shifting it in time [Fig. 4.1(C)]. This prevented the gerbils from using a possible offset cue and the overall duration of the overlap was identical to the synchronous case (400 ms). At an onset asynchrony of 100 ms, no substantial differences were found between the FDL of the control measurement with the prolonged component and the experiment with the time-shifted component for both the 2nd and the 32nd harmonic [Fig. 4.4(B)]. All four FDL values (of the control and for the time-shifted stimuli) were similar to the FDL in the synchronous condition. At an onset asynchrony of 300 ms, however, the FDLs for the prolonged stimuli (control) were clearly lower than the FDLs for the time-shifted stimuli [Fig. 4.4(B)]. The FDLs for the 2nd and 32nd harmonic in the control measurement were now similar to those found in the synchronous condition. The results of the second control experiment indicate that there was no effect of onset asynchrony alone, provided that the duration of the overlap was not shortened by more than 200 ms.

4.5 Discussion

4.5.1 Detection of mistuning in harmonic complexes with different fundamentals

The first experiment was conducted to examine the influence of different fundamental frequencies on the FDLs of mistuned components in harmonic complexes with all harmonics starting synchronously. Previous studies in humans showed that the harmonic number and the fundamental frequency have an effect on the FDL as both parameters determine how many harmonics will fall within the limits of an auditory filter. Moore *et al.* (1985) found a decrease of thresholds with increasing F0 and increasing harmonic number. A similar response pattern could be observed in gerbils in the current study

[Fig. 4.2(A)], but the gerbils generally detected mistuned components in a harmonic complex approximately four to ten times better than humans.

The physiologically relevant feature of the stimulus may be the representation of the temporal waveform of the stimulus that changes depending on the frequency of the fundamental and the bandwidth of the auditory filters. The fundamental frequency determines the duration of the period of the harmonic complex and the number of harmonics that fall within one auditory filter. As discussed below, a longer period duration in the 200 Hz complex compared to an 800 Hz complex may be advantageous for detecting changes in the temporal pattern of the stimulus. Furthermore, a lower fundamental increases the number of harmonics interacting within an auditory filter. Thus, FDLs at matching frequencies of the mistuned harmonics [800, 1600, or 6400 Hz, Fig. 4.2(B)] were compared for the 200 Hz and the 800 Hz complex. Based on the considerations above, FDLs for mistuning detection in a 200 Hz complex should be lower than in an 800 Hz complex when comparing thresholds at the same harmonic frequencies.

Mongolian gerbils are two to three times more sensitive in detecting a mistuned harmonic at a given frequency in a 200 Hz complex compared to an 800 Hz complex. For high frequencies at which neighboring frequency components may interact within one auditory filter, it is generally assumed that the detection of mistuning in harmonic complexes is based on temporal cues, like for example, beating or roughness (e.g., Moore *et al.*, 1985; de Cheveigné, 1998). Also in the gerbil, the mistuning detection in unresolved harmonics (i.e., the 6400 Hz component of both the 800 Hz and the 200 Hz complex, and the 1600 Hz component of the 200 Hz complex) could be explained by temporal cues such as changes in the fast fluctuations of the envelope of the complex waveform caused by the interacting harmonics within one auditory filter. For low resolved harmonics, it has been disputed if the human auditory system uses alternate mechanisms or if only a single mechanism is needed to detect a mistuning or perceive a pitch (Gockel *et al.*, 2004; Le Goff and Kohlrausch, 2005). The alternate hypotheses used to explain the detection of mistuning or the perception of pitch at low resolved harmonics by mechanisms that rely on the frequency analysis of individual components in separate filters based on the pattern of excitation (e.g., Duifhuis *et al.*, 1982) or that rely on autocorrelation analysis (Meddis and Hewitt, 1991a; Lin and Hartmann, 1998). However, changes in the excitation pattern or other mechanisms of frequency analysis of individual components of the harmonic complex to detect mistuning at low resolved harmonics are unlikely to be exploited by the gerbil auditory system. The poor performance in pure tone frequency discrimination suggests that the gerbil auditory system cannot detect such small frequency shifts using a place mechanism (Sinnott *et al.*, 1992; Klinge and Klump, 2009).

Distortion products as another cue for detecting mistuning have already been ruled out for the 800 Hz complex in a previous study (Klinge and Klump, 2009). The estimated level of the most prominent cubic distortions at the place of the nominal frequency of the mistuned harmonic would be approximately 20 dB lower than the level of each of the harmonics. The interaction between the mistuned harmonic and the distortion product would result in an amplitude modulation (AM) with a modulation depth of approximately 1.5 dB (peak to trough) which is much lower than the detection threshold for amplitude modulation in gerbils (approx. 5.5 dB peak to trough with a white noise carrier at all tested modulation frequencies (62.5 to 1000 Hz); Wiegrebe and Sonnleitner, 2007; approx. 5.5 dB at a modulation frequency of 5 Hz and a carrier of 4 kHz, Santoso and Klump, unpublished data), and is therefore unlikely to be detectable. When mistuning the first or the second harmonic of the 200 Hz complex, however, the estimated levels of developing distortion products may actually reach or exceed the level of the nearest (regular or mistuned) harmonic which results in a large depth of amplitude modulation due to beating. Thus, modulation depths could be high enough to allow the detection of the mistuning. The modulation frequency of the AM, however, is mostly below 1 Hz and therefore, less than half a modulation cycle has occurred before the stimulus (400 ms) ends. Furthermore, the level roving of the stimuli should make level differences that are due to AM less usable. Estimated distortion products developing with the fourth harmonic being mistuned are probably not high enough to produce audible beatings.

To assess which of the cues such as slow or fast fluctuations of the envelope of the temporal waveform or changes in the excitation pattern might be used for detecting a mistuned component in a harmonic complex, we simulated the signal processing of auditory filters of the gerbil when excited by the 200 Hz and the 800 Hz complex. The width of the gammatone filters applied in the simulation was estimated based on the Greenwood function (Greenwood, 1961) using parameters that provided a best fit to psychophysical and physiological data of the gerbil (Kittel *et al.*, 2002). The results of the simulation shown in Fig. 5 were obtained with a frequency shift of 0.05%. This value is close to the threshold for detecting the mistuning of the 1600 Hz component in the 800 Hz complex (upper right panel of Fig. 4.5) but above threshold for the other examples. The first column of each panel in Fig. 4.5 shows the simulated filter of the gerbil and the spectral lines corresponding to the harmonic frequencies in the range of the filter. Generally, the temporal structure of the output signals of the various filters differs between a harmonic complex in sine phase and the same complex with one of the harmonics mistuned ("mistuned complex"). The differences in the temporal pattern concern the slowly fluctuating envelope (evident in the 2nd and 4th column of each panel showing the filter output to the complete 400 ms stimulus) as well as the fast fluctuations of the envelope at modulation frequencies corresponding to the period of the fundamental frequency (200 Hz and 800 Hz, respectively; evident in the 3rd and 5th column of each panel showing the filter output at a zoomed in time scale). The slow fluctuations of the envelope, however, are unlikely to produce usable cues to the gerbils, especially since the fluctuations that are evident for the 6400 Hz frequency region (bottom panels, Fig. 4.5) are much reduced if the mistuning is at threshold. Amplitude modulation detection in gerbils has been shown to be much worse than in humans (Stein *et al.*, 2005; Wiegrebe and Sonnleitner, 2007; Santoso and Klump, unpublished data). A cue that might be used by the gerbil auditory system are the fast fluctuations of the envelope at the filter output as evident in the 3rd and 5th column of the upper left panel and the two lower panels of Fig. 4.5 which show two F0 periods of the response to the harmonic or mistuned complex stimulus, respectively. In the case of a harmonic complex, the shape of the fast fluctuations of the envelope of the stimulus waveform at the filter output does not change during the ongoing stimulus. In the case of the mistuned complex, the shape of the fast fluctuations of the envelope does gradually change with a period corresponding to the reciprocal of the amount of the frequency shift of the mistuned harmonic. At the beginning of the stimulus, a mistuned complex will show a temporal pattern similar to that found for a harmonic complex whereas a strongly deviating pattern is found when the mistuned component is out of phase with the other harmonics (the fifth column in each panel shows the waveform at the time when the mistuned component is maximally out of phase, for details see figure caption of Fig. 4.5).

To distinguish between a harmonic complex and a mistuned complex, the auditory system can make sequential comparisons within one auditory filter and simultaneous

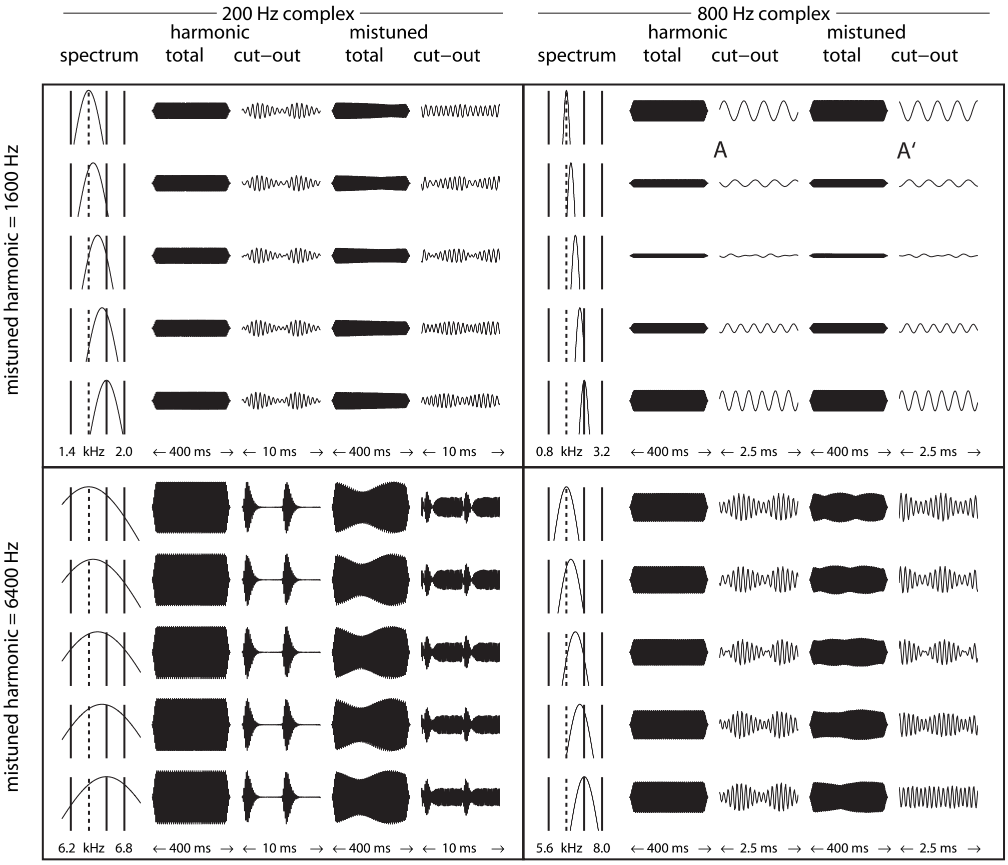

Figure 4.5: Four examples of the simulated processing of harmonic complexes with and without mistuned harmonics by auditory filters. The two left and the two right panels demonstrate changes in the waveform of the filter output due to mistuning of a component in a 200 Hz and an 800 Hz complex, respectively. The upper and lower panels show the effects of mistuning the 1600 Hz and 6400 Hz component, respectively. Within each panel the left-most column shows some spectral lines (solid lines) of the complex, the mistuned harmonic (dashed line), and the simulated auditory filter. The second and third column show the filter outputs over the total duration and a cut-out of the stimulus if excited by a harmonic complex. The fourth and fifth column show the the equivalent output if one of the harmonics in the complex is mistuned by 0.05% (Weber fraction). The cut-out for all four mistuning situations was made at the point at which the mistuned harmonic was maximally out of phase (upper two panels: at 365 ms; lower two panels: at 156.25 ms).

comparisons between auditory filters. The only condition in Fig. 4.5 in which the auditory filters are so narrow that the output of the filter reflects almost exclusively one harmonic of the complex is the frequency region around the 1600 Hz component of the 800 Hz complex (upper right panel of Fig. 4.5). In this condition, the most prominent cue useful for a sequential comparison may be provided by a gradual phase shift in those auditory filters that contain the mistuned harmonic (compare phases in the cut-out between A and A' in upper right panel of Fig. 4.5). One has to differentiate between this gradual phase shift and the difference in periodicity between filters containing only harmonic components. While the relationship between waveforms of harmonic filter outputs repeats itself with a period corresponding to the fundamental, the relationship between the waveform of a mistuned harmonic filter output and a harmonic filter output gradually changes over the course of the 400 ms stimulus. This gradual phase shift of the filter output might then be detected by comparing the filter outputs between filters containing the mistuned harmonic and filters containing only in-tune harmonics of the complex (e.g., the 800 Hz harmonic). Such a comparison which must occur above the level of the cochlea requires that the phase is represented in the neural response corresponding to the filter output.

The 1600 Hz component of the 200 Hz complex (upper left panel of Fig. 4.5) as well as the 6400 Hz component of the Hz complex (lower right panel of Fig. 4.5) are located in frequency regions where the waveform of at least two harmonics interact with each other within an auditory filter. In the frequency region around the 6400 Hz component of the 200 Hz complex many harmonics of the complex contribute to the filter output (lower left panel of Fig. 4.5). Due to the interaction of the harmonics, fast fluctuations of the envelope with a period corresponding to that of the fundamental frequency can be observed in all three conditions. A gradual change of the waveform from a temporal pattern similar to that of the harmonic condition to the pattern shown in the fifth column of the three panels of Fig. 4.5 could provide the cues for a sequential comparison. Furthermore, simultaneous comparisons of the waveform across different filter outputs could be made. Both differences between temporal modulation patterns evident in the fast fluctuations of the envelope of the various filter outputs as well as phase differences of the carrier in different filters can be found, although the latter may only provide cues at low frequencies (e.g., at 1600 Hz but not at 6400 Hz). The changes in the shape of the fast fluctuation of the envelope may result in differences of the neural response allowing successive stages of the auditory system to detect the mistuning. For example, it has been shown that neurons from the ventral cochlear nucleus and the inferior colliculus (IC) of guinea pigs respond differently to ramped and damped sinusoids (Pressnitzer *et al.*, 2000; Neuert *et al.*, 2001). Krishna and Semple (2000) showed that the response (rate modulation transfer functions) of neurons of the IC in gerbils differed with varying modulation depths of a sinusoidally amplitude modulated (SAM) tone. These studies indicate that changes in the fast fluctuations of the envelope may provide cues that are suitable for mistuning detection. The significantly lower FDLs of mistuned harmonics of the 200 Hz complex at a given frequency compared to those of the 800 Hz complex [Fig. 4.2(B)] could also be explained on the basis of neuronal modulation transfer functions (MTFs) (e.g., Joris *et al.*, 2004). Joris *et al.* (2004) presented studies which showed that neurons at different stages along the auditory pathway are able to follow modulation frequencies of up to approximately 1000 Hz. However, best modulation frequencies (BMFs) of neurons mostly ranged between 7 and about 500 Hz (Joris *et al.*, 2004). In the 200 Hz complex, the fast fluctuations of the envelope have a modulation frequency of 200 Hz (which corresponds to a period of 5 ms) in comparison to the 800 Hz complex with a modulation frequency of 800 Hz (a period of 1.25 ms). Changes in the modulation depths or in the shape of the fast fluctuations of the envelope as seen in Fig. 4.5 may thus be better represented in the response of the neuron to a 200 Hz compared to an 800 Hz complex.

4.5.2 FDLs of mistuned components with an onset asynchrony

The synchrony of onset times of frequency components is an important grouping cue and may interact with the harmonicity cue in mistuning detection (e.g., Darwin and Ciocca, 1992; Darwin *et al.*, 1994; Darwin and Hukin, 1998; Ciocca and Darwin, 1999; Gockel *et al.*, 2005; Bleeck *et al.*, 2008). For example, signals can provide conflicting evidence in which harmonicity in a harmonic complex is a strong cue to group frequency components, and onset asynchrony is a strong cue to segregate frequency components. An onset asynchrony applied to the mistuned component of a harmonic complex may affect the detection of the mistuned harmonic by either diminishing the integration of the mistuned harmonic into the remaining complex or by affecting temporal processing mechanisms due to a shortening of the temporal overlap between the mistuned harmonic and the remaining complex [if onset asynchrony is realized by shifting the mistuned harmonic in time as done in the present experiment, see Fig. 4.1(A)]. If we assume that mistuning detection in the gerbil is mediated by a temporal cue, then onset asynchrony may be more important by changing the duration of the temporal overlap of the mistuned component with the remaining complex than by diminishing the integration of the mistuned harmonic into the complex.

The purpose of the second experiment was to examinebhow an increasing onset asynchrony and decreasing temporal overlap between the mistuned harmonic and the remaining harmonics of the complex affects mistuning detection. The results of this experiment were compared with FDLs obtained for mistuned harmonics without onset asynchrony as well as with pure tone FDLs determined in a previous study (Klinge and Klump, 2009). The large difference between pure tone FDLs and FDLs for mistuned components in a harmonic complex of up to two orders of magnitude should provide an objective measure to determine the amount of onset asynchrony necessary to deteriorate the ability of the gerbil to detect the mistuned component in a harmonic complex. The results of the second experiment show this transition point quite clearly (Fig. 4.3). Up to an onset asynchrony of 200 ms, none of the obtained FDLs were significantly different from thresholds obtained for the synchronous condition. At an onset asynchrony of 300 ms, however, the perception of the mistuned harmonic changed. For the second harmonic at the 300 ms onset asynchrony condition all gerbils seemed to switch to a pure tone frequency discrimination. They may not have been able to use the 100 ms temporal overlap between mistuned harmonic and remaining complex to detect the frequency shift using temporal cues. The large differences between the individual gerbil data for the 6400 Hz component at 300 ms onset asynchrony (Table 4.2) suggests that the gerbils detected the frequency shift in the asynchronous harmonic in two different ways. Only one gerbil seemed to switch to a pure tone frequency discrimination whereas two of the three gerbils still may have been able to use the mistuning information in the mistuned harmonic that was temporally overlapping with the remaining complex. Thus it seems that the temporal overlap rather than the duration of the mistuned component leading the complex may be crucial for the detection of the mistuning. If the temporal overlap is too short, the auditory system may not be able to use it for detecting the mistuning and the gerbil has to switch to a pure tone frequency discrimination which points to separate processing mechanisms for the mistuned harmonic leading the complex and for the mistuned harmonic within the remaining complex.

This switching between separate processing mechanisms for the mistuned harmonic leading the complex and for the mistuned harmonic within the remaining complex seems to be different to the effect that onset asynchrony has on a mistuned component in humans. In human psychophysical pitch-matching studies, Darwin and Ciocca (1992) determined the perceptual pitch shift of a harmonic complex (compared to a reference complex) while one of the harmonics was shifted in frequency (mistuned). For a frequency increase of the shifted component of up to 3%, the pitch of the complex increased. In comparison to the mistuned harmonic complex with synchronous components, Darwin and Ciocca (1992) found a significantly reduced pitch shift in the mistuned harmonic complex at an onset asynchrony of 300 ms (i.e., the mistuned harmonic started earlier and continued throughout the 410 ms long complex). Smaller amounts of onset asynchrony reduced the effect gradually and were significant up to an onset asynchrony of 80 ms (only tested for a 90 ms long complex though). Pitch shifts for complexes with a 30 ms asynchronous mistuned component were almost identical to those observed in the synchronous condition. These results suggest that a sufficiently large amount of onset asynchrony between the mistuned harmonic and the remainder of the complex influences the effect this mistuned harmonic has on the pitch of an otherwise harmonic complex. Although the data of the study from Darwin and Ciocca (1992) show some parallels to the present results in the gerbil, one could make the caveat that matching the pitch of two complexes and detecting a mistuning in a complex might be different tasks for the auditory system. In other studies (e.g., Darwin, 1984; Roberts and Moore, 1990) it has been shown that already 30 ms onset asynchrony are sufficient to perceptually remove a harmonic from the computation of the first formant frequency of a vowel.

The observed deterioration of performance with very large onset asynchronies in the current experiment is also consistent with the idea of “overintegration”, i.e., the analysis of cues with a time window that is longer than the time period during which the cues are presented. An overintegration mechanism has been suggested in several studies (Sheft and Yost, 1990; Carlyon 1996; Ciocca and Darwin, 1999) to explain the deterioration of discrimination or detection thresholds in the presence of forward and/or backward fringes of signals not containing the relevant cues. However, two observations in the current experiment render it less likely that the remainder of the complex deteriorates mistuning detection by overintegration: (1) the extreme differences between the thresholds of individual gerbils for the 300 ms onset asynchrony condition in the 32nd harmonic would require vastly different individual temporal integration windows, (2) overintegration makes no specific prediction that thresholds deteriorate to a value close to the pure tone FDL.

We have argued above that one possible cue used in detecting the mistuning is the change in the shape of the fast fluctuations of the envelope of the waveform at filter output during the ongoing stimulus. Such a change accumulates over the duration of the overlap. Therefore, the duration of the overlap between the mistuned harmonic and the rest of the complex per se may be critical for the FDL. To test this hypothesis, two control measurements were carried out which assessed the influence of the duration of a mistuned harmonic complex on the FDLs. In the first control, the duration of the complex was shortened to 100 ms and all harmonics started synchronously in order to match the duration of the control with the duration of the overlap in the 300 ms onset asynchrony condition [Fig. 4.1(B)]. A comparison of thresholds [Fig. 4.4(A)] revealed lower FDLs for the 100 ms long complex (control) than for the latter condition (overlap of 100 ms). Thus, in addition to the possible dependence of the mistuning detection on the duration of the overlap, other effects such as a reduction of the neural response of the leading mistuned harmonic due to peripheral short-term adaptation processes may play a role (Ciocca and Darwin, 1993; Bleeck *et al.*, 2008). In the second control experiment, possible offset cues were eliminated by starting the mistuned component earlier than but ending it together with the remaining complex to introduce an onset asynchrony [Fig. 4.1(C)]. The overlap between the mistuned harmonic and the rest of the complex thus remained at 400 ms as for the synchronous case. Both gerbils showed a substantial reduction of FDLs from the condition in which the mistuned component was time-shifted by 300 ms to the condition in which the mistuned component was prolonged [control; Fig. 4.4(B)]. The FDLs for the 2nd and 32nd harmonic in the control measurement were now similar to those found in the synchronous condition. Thus, increasing the duration of the overlap from 100 to 400 ms reduced the FDL to that observed in the synchronous condition. Onset asynchrony of up to 300 ms when the mistuned component continues throughout the harmonic complex seems not to affect the detection of the mistuning by the gerbil.

The results indicate that the suggested temporal processing mechanisms for detecting a mistuned component in a harmonic complex in gerbils depends on the time period during which the mistuned harmonic overlaps with the other harmonics and thus, temporal cues in the fast fluctuations of the envelope within or in the phase shifts between auditory filters are available. Similar conclusions were reached in human psychophysical studies on pitch perception and mistuning detection (Moore *et al.*, 1985; Gockel *et al.*, 2007). Ciocca and Darwin (1999) suggested a temporal integration window in virtual pitch processing that is updated as long as the stimulus is presented. For the current mistuning experiment this proposed mechanism could imply that if the time period during which the mistuned and the other harmonics overlap is longer than the temporal integration window, the auditory system can fully exploit the temporal cues in the envelope. If the time period of overlap is shorter than the temporal integration window, the accurate frequency analysis based on the harmonic relations may be compromised. The present data suggest that the gerbil applies such a time window in the range of 100 to 200 ms. The sharp transition of the FDL at a duration of overlap of between 100 and 200 ms indicates that the slopes of the gerbil's time window may be rather steep.

4.6 Conclusion

The present experiments confirm previous results indicating that the gerbil shows a high sensitivity in detecting a mistuning in components of a sine phase harmonic complex (Klinge and Klump, 2009). The change in sensitivity of mistuning detection in components of a harmonic complex in relation to frequency is similar between gerbils and humans. FDLs are higher at low frequencies and decrease with increasing frequency of the harmonic. Compared to humans, however, the gerbils' sensitivity for detection of the mistuned component in a complex is about four to ten times higher. This is unexpected if the spatial representation of frequencies in the cochlea affects the FDL for mistuning and led us to suggest that the gerbils use temporal cues to detect mistuning. The results of the experiment investigating the role of onset asynchrony on mistuning detection are consistent with the hypothesis that particularly changes in the fast fluctuations of the envelope within and gradual phase shifts between auditory filters provide the most important cues in mistuning detection. A simulation of the processing of complexes with mistuned harmonics by auditory filters of the gerbil indicates that for such stimuli the waveform of the filter output differs considerably between auditory filters as well as within one auditory filter over time. The changes in the shape of the fast fluctuations of the envelope within an auditory filter and in the gradual phase shift between auditory filters that are due to the mistuning may elicit correlated neural responses that can be processed at stages of the auditory system beyond the cochlea. We propose that temporal processing mechanisms are not only employed in the high-frequency region where harmonics are unresolved but also in the low-frequency region where harmonics are resolved. The gerbil may thus provide a suitable model for studying temporal mechanisms of integration of harmonics in sounds that are uniformly applicable throughout the frequency range of perception.

4.7 Acknowledgments

We thank Susanne Groß for helping with the experiments, Rainer Beutelmann for providing Fig. 4.5, and Simon Jones and two anonymous reviewers for helpful comments on earlier versions of the manuscript. This study was supported by the DFG within the Sonderforschungsbereich SFB/TRR31 “The Active Auditory System” and the International Graduate School GRK 591 “Neurosensory science, systems, and applications”.

4.8 References

Bleeck, S., Ingham, N. J., Verhey, J. L., and Winter, I. M. (**2008**). “Rebound depolarization in single units of the ventral cochlear nucleus: A contribution to grouping by common onset?” Neuroscience **154**, 139-146.

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

Carlyon, R.P. (**1996**). “Masker asynchrony impairs the fundamental-frequency discrimination of unresolved harmonics,” J. Acoust. Soc. Am. **99**, 525-533.

Carlyon, R. P., and Shackleton, T. M. (**1994**). “Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms?” J. Acoust. Soc. Am. **95**, 3541-3554.

Ciocca, V., and Darwin, C.J. (**1993**). “Effects of onset asynchrony on pitch perception: Adaptation or grouping?” J. Acoust. Soc. Am. **93**, 2870-2878.

Ciocca, V., and Darwin, C. J. (**1999**). “The integration of nonsimultaneous frequency components into a single virtual pitch,” J. Acoust. Soc. Am. **105**, 2421-230.

Darwin, C. J. (**1984**). “Perceiving vowels in the presence of another sound: Constraints on formant perception,” J. Acoust. Soc. Am. **76**, 1636-1647.

Darwin, C. J. (**1997**). “Auditory grouping,” Trends Cogn. Sci. **1**, 327-333.

Darwin, C. J., and Ciocca, V. (**1992**). “Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of a mistuned component,” J. Acoust. Soc. Am. **91**, 3381-3390.

Darwin, C. J., Ciocca, V., and Sandell, G. J. (**1994**). “Effects of frequency and amplitude modulation on the pitch of a complex tone with a mistuned harmonic,” J. Acoust. Soc. Am. **95**, 2631-2636.

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

de Cheveigné, A. (**1998**). “Cancellation model of pitch perception,” J. Acoust. Soc. Am. **103**, 1261-1271.

Duifhuis, H., Willems, L. F., and Sluyter, R. J. (**1982**). “Measurement of pitch in speech: An implementation of Goldstein’s theory of pitch perception,” J. Acoust. Soc. Am. **71**, 1568-1580.

Gockel, H. E., Carlyon, R. P., and Plack, C. J. (**2004**). “Across-frequency interference effects in fundamental frequency discrimination: Questioning evidence for two pitch mechanisms,” J. Acoust. Soc. Am. **116**, 1092-1104.

Gockel, H. E., Moore, B. C., Carlyon, R. P., and Plack, C. J. (**2007**). “Effect of duration on the frequency discrimination of individual partials in a complex tone and on the discrimination of fundamental frequency,” J. Acoust. Soc. Am. **121**, 373-382.

Gockel, H. E., Plack, C. J., and Carlyon, R. P. (**2005**). “Reduced contribution of a nonsimultaneous mistuned harmonic to residue pitch,” J. Acoust. Soc. Am. **118**, 3783-3793.

Goldstein, J. L. (**1973**). “An optimum processor theory for the central formation of the pitch of complex tones,” J. Acoust. Soc. Am. **54**, 1496-1516.

Greenwood, D. D. (**1961**). “Critical bandwidth and the frequency coordinates of the basilar membrane,” J. Acoust. Soc. Am. **33**, 1344-1356.

Hall, J. W., and Grose, J. H. (**1991**). “Some effects of auditory grouping factors on modulation detection interference (MDI),” J. Acoust. Soc. Am. **90**, 3028-3035.

Hamann, I., Gleich, O., Klump, G. M., Kittel, M. C., and Strutz, J. (**2004**). “Age-Dependent Changes of Gap Detection in the Mongolian Gerbil (*Meriones unguiculatus*),” J. Assoc. Res. Otolaryngol. **5**, 49-57.

Hartmann, W. M., McAdams, S., and Smith, B. K. (**1990**). “Hearing a mistuned harmonic in an otherwise periodic complex tone,” J. Acoust. Soc. Am. **88**, 1712-1724.

Houtsma, A. J., and Smurzynski, J. (**1990**). “Pitch identification and discrimination for complex tones with many harmonics,” J. Acoust. Soc. Am. **87**, 304-310.

Joris, P. X., Schreiner, C. E., and Rees, A. (**2004**). “Neural processing of amplitude-modulated sounds,” Physiol. Rev. **84**, 541-577.

Kittel, M., Wagner, E., and Klump, G. M. (**2002**). “An estimate of the auditory-filter bandwidth in the Mongolian gerbil,” Hear. Res. **164**, 69-76.

Klinge, A., and Klump, G. M. (**2009**). “Frequency difference limens of pure tones and harmonics within complex stimuli in Mongolian gerbils and humans,” J. Acoust. Soc. Am. **125**, 304-314.

Krishna, S. B., and Semple, M. N. (2000). “Auditory temporal processing: Responses to sinusoidally amplitude-modulated tones in the inferior colliculus,” J. Neurophysiol. **84**, 255-273.

Le Goff, N., and Kohlrausch, A. (**2005**). “The use of temporal cues for frequency discrimination of the fundamental component in a complex tone,” in Proceedings of the Forum Acusticum, Budapest, pp. 1503-1507.

Licklider, J. C. (**1951**). “A duplex theory of pitch perception,” Experientia **7**, 128-134.

Lin Jian-Yu, and Hartmann, W. M. (**1998**). “The pitch of a mistuned harmonic: Evidence for a template model,” J. Acoust. Soc. Am. **103**, 2608-2617.

Meddis, R., and Hewitt, M. J. (**1991a**). “Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification,” J. Acoust. Soc. Am. **89**, 2866–2882.

Meddis, R., and Hewitt, M. J. (**1991b**). **“**Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: Phase sensitivity,” J. Acoust. Soc. Am. **89**, 2883–2894.

Moore, B. C., Glasberg B. R., and Peters, R. W. (**1984**). “Frequency and intensity difference limens for harmonics within complex tones,” J. Acoust. Soc. Am. **75**, 550-561.

Moore, B. C., Peters, R. W., and Glasberg, B. R. (**1985**). “Thresholds for the detection of inharmonicity in complex tones,” J. Acoust. Soc. Am. **77**, 1861-1867.

Neuert, V., Pressnitzer, D., Patterson, R. D., and Winter, I. M. (**2001**). “The responses of single units in the inferior colliculus of the guinea pig to damped and ramped sinusoids,” Hear. Res. **159**, 36-52.

Oxenham, A. J., Bernstein, J. G., and Penagos, H. (**2004**). “Correct tonotopic representation is necessary for complex pitch perception,” Proc. Natl. Acad. Sci. USA **101**, 1421-1425.

Patterson, R. D., Allerhand, M. H., and Giguere, C. (**1995**). “Time domain modelling of peripheral auditory processing: A modular architecture and a software platform,” J. Acoust. Soc. Am. **98**, 1890-1894.

Pressnitzer, D., Winter, I. M., and Patterson, R. D. (**2000**). “The responses of single units in the ventral cochlear nucleus of the guinea pig to damped and ramped sinusoids,” Hear. Res. **149**, 155-166.

Roberts, B., and Moore, B. C. (**1990**). “The influence of extraneous sounds on the perceptual estimation of first-formant frequency vowels,” J. Acoust. Soc. Am. **88**, 2571-2583.

Ryan, A. (**1976**). “Hearing sensitivity of the mongolian gerbil *Meriones unguiculatis*”, J. Acoust. Soc. Am. **59**, 1222-1226.

Sheft, S., and Yost, W.A. (**1990**). “Temporal integration in amplitude modulation detection,” J. Acoust. Soc. Am. **88**, 796-805.

Shinn-Cunningham, B. G. (**2008**). “Object-based auditory and visual attention,” Trends Cogn. Sci. **12**, 182-186.

Sinnott, J. M., Brown, C. H., and Brown, F. E. (**1992**). “Frequency and intensity discrimination in Mongolian gerbils, African monkeys and humans,” Hear. Res. **59**, 205-212.

Stein, A., Ewert, S.D., and Wiegrebe, L. (**2005**). “Perceptual interaction between carrier periodicity and amplitude modulation in broadband stimuli: A comparison of the autocorrelation and modulation-filterbank model,” J. Acoust. Soc. Am. **118**, 2470-2481.

Terhardt, E. (**1974**). “Pitch, consonance and harmony,” J. Acoust. Soc. Am. **55**, 1061-1069.

Wiegrebe, L., and Sonnleitner, V. (**2007**). “Perceptual interaction between carrier periodicity and amplitude-modulation in the gerbil (*Meriones unguiculatus*),” J. Comp. Physiol. [A] **193**, 305-312.

Wightman, F. L. (**1973**). “The pattern-transformation model of pitch,” J. Acoust. Soc. Am. **54**, 407-416.

5 Effect of harmonicity on the detection of a signal in a complex masker and on spatial release from masking[1]

---

[1] This chapter has been submitted (25.06.2010) in the present form to JARO.

5.1 Abstract

The masking of sounds from one source (target) by sounds from a competing source (maskers) can be reduced by spatial separation (spatial release from masking, SRM). A free-field simultaneous masking experiment was conducted to evaluate the effects of harmonicity and inharmonicity of a complex masker and of the spatial segregation of target and masker on the detection of a pure tone target. Thresholds were determined for combinations of two target frequencies (1 and 8 kHz), two spatial configurations (co-located, masker and target spatially separated by 90° azimuth), and five different masker types (four complex stimuli, one noise masker). The amount of masking significantly depended on the masker type. The inharmonic relation between a harmonic complex masker and a mistuned target improved the masked thresholds by about 7.5 and 17 dB (1 and 8 kHz targets, respectively) compared to thresholds estimated for purely energetic masking. Thresholds were additionally decreased if target and masker were spatially separated. A harmonic relation between target and a harmonic complex masker had no effect on the amount of masking nor the amount of SRM in the 1 kHz condition; masked threshold was improved in the 8 kHz condition. An inharmonic masker with a random but constant frequency composition throughout a session resulted in thresholds similar to those obtained with the harmonic complex masker in all conditions. An inharmonic complex masker with a constantly varying random frequency composition resulted in an elevated threshold in the 1 kHz condition compared to the predicted energetic masking. It is suggested that additional informational masking occurred in this condition. The results indicate that harmonicity cues affect the detectability of a tonal target in a complex masker, but (except for the informational masker in the 1 kHz condition) the influence of harmonicity on spatial unmasking is generally negligible.

5.2 Introduction

In an acoustically complex environment, the perception of behaviorally important target sound is often compromised by interfering masking stimuli. The amount of masking depends on the spectral and temporal characteristics of masker and target as well as on their relative spatial location. Thus, segregating the target from competing sounds involves a variety of cues, such as differences in harmonicity, spectral content, modulation patterns, common onset and binaural differences in time of arrival and intensity. In the natural environment these cues rarely occur in isolation. So far, only few studies focussed on the mechanisms underlying the perceptual interaction of these cues (e.g. Carhart *et al.*, 1969; Best *et al.*, 2005; McDonald and Alain, 2005; Gallun *et al.*, 2008; Kopco and Shinn-Cunningham, 2008; Epp and Verhey, 2009).

How well a target is segregated from a masking background is indicated by the detection threshold. It is not only due to energetic masking that arises from an overlap of excitation patterns in the cochlea caused by the target and the masker but also non-energetic effects influence the amount of masking (e.g. Carhart *et al.*, 1969; Neff *et al.*, 1993; Kidd *et al.*, 1998; Durlach *et al.*, 2003). These non-energetic masking effects that result in an additional threshold elevation (Lutfi, 1990) have been merged into the concept of "informational masking" and occur when target and masker share stimulus characteristics or if there is a high stimulus uncertainty. On the contrary, grouping cues can improve the detection of a signal in a masker (e.g. as in comodulation masking release, e.g. Hall and Grose, 1990).

The detectability of a target in a masker can be considerably improved by separating both spatially (e.g. Cherry, 1953; Saberi *et al.*, 1991; McDonald and Alain, 2005). This spatial release from masking (SRM) is defined as the difference between the masked threshold in the co-located configuration (masker and target from the same direction) and the masked threshold in the spatially separated configuration (masker and target coming from different directions).

The aim of the present free-field study is threefold. First, we wish to explore the influence of harmonicity cues on the detection of a sinusoidal signal in different harmonic and inharmonic complex maskers. It has been shown that harmonicity is an important grouping cue and that a mistuned component of a harmonic complex can easily be detected or even be segregated from the remaining complex (Moore *et al.*, 1985; Klinge and Klump, 2010). Secondly, we want to examine the extent to which the processing of harmonicity cues interacts with the processing of spatial cues in determining the spatial release from masking for the different complex masker types. How such a combination of cues is processed in the auditory system has rarely been examined (Neff *et al.*, 1993; McDonald and Alain, 2005). Third, by comparing perception in a low and a high frequency region we want to examine whether the auditory system uses different mechanisms for signal detection in complex maskers when tonal masker components are resolved (i.e., only one component falling into an auditory filter) or unresolved (i.e., more than one component falling into an auditory filter). Furthermore, the frequency region affects monaural and binaural cues that can be exploited by mechanisms for spatial unmasking.

5.3 Materials and methods

5.3.1 Subjects

Five listeners – one male, four females, including the first author – participated in the experiments. They were between 21 and 31 years old (average of 25.6 years) and had normal hearing between 250 Hz and 8000 Hz within 15 dB HL. Two subjects were experimentally naïve. Three subjects had previously participated in psychoacoustic experiments, and one of these subjects (the author) obtained prior experience with the stimuli in a pilot experiment preceding the present study. Each listener received at least two hours of training before data collection began using the same paradigm as in the actual experiment. The subjects took between 5 to 15 days to complete the experiments.

5.3.2 Apparatus and stimulus generation

The free-field experiment took place in the anechoic chamber of the University of Oldenburg which fulfills the requirements for free-field measurements down to a lower cut-off frequency of 50 Hz. Subjects were placed on a seat in the middle of the chamber with their head position fixed by a head-rest mounted on the backrest of the chair. Two loudspeakers (Canton XS Plus) were placed 1.5 m from the subject's head, one at 0° azimuth and the other at 90° to the right. They were adjusted in height for each individual to maintain an elevation of 0°. For visual feedback, a 15" flat-panel display was placed next to but slightly behind the plane of the front loudspeaker. Subjects controlled their experiment via a standard keyboard.

On each day of experiment, the setup was calibrated using a sound level meter (2238 Mediator, Brüel&Kjær) with the microphone placed at the position of the subject's head and facing the loudspeaker that was to be adjusted.

Stimuli were generated using a mobile Linux workstation and an RME sound card (Hammerfall DSP Multiface II connected via PCMCIA card). The output of the sound card was passed through a stereo amplifier (Harman Kardon HK 6350R) and sent to the two loudspeakers.

The target stimulus was a pure tone of 1 or 8 kHz, respectively, starting in sine phase and a duration of 125 ms including 25 ms Hanning ramps at stimulus onset and offset. The masker was switched on simultaneously with the target and had the same duration and on- and offset. We examined the effect of five different masker types on the detection of the target. Four of the five maskers were complex stimuli

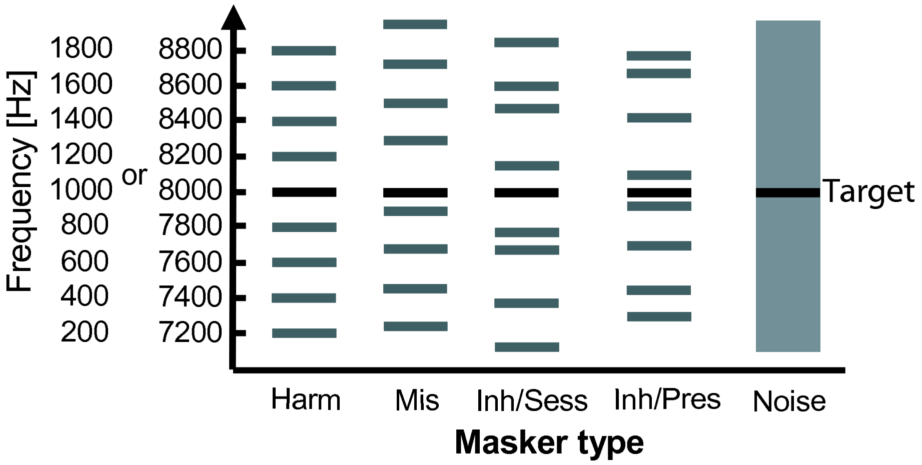

Figure 5.1: Schematic drawing of the five masker types (grey) in the two frequency regions used in the experiment as indicated on the ordinate. "Harm" = harmonic, "Mis" = mistuned compared to target, "Inh/Sess" = inharmonic, frequencies varied per session, "Inh/Pres" = inharmonic, frequencies varied per stimulus presentation, "Noise" = bandpass noise. See text for further explanations of the maskers. Target signals were 1 or 8 kHz pure tones, respectively (black).

composed of tones starting in sine phase and the fifth masker was a bandpass noise masker (Fig. 5.1). The first masker type (“Harm”) was a harmonic complex with a fundamental frequency (F0) of 200 Hz comprised of four harmonics below and four harmonics above the target frequency (harmonic frequencies from 200 to 1800 Hz and from 7200 to 8800 Hz for the 1 and the 8 kHz target frequency, respectively). The second masker type (“Mis”) was a harmonic complex that also consisted of four harmonics below and four harmonics above the target frequency. The F0 for the target frequency of 1 kHz was 211 Hz (harmonic frequencies from 211 to 1899 Hz) and for the target frequency of 8 kHz it was 209 Hz (harmonic frequencies from 7106 to 8778 Hz). Due to the different F0 compared to the first masker type, the target tone was mistuned to the otherwise harmonic complex. The F0s for the low and the high target frequency were slightly different to ensure a sufficient spacing between the target frequency and the adjacent harmonics of the masker. The third and fourth masker type (“Inh/Sess”: inharmonic per session, and “Inh/Pres”: inharmonic per presentation) were both maskers with inharmonic components that randomly deviated in frequency from the four harmonics below and above the target frequency presented in the harmonic complex masker. The frequencies for the “inharmonic per session” masker type were selected according to three conditions: (1) to be inharmonic to the target frequency, (2) to be in no harmonic relation to any other component at any F0, and (3) to lie within a pre-selected frequency of ±100 Hz around the nominal harmonic frequency. For every session that was conducted, a new composition of 8 different frequencies was used. The frequencies for the “inharmonic per presentation” masker type were randomly selected within a frequency range of ±75 Hz around the nominal harmonic frequency without the restrictions mentioned above and a new frequency composition was used for each masker stimulus that was presented to the subject (see

Procedure). The frequency range for the “Inh/Sess” from which the frequencies for the components were drawn by a customized computer script was increased to ±100 Hz to provide enough variability in the selected frequencies. The sound-pressure level of each harmonic component of the four complex masker stimuli were set to 60 dB SPL which resulted in an overall masker level of 69 dB SPL. The fifth masker was a bandpass noise with a frequency range from 177 to 1910 Hz for the 1 kHz target frequency condition and from 6799 to 9287 Hz for the 8 kHz target frequency condition. The lower cutoff frequency was chosen to be at half an auditory filter bandwidth below the frequency of the first harmonic of the harmonic masker and the higher cutoff frequency was chosen to be at half a filter bandwidth above the frequency of the highest harmonic. The spectral density was set to 36 dB/Hz which resulted in an overall noise masker level of 69 dB SPL. There was no spectral notch in the bandpass noise. At the beginning of each session a bandpass noise with a duration of 30 s was generated and for each stimulus presentation an 125-ms token was cut out at a random position to ensure that subjects would not learn a “frozen” noise token.

The maskers were always presented from the front loudspeaker. The pure tone target was either presented from the front (co-located configuration) or presented from 90° from the right (spatially separated configuration). The resulting 20 conditions (2 target frequencies × 2 spatial configurations × 5 masker types) were divided into two experimental series – one for the 1 kHz target frequency and one for the 8 kHz target frequency. The presentation order of the experimental series as well as the order of the conditions within each series were randomized for each subject.

5.3.3 Procedure

A Go/NoGo paradigm with a continuously repeating background stimulus was used to determine the masked threshold of the sinusoidal signal. The masker alone was presented every 1.3 s forming the repeating background stimulus. Pushing a button on the keyboard started a trial and initiated a random waiting interval of between one and seven seconds. After the waiting interval the test stimulus was presented which could either be the masker plus the simultaneously presented pure tone target signal (Go-stimulus) or the masker alone (NoGo-stimulus). The subject indicated the detection of the signal in the masker by pressing a second button on the keyboard. For each correct response ("Hit") a visual feedback was given and the subject pushed the first button again to start the next trial. If the subject did not detect the signal in the masker, the trial was counted as a "Miss" and the next trial was initiated automatically. Trials in which the test stimulus was the masker alone ("catch trials") occurred with a probability of 30%. A response of the subject within the response time during a catch trial was counted as a "False Alarm". "Hit" rates and "False Alarm" rates were used to calculate the sensitivity measure d'.

A session consisted of 11 blocks of ten trials each with the first ten trials serving as warm-ups. Within each block seven different test trials and three catch trials were presented in a randomized order. Target levels (step size 3 dB) were chosen according to the method of constant stimuli. The range from the lowest to the highest target level was adjusted before each session to provide both sub-threshold and supra-threshold stimuli. During the warm-up block at the beginning of each session only the most salient test stimuli with the largest target-to-masker ratio and three catch trials were presented.

5.3.4 Data analysis

A session was accepted as being valid based on two conditions: (1) the two most salient stimuli must have an average hit rate of 80%, and (2) the false alarm rate must not exceed 20%. The threshold in a valid session was determined as the level of the target resulting in a d' of 1.8. Data from two consecutive valid sessions in which thresholds differed no more than 3 dB from each other were combined to determine the final masked threshold at a d' of 1.8. After determining the last threshold in an experimental series, the subject had to repeat the first threshold to ensure that training effects are excluded. If the threshold obtained for a second time differed by more than 3 dB from the threshold obtained the first time, then the next threshold in series had to be repeated until the repeated threshold matched the threshold obtained the first time (difference $\leq$ 3 dB). Data were always taken from the last threshold measurement. The statistical software packages SPSS 17.0 and SigmaStat 2.03 were used to analyze the data.

5.4 Results

5.4.1 Effect of masker type on the masked thresholds

A repeated measures ANOVA revealed significant main effects of the within-subject factors target frequency, spatial configuration and masker type on the masked thresholds (target frequency: $F_{1,4} = 22.71, p < 0.01$; spatial configuration: $F_{1,4} = 101.00, p < 0.01$, masker type: $F_{4,16} = 61.35, p < 0.001$). All two-way interactions were significant ($p < 0.05$).

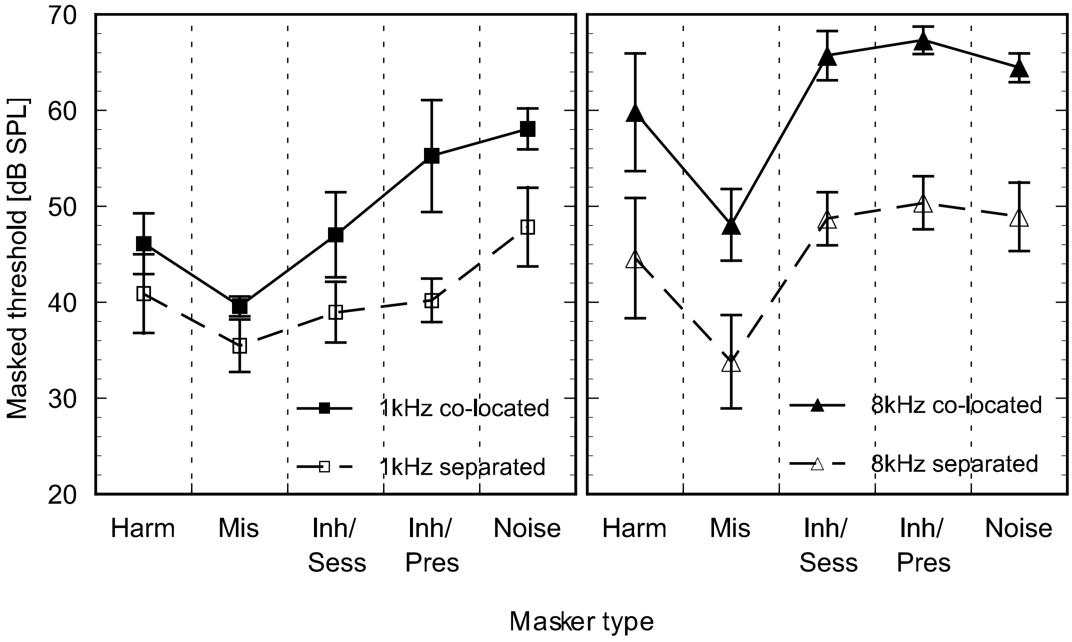

Figure 5.2: Mean masked thresholds in absolute values separated by target frequency (left panel = 1 kHz, right panel = 8 kHz). Masker types are: "Harm" = harmonic, "Mis" = mistuned, "Inh/Sess" = inharmonic per session, "Inh/Pres" = inharmonic per presentation, "N" = bandpass noise. Error bars represent two times the standard error (2SE).

Fig. 5.2 shows that the difference between masked thresholds of the co-located configuration (filled symbols and continuous line) and the spatially separated configuration (open symbols and dashed line) were always higher for the high-frequency condition (8 kHz target frequency, right panel) than for the low-frequency condition (1 kHz target frequency, left panel). Thresholds for the noise masker in the 1 kHz condition are the highest compared to those of the complex stimuli maskers. Thresholds for the mistuned masker tended to be the lowest from all masker types for every condition and spatial configuration.

To evaluate which masker type differed significantly from each other, a RM ANOVA with planned contrasts based on a priori hypotheses was performed (Table 5.1). Prior to the planned comparisons, however, the data were divided into four subgroups (1 kHz co-located, 1 kHz separated, 8 kHz co-located, and 8 kHz separated) as the main statistical analysis revealed a significant main effect of the target frequency and the spatial configuration on the masked thresholds.

For the first planned contrast, the bandpass noise masker was compared to the group of the four complex stimuli. The bandpass noise masker lacked many of the cues present in the complex stimulus maskers. All complex stimulus maskers had distinct spectral peaks that were invariant throughout the presentation of each stimulus, and the “harmonic”, “mistuned”, and “inharmonic per session” maskers had an invariant temporal pattern over the course of a session which might influence the detectability of the pure tone in the masker. In addition, components of “harmonic” and “mistuned” maskers had a common fundamental, a feature that is not found in the noise masker.

Table 5.1: Results of the four planned contrasts for each of the four subgroups.

subgroups		Contrast 1 noise vs. complex	Contrast 2 harmonic vs. inharmonic	Contrast 3 “Harm” vs. “Mis”	Contrast 4 “Inh/Sess” vs. “Inh/Pres”
1 kHz	co-located	$F = 129.6$	$F = 12.8$	$F = 17.9$	$F = 17.9$
		$p < 0.001$	$p < 0.05$	$p < 0.05$	$p < 0.05$
		$\eta = .97$	$\eta = .76$	$\eta = .82$	$\eta = .82$
	separated	$F = 31.7$	$F = 2.5$	$F = 4.9$	$F = 2.2$
		$p < 0.01$	$p = 0.19$	$p < 0.05$	$p = 0.211$
		$\eta = .89$	n.s.	$\eta = .55$	n.s.
8 kHz	co-located	$F = 5.5$	$F = 51.8$	$F = 34.8$	$F = 2.2$
		$p = 0.079$	$p < 0.01$	$p < 0.01$	$p = 0.208$
		n.s.	$\eta = .93$	$\eta = .90$	n.s.
	separated	$F = 4.3$	$F = 46.4$	$F = 29.6$	$F = 1.9$
		$p = 0.106$	$p < 0.01$	$p < 0.01$	$p = 0.243$
		n.s.	$\eta = .92$	$\eta = .88$	n.s.

Furthermore, the spectral energy in the filter centered at the target frequency in the 1 kHz condition was higher in the noise masker compared to the complex maskers which should result in higher masked thresholds. The planned contrasts revealed that the masked thresholds for the noise masker were only significantly different from the masked thresholds of the four complex maskers in the 1 kHz target frequency condition (Table 5.1) which would support the prediction based on the spectral energy in the filter.

In a second planned contrast, the four remaining complex maskers were divided into a harmonic complex stimulus group ("Harm" and "Mis" with components having a common fundamental) and an inharmonic complex stimulus group ("Inh/Sess" and "Inh/Pres"), and the masked thresholds of both groups were compared with each other. Without harmonicity as a grouping cue in the inharmonic masker types it was hypothesized that subjects perform worse compared to the harmonic complex maskers. The second contrast revealed significant differences between the two harmonic and the two inharmonic maskers in all but the subgroup "1 kHz spatially separated" (Table 5.1).

The third contrast (between the harmonic and the mistuned masker) was based on two hypotheses. First, the inharmonicity between the mistuned masker and the target should provide an additional segregation cue and result in an increased detectability of the target in the co-located masker as well as in an improved segregation in a spatially separated configuration. Secondly, a harmonic relation between the harmonic complex masker and the target should have a strong grouping effect and therefore, should decrease the detectability of the target. The planned contrasts revealed that the thresholds were significantly higher for the harmonic compared to the mistuned masker for all four subgroups (Table 5.1).

The fourth contrast was designed to compare the masked thresholds between the "Inharmonic per session" ("Inh/Sess") and the "Inharmonic per presentation" ("Inh/Pres") masker. The high stimulus uncertainty due to a constantly varying frequency composition with every stimulus presentation in the latter masker should result in increased masked thresholds compared to the "Inh/Sess" masker in which the frequency composition only changed with every session. According to several studies that used highly uncertain masker types (informational maskers), also an increased spatial release from masking at least in low-frequency regions can be expected with spatial separation between target and masker (e.g. Neff and Green, 1987; Kidd *et al.*, 1994; Neff and Dethlefs, 1995; Best *et al.*, 2005). However, the planned comparison only revealed significant differences between thresholds of both masker types for the subgroup "1 kHz co-located".

5.4.2 Effect of masker type on the spatial release from masking

The spatial release from masking was calculated by subtracting the thresholds of the spatially separated configuration from the thresholds of the co-located configuration for each masker type and at each target frequency. Figure 5.3 shows the mean thresholds for the two target frequencies at each of the five maskers. A RM ANOVA revealed significant main effects of the target frequency ($F_{1,4} = 352.984, p < 0.001$) and the masker type ($F_{4,16} = 3.27, p < 0.05$) on the amount of SRM, but there was no significant interaction ($F_{4,16} = 2.79, p = 0.062$). As the target frequency had a significant main effect on the amount of SRM, the data were divided into two subgroups (1 kHz and 8 kHz target frequency) for further statistical analysis. A RM ANOVA with a Tukey's HSD *post hoc* test was performed within each of the two subgroups to evaluate which of the masker types differed significantly from each other in their amount of SRM. In the 1 kHz subgroup, the main effect of masker type was significant ($F_{4,16} = 5.3, p < 0.01$). However, the only significant differences in the pair-wise comparisons were found for the "Inharmonic per presentation" masker compared to the harmonic masker ($p < 0.05$) and to the mistuned masker ($p < 0.01$). In the 8 kHz subgroup, the RM ANOVA revealed no significant main effect of the masker type on the amount of SRM and no *post hoc* tests were made.

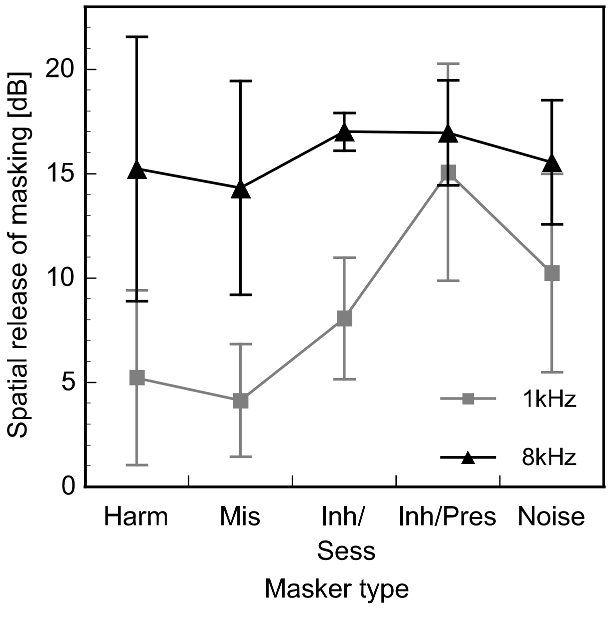

Figure 5.3: Amount of spatial release from masking (difference between co-located and spatially separated configuration) for the 1 kHz target frequency (grey, squares) and for the 8 kHz target frequency (black, triangles) for each masker. Error bars represent two times the standard error (2SE).

5.5 Discussion

5.5.1 Influence of masker type on the detection threshold

The current experiment showed that the detection of the sinusoidal target in the masker depended significantly on the type of the masker. Comparable data for some of the conditions of the present study can be found in studies by Oh and Lutfi (2000) and McDonald and Alain (2005). In a study using monaural headphone stimulation, Oh and Lutfi (2000) showed that the amount of masking depended on whether the pure tone signal (1000 or 1047 Hz) was presented in a harmonic, a mistuned or an inharmonic masker with a constantly varying frequency composition. Similar to the results in this study, the highest masked thresholds were obtained for inharmonic maskers, a lower masking effect was found for harmonic and the least amount of masking was found for the mistuned maskers. McDonald and Alain (2005) showed in a free-field study that the probability of perceiving two sounds depended on the harmonic relation and the spatial location between a tone complex (F0 200 Hz, 10 harmonics) and its 3rd component (harmonic or mistuned by 2% or 16%). Subjects were better at segregating the component from the complex if it was inharmonic to the complex rather than harmonically related. Spatial location also had an effect, but there was no interaction.

One factor determining the masked threshold that will be discussed below is energetic masking. Besides energetic masking, however, other effects might account for the observed differences in masked thresholds. It has been shown, for example, that informational masking effects increase the amount of masking (e.g. Carhart *et al.*, 1969) and comodulated maskers decrease the amount of masking (e.g. Hall *et al.*, 1984). Grouping by harmonicity might be another factor that elevates the detection threshold for the target in a harmonic complex masker and mistuning might enhance the detectability of the target in the mistuned masker. To be able to differentiate between these effects, we estimated the amount of purely energetic masking and compared the results to the observed masked thresholds (Table 5.2). The calculation of each of these energetic masking thresholds (EMTs) was carried out in a two-stage model. The first stage included monaural peripheral preprocessing and binaural processing. The intensities of the signal and the masker for each target frequency (1 and 8 kHz) and each spatial configuration (co-located and separated) were estimated under the assumption that an equalization-cancellation (EC) process was performed by the auditory system (Durlach, 1963; Beutelmann *et al.*, 2010). The EC processing takes advantage of the fact that signals from different directions result in different interaural time and level differences. However, if the estimated output of the EC process for binaural processing was smaller than the estimated monaural target-to-masker ratios (TMRs), it was assumed that the auditory system bases its decision on these monaural effects. The second stage implemented a signal detection mechanism and the masked threshold was determined using the simultaneous masking model provided by Buus (1997, Eq. 6) and a d' of 1.8.

Table 5.2: Comparison between observed thresholds and estimated energetic masking thresholds (EMTs, in dB SPL) for all target frequencies and spatial conditions. Last two columns show the estimated and the observed amounts of spatial release from masking (SRM) in dB. EMTs were estimated using an equalization-cancellation (EC) model (Beutelmann *et al.*, 2010) and an energy detector model (Buus, 1997). The left- and right-ear-signals were processed through HRTFs (data from a symmetric, large pinna KEMAR manikin of the MIT) to generate the spatialized stimuli. The calculation was repeated 25 times in a Monte-Carlo-simulation to account for the varying spectro-temporal characteristics of the noise, the inharmonic per session (“Inh/Sess”) and the inharmonic per presentation (“Inh/Pres”) masker and the mean thresholds and their standard deviations are displayed for these maskers.

Target frequency	Masker type	Predicted Energetic Masking		Observed Threshold		Predicted SRM	Observed SRM
		co-located	separated	co-located	separated		
1 kHz	Bandpass noise	60.7 ± 0.7	56.2 ± 0.6	58.1	47.9	4.5	10.2
	Harmonic	45.8	41.4	46.1	40.9	4.4	5.2
	Mistuned	47.2	42.8	39.6	35.5	4.4	4.1
	“Inh/Sess”	50.1 ± 4.5	44.0 ± 5.4	47.0	39.0	6.1	8.1
	“Inh/Pres”	46.6 ± 3.5	41.9 ± 4.1	55.3	40.2	4.7	15.1
8 kHz	Bandpass noise	66.0 ± 0.3	51.5 ± 0.3	64.5	48.9	14.5	15.5
	Harmonic	65.8	51.3	59.8	44.6	14.5	15.2
	Mistuned	65.4	50.8	48.1	33.8	14.6	14.3
	“Inh/Sess”	66.0 ± 0.4	51.4 ± 0.5	65.7	48.7	14.6	17.0
	“Inh/Pres”	66.0 ± 0.3	51.4 ± 0.2	67.3	50.4	14.6	16.9

For the noise masker, the EMT in the co-located condition was similar to the observed threshold for both target conditions (1 and 8 kHz) suggesting a purely energetic masking. For the spatially separated configuration in the 1 kHz condition, however, the observed threshold was about 8.5 dB lower than the EMT and yielded about 10 dB SRM. The results for the 1 kHz condition of the present study in terms of the masking and the spatial/binaural release from masking are supported by results obtained in a free-field study by Gilkey and Good (1995) and a headphone experiment by van de Par and Kohlrausch (1999).

For the harmonic masker ("Harm"), the harmonic relation between the components and between the masker and the target was hypothesized to be a strong cue to group the target to the masker resulting in an increased masked threshold. The results showed that the deviation from the EMT was different for the two target frequencies. In the 1 kHz condition, the observed masked thresholds for the harmonic masker were similar to the EMT for both spatial configurations and similar to the observed threshold for the "inharmonic per session" masker. Thus, the assumption that the detectability of the target in the masker decreases due to the strong grouping effect of the harmonicity cues could not be confirmed. In the 8 kHz condition, the observed thresholds were about 6 dB lower than the EMT in both spatial configurations suggesting that an additional cue resulting in a release from masking could be exploited by the subjects in the harmonic condition. Such an additional cue in an auditory filter with more than two harmonics interacting in a filter could be a change in the otherwise constant envelope of the temporal waveform when adding the target to the masker during a test stimulus. The ability to detect such a change in the temporal waveform has been suggested before (e.g. Moore and Glasberg, 1989; Klinge and Klump, 2010).

The masked thresholds for the mistuned masker yielded, as predicted, the lowest thresholds compared to the other maskers (about 7.5 and 17.5 dB lower than the EMT in each target frequency condition and spatial configuration). These large improvements in masked thresholds indicate that mistuning provided a strong additional cue to detect the mistuned target in the otherwise harmonic masker which is also supported by the significant difference in thresholds between the harmonic and the mistuned masker for all four subgroups (Table 5.1). For the low frequency region in which the harmonics of the complex masker are resolved, the enhanced detectability may have been mediated by the segregation of the mistuned target from the harmonic complex (e.g. Moore *et al.*, 1985). Roberts and Brunstrom (2001) proposed that the perceptual segregation of a low-frequency mistuned component relies on an across-channel comparison of periodicity information. In the 8 kHz condition in which harmonics are unresolved and the pitch of the complex masker is weak, the predominant cue might have been the change of the envelope of the temporal waveform by adding the target to the masker or a gradually changing waveform due to the mistuning of the target in relation to the otherwise harmonic masker (e.g. Moore and Glasberg, 1989; Klinge and Klump, 2010).

The inharmonicity between all components and the target in the "Inh/Sess" masker did not seem to impair the masked thresholds as the observed thresholds of this masker were similar to the EMT and to the masked thresholds of the harmonic masker in the 1 kHz conditions. Instead, the constant temporal waveform over the course of one session which only changed when the target was added to the masker might have aided the development of a template during the presentation of the masker-alone-stimulus in the 1 kHz condition. The masker-plus-target stimulus could then have been compared to a stored spectral or temporal image (e.g. Agus *et al.*, 2010). In the 8 kHz condition, however, the similar results between observed and predicted thresholds indicate that a template could not be learned or would not have improved the detectability of the pure tone target.

The assumption that the constantly varying random frequency composition in the "Inh/Pres" masker imposes a high stimulus uncertainty which leads to a high amount of informational masking and increased thresholds was supported only in the "1 kHz co-located" condition. In this condition, the observed threshold was 8.5 dB higher than the EMT and the threshold between the "Inh/Sess" and the "Inh/Pres" masker was significantly different (Table 5.1). The spatial separation of the target from the informational masker yielded observed thresholds that were similar to the EMT. Shinn-Cunningham *et al.* (2005) suggested that such a spatial separation increases the ability of subjects to focus on the desired location and thus, reduces informational masking. For the 8 kHz co-located condition, the observed thresholds were unexpectedly similar to the EMT, suggesting that the permanently varying frequency composition of the "Inh/Pres" masker has no additional informational masking effect in this frequency region. Another possibility would be that the frequency variation of ±75 Hz around the nominal harmonic frequency was not large enough to induce a high stimulus uncertainty as it lies below the pure tone frequency difference limens of about 96 Hz observed at 8 kHz (Moore, 1973).

5.5.2 Influence of masker type on spatial release from masking

In the present study, we wanted to examine the extent to which the processing of harmonicity cues interacts with the processing of spatial cues in determining the SRM for the different complex masker types. The audibility of a target in a spatially separated masker is determined by two main spatial effects: (1) monaural effects such as the "better ear" effect in which the TMR is increased in one of the ears when the target comes from a different spatial location than the masker (Shinn-Cunningham *et al.*, 2005), and (2) binaural effects that result from interaural time or level differences (ITD and ILD). Binaural models (e.g., the EC model or the lateralization model, see, e.g. Hafter, 1971) were traditionally used to estimate the amount of spatial or binaural release from masking. However, Shinn-Cunningham *et al.* (2005) pointed out that they fail to accurately predict spatial or binaural release from masking in situations in which masker and target are similar in their spectro-temporal characteristics and additional informational masking occurs.

The EC model that was used in the present experiment to estimate masked thresholds for the co-located and the spatially separated configuration was relatively accurate in predicting the amount of SRM for all masker types for the 8 kHz target frequency and for the “Harm”, “Mis”, and the “Inh/Sess” masker for the 1 kHz target frequency. For the “Harm” masker this result indicates that harmonicity as a grouping cue was irrelevant in determining the SRM at 1 kHz (Table 5.2). This is consistent with the observation by McDonald and Alain (2005) that the perceptual segregation of a component in a complex by differences in spatial cues is similar for harmonics and mistuned components of a tone complex.

As hypothesized for the mistuned masker, the mistuning cue could be exploited in both spatial configurations for improving masked thresholds but the mistuning did not affect the amount of SRM. This “summation” of masking releases, in the current experiment due to mistuning and spatial separation, has also been observed for comodulation and dichotic listening conditions in other studies and a serial processing of these cues in the auditory system was proposed (Cohen and Schubert, 1991; Epp and Verhey, 2009).

The “Inh/Pres” masker in the 1 kHz condition was the only condition in which the SRM significantly differed from that obtained with the harmonic masker or the mistuned masker. The large difference between the predicted SRM and the observed SRM in this condition indicates that the “Inh/Pres” masker with its high stimulus uncertainty yields an additional release from masking that cannot be accounted for by purely energetic masking (e.g., see also Shinn-Cunningham *et al.*, 2005). In contrast to the results of our study, Gallun *et al.* (2008) could not find a clear difference in the processing of binaural cues for an informational masker with high stimulus uncertainty (random-frequency multitone masker) and noise maskers.

The SRMs predicted by binaural processes implemented in the EC-model were generally similar or smaller than the SRMs predicted purely on the TMR in the “better ear”. This suggests that a monaural mechanism determines the SRM for those masker types for which similar observed and predicted SRMs were found.

5.6 Acknowledgements

We thank Volker Mellert for providing access to the anechoic chamber, Bastian Epp for help regarding the setup, Bastian Epp and Stefan Strahl for helpful comments on earlier versions of the manuscript, and the four subjects taking part in the experiment. This study was supported by the DFG within the Sonderforschungsbereich SFB/TRR31 “The Active Auditory System”.

5.7 References

Agus, T. R., Thorpe, S. J., and Pressnitzer, D. (**2010**). “Rapid formation of robust auditory memories: insights from noise”, Neuron **66**, 610–618.

Best, V., Ozmeral, E., Gallun, F. J., Sen, K., and Shinn-Cunningham, B. G. (**2005**). "Spatial unmasking of birdsong in human listeners: energetic and informational factors", The Journal of the Acoustical Society of America **118**, 3766–3773.

Beutelmann, R., Brand, T., and Kollmeier, B. (**2010**). "Revision, extension, and evaluation of a binaural speech intelligibility model", The Journal of the Acoustical Society of America **127**, 2479–2497.

Buus, S. (**1997**). "Auditory masking", in *Encyclopedia of Acoustics*, edited by M. Crocker, volume 3, 1427–1445 (Wiley, New York).

Carhart, R., Tillman, T. W., and Greetis, E. S. (**1969**). "Perceptual masking in multiple sound backgrounds", The Journal of the Acoustical Society of America **45**, 694–703.

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**, 975–979.

Cohen, M. F. and Schubert, E. D. (**1991**). "Comodulation masking release and the masking-level difference", The Journal of the Acoustical Society of America **89**, 3007–3008.

Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural Masking-Level differences", The Journal of the Acoustical Society of America **35**, 1206–1218.

Durlach, N. I., Mason, C. R., Kidd, G., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (**2003**). "Note on informational masking", The Journal of the Acoustical Society of America **113**, 2984–2987.

Epp, B. and Verhey, J. L. (**2009**). “Superposition of masking releases”, Journal of Computational Neuroscience **26**, 393–407.

Gallun, F. J., Durlach, N. I., Colburn, H. S., Shinn-Cunningham, B. G., Best, V., Mason, C. R., and Kidd, G. (**2008**). “The extent to which a position-based explanation accounts for binaural release from informational masking”, The Journal of the Acoustical Society of America **124**, 439–449.

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

Hafter, E. R. (**1971**). “Quantitative evaluation of a lateralization model of Masking-Level differences”, The Journal of the Acoustical Society of America **50**, 1116–1122.

Hall, J. W. and Grose, J. H. (**1990**). “Comodulation masking release and auditory grouping”, The Journal of the Acoustical Society of America **88**, 119–125.

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**, 50–56.

Kidd, G., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colburn, H. S. (**1994**). “Reducing informational masking by sound segregation”, The Journal of the Acoustical Society of America **95**, 3475–3480.

Kidd, G., Mason, C. R., Rohtla, T. L., and Deliwala, P. S. (**1998**). “Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns”, The Journal of the Acoustical Society of America **104**, 422–431.

Klinge, A. and Klump, G. M. (**2010**). “Mistuning detection and onset asynchrony in harmonic complexes in mongolian gerbils”, The Journal of the Acoustical Society of America **128**, 280–290.

Kopco, N. and Shinn-Cunningham, B. G. (**2008**). “Influences of modulation and spatial separation on detection of a masked broadband target”, The Journal of the Acoustical Society of America **124**, 2236–2250.

Lutfi, R. A. (**1990**). “How much masking is informational masking?”, The Journal of the Acoustical Society of America **88**, 2607–2610.

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**, 1593–1604.

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

Moore, B. C. and Glasberg, B. R. (**1989**). “Difference limens for phase in normal and hearing-impaired subjects”, The Journal of the Acoustical Society of America **86**, 1351–1365.

Moore, B. C., 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**, 1861–1867.

Neff, D. L. and Dethlefs, T. M. (**1995**). “Individual differences in simultaneous masking with random-frequency, multicomponent maskers”, The Journal of the Acoustical Society of America **98**, 125–134.

Neff, D. L., Dethlefs, T. M., and Jesteadt, W. (**1993**). “Informational masking for multicomponent maskers with spectral gaps”, The Journal of the Acoustical Society of America **94**, 3112–3126.

Neff, D. L. and Green, D. M. (**1987**). “Masking produced by spectral uncertainty with multicomponent maskers”, Perception & Psychophysics **41**, 409–415.

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

Roberts, B. and Brunstrom, J. M. (**2001**). “Perceptual fusion and fragmentation of complex tones made inharmonic by applying different degrees of frequency shift and spectral stretch”, The Journal of the Acoustical Society of America **110**, 2479–2490.

Saberi, K., Dostal, L., Sadralodabai, T., Bull, V., and Perrott, D. R. (**1991**). “Free-field release from masking”, The Journal of the Acoustical Society of America **90**, 1355–1370.

Shinn-Cunningham, B. G., Ihlefeld, A., Satyavarta, and Larson, E. (**2005**). “Bottom-up and top-down influences on spatial unmasking”, Acta Acustica united with Acustica **91**, 967–979.

van de Par, S. and Kohlrausch, A. (**1999**). “Dependence of binaural masking level differences on center frequency, masker bandwidth, and interaural parameters”, The Journal of the Acoustical Society of America **106**, 1940–1947.

Danksagungen

Ich möchte allen Personen, die an der Entstehung meiner Doktorarbeit beteiligt waren, ganz herzlich für ihre Hilfe und Unterstützung danken.

An erster Stelle möchte ich meinem Betreuer Prof. Dr. Georg Klump danken - dafür, dass er von Beginn an an mich geglaubt hat, dass er mich jederzeit unterstützt hat und dass er mich immer ein wenig unter Druck gesetzt hat. Sein enormes Wissen über die vorhandene Literatur, über Elektrotechnik, über das auditorische System und über eigentlich sonst alles beeindruckt mich immer noch und ich danke ihm dafür, dass ich von diesem Wissen und von seiner wissenschaftlichen Exzellenz profitieren durfte.

An zweiter Stelle möchte ich Prof. Dr. Christine Köppl danken, dass sie das Zweitgutachten übernommen hat.

Des Weiteren bin ich dem Internationalen Graduiertenkolleg (InterGK 591) „Neurosensory science, systems and applications" unter Leitung von Prof. Dr. Dr. Birger Kollmeier zu Dank verpflichtet. Als Stipendiatin durfte ich unheimlich viel „Sensorisches" in Seminaren, Workshops, Summer Schools und Vorträgen lernen. Auch sind aus dieser Gruppe viele Freundschaften erwachsen und ein Netzwerk, was von unschätzbarem Wert sein wird. Insbesondere möchte ich dabei Dr. Bastian Epp danken, der mich in meinen ersten Tagen als Hörforscherin in das Programm Matlab eingeführt hat und seitdem haben wir unzählige Stunden mit Diskussionen, gemeinsamen Organisationen und Unternehmungen verbracht. Susanne Garre möchte ich in diesem Zusammenhang für ihre Organisation, Unterstützung und Verwaltung als Sekretärin des InterGK danken.

Großer Dank geht auch an Prof. Dr. Jesko Verhey, der mir jederzeit, unter anderem als Mentor, mit Rat und Diskussion zur Seite stand.

Ich danke Prof. Dr. i.R. Volker Mellert dafür, dass ich seinen reflexionsarmen Raum in der Physik nutzen durfte und danke allen Beteiligten, die mir dort mit Material und Wissen oder der Bereitstellung von Messzeit geholfen haben. Vielen Dank auch an Anita Gorges für ihre Hilfsbereitschaft beim Erstellen der Audiogramme all meiner Versuchspersonen.

Ganz herzlich möchte ich Susanne Groß danken, die meine Renner exzellent am Laufen gehalten hat und unzählige Male kurzfristig ausgeholfen hat.

Ich danke Dr. Ulrike Langemann, weil sie unter anderem für alles Organisatorische in der Arbeitsgruppe zuständig war und ist und somit ein reibungsloser Ablauf meiner Doktorarbeit erst möglich war.

Ich möchte natürlich auch der gesamten Arbeitsgruppe Zoophysiologie und Verhalten ganz herzlich danken. Ihr habt mich herzlich in eurer Mitte aufgenommen und ich habe mich die ganzen Jahre über sehr wohl hier gefühlt. Vielen Dank für all die kleinen und großen Hilfestellungen, Diskussionen, Gespräche, Mittagessen, Freundschaften etc.

Vor allem möchte ich dabei Lena-Vanessa Dollezal danken, die mir eine enge Freundin geworden ist und die so viel organisiert und mich unterstützt hat und die immer ein offenes Ohr für meine Probleme hat. Außerdem möchte ich Holger Dierker danken, durch den ich sehr viel über Computer und über Linux und über all die am Anfang sehr verwirrenden Programme gelernt habe. Großer Dank geht auch an Dr. Rainer Beutelmann, der uns mit seinen Fähigkeiten in Signalverarbeitung, Programmierung und Modellierung unschätzbare Dienste geleistet hat und leistet und mit dem ich sehr viele fruchtbare auditorische Diskussionen hatte und habe. Mein Dank gilt weiterhin Dr. Melanie Zokoll-van der Laan, Dr. Julia Maier-McAlpine und Dr. Karin Klink, die mir von Beginn meiner Oldenburger Zeit an ihre Unterstützung und ihre Freundschaft geboten haben.
Ich möchte hier auch nochmal meinen Diplomanden und Bachelor-Studenten danken, allen voran Anne Vellage, die an der Entstehung vieler Ergebnisse beteiligt waren.
Ich bedanke mich bei all meinen Versuchspersonen und gedenke hier nochmal all meiner noch lebenden bzw. schon verstorbenen Wüstenrennmäuse, ohne die diese Arbeit ja auch nicht hätte entstehen können.

An dieser Stelle möchte ich nun meiner Familie, insbesondere meinen Eltern von ganzem Herzen für ihre bedingungslose Unterstützung und ihren Glauben an mich, oder eigentlich ihrem Wissen, dass ich es schaffe, danken.
In tiefer Liebe und von ganzem Herzen danke ich Dr. Stefan Strahl! Er ist die wichtigste Person in meinem Leben und ohne ihn hätte ich nicht die seelische und psychische Ausgeglichenheit, mit der ich auch schwierige Phasen der Promotionszeit besser überstehen konnte. Seine grenzenlose Unterstützung, tiefe Liebe und sein Glaube an meine Fähigkeiten sind eine unschätzbare Quelle der Stärkung für mich. Er hat alle Arbeitsabende, -nächte und -wochenenden nicht nur klaglos sondern auch mit Freude mitgemacht, hat mir unablässig mit Computer- und Programmierproblemen geholfen und unermüdlich fachliche Diskussionen mit mir geführt.

Finanziell wurde ich von der Deutschen Forschungsgemeinschaft (DFG) durch das InterGK 591 und den SFB „Das aktive Gehör” (SFB/TR 31) unterstützt.

Lebenslauf

Dipl.-Biol. Astrid Klinge

geboren am 27. Oktober 1978
in Jena

Staatsangehörigkeit: deutsch

September 2005 bis August 2010	Promotion bei Prof. Dr. Georg Klump in der Arbeitsgruppe “Zoophysiologie und Verhalten” an der Carl-von-Ossietzky Universität Oldenburg im Rahmen des Sonderforschungsbereichs “Das Aktive Gehör"(SFB/TR 31).
September 2006 bis August 2009	Stipendiatin des internationalen Graduiertenkollegs “Neurosensory science, systems and applications”, danach assoziiertes Mitglied.
Oktober 1998 bis August 2005	Diplomstudium der Biologie mit den Nebenfächern Humananatomie und Geologie an der Friedrich-Schiller Universität Jena.
	Titel der Diplomarbeit: “Lokomotion von Trichosurus vulpecula (Marsupialia) - Analyse metrischer, kinetischer und morphologischer Parameter”.
November 2002 bis Mai 2003	DAAD Stipendium für einen Forschungsaufenthalt am “Department of Environmental Management & Ecology” an der LaTrobe University, Albury, Australien.
August 1998	Abitur am Gymnasium “Am Weißen Turm”, Pößneck.

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbständig verfasst und nur die angegebenen Hilfsmittel verwendet habe. Die Dissertation hat weder in Teilen noch in ihrer Gesamtheit einer anderen wissenschaftlichen Hochschule zur Begutachtung in einem Promotionsverfahren vorgelegen. Teile der Dissertation wurden bereits veröffentlicht bzw. sind zur Veröffentlichung eingereicht, wie an den entsprechenden Stellen angegeben. Der Anteil der Koautoren an den Veröffentlichungen bestand in der Betreuung der Arbeit und Korrektur der Manuskripte. Die Entwicklung, Durchführung und Auswertung der Experimente lagen in meiner Hand. Die Durchführung der Experimente wurde teilweise von wissenschaftlichen Hilfskräften unter meiner Überwachung übernommen.

Oldenburg, den 25. Juni 2010

....

Astrid Klinge