Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by
Dipl.-Ing. Steffen Ritter
born in Eberbach
Oral examination: 13.04.2005

Magnetoencephalographic Study on the Representation of Huygens Noise in the Auditory Cortex of Humans

Referees:
Prof. Dr. Hans Günter Dosch / Prof. Dr. Hans-Joachim Specht
Prof. Dr. Michael Scherg

Magnetoenzephalographische Untersuchung der Repräsentation von Huygens Rauschen im auditorischen Kortex des Menschen

Die Verarbeitung und die Repräsentation der Tonhöhe wurde anhand von Huygens Rauschen (regular interval sounds–RIS) untersucht. Die Tonhöhe von RIS wird durch zeitlich verzögerte Überlagerung von weißem Rauschen mit sich selbst hervorgerufen. Mittels Magnetoenzephalographie (MEG) wurde die Beziehung zwischen der spezifischen Antwort auf den Beginn eines Tones (pitch onset response–POR) und der perzipierten Tonhöhe hergestellt. Die Quellenanalyse zeigte Aktivität im lateralen Heschl Gyrus des auditorischen Kortex beider Hemisphären. Die Latenz der POR war hochkorreliert mit der Tonhöhe, so dass sich bei tieferer Tonhöhe die Latenz der POR verlängerte. Im zweiten Experiment wurde der Einfluß der Klangfarbe sowohl auf die Wahrnehmung als auf die POR untersucht. Mit dem Anstieg der Tonhöhe, bedingt durch das Verschieben eines Bandpassfilters im Spektrum, verlängerte sich die Latenz. Die Ergebnisse weisen darauf hin, dass die POR sowohl die Tonhöhe als auch die Klangfarbe widerspiegelt und damit eine objektive neurophysiologische Repräsentation der Wahrnehmung eines Tones darstellt. Simulationen der Tonhöhenverarbeitung zeigten, dass zeitliche Modelle die perzipierte Tonhöhe von RIS erklären können, aber bei einer Veränderung der Tonhöhe, hervorgerufen nur durch eine Klangänderung bei gleicher Grundfrequenz, versagen. Die Resultate der POR zeigen, dass mit Hilfe des MEG wertvolle Parameter gewonnen werden können, die in Modellen der Tonhöhenverarbeitung eingebunden werden könnten.

Magnetoencephalographic study on the representation of Huygens noise in the auditory cortex of humans

Processing and representation of pitch in human was studied by using regular interval sounds (RIS–Huygens noise). The distinct pitch of RIS is introduced by delaying random noise and adding it back to the same noise. The relation between the psychoacoustically measured pitch of RIS and the auditory evoked neuromagnetic responses was investigated. The pitch specific onset response (POR) was derived by means of magnetoencephalography (MEG). Source analysis revealed the center of this activity in the lateral part of Heschl's Gyrus in the auditory cortex of both hemispheres. The latency of the POR correlated highly with the perceived pitch. When the perceived pitch of RIS decreased, the latency of the POR increased. In a second experiment, the influence of timbre on perception and the POR was investigated. An increase of the perceived pitch resulted in an increase of the POR-latency, when the center frequency of the bandpass filter was increased. These results indicate that the POR integrates both, pitch and timbre and, thus, POR is an objective neurophysiological representation of the perception of a sound. Simulations of pitch processing showed, that temporal pitch models can account for the perceived pitch of RIS but fail to simulate timbre-induced pitch changes. Hence, MEG provides valuable parameters which could be integrated in models of pitch processing.

Für meinen Vater

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Common sounds with a well defined pitch are produced e.g. by our vocal chords or by musical instruments. The processing of sounds in the human auditory system, that lead to a specific sensation of pitch and timbre (e.g. understanding of speech) is one of the major topics in the field of auditory research. Knowledge, collected from two entirely different sources of experimental information are combined to reveal the mechanisms of auditory processing. First, from findings of anatomy and physiology of the central auditory system and secondly from the perception of pitch revealed by psychoacoustic measurements. Joining together the results of both fields, the goal is to explain the findings in terms of some model, law, hypothesis or theory. Thus, the pattern represents the triangle in the research of human perception, illustrated in Figure 1.1.

Investigations concerning the perception of pitch have concentrated on two questions: (1) Which physical parameters determine pitch and (2) How are these parameters processed in the auditory nervous system.
So far, psychoacoustic experiments have elucidated many aspects regarding the first question. They showed that pitch is related to the repetition rate of a sounds' waveform and that the perception is linked to the fundamental frequency (f_0), when the frequencies of the partials are integer multiples of a common basic at f_0. The second question is still unexplained and much remains unknown about the processing of pitch. Seebeck (1841) and Ohm (1843) had the discussion on the 'missing fundamental': Seebeck reported that the pitch of periodic sounds always seemed to follow f_0, even when this component was attenuated. Ohm replied that the observations must be based on an illusion and argued that the ears perform a real-time frequency analysis similar to the mathematical formulation of Fourier. He, and later Helmholtz (1863) concluded that the lowest spectral component determines the pitch of the tone, whereas higher harmonics determine the sounds'

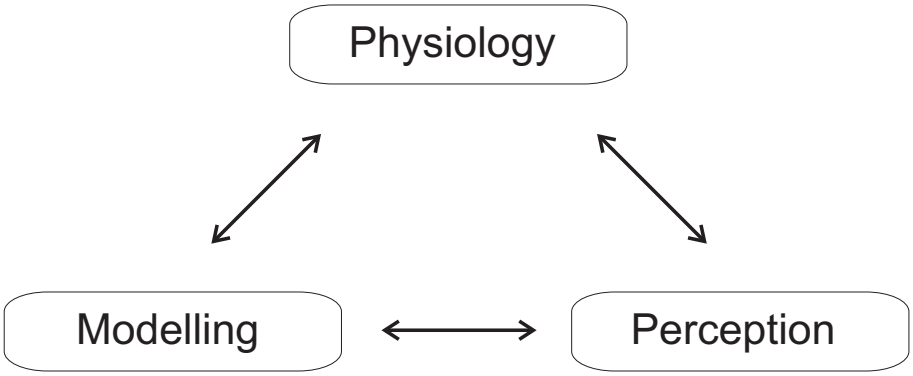

Figure 1.1: *The triangle in the research of human perception. Experimental evidence from anatomy and physiology combined with psychoacoustic results (perception) must correspond with models and theory.*

timbre.
But even more than a century of research after this initial discussion about the phenomenon of the missing fundamental, have not revealed the specific neural mechanisms subserving the perception of an acoustic stimulus. Although a great deal is known about neural response properties at many levels of the auditory system, we presently have only a very rudimentary understanding of which particular response properties are responsible for the sensation of a specific pitch or timbre. One of the most heavily debated topics in hearing research is the question, how populations of neurons represent and convey information through trains of spikes in the ascending auditory system.
There are fundamentally two basic ideas about how information in auditory stimuli is encoded: the spectro-temporal theory and the autocorrelation theory of pitch processing. Historically, the theory of coding by spatial patterns of neural excitation can be ascribed to the formulations of the resonance-place theory of auditory representation (Helmholtz, 1863). It is based upon the tonotopically organized auditory periphery, in which the single frequencies of a sound are processed in spatially-organized auditory channels. In contrast, the idea of temporal coding was initiated by Rutherford (1886), with the development of his 'telephone theory' of hearing. Licklider (1951) proposed that the auditory system performs an autocorrelation process of the temporal response across all frequency channels. The pitch information is extracted by temporal patterns between spikes in a spike train or by the time-of-arrival of spikes relative to a reference event.
But until today, experimental research could not provide conclusive evidence, which theory can account for the perception and the processing of sounds.

The present work attempts, to shed some light on the unresolved controversy about perception, processing, and representation of pitch in the human auditory cortex. With respect to the interdependency of perception, physiology and modelling, illustrated in Figure 1.1, the major focus lays on a framework that connects all domains. The thesis contains psychoacoustic experiments to reveal the individual sensation of pitch. To understand the relation between perception and neural responses, there are different types of modalities with which one can look at the human brain. Magnetic resonance imaging (MRI) and computer tomography (CT) scanners give information about the macroscopic structure of the brain; positron emission tomography (PET) and functional MRI units provide information on the blood flow, oxygen consumption and metabolic activity in the brain. But the only two methods that provide information with a high temporal resolution are magnetoencephalography (MEG) and electroencephalography (EEG). Thus, neurophysiological measurements of the evoked magnetic fields are able to give information about the signal processing of sounds and their representation in the brain in sufficiently good time resolution. The simulation and modelling part is comprised by applying the auditory image model of pitch processing provided by Patterson et al. (1995) to find out if temporal pitch models can account for the perceived pitch.

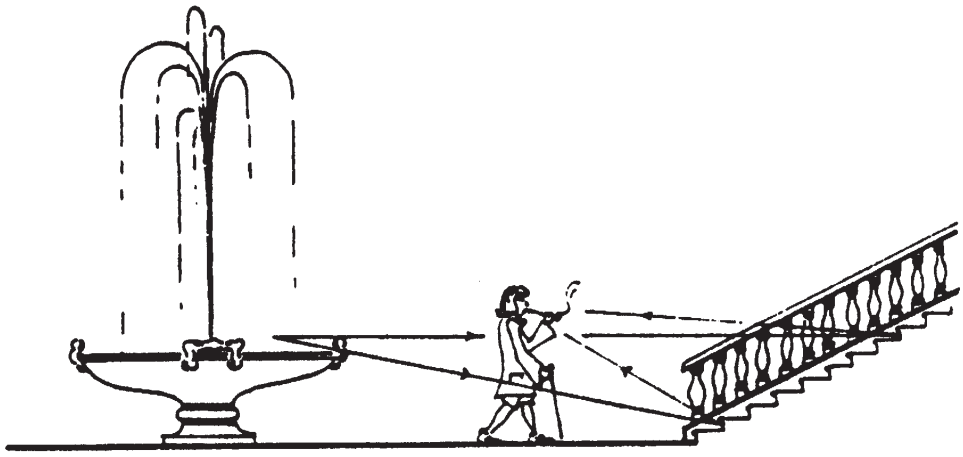

Figure 1.2: *The discovery of RIS in 1693: Huygens in the courtyard as imagined by Bilsen and Ritsma (1969/70). The noise coming from the fountain is periodically reflected by each step of the staircase and introduces a pitch.*

Several investigators have demonstrated, that the auditory system can extract pitch information from complex stimuli (e.g. a tone played by a piano) whose spectral components are not resolved by the auditory periphery (Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994). The stimuli are typically highpass filtered harmonic sounds but for many, if not most, of these stimuli the complex pitch occurs along with other perceptual attributes.
A class of sounds that offers a potential advantage for studying the processing of pitch are Regular Interval Sounds. The discovery of this naturally occurring sound goes back to the 17^{th} century, when the Dutch physicist and astronomer Christiaan Huygens (1629–1695) discovered a form of pitch, while visiting the French castle of Chantilly de la Cour near Paris. As illustrated in Figure 1.2, Huygens was standing outdoors between a garden fountain and a marble staircase, when he noticed that the noise, coming from a fountain, was transformed by the staircase and produced a noisy sound with a distinct musical pitch. Huygens described the situation in a letter to de la Hire (see Appendix A). In general, the noise is delayed and added back to the undelayed version (delay-and-add). The introduced pitch varies with the reciprocal of the delay time d, the pitch strength increases with n, the cascaded delay-and-add iterations (Bilsen, 1966). In contrast, when the delayed noise is first multiplied by a factor of minus one before it is added back to the original (delay-and-subtract), pitch differs significantly compared to that, generated with the delay-and-add condition. When the number of iterations is one, Yost (1996a) reported that the pitch is ambiguous, having two pitches in the region of $\frac{1}{0.9d}$ and $\frac{1}{1.1d}$. He stated, when n was increased to eight, pitch shifts to $\frac{1}{2d}$. However, Raatgever and Bilsen (1992) found that pitch values depend on the delay time. They concluded, that pitch is equal to $\frac{1}{2d}$ for delay times of less than about 6 ms and pitch shifts to ambiguous values of $\frac{1}{0.9d}$ and $\frac{1}{1.1d}$ when the delay time increased. However, pitch matching is very difficult and often only musically trained listeners can perform the task reliably. In the last years, the name for these sounds changed from Huygens noise, rippled noise to Regular Interval Sounds (RIS), which is today commonly used. In order to avoid any further confusion in the present work, the sounds produced by feedback of the delayed noise are called RIS.
An understanding of the pitch shift of RIS is important for solving the problem of pitch processing in human. Models and theories of pitch must be able to account for the pitches of these stimuli.

The aim of the present work is to investigate the relationship between the perceived pitch and the neuromagnetic responses evoked by the onset of pitch.

MEG is applied in the experiments to measure the auditory evoked magnetic field on the scalp. By presenting a continuous stimulation, responses to the simple onset of energy are avoided and the specific pitch onset response (POR) is isolated. The center of activity is determined, using an equivalent current dipole model that fits the evoked field.
The controversial results, as described above, reported on the perceived pitch elicited by RIS are examined and find particular attention in the psychoacoustic parts of this work. In addition, the pitch of the applied RIS is compared with the sensation evoked by complex sounds. All applied sounds are simulated with the auditory image model to test if the sensation and the neuromagnetic representation in the auditory cortex can be predicted with the model.

The following Chapter 2 provides the indispensable background to the field of hearing research. Psychoacoustic results on the pitch of sinus tones, complex harmonic tones and RIS are presented for an overview of the perception of these sounds. Anatomical and physiological basics of the ascending auditory system are given briefly, to build the connection to the neurophysiological representation in the human auditory cortex. The last section of this Chapter introduces the formalism underlying the auditory image model, that is used to simulate the processing of sounds to extract a specific pitch. Chapter 3 provides properties on the generation of RIS and shows how changes of the parameters affect the sound. The algorithm of the Bradley-Terry-Luce (BTL) method is shortly sketched and is used in the psychoacoustic experiments to derive a relative scale of the perceived pitch. The third block covers the physiological, physical, and technical basics which are mandatory to derive the neuromagnetic activity of the human brain. With the knowledge of Chapters 2 and 3, the perception and the neurophysiological correspondence of RIS is examined in Chapter 4. The representation of the POR in the auditory cortex is determined and the correlation between the evoked responses and the perceived pitch of RIS is built up. Finally, the auditory image model is applied to find out, if the perception and the physiological representation of RIS can be predicted with simulations. In Chapter 5, the sensation of bandpass filtered harmonic sounds is investigated. It is tested, if harmonic complex sounds evoke a pitch equal to that of RIS. The bandpass filtered harmonic complex sounds are used for a MEG experiment to investigate the dependence of the POR to the perceived pitch when timbre changes due to different filter conditions. These sounds were also analyzed with the temporal pitch model.

Chapter 2

Pitch—Perception, Physiology and Simulation

This Chapter provides an overview over psychoacoustic results on the perception of pitch, beginning with the scaling of the perceived pitch of pure tones and ending with that of Regular Interval Sounds (RIS). The processing of sounds in the mammalian central auditory pathway is described with emphasis on the physiology of the human auditory system, although results of invasive research mostly depend on studies of nonhuman animals. Today, most of the neural function of the different ascending processing stages and the complicated connections among each other that we know, have been obtained primarily from cats, gerbils and non-human primates. The last part of this Chapter covers the modelling of pitch processing in general, and gives a detailed description of the Auditory Image Model (Patterson et al., 1995) that is used for simulation of pitch processing in the Chapters 4 and 5.

2.1 Perception of Pitch

In 1973 the American National Standards Institute (ANSI) defined *pitch* as "...that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from high to low." Pitch is an essential element for features like melody or harmony in music, and not to forget, pitch conveys information important for the perception of speech. Like loudness or timbre, it is a subjective measure that cannot be expressed by physical means. The primary objective correlate to pure tones is the physical attribute frequency. However, the intensity, duration or the temporal envelope influences the pitch of a sound,

too. In terms of complex tones, like a tone played by a piano, its sound can be perceived as a single pitch or as a cluster of different pitches. Even individual sinusoidal partials can be heard simultaneously, all of them with a different pitch sensation. The third class of sounds that are discussed, have a continuous spectrum with a spectral or temporal regularity. The pitch sensation evoked by these sounds can either correspond to the regularity of the spectral ripples or to their temporal properties (details are given in Chapter 3).

2.1.1 Pure Tones

The easiest way to evoke a specific pitch is the use of a tone, that consists of only one distinct frequency. But even the sensation evoked by a pure tone is difficult to measure since "perceived pitch" is not a physical unit but rather a relative scale. One way to overcome this problem is to present a pure tone of the frequency f_1 and to adjust the frequency of a second pure tone in a way until it sounds twice or half as high as the pitch of the reference tone. The classical result of such experiments leads to the *mel scale* introduced by Stevens et al. (1937). The pitch, expressed in *mels* (derived from melody scale), is neither identical nor even linear to frequency and has an arbitrary pitch reference of 1,000 mels at a frequency of 1,000 Hz. At low frequencies the *mel scale* is proportional to the frequency scale, but for higher frequencies, it bends more and more (e.g. a frequency of 10 kHz corresponds only to 4,500 mels).
But also other parameters, like the sound pressure level (SPL), influence the pitch of pure tones. For tones below 1,000 Hz, pitch decreases with increasing intensity (up to 10% for SPL levels of 90 dB). The pitch of pure tones with a frequency between 1,000 and 2,000 Hz remains rather constant, and for frequencies above 2,000 Hz, it rises up to 20% (f=8 kHz, 90 dB SPL) with increasing intensity (Morgan et al., 1951). Hence, frequency alone is not sufficient for a valid description of pitch, even when the sensation is evoked by pure tones.
A pitch shift of a test tone can also be caused by the simultaneous presentation of other tones or noise which mask the tone. Partial masking produced by a sound that is lower in frequency than the test tone yields positive pitch shifts, whereas masking sounds of higher frequency causes a decreased pitch. Altogether, the pitch of pure tones is shifted away from the spectral slope of the partial masking sound up to a semitone (Terhardt and Fastl, 1971).

2.1.2 Complex Tones

Complex tones comprise several pure tones and occur much more frequently in daily life than sounds that consist of only one frequency. The vowels of human speech for example, or the sounds produced by musical instruments are complex tones. If the frequencies of the components are integer multiples of a common basic or fundamental frequency (f_0), the resulting complex tone is called a harmonic complex tone or a harmonic series of a complex tone.

The pitch of complex tones is a special topic in hearing research since 150 years ago, Seebeck (1841) and Ohm (1843) discussed the phenomenon of the missing fundamental. Twenty years later, Helmholtz (1863) argued that the pitch sensation associated with the missing fundamental can be explained as a nonlinear difference tone generated in the auditory periphery. Schouten (1940) published his, *residue theory of pitch* and claimed that the sensation of the fundamental frequency (f_0) is caused by periodic fluctuations in the envelope pattern of clusters of harmonics that the ear fails to resolve. Thus, the output of a cochlear filter consists of a summed signal of two or more harmonics. The periodicity of the envelope is the same as f_0 and probably determines the perception of the pitch, even if this fundamental is physically absent. Some years later, new experiments showed that Schouten's residue theory failed to explain newer results. Ritsma (1962), e.g. reported a clear upper limit of the harmonic order beyond which no pitch of a tonal residue can be perceived anymore. Ritsma (1967) and Plomp (1967) showed that the region of harmonics that convey a pitch sensation of the missing fundamental is to be best in the order of 3, 4, and 5. In this so-called dominant region, harmonics differ by 25% or more, and are well resolved in the cochlea. Houtsma and Goldstein (1972) concluded that the perceived pitch of complex sounds is mediated primarily by a central mechanism that operates on neural signals derived from the spectrally resolved harmonics in the cochlea. The importance of the resolved frequency components in the processing of pitch resulted in the development of different harmonic pattern recognition models (Goldstein, 1973; Wightman, 1973; Terhardt, 1979). All models combine the basic idea that the tonotopically organized auditory nerve fibers are stimulated by different resolved harmonics. The fundamental pitch is extracted not in the auditory periphery but in the central nervous system by combining the activity of several groups of auditory nerve fibers.
On the other hand, psychoacoustic experiments conducted by Moore and Rosen (1979) showed that periodic pulse-trains retain a certain pitch quality even in the total absence of resolved harmonics. Houtsma and Smurzynski (1990) in-

vestigated the perceived pitch of complex sounds as well as the discrimination performance between sounds in dependence of the lowest harmonics. They found that the identification performance of a fixed f_0 dropped progressively from nearly perfect to a low but significantly above chance level with increasing the order of the lowest harmonic. The just noticeable difference between two pitches also increased from 0.5 Hz to 5 Hz with increasing the harmonic order. However, they reported that the discrimination performance remained constant, if the order of the lowest harmonic was on the order of 12 or higher.

Thus, the sum of these results is that not only low-order harmonics, that are resolved in the cochlea, determine the pitch of a complex tone, but also higher, unresolved harmonics contribute to the sensation. But their degrees of contribution to the perception is quite different. The resolved harmonics evoke a strong, sharply defined pitch, whereas the unresolved components are characterized by much higher identification scores. So, the sensation of the perceived pitch of a broadband stimulus that consists of resolved and unresolved components will be dominated by the lower, resolved harmonics. An exception was only found for complex tones with very low fundamental frequencies (Moore and Peters, 1992). But psychoacoustic experiments also showed, that the sensation evoked by a complex tone in terms of a musical note can be ambiguous and is far from a distinct pitch. It is especially the case when the low-order harmonics are weak or totally absent or if only a few harmonics are present. Schouten et al. (1962) used frequency components of three tone complexes with a fundamental frequency of 200 Hz and shifted them equidistantly in the frequency domain. The results of a similar experiment conducted with three tone complexes with a fundamental frequency of 300 Hz (Zwicker and Fastl, 1999) is shown in Figure 2.1. The perceived pitch of these inharmonic complex sounds rises, if the frequency components are shifted upwards, and falls, if they are shifted downwards. Even for a purely harmonic complex with the lowest component shifted between 1,400 Hz and 1,600 Hz, a pitch is perceived within an interval of 270 Hz and 310 Hz.

Another source of ambiguity, discussed by Helmholtz (1863) is the fact, that the auditory system can perceive a sound complex in two different ways. Helmholtz distinguished between a single pitch of the fundamental frequency and called it a *synthetic* perception (or in the words of Helmholtz: "perzipiert"). The second possibility is the dominant perception of the single pitches of the overlapping individual harmonics or partials, leading to a *analytic* ("aperzipiert") perception. The conditions, that decide whether the pitch of a sound complex is derived analytically or synthetically is completely unexplained. So far, experimental attempts failed to measure and control the mode of perception in single individuals (Smoorenburg, 1970; Houtsma and Fleuren, 1991). Some listeners show a strong

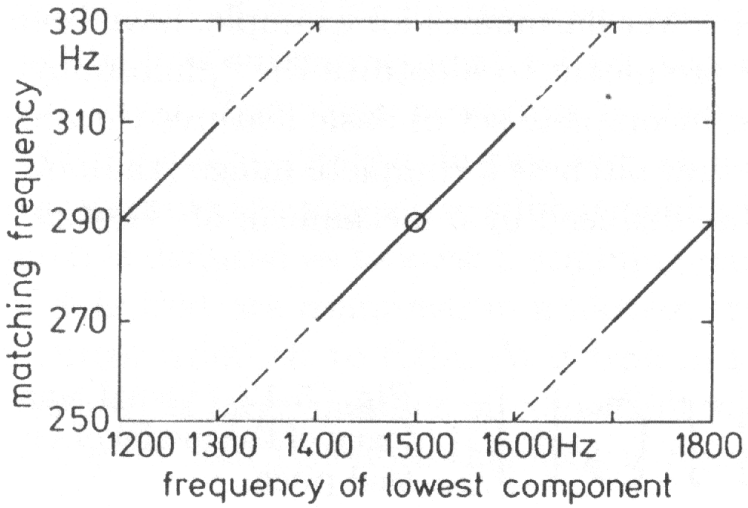

Figure 2.1: *Pitch shift of complex tones in dependence of the lowest harmonic. Frequency components of three tone complexes with a frequency spacing of 300 Hz between adjacent harmonics are shifted in the frequency domain. The perceived pitch of these inharmonic complex sounds rises, if the frequency components are shifted upwards, and falls, if they are shifted downwards. Figure adopted from Zwicker and Fastl (1999).*

tendency towards analytic perception behavior, others have a tendency towards synthetic perception, and still others show an inconsistent behavior. Only group-averaged behavior under conditions where synthetic and analytic processing leads to opposite responses shows some definite tendencies (Schneider et al., 2003). Houtsma (1995) reported for two-tone complexes with high-order harmonics the listeners' responses are divided about half-and-half into synthetic and analytic responses. Lowering the order of the presented harmonics caused a dominance of the analytic responses.

2.1.3 Pitch of Regular Interval Sounds (RIS)

The first observation on the pitch of rippled noise was made by Huygens (1693) while visiting the French castle of Chantilly de la Cour. He described the pitch as corresponding with the sound of an open organ pipe of a length matching the depths of the stairs. For reference, parts of the original work of Huygens can be viewed in Appendix A. The main difference in comparison to complex sounds is, that the basis of rippled noise or regular interval sounds is noise and therefore, the envelope of the spectrum is not a line spectrum but rather a flat continuum

comparable to the spectrum of white noise, but with spectral ripples depending on different parameters such as the delay time, the number of iterations and the gain factor. A detailed description of how exactly these sounds are generated and how their characteristic spectra are affected by changing the parameters is given in section 3.1.

In general, if an arbitrary sound $x(t)$ and its echo $x(t-d)$ are added (delay-and-add), a repetition pitch is introduced that is perceived with the frequency corresponding to a pure tone frequency of $\frac{1}{d}$ (Bilsen, 1966). The pitch of RIS has been studied systematically for both, monotic and diotic conditions (Yost and Hill, 1978) and pitch effects are typically found for delay times between 1 and 20 ms, corresponding to pitches between 1,000 and 50 Hz. With an increase of the number of the delay-and-add stages, the salience of the pitch increases, but the perceived pitch itself is unaffected (Yost, 1996b). Patterson (1996) stated that the dominant perception of RIS produced with two or less iterations is noise. He observed with an increase of the number of iterations, that the tonal component of the perception grows stronger, whereas the noise component weakens with the greatest rate of change in the region of four iterations. With eight iterations or more, the dominant perception is that of a buzzy tone. In the study of Krumbholz et al. (2003), the pitch of RIS is described as that of a 'cracked' bassoon. Patterson (1996) found, that doubling the number of iterations causes a 3.8 dB increase of the tone/noise ratio. With increasing the number of iterations, the pitch of RIS approaches a line spectrum, close to that of a complex sound.

Rippled sounds that are generated with the delay-and-subtract process (that is the echo $x(t-d)$ is first multiplied by a factor of minus one before it is added to the original noise $x(t)$) have a different pitch compared to the sounds produced with the delay-and-add process. The perceived pitch was examined in different psychoacoustic studies and is still a controversy. Bilsen (1966) generated the sounds in his experiment with one iteration and claimed that these stimuli produced with the delay-and-subtract process have two pitches equal to approximately $\frac{0.88}{d}$ and $\frac{1.14}{d}$. Yost et al. (1978) reported for RIS generated with the delay-and-subtract process and more than four iterations a pitch shift of an octave below the corresponding RIS produced with the same delay time but positive feedback. But for RIS generated with less than four iterations, they reported the perceived pitch not in the expected region of $f = \frac{1}{2d}$, but around $\frac{1}{0.9d}$ and $\frac{1}{1.1d}$, independent of the delay time d. Raatgever and Bakkum (1986) found different results for their pitch matching experiment, using an infinite number of iterations. The change in the perception of RIS produced with the delay-and-subtract process only depended on the delay d. Figure 2.2 shows the results of their experiment. Pitch matches at $\frac{1}{2d}$ only occurred for delay times of less than 6 ms, but with increas-

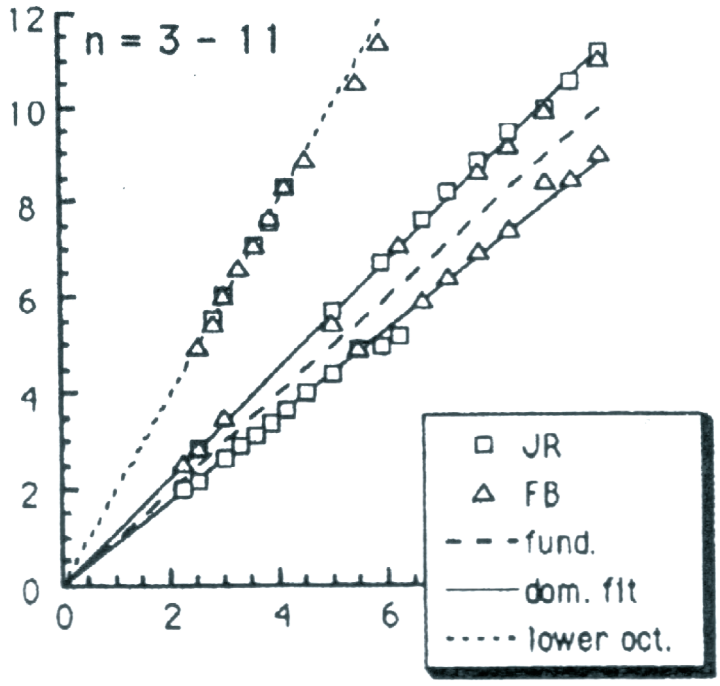

Figure 2.2: *Pitch of RIS generated with the delay-and-subtract process. The y-axis represents the pitch of RIS generated with the delay-and-add process, matched to the test RIS generated with the delay-and-subtract process on the x-axis. Both axes show the delay time (in ms) of RIS. The test- and the matching-RIS consisted of the third to the eleventh harmonics; the pitch matching was done by the two authors (JR and FB), where the Figure is adopted from (Raatgever and Bilsen, 1992).*

ing delay, pitch matched around $\frac{1}{0.9d}$ and $\frac{1}{1.1d}$ compared to RIS generated with the same delay time but positive gain. Yost (1996a) reported that the perceived pitch is determined only by the number of iterations n. He found that the transition from the perception of ambiguous pitches in the region of $\frac{1}{0.9d}$ and $\frac{1}{1.1d}$ to pitch matches at $\frac{1}{2d}$ is between three and five iteration processes, independent of the delay time. Recently, in a similar pitch matching experiment, Wiegrebe and Winter (2000) used high-pass filtered RIS and found the transition for the pitch shift depending on both, the delay and the cut-off frequency. The results of their study show (see Figure 2.3) that, e.g. for a cut-off frequency of 625 Hz an octave shift occurs for stimuli with delay times of 2 and 4 ms. For delays of 8 and 16 ms the deviation of RIS with negative g was found to be differing by 10% compared to RIS generated with the delay-and-add process. In general, these results show that whether listeners hear an octave shift between the delay-and-add and the delay-and-subtract processed sounds, depends on the delay time as well as on

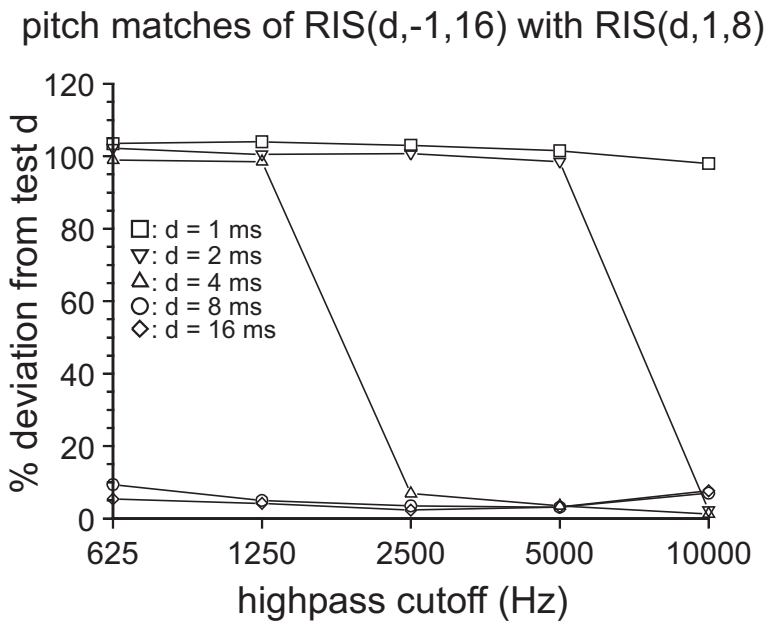

Figure 2.3: *Pitch matches of test RIS produced with delay-and-subtract processes to RIS generated with delay-and-add. The adjusted pitch with about 100% deviation of the delay time, corresponds to an octave shift between the two stimuli. 10% deviation represents ambiguous pitches (adopted from Wiegrebe and Winter (2000).)*

the high-pass cut-off frequency. When the delay time was 1 ms, the octave-down shift was still perceived even when the stimuli were high-pass filtered at 10 kHz. With a delay time of 2 ms, the perception switched when the cut-off frequency was raised from 5 kHz to 10 kHz. The sensation of an octave shift vanished and a pitch near the matching stimulus was perceived.

2.2 Neuronal Representation of Pitch

2.2.1 Neurophysiology of the Auditory Pathway

In this section, a short review of the specific properties of the human auditory pathway, as illustrated in Figure 2.4, is given, starting with the auditory periphery and ending in the auditory cortex. The knowledge of the generation of auditory evoked cortical activity in its different steps is a crucial basis for understanding the neuromagnetic measurements of Chapter 4 and 5.

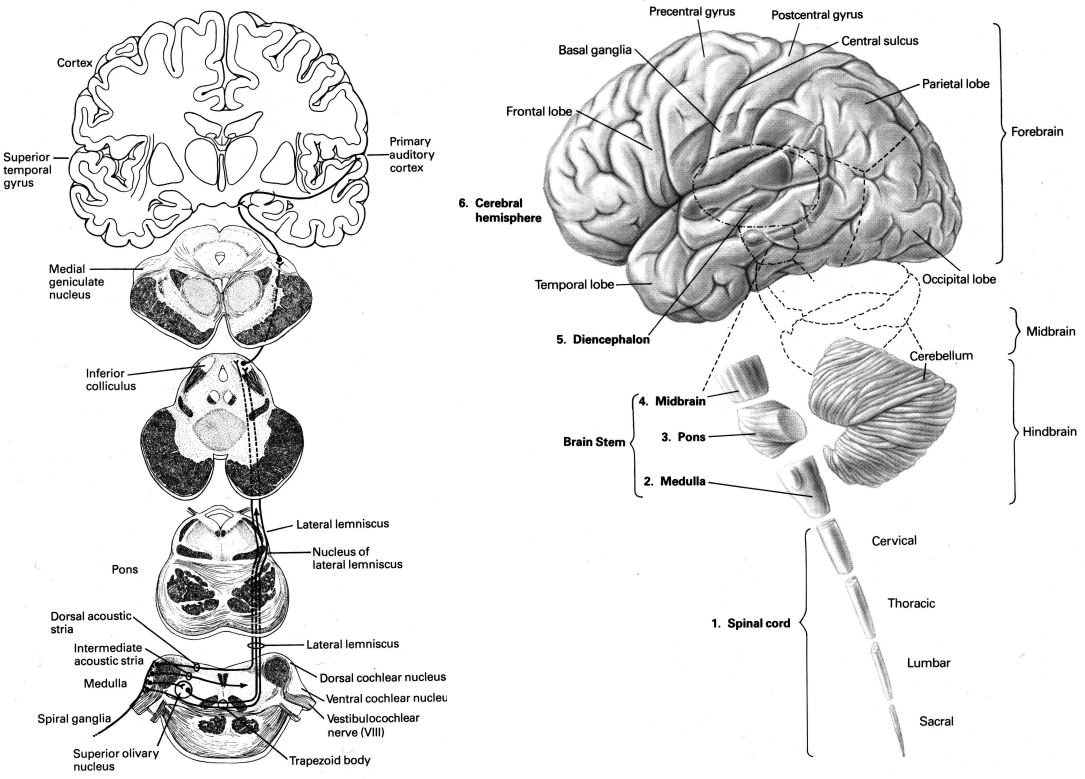

Figure 2.4: *Left: Schematic overview of the afferent connections in the central auditory pathway, starting from the brainstem (Superior olivary nucleus) up to the primary auditory cortex. Right: Lateral view of the human cortex. After Kandel and Schwartz (1985), page 406.*

The Auditory Periphery and the Cochlea

The preprocessing of sounds starts in peripheral regions where the mechanical oscillations are converted in receptor potentials for the neural activation. The function of the outer ear is to collect sound oscillations and to transmit them through the outer ear canal to the ear drum. The outer ear canal of humans acts like an open pipe with a length of 2 cm corresponding to a frequency of about 4 kHz. So, the outer ear canal is responsible for the high sensitivity of humans in this frequency region. The middle ear (see Figure 2.5) converts the air-oscillations transmitted through the outer ear into salt water-like fluid motions of the inner ear. To avoid large losses of energy through reflections at the air-fluid border, a transformation has to occur in the middle ear. Like in mechanical systems the impedance matching can be achieved using levers. At the end of the outer ear canal a light but sturdy funnel-shaped tympanic membrane (eardrum) acts as a pressure receiver over a wide frequency range. It is firmly attached to the large arm of the malleus (hammer). The motions of the eardrum are transmitted over the incus (anvil) to the footplate of the stapes (stirrup) by the so called three middle ear ossicles (malleus, incus, and stapes). A ring-shaped membrane (oval window) forms together with the stapes footplate the entrance to the inner ear. Through the levers, an almost perfect match between the impedances is reached in human at a middle frequency region around 1 kHz.
The inner ear is embedded in the extremely hard temporal bone and is basically the same for all mammals. The snail-shaped cochlea of humans forms $2\frac{1}{2}$ turns, is filled with two electrochemically different fluids (perilymph and endolymph)

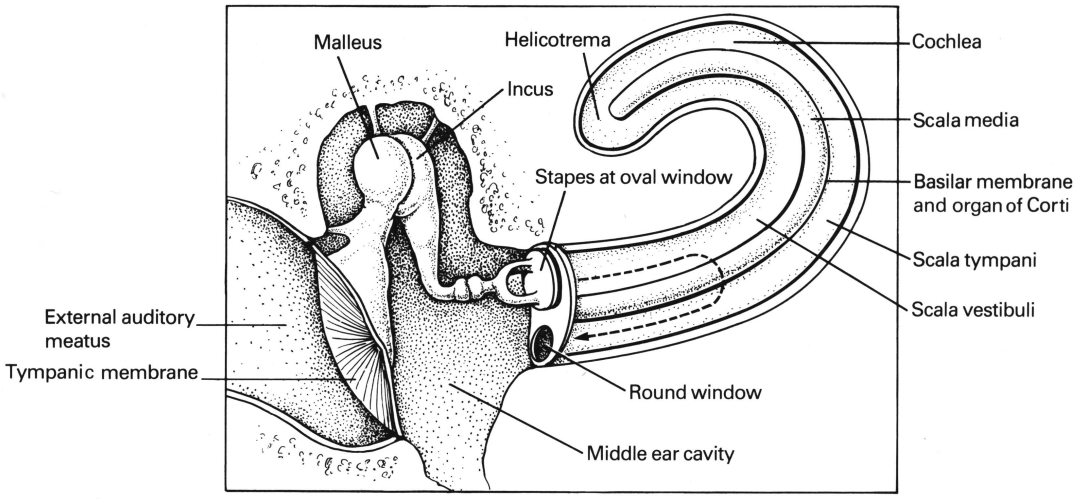

Figure 2.5: *Schematic drawing of the auditory periphery.*

and consists of three different channels of scalae, which run together from the base to the apex. The footplate of the stapes is in direct contact with the fluid in the scala vestibuli which is separated from the scala media only by the very thin and light Reissner's membrane. The fluid oscillations are transmitted to the basilar membrane which separates the scala media from the the scala tympani and supports the organ of Corti with its sensory cells. The fluids can be considered as incompressible, and so the movement of the stapes must be equalized through the basilar membrane at the round window, which closes off the scala tympani at the base of the cochlea. For very low frequencies, the equalization occurs through a connection between the scala tympani and vestibuli at the apex of the cochlea called helicotrema. The basilar membrane of humans has a total length of 32 mm and widens with a factor of about three from the base to the apex. It carries the organ of Corti with about 3,000 inner and 16,000 outer haircells, which transform the mechanical oscillations in the inner ear into a signal (action potentials), that can be processed by the nervous system.

The haircells are arranged in one row of inner haircells on the inner side of the organ of Corti, and three rows of outer haircells near the middle of the organ of Corti. The organ of Corti is partly covered by the tectorial membrane, that is made up of two kinds of hydrated protofibrils. Although the functions and their neural innervation of the inner and outer haircells are not clear in single detail, it is assumed that the outer haircells only effect the organ of Corti and that they exert large influence on the inner haircells. The afferent synapses of the inner haircells show a characteristic chemical behavior, whereas those of the outer haircells are atypical. More than 90% of the fibers contact the inner haircells, with each fiber normally connecting only one inner haircell. In this way, each inner haircell is contacted by 10–20 afferent dendrites of primary auditory neurons. In contrast, the outer haircells are mainly innervated by efferent fibers coming from the brain. The inner haircells themselves rarely receive efferent terminals. The few efferent fibers contact only the afferent fibers and do not have direct synaptic connections with the inner haircells.

Since the basilar membrane widens from the oval window to the apex and is more flexible near the apex of the cochlea, different places along the basilar membrane resonate with different sound frequencies. Békésy proved that low frequencies producing oscillations on the basilar membrane near the helicotrema whereas high frequencies resonate near the oval window. Figure 2.6 shows a schematic drawing of the discoveries of von Békésy: a wave originates with a small amplitude at the oval window, travels along the basilar membrane, grows slowly, reaches its maximum at a certain location and rapidly dies out in the direction of the helicotrema. The envelope of the stimulus oscillations according

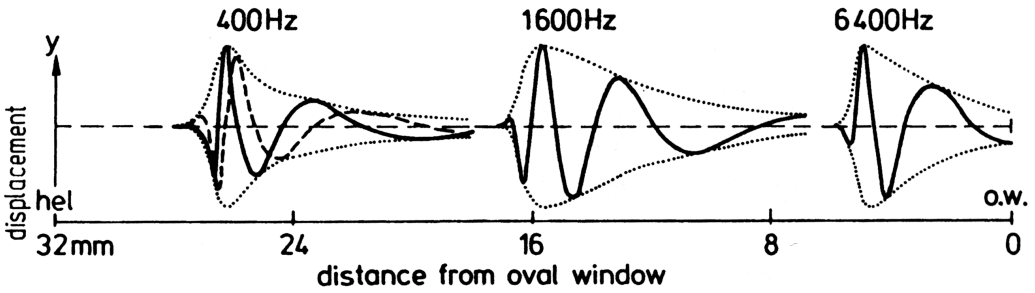

Figure 2.6: *Schematic drawing of the unwound basilar membrane from the helicotrema (hel) to the oval window (o.w.). Three simultaneously presented frequencies produce travelling waves with their maxima at different places along the basilar membrane, from Zwicker and Fastl (1999).*

to the different regions of the three maxima (indicated by the dotted lines) is invariant in its position along the basilar membrane. Each frequency has its peak amplitude at a different position with a logarithmic representation of the frequencies along the basilar membrane. The tonotopical organization of the inner ear performs a spectral analysis and separates the different frequencies of the sound according to a place principle. The delay time between the incoming sound at the oval window and the response of the basilar membrane increases with increasing distance along the basilar membrane, respectively increases with decreasing center frequency. Sounds with low-frequency-content that reach their maximum near the end of the cochlea show delay times up to 11 ms.

Auditory Nerve

In human, the entire flow of information from the inner ear to the brain runs through approximately 30,000 afferent neurons of the spiral ganglion that contact the hair cells of the cochlea and end centrally in the cochlear nucleus. Two different types of primary auditory neurons form the auditory nerve: type I neurons are bipolar and myelinated. They represent about 90–95% of the neurons in the auditory nerve and contact only one inner hair cell. Type II neurons (5–10%) are smaller, pseudo-unipolar, and unmyelinated. In contrast, the dendrites of the type II neurons are ramified and contact up to ten outer hair cells. The specific role of the type II neurons is still unclear, since no nerve impulse could be recorded so far. It is even not certain if they conduct action potentials.
The fact that a frequency analysis is performed in the cochlea and the one-to-one connection of the type I auditory nerve fibers with the inner hair cells makes it obvious that the single auditory nerve fibers are sensitive only to a restricted

frequency range. The threshold of the responses to pure tones of different frequencies is described by tuning curves of the single auditory nerve fibers. The frequency to which the fiber is most sensitive is the characteristic frequency and is directly related to the longitudinal position along the cochlea (Liberman, 1978, e.g.). Spiral ganglion neurons showing a low characteristic frequency, contact with inner hair cells in the apical part of the cochlea, whereas fibers with a high characteristic frequency innervate the cochlea at its base. In absence of a sound, most of the type I neurons show an irregular, spontaneous activity. With the onset of a tone above threshold of the neuron, the discharge pattern changes. The peristimulus time histogram (PSTH) of the fiber shows an initial phasic increase in the number of discharges, followed by a maintained tonic firing rate at a steady level that persists for the duration of the tone. At the offset of the tone, the discharge rate is depressed below the spontaneous level before returning to its resting state. In response to low characteristic frequencies up to 4–5 kHz Cariani (1999), the firing of the individual auditory nerve fibers is synchronized with the phase of the tonal stimulation (phase locking). The distribution of the intervals between successive spikes (first-order intervals) on an interspike interval histogram is characterized by peaks corresponding to integer multiples of the period of the pure tone. When both, the distribution of successive and nonsuccessive spikes is counted, the all-order spike interval histograms of each fiber can be summed and weighted according to their characteristic frequencies. The result is an all-order interval distribution of the entire population of auditory nerve fibers (Figure 2.7).

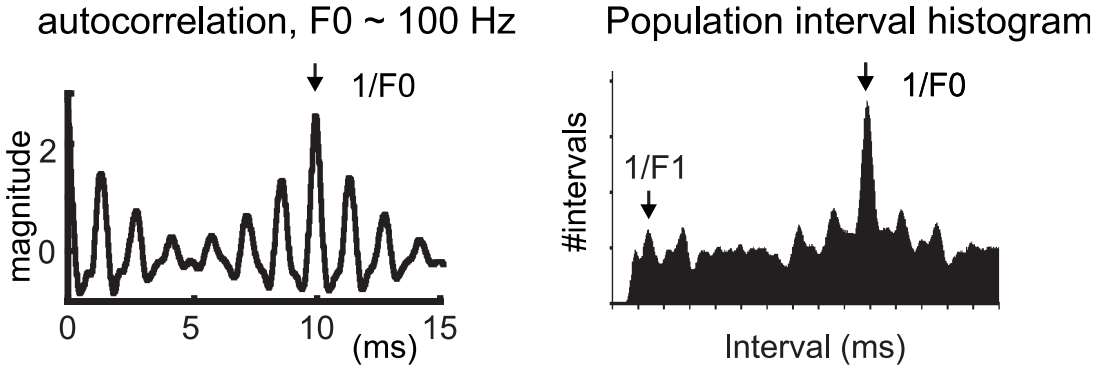

Figure 2.7: *Autocorrelation like responses of the auditory nerve fibers. Left: short term autocorrelation function of a voice like stimulus with fundamental frequency f_0. Right: Population interval histogram recorded from the auditory nerve in cat, after Cariani (1999)*

Cochlear Nucleus

All primary auditory nerve fibers terminate in the two cochlear nuclei, located on the dorsolateral side of the brainstem between the spinal cord and pons. At the first stage of the ascending auditory pathway, the rather homogenous incoming signals from the auditory nerve fibers are re-encoded into various functionally different channels. Shortly after entering the brainstem, the spiral ganglions all bifurcate into an anterior ascending branch that goes to the anterior ventral cochlear nucleus (AVCN) and in a posterior descending branch going first to the posterior ventral cochlear nucleus (PVCN) and then to the dorsal cochlear nucleus (DCN). Thereby, the tonotopy of the cochlea is preserved: fibers from the cochlear apex (representing low frequencies) bifurcate as soon as they enter the cochlear nucleus complex. The ascending branch stays in the ventrolateral portion of AVCN, whereas the descending branch stays in the ventrolateral portions of PVCN and DCN. Spiral ganglions carrying high frequency information go to the dorsomedial portion of the VCN before they bifurcate. Both branches project to regions more medial of the AVCN, PVCN and DCN. This cochleotopic organization of the two cochlear nuclei is maintained throughout the ascending auditory pathway. The cochlear nuclei themselves show a large variety of neuronal cell types, like bushy-, stellate-, octopus- or fusiform-cells. In correlation with the different cell types, various responses have been derived from single unit recordings especially in the cat. An overview over the different response types is given in Figure 2.8. Three different types of the *primary-like* (PL) neurons are described in the VCN. They show a similar response compared to the PSTH of the auditory nerve fibers and differ only in their onset response: the *sustained primary-like* neurons (PL_s) do not show an initial peak, whereas the *primary-like with notch* (PL_N) exhibit a short pause after the onset. *Onset units* respond particularly to sound onsets. The two most common types are *onset-lockers* (O_L) and *onset-chopper* (O_C) units. Cell types that discharge at regular intervals, independent of stimulus frequency and phase, are called chopper units (C). They are subdivided in sustained (C_S) and transient (C_T) chopper units and differ only in their discharge regularity.

The ascending connections of the cochlear project to several brain stem auditory nuclei in three main output pathways, as illustrated in Figure 2.4. The first, the ventral acoustic stria (VAS), or trapezoid body, contains axons originating mainly from bushy cells of the AVCN and stellate cells of the PVCN. The bushy cell axons terminate essentially in the three principal nuclei of the superior olivary complex: the lateral and the medial superior olivary nucleus, and the medial nucleus of the trapezoid body. Other axons of the VAS project to the

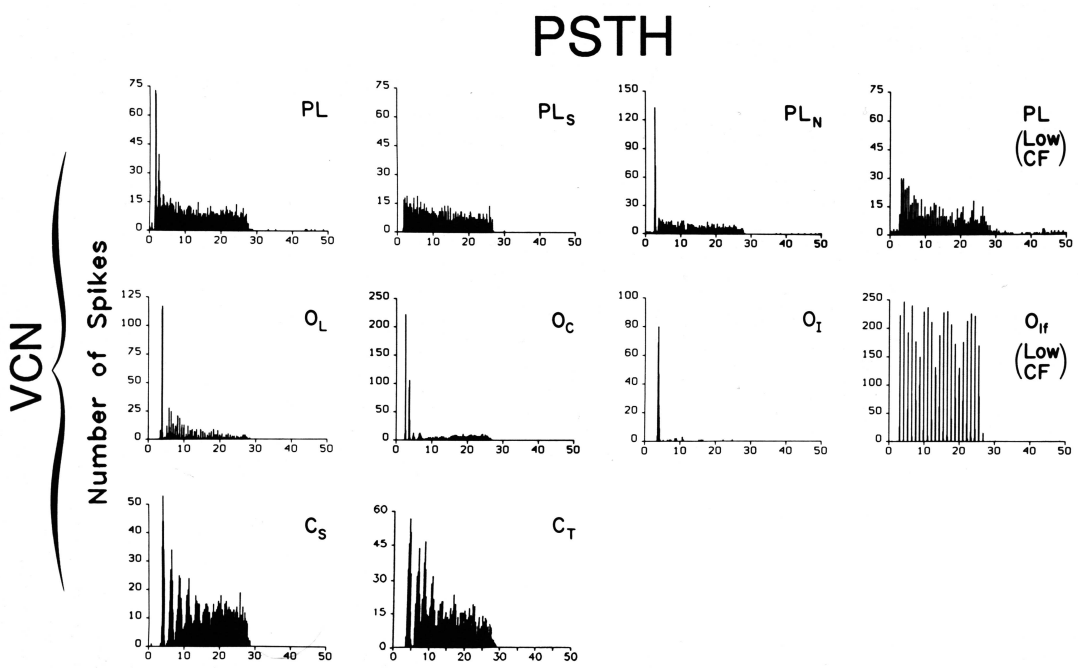

Figure 2.8: *Peristimulus time histogram (PSTH) of different cell types from recordings of the ventral cochlear nuclei (VCN). Primary-like sustained (PL_S) and with notch (PL_N). Onset units that lock (O_L), chopper after onset (O_C) or are inhibitory after onset (O_I). Chopper units, either with sustained (C_S) or transient responses (C_T). After Popper and Fay (1992b).*

lateral lemniscus (LL) and the inferior colliculus. The second output pathway, the intermediate acoustic stria (IAS), or stria of Held, includes axons originating mainly from the octopus cells of the PVCN and project to the preolivary nuclei of the superior olivary complex, higher to the lateral lemniscus and to the inferior colliculus. These two latter structures also receive ascending inputs via the third output pathway, the dorsal acoustic stria (DAS), or stria of Monakow, containing axons mainly from the DCN. In addition to the projections to the more central auditory nuclei, there is also a weak connection between the two cochlear nuclei.

Superior Olivary Complex and Lateral Lemniscus

The superior olivary complex consists of several nuclei and is located in the ventral aspect of the brainstem, between the cochlear nucleus and midline. The three main nuclei–lateral (LSO) and medial superior olivary nuclei (MSO) together with the medial nucleus of the trapezoid body (MNTB)–represent the first stage of the auditory pathway where a massive convergence of the information of both ears takes place. Especially the MSO is thought to be concerned with sound localization on the basis of time differences. The nucleus is composed of spindle-shaped cells with two dendrites. The lateral end receives input from both, the contralateral and ipsilateral cochlear nuclei. In contrast to the MSO, the LSO uses intensity differences to localize sounds in space. Nearly all binaural neurons of these nuclei are excited by inputs from the ipsilateral cochlear nucleus and inhibited, via an interneuron, by inputs from the contralateral cochlear nucleus. When this inhibition is reduced, either because of differences in time or intensity, the cells give a maximum response. Thus, from the outset of the superior olivary complex, there is an extensive bilateral representation of each ear in the auditory system. The axons projecting from the cochlear nucleus together with the output from the superior olivary complex form the lateral lemniscus that mainly projects to the mesencephalon (inferior colliculus). A relatively narrow band of neurons that is oriented vertically, constitutes the nucleus of the lateral lemniscus. This nucleus exhibits again an extensive crossing between both ascending neurons via the Probst commissure.

Inferior Colliculus

The inferior colliculus represents the principal source of ascending inputs to the auditory thalamus. The main auditory structure of the mesencephalon receives its inputs from the axons originating in the cochlear nucleus, superior olivary complex and lateral lemniscus. According to functional differences in the response types, the inferior colliculus can be divided in three main subregions: the central nucleus (ICC), the pericentral nucleus (ICP) and the external nucleus (ICX). Single unit recordings in the ICC exhibited two main response types. The first type showed a transient onset response to the beginning of a tone, whereas the second type is characterized by a sustained response lasting during the whole duration of the stimulus. However, all response types of the Cochlear Nucleus as described in Figure 2.8, are also observed in the ICC. The majority of the recordings in the ICC showed a good tonal selectivity of the single units, indicated by a narrow, V-shaped tuning curve. The good tonal selectivity represents the

basis of the tonotopic organization of the inferior colliculus. Single units of the ICC are arranged according to their increasing characteristic frequency along the dorsoventral axis. In contrast, single units in the ICP and ICX exhibited a poorer tonal selectivity, characterized by broad and irregular tuning curves. Single units in the ICP and ICX show adaptation to repetitive stimuli, which is not the case in the ICC. Furthermore, some single units in the ICX seem to be particularly sensitive to complex acoustic stimuli.

Thalamus and Auditory Cortex

The auditory thalamus receives input mainly from the inferior colliculus and includes three distinct regions where sounds evoke activity: the medial geniculate body (MGB), the lateral part of the posterior nucleus of the thalamus (PO), and the auditory sector of the reticular nucleus of the thalamus (RE). The tonotopic organization observed in the inferior colliculus is maintained throughout the thalamical projections to the auditory cortex. Tracing studies in cats (Rouiller and Ribaupierre, 1989) have shown that the principal thalamic region (MGB) is organized tonotopically and that neurons project to the anterior, primary, and posterior auditory fields. A clear tonotopic organization was also observed for the PO, which mainly projects to the primary and anterior auditory cortex. In contrast, the sector of RE which is sensitive to auditory stimuli receives input from the MGB as well as descending projections from the cerebral cortex and can be understood as a feedback control.
The auditory cortex in mammals is located in the caudal part of the Planum supratemporale. In human, the primary auditory cortex or koniocortex is located on the superior surface of the temporal lobe, adjacent to the superior temporal gyrus, in the depth of the Sylvian fissure where it occupies most of Heschl's Gyrus. Cytoarchitectonically this part is defined as Brodman area 41. The size and exact location of the arial borders and the distribution of these areas differ between the studies, since a large interindividual as well as an interhemispherical variability can be found. There may be one to three–rarely four–transverse gyri of Heschl in each hemisphere (Campain and Minckler, 1976). Galaburda and Sanides (1980) described the primary auditory cortex almost entirely within the first transverse gyri of Heschl, whereas other authorities include the parakoniocortex known as area 42 (Kulynych et al., 1994; Leonard et al., 1998)), which surrounds the koniocortex and extends a variable distance beyond the transverse gyri of Heschl onto Planum temporale. It is described as the posterior part of the temporal gyrus and comprises area 22. Like most of the other cortical areas, the auditory cortex has a classical cytoarchitecture formed by six layers. The lamination is strongly

related to the topography of the interconnections of the auditory cortical areas. Afferent input originating from the thalamus, terminates mainly in layer III and IV, whereas layer I receives preferential input from the MGB. The deep layers V and VI are source of the descending projections that terminate in the inferior colliculus and in the auditory thalamus. Transcallosal projections terminating in the opposite auditory cortical fields originate from layers III and V (Rouiller and Welker, 1991).

2.2.2 Pitch–Specific Evoked Responses in Human

The great benefit of scalp-recorded electric potential and magnetic field studies is the possibility to measure noninvasively in awake, behaving humans and can be directly correlated with aspects of perception. The sum of active sources (sites of inhibitory postsynaptic potentials – IPSP) and active sinks (excitatory postsynaptic potentials – EPSP), as well as their passive return currents, weighted by strength, synchrony, orientation, and distance from the recording site is represented in the topography of the field on the scalp as can be seen from section 3.3.

Merzenich and Brugge (1973) recorded the electric potential of anesthesized macaques evoked by sinus tones of different frequencies and localized three tonotopic organized auditory areas. They reported a representation of high frequencies in the caudal part of the koniocortex, whereas low frequencies caused activity in the more rostral located part. Multiple auditory representations, with some of the areas exhibiting a specific tonotopic organization have been reported in other mammmals such as gerbil (Scheich et al., 1993) or monkey (Morel et al., 1993). The first study on auditory evoked fields in humans was published by Reite et al. (1978). Romani et al. (1982) could show in humans that the magnetic field evoked by the presentation of amplitude modulated sinus tones can be modelled with tonotopically organized equivalent source dipoles in the primary auditory cortex. High carrier frequencies were located medial, whereas the center of activity evoked by lower carrier frequencies was found more lateral.

Scalp-recorded auditory evoked potentials (AEPs) or auditory evoked magnetic fields (AEFs) are often onset responses (see Figure 2.9, left). The response could be generated from neurons that respond specifically to the onset of the stimulus but it could also be that only the change in the stimulus synchronizes sufficiently enough neurons to generate a field potential. Subsequent activity is cancelled out, since it contains both, negative and positive components. However, there is no total cancellation effect because of the differences in the rise and fall times of the

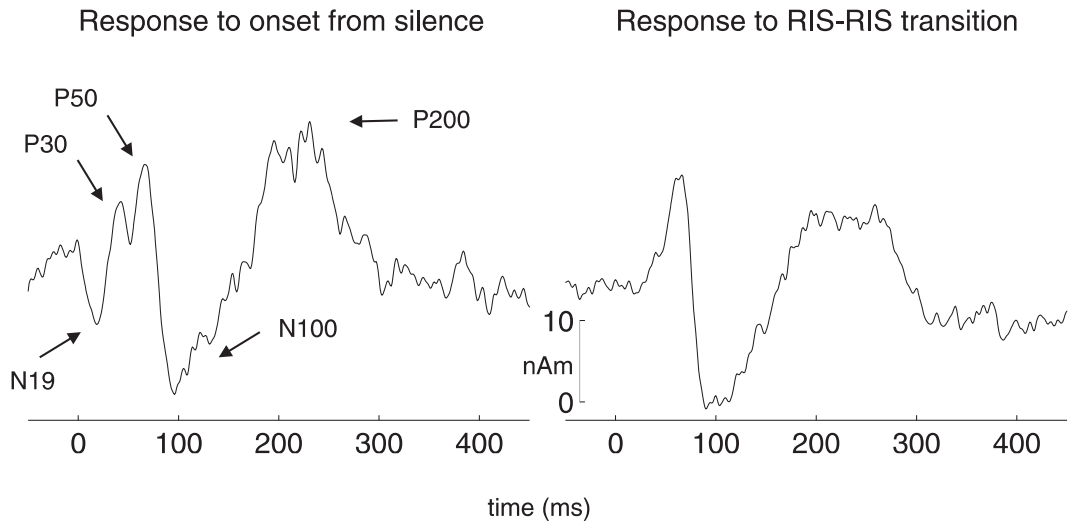

Figure 2.9: *Typical auditory evoked response by the presentation of RIS. Left: stimulus onset from silence, right: response to the transition from RIS to RIS. The nomenclature describes the components according to their polarity either with* N *(negative) or* P *(positive) and the ideal latency of the component after stimulus onset.*

synaptic potentials. The remaining steady-state response is known as sustained field. The transient responses are generally classified following the categorization suggested by Picton et al. (1974): (1) early auditory evoked fields (EAEF) with latencies up to 8 ms, (2) responses in the middle latency (MAEF) interval between 8 ms and 40 ms, and (3) late auditory evoked fields (LAEF) that occur later than 40 ms after stimulus onset. EAFS mirror activity beginning in the auditory nerve, following the ascending auditory pathway up to the brainstem. For a review of the generators of the EAFs, see Scherg (1991). The nomenclature, that is used in the present work describes the components according to their polarity either with *N* (negative) or *P* (positive) and the ideal latency of the component after stimulus onset. For example, P30 stands for the positive peak, approximately 30 ms after the onset of a sound. Additionally, an 'm' for neuro*magnetic* is added to distinguish it from a response measured by EEG. The single peaks of the MAEFs are ordered with increasing latency (8 ms, 12 ms, 19 ms, 30 ms, and 40 ms). The two early components at latencies of 8 and 12 ms seem to be the result of subcortical sources. Because of the radially distribution, both components can not be detected with MEG and occur in the EEG only (see section 3.3). A typical time course of an equivalent dipole source of the magnetic field, located in the auditory cortex is shown in the left part of Figure 2.9. A different waveform is observed when the magnetic fields were evoked by the transition from one tone to another. As illustrated in the right part of Figure 2.9, the N19m and P30m components

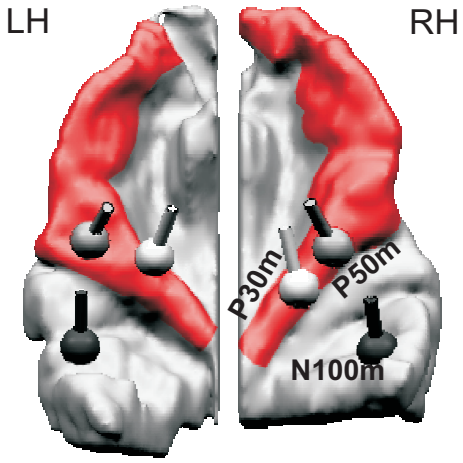

Figure 2.10: *Equivalent source dipoles of MEG components P30m, P50m and N100m fitted on the single components. Figure adopted from Schneider et al. (2004).*

vanished and the first positive peak is observed approximately 50–60 ms after the transition. The source of the N19–P30 complex could be explained by Scherg and von Cramon (1986) with a two dipole model, suggesting a single source in each auditory cortex. It was validated with intracranial measurements by Liégeois-Chauvel et al. (1991) who located the N19–P30 generator in the posterior part of medial Heschl's Gyrus (see also Figure 2.10).
The late auditory evoked response of Figure 2.9 shows the typical P50–N100–P200 complex, arising from different areas of the auditory cortex. Reite et al. (1988) identified the source of the P50 component as most likely in Planum temporale. Liégeois-Chauvel et al. (1994) reported the generator of the P50 to be more lateral compared to the N19–P30 complex, which was validated in MEG measurements by Mäkelä et al. (1994) and can be seen in Figure 2.10 from Schneider et al. (2004).
The prominent N100 peak is composed of multiple, partially temporally overlapping, independent components (Näätänen and Picton, 1987; Lütkenhöner, 2001). Pantev et al. (1988) and Roberts and Poeppel (1996) showed with MEG the dependence of the transient N100m deflection to sinusoidal tones of different frequencies. Both studies reported a correlation between frequency and latency of the N100. Ragot and Lepaul-Ercole (1996) used EEG and showed that for harmonic series with a varying spectrum the latency of the N100 only depends on the fundamental frequency and not on details of the spectrum. Other studies (Stufflebeam et al., 1998, e.g.) revealed that the N100 is also sensitive to other stimulus features as the intensity of the stimuli. The multiple sources of the N100 component are probably generated over a wide region of the supratemporal plane.

Magnetic field recordings by Lütkenhöner and Steinsträter (1998) suggested that the N100 arises from Planum temporale and that the P200 seems to have its center of activity in Heschl's Gyrus.
To isolate the specific response to the onset of pitch, Mäkelä et al. (1988) used a continuous stimulation with transition from noise to square waves. In this way, responses to the simple onset of energy flux are avoided and responses to the pitch onset can be extracted. About 100 ms after the transition, a prominent deflection (N100m') was found to be sensitive to pitch height and pitch salience. In a recent MEG study, Krumbholz et al. (2003) applied transitions from noise to RIS to investigate the POR to RIS with positive gain. They reported the specific N100m' response to be involved in pitch processing and found the center of activity for this component in the medial Heschl's Gyrus.
Gutschalk et al. (2002) applied a four dipole model to describe the relationship between click train of different pitch strength at different sound intensities. The dipole pair that accounted for the pitch strength was located in the lateral part of Heschl's Gyrus. The second pair of dipoles was fitted posterior to the first in Planum temporale. It was found to be sensitive to the intensity of the sounds. Thus, they concluded, that the anterior dipole pair is involved in the perception of pitch. In a positron emission tomography (PET) study from Griffiths et al. (1998), activation in Heschl's Gyrus increased with the temporal regularity of RIS. The results of a functional magnetic resonance imaging (fMRI) experiment conducted by Warren et al. (2003) showed, that the medial Heschl's Gyrus is activated similarly when processing either pitch or noise. But they reported an increased activity evoked by pitch changes in a small area of the lateral Heschl's Gyrus. The results of another fMRI experiment conducted by Patterson et al. (2002) revealed activity in the lateral aspect of Heschl's Gyrus for a melody of RIS versus a fixed pitch condition, which gives reason to the fact, that the lateral Heschl's Gyrus is involved in detecting pitch changes.

2.3 Modelling of Pitch Processing

The first theory dealing with the "pitch of the missing fundamentals" was introduced by Schouten (1940), who claimed that the residual pitch is extracted by using the temporal activity pattern of nerve fibers, stimulated by unresolved frequency components. Licklider (1951) introduced the first model of the cochlea that described the perceived pitch in two dimensions. Figure 2.11 shows the tonotopy of the cochlea in x–direction. The τ–dimension represents peaks in the autocorrelation function calculated with interspike intervals of the auditory

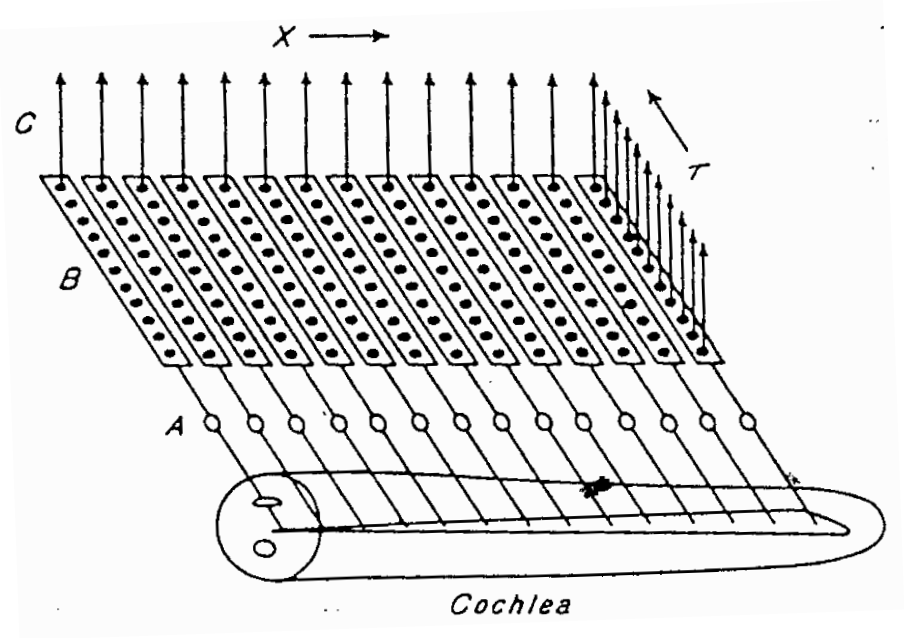

Figure 2.11: *Schematic drawing of Licklider's duplex theory. The tonotope* x*–dimension is along the uncoiled cochlea. The signal of the activated neurons of the auditory nerve (A) is fed to autocorrelators (B), whose delay– or* τ*–dimension is orthogonal to* x*. Figure adopted from Licklider (1951).*

nerve.

A few years later, experimental findings (Houtsma and Goldstein, 1972) suggested, that pitch is derived at a more central processing stage of the auditory system. Thus, most of the today known pitch models contain elements of central pitch processing.

Goldstein (1973) assumed in his so called optimum processor theory that the frequencies f_i of spectrally resolved harmonics are transformed into Gaussian random variables x_i, with means equal to f_i and variances that depend on f_i only. According to this theory, all amplitude and phase information is ignored. It is assumed that the input numbers x_i are noisy representations of harmonic frequencies and the central processor makes an estimate of the unknown harmonic numbers and the corresponding fundamental frequency. The probability of an incorrect estimate of the harmonic order can cause mismatch in the harmonic numbers, which results in ambiguous pitches. The virtual pitch theory of Terhardt (1979) is an alternative attempt to explain the perception of a distinct pitch. Unlike the optimum processor theory, it is formulated in a deterministic manner and assumes that the single frequencies are transformed in the auditory periphery into special pitch cues. The output of the model are different possible

pitches, each with an associated strength, derived from the stimulus properties. Another computational model has been proposed by Meddis and Hewitt (1991a). It is composed of a linear bandpass filter representing the outer and middle ear, followed by a bank of 128 overlapping critical-band (gammatone) filters mapping the basilar membrane motion. The output of each channel is transformed with the hair-cell simulation model (Meddis, 1988), which includes a simple refractory-repository period model for the auditory nerve fibers. Finally, an interspike interval autocorrelation process according to the formulations of Licklider (1951) is computed and averaged across the channels (see Figure 2.11). The perceived pitch is predicted by the location of the first peak in the windowed autocorrelation function, the pitch strength by the relative height of this peak.

2.3.1 Auditory Image Model

A model related to the formulations of Meddis (1988) was introduced by Patterson et al. (1995) is applied in the present work to test if the perceived pitch of RIS and bandpass filtered complex sounds can be explained with the model.
It mainly consists of three different stages of auditory processing: spectral analysis, neural encoding and a temporal integration stage. Briefly, the spectral analysis converts the sound wave into a representation of the Basilar membrane motion (BMM). The neural encoding stage stabilizes the BMM in level and sharpens special features like vowel formants to produce a simulation of the neural activity pattern (NAP) evoked by the sound in the auditory nerve. The temporal integration stage stabilizes the periodic structure in the NAP and produces a simulation of the perceived pitch, referred to as the auditory image.

Spectral Analysis

The first step of the model is to perform a spectral analysis of the incoming sound. The digitized wave is transformed into an array of filtered waves, which is the model's multichannel representation of the BMM. The single filters a linearly distributed along the frequency scale, each of their bandwidths corresponding to a fixed distance on the basilar membrane. It is essentially the same function originally described by Greenwood (1961) as the 'cochlear frequency position' function. The simulated human filter shape relating the equivalent rectangular bandwidth (ERB) to its center frequency (f_0) is

$$ERB = 24.7\,Hz + 0.108 f_0. \tag{2.1}$$

The conversion of the multichannel representation of the basilar membrane motion to the approximation of the neural activity pattern (NAP) that flows from the cochlea up the auditory nerve to the cochlear nucleus can be performed (1) by a gammatone filterbank or (2) a physiological auditory filter for generating the BMM. The description of AIM is restricted to the physiological version, since these settings were applied in the present work to simulate the pitch of RIS and bandpass filtered complex sounds.
A non-linear transmission line filter is used to implement the classical, one-dimensional, approximation to cochlear hydrodynamics (Giguère and Woodland, 1994a). According to (Giguère and Woodland, 1994b), a feedback circuit representing the fast motile response of the outer haircells generates level-dependent BMM. The filterbank generates combination tones of the type $f_1 - n(f_2 - f_1)$ which propagate to the appropriate channels.

Simulation of the Neural Activity Pattern

In the second stage of AIM is the conversion from the mechanical BMM to the afferent neural activity, at the level of the auditory nerve, realized. The transduction process itself is performed by inner haircells. Again, two alternative simulations are provided for generating the neural activity pattern (NAP): a bank of two-dimensional adaptive thresholding units (Holdsworth and Patterson, 1993) and a physiologically adopted bank of inner haircell simulators (Meddis, 1988). The module provided by Meddis simulates the operation of an individual inner haircell; specifically it simulates the flow of neurotransmitters across three reservoirs that are postulated to exist in and around the haircell. The module reproduces important properties of single afferent fibers, such as two component time adaption and phase locking. The solution of the transmitter flow equations is realized with the wave-digital-filter algorithm provided by Giguère and Woodland (1994a). In this way is one haircell coupled to each channel of the filterbank. Figure 2.12 shows the typical simulation output of a NAP in response to a complex harmonic sound with a fundamental frequency of 250 Hz. The ordinate of the upper panel of the figure is ERB-scaled. Thus, each line represents the output of an auditory channel over the time. The lower panel is the sum of all channels.

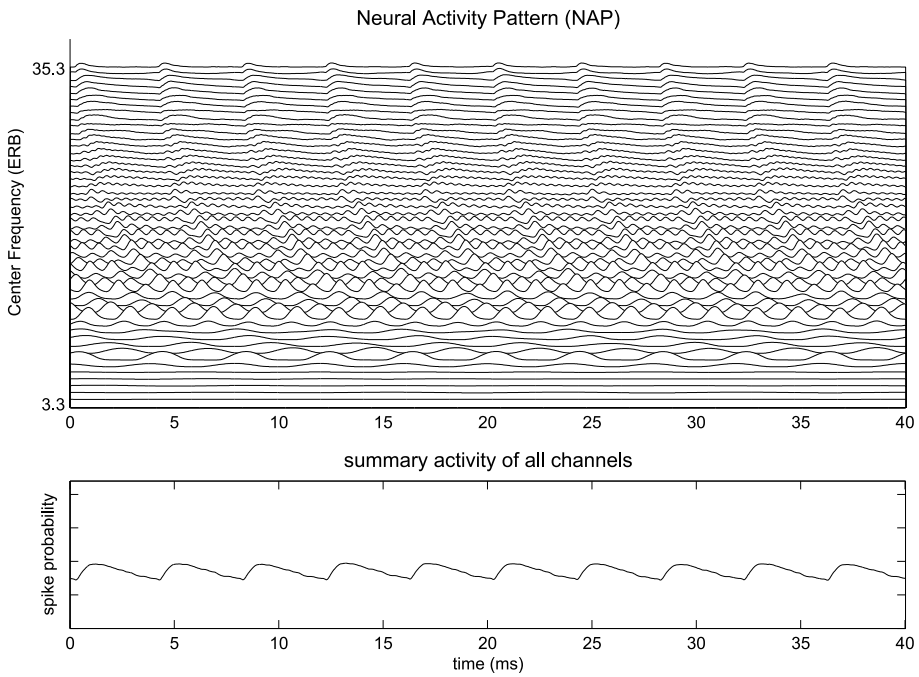

Figure 2.12: *Simulation of the Neural Activity Pattern of a harmonic complex sound with a fundamental frequency of 250 Hz. The ordinate of the upper panel is ERB-scaled. Each line represents the output of an auditory channel over time. The sum of activity in all channels is given in the lower panel.*

The Auditory Image

It is often assumed that the peripheral auditory processing ends at the output of the cochlea and that the pattern of activity recorded at the level of the auditory nerve is in some sense what we hear. But the simulation of the NAP is not the representation of human auditory sensation because phase differences are still included which we do not hear. Furthermore, at this level the model does not include auditory temporal integration. For example, when the processing of a periodic sound is simulated, the NAP oscillates. But the sensation of periodic sounds does not flutter. In fact, periodic sounds produce the most stable auditory images. The first models suggested to integrate the NAP over time using a sliding temporal window. However, the results of these early models smear out fine-grain temporal detail that we hear (Patterson and Moore, 1986). So, the problem is to determine how the auditory system can integrate over cycles to form a stable image, but without losing fine-grain temporal information. In the auditory image model a mechanism is implemented that monitors the activity in the single channels of the NAP (see Figure 2.12). To illustrate the mechanism,

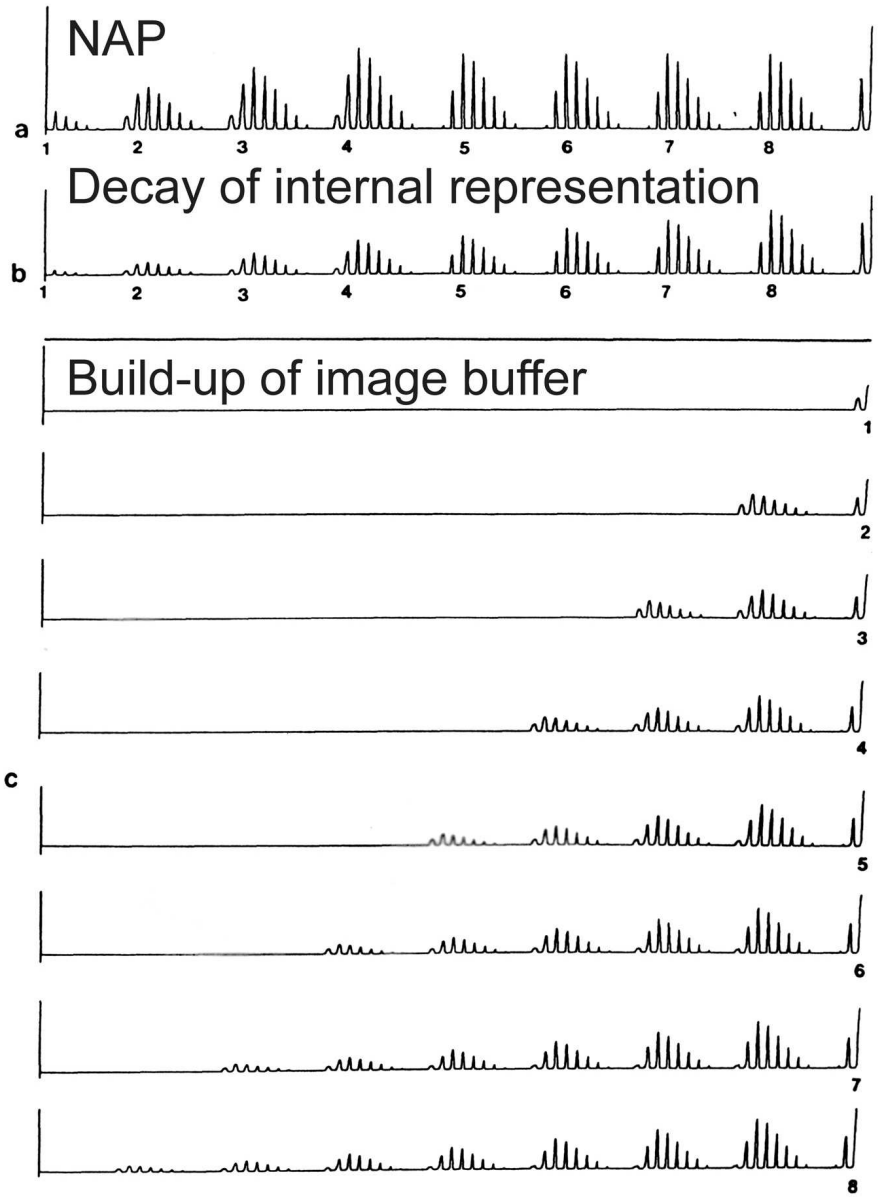

Figure 2.13: *Mechanism of strobed temporal integration. The top part (a) shows activity in a single channel of the NAP, enumerated from one to eight according to the occurrence in time. The linear decay of the activation in the buffer of the NAP is represented in the second row (b). When a pulse in the NAP reaches the threshold level, the whole information of the NAP is transferred to the corresponding channel of the static image buffer (entitled 'Build-up of image buffer') and summed with the information that is already there (c). The rows c1-c8 show the auditory image just after each successive trigger pulse.*

underlying the construction of the auditory image, Figure 2.13a shows a single channel of the NAP. It is assumed that there is a buffer store in the auditory system through which the neural activity flows in a first-in first-out (FIFO) fashion. At the same time the level decays linearly in time (Figure 2.13b). A large pulse in the channel triggers temporal integration in that channel; that is a copy of all the current information in the corresponding NAP-channel (FIFO-buffer) is transferred to the corresponding channel (Build-up of image buffer in Figure 2.13) of a static image buffer and summed point for point with the information that is currently there. At this point the temporal code is converted to a spatial code. The strength of the auditory image decays exponentially in time but the position of the information does not change once it occurs in the image buffer (Figure 2.13c).

The triggering mechanism itself is an adaptive threshold value for each chan-

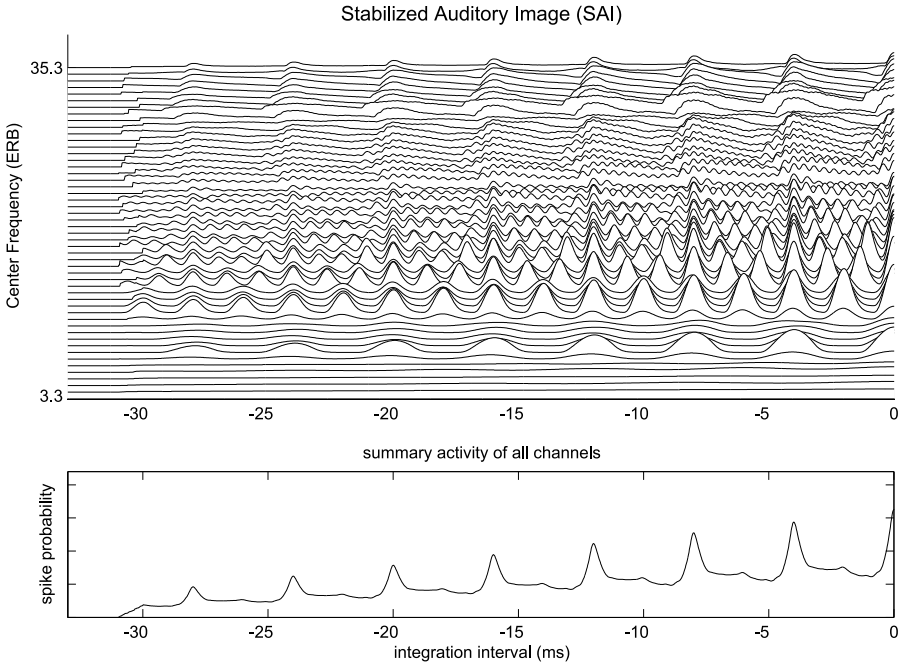

Figure 2.14: *Simulation of the Stabilized Auditory Image from the NAP shown in Figure 2.12. The abscissa represents the time interval and corresponds to the decaying image buffer with a half life time of 20 ms. In analogy to the NAP, the ordinate is the center frequency of each channel. The SAI is computed from the NAP of a complex sound with* $f_0 = 250\,Hz$*. The activity of each frequency channel can be summed up and results in the summed stabilized auditory image (lower part of the Figure). The distance from 0 to the location* τ *of the first peak in the summarized SAI corresponds to a predicted pitch of* $\frac{1}{\tau} = 250\,Hz$*.*

nel. After the NAP pulse exceeds threshold, a trigger pulse is generated at the time associated with the maximum of the peak. The threshold value is then reset to a value above the height of of the current NAP peak. The value decays linearly in time until another peak is encountered. The amount of threshold elevation is determined by the height of the most recent peak and the time between the last pair of trigger pulses in a way that for periodic sounds the mechanism turns to generate one trigger pulse per period at the time of the largest peak. The period-synchronous integration causes periodic information to accumulate whereas aperiodic information is faded out. The triggering is done on a channel by channel basis and since the major peaks in the single channels occur at different times this mechanism is asynchronous across channels and it is this aspect that causes the alignment of the auditory image and which, in turn, accounts for the loss of phase information in the auditory system (Patterson, 1987). The quantized temporal integration of the periodic information generates a stabilized auditory image (SAI). Figure 2.14 shows the SAI computed from the NAP of a complex sound with an f_0 of 250 Hz. The perceived pitch is represented in the first peak of the summarized SAI (lower part of Figure 2.14). As can be seen, the first peak is at τ=4 ms which corresponds to a predicted pitch of $\frac{1}{\tau}$ =250 Hz.

Chapter 3

Methods

This Chapter gives a brief description about the applied methods and is divided in three parts. In the first part, a review of how noise can be processed to introduce a pitch is presented. The special characteristics of the applied stimuli in the experiment described in Chapter 4 are shown.
In the second part, an introduction on the method of paired comparison according to Bradley and Terry (1952), and Luce (1959) is given, together with the applied algorithm. Using the method they developed, the known problems of subjects who have to judge pitch, is solved for the experiments that follow.
In the last part of the Chapter, the technical background for the magnetoencephalographic measurements is provided, a crucial cue to understand the investigations, presented in Chapters 4 and 5.

3.1 Huygens Noise—Regular Interval Sounds

Different and also controversy definitions of the sound that Huygens described initially in 1693 exist in the literature about regular interval sounds. Bilsen and Ritsma (1969/70) introduced the term 'rippled noise' for the phenomenon observed in the garden of Chantilly de la Cour. In their experiment, the noise was processed with only one iteration and the name was derived from the rippled power spectrum (see Figure 3.1, where the spectrum is shown for RIS generated with two iterations for better illustration). However, this type of sound has also been referred to as 'ripple noise' (Yost and Hill, 1978), 'repetition noise' (Fay et al., 1983), 'cosine-noise' (Bilsen et al., 1975) and 'comb-filtered noise' (Pick, 1980). With an increase of the number of iterations, Yost et al. (1993) suggested that the stimuli produced by the circuits shown in Figure 3.2 be referred to as

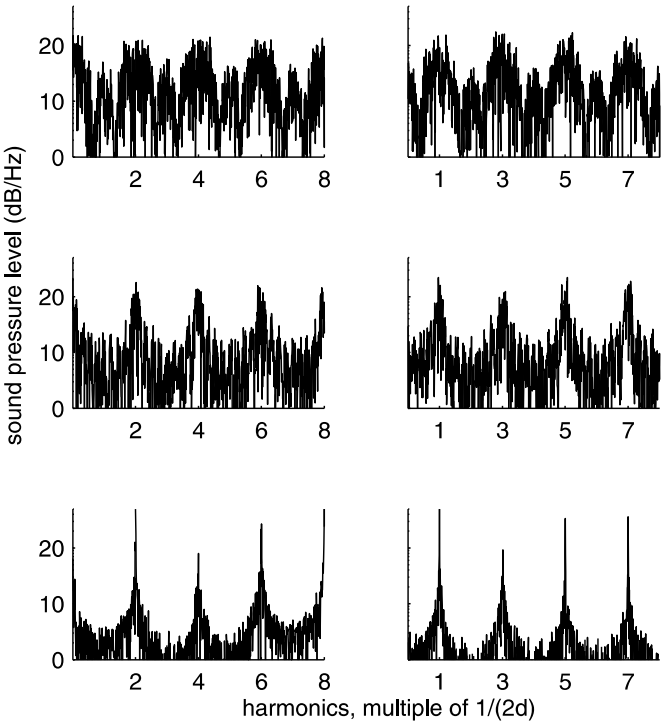

Figure 3.1: *Power spectra of RIS, generated with positive (left) and negative gain (right). Sounds with increasing pitch salience are depicted from top to bottom. The number of iterations ($n = 2, 8,$ and $4,096$ - top to bottom) determines the peak-to-valley ratio.*

'iterated rippled noise (IRN)' since the first stage of the process (Figure 3.2a) introduces ripples in the power spectrum and subsequent stages (see Figure 3.1) iterate the process and only sharpen the ripples. Griffiths et al. (2001) introduced the term 'regular interval noise' to give more emphasis on the rippled spectrum; Patterson et al. (2002) extended it to 'regular interval sounds'. In the past, RIS generated with infinite numbers of iterations (see Figure 3.2b) has also been referred to as 'comb-filtered noise' (Raatgever and Bilsen, 1992) and 'peaked rippled noise' (Fastl, 1988). To avoid any confusion on the notation, we use the term Regular Interval Sounds (RIS) throughout the present work, independent of the number of iterations. The power spectra of RIS (Figure 3.1, top to bottom) are combed and come close to a line spectrum for a high number of iterations n. Increases of the peak-to-valley ratio in the spectrum are often used to explain the increase in the pitch strength with increasing iterations. For a positive sign of g (Figure 3.1, left), the spectra peak at integral multiples of the inverse of the delay ($f = \frac{1}{d}, \frac{2}{d}, \frac{3}{d}, \ldots$).
If the gain is negative (Figure 3.1, right), the spectra peak at odd multiples of

$f = \frac{1}{2d}$ ($f = \frac{1}{2d}, \frac{3}{2d}, \frac{5}{2d}, \ldots$). Independent of the set of the parameters, there are also different possibilities to process and introduce the pitch in the noise. The basis of RIS is always noise that is periodically reflected against the staircases. The noise signal $\Phi(t)$ is multiplied by a gain factor $|g| \leq 1$, delayed by a time d and added back to the original noise. This process is repeated n times. The amplitude $y(t)$ of the original Huygens noise is given by

$$y(t) = \Phi(t) + g \sum_{k=1}^{n} \Phi(t - kd). \tag{3.1}$$

In the hearing experiments of our days, the output $y(t)$ is derived slightly different:

$$y(t) = \Phi(t) + \sum_{k=1}^{n} b_k \Phi(t - kd). \tag{3.2}$$

In contrast to the original Huygens way, where the coefficients b_k would be simply

$$b_k(H) = g, \tag{3.3}$$

the noise signal is processed as shown in Figure 3.2c. The delayed and attenuated component is added to the original noise, giving b_k the form

$$b_k(O) = g^k \tag{3.4}$$

In the literature, it is often called IRNO (Yost, 1996b), and stands for the add-original process. The pitch that is derived using the network of Figure 3.2c is stronger, since the spectral peaks grow faster.
A second method to create RIS is shown in Figure 3.2d. In this special network that was initially used by Yost et al. (1996), the delayed and attenuated noise is added, not to the original noise, but rather to the same signal that enters the current delay-attenuate-add stage of the circuit (IRNS). The coefficients b_k in the equation 3.2 would be

$$b_k(S) = g^k \binom{n}{k} \tag{3.5}$$

This version of processing the noise does not have an analog in the natural environment. It does, however, provide an important perceptual comparison for the naturally occurring IRNO stimulus because the two are perceptually similar but can exhibit significant spectral differences. The envelope of RIS generated with a delay of 2 ms, positive gain and six iterations is shown exemplarily for the add-original and the add-same network in Figure 3.3. The Hilbert transformation of the resulting sound represents the envelope in time, which is shown in the

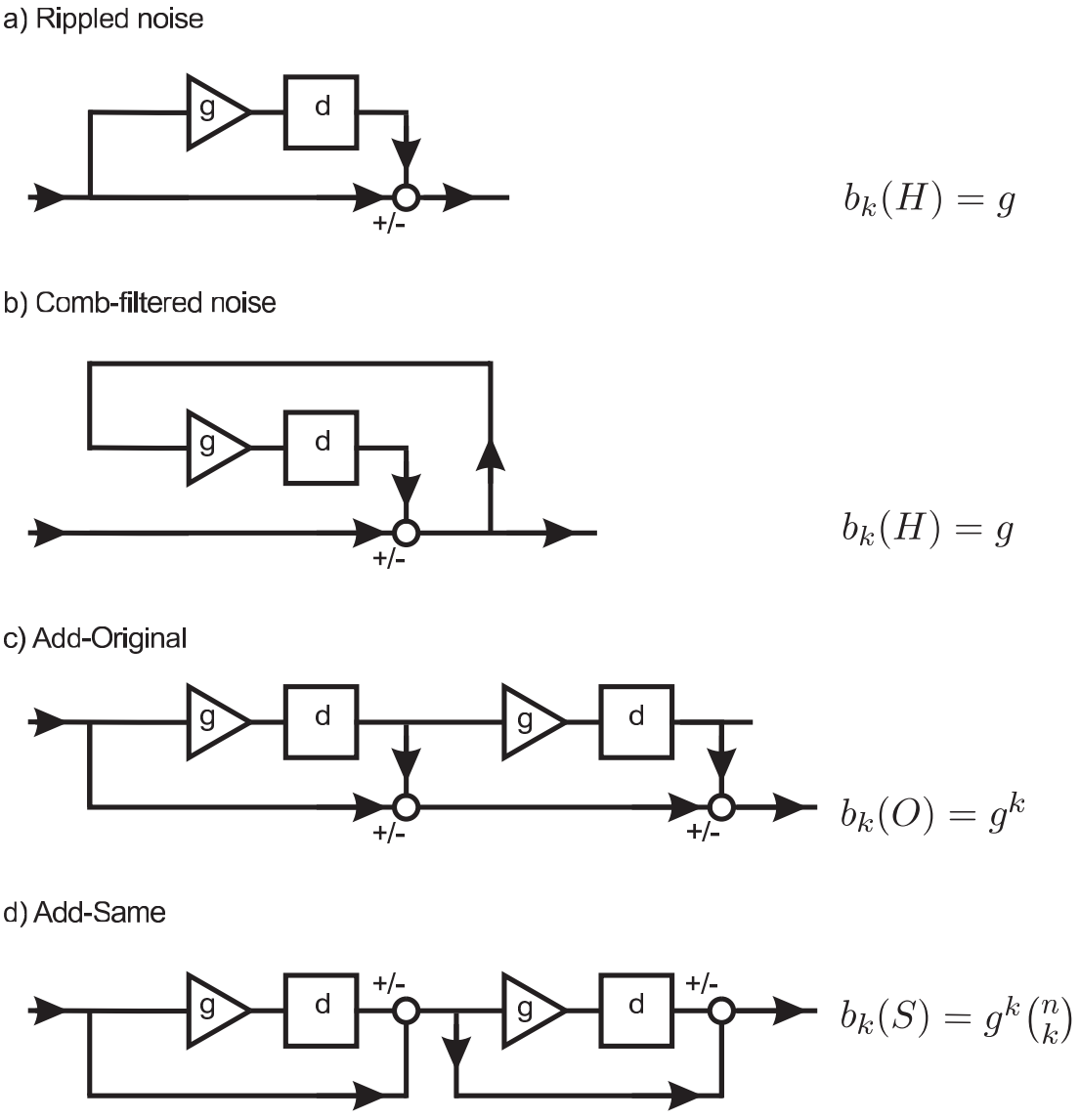

Figure 3.2: *Different possibilities of delay-and-add circuits to generate: a) rippled noise,with one iteration, b) comb-filtered noise (infinite number of iterations), c) IRNO add-original, the network that was used in the present work to generate RIS, and d) IRNS add-same network.*

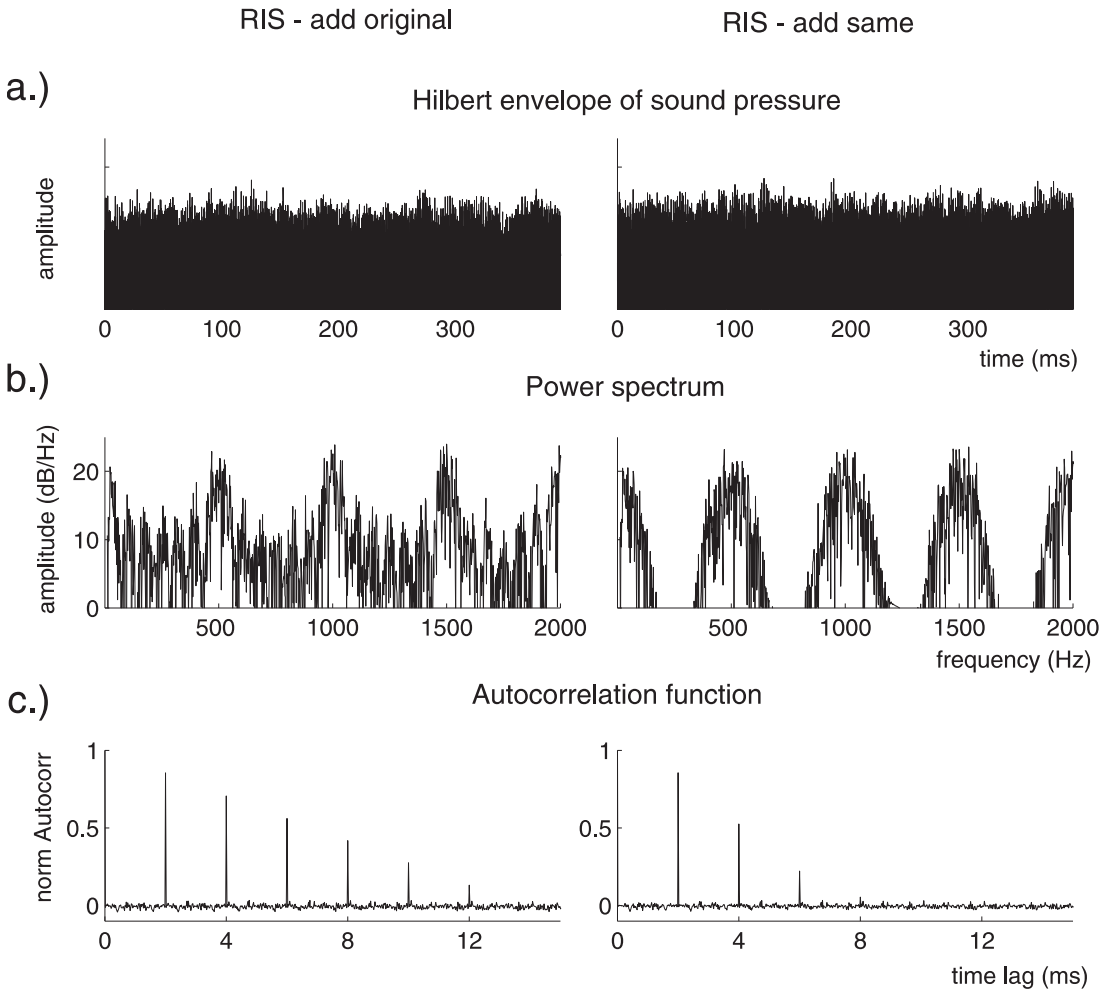

Figure 3.3: *RIS, generated with 2 ms delay, positive gain and six iterations. Figure a.) shows the Hilbert envelope of the time signal. No difference between RIS generated with the IRNO (left) and the IRNS (right) can be observed. In contrast, the spectra of Figure b.) demonstrate the differences. The peaks of the IRNO circuit are sharper, but the peak-to-valley ratio is higher for the IRNS condition. Figure c.) shows the normalized autocorrelation functions: The first peak at lag 2 ms, which commonly approximates the perceived pitch is identical for both circuits. The difference in the autocorrelation function is in the slope of decrease, which is larger for the IRNS condition.*

upper panel of the Figure. The fine structure of both signals exhibit no periodic fluctuations and does not allow to distinguish the sounds from each other. Figure 3.3b clearly demonstrates the differences in the power spectrum of the sounds. When using the IRNO network to generate RIS, several smaller peaks appear in the valleys between the major peaks. The number of these peaks increase and their amplitudes decrease as the number of iterations is increased. In contrast, no smaller spectral peaks between the main peaks are observed, when the IRNS circuit is used to generate RIS.
The Wiener-Khintchine relation says, that the inverse Fourier transform of the power spectrum is the autocorrelation function. In the lower part of Figure 3.3, the autocorrelation functions of both signals are computed. In each case, the peaks in the functions appear at time lags corresponding to the delay time d and its integer multiples. The first peak in the autocorrelation function at time lag τ results in a perceived pitch with a frequency of $\frac{1}{\tau}$. The height of the first peak increases with n for both, IRNO and IRNS stimuli. A comparison of the autocorrelation functions (Figure 3.3c) for IRNO and IRNS reveals that the remaining first peak in the autocorrelation function has the same height, when the stimuli are generated with the same gain g and the same number of iterations n. The magnitude of the remaining peaks decrease as the lag increases. As can be seen, the remaining peaks in the autocorrelation are greater for the IRNO-condition. However, (Yost et al., 1996) reported, that listeners were not able to distinguish between RIS generated either with the IRNO and the IRNS network. Thus, it can be concluded that the time lag of the first positive peak of the autocorrelation function is related to the perceived pitch of the sound and the height of this peak corresponds to the perceived pitch strength. The experiments, conducted in the present work were all produced with the IRNO network since an analog in nature exists.
If the gain is negative, the autocorrelation function of RIS is different, compared to RIS generated with a positive gain factor: the peaks in the autocorrelation function at τ and its even integral multiples are negative. The peaks at odd multiples are positive. Thus, the first positive peak in the autocorrelation function of RIS generated with a negative gain factor g is located at time lag $\tau = \frac{1}{2f_0}$.

3.2 Scaling of Pitch

The task of pitch matching is very difficult and subjects often have problems to perform the task. Yost (1996a) reported that even musicians have problems to rate the pitch and had to exclude four of six musically trained subjects in his

experiment. Thus, it is necessary to introduce a procedure to circumvent the known problems on pitch matching. The method of paired comparison, where subjects are simply asked to decide between the pitch of two sounds, allows to derive a relative pitch scale reliably for all subjects.
The Bradley-Terry-Luce (BTL) model is such a method of paired comparisons and is used in the present work to derive relative pitch scales. It was developed by Bradley and Terry (1952) and extended by Luce (1959). It is typically applied in tasks when objects to be compared can be judged only subjectively (e.g. taste, art). The method was initially introduced by Fechner (1860/1965) and became popular in 1927 when Thurstone published his "law of comparative judgement". In the method of paired comparisons objects or in our case auditory stimuli are presented in pairs. The basic experimental unit is the comparison of two stimuli, A_1 and A_2, by a judge who must choose one of them. If more than two stimuli are under consideration every judge has to perform every possible paired comparison. The setup is called "balanced paired-comparison experiment" and for n stimuli and t judges the numbers of paired comparisons will be

$$t\binom{n}{2}. \tag{3.6}$$

The goal is to determine the probability that the stimulus A_i is chosen ($P(A_i)$) and to use this value to derive a relative scale of pitch. In our experiments, the judges have to decide which tone is of higher pitch ($A_i \succ A_j$). Ties or indifferent judgements are not permitted, so a listener must claim one of the two stimuli to be of higher pitch. To avoid judging effects, depending on the order of representation, each pair was presented twice. With random variation, once in the order A_i, A_j and in reversed order A_j, A_i. The mean of both comparisons was used to get the raw data of the number of times each stimuli was judged by the t judges to have a higher pitch than each of the other stimuli. From these raw data, a n square matrix $\mathbf{A}$ is formed as

$$\mathbf{A}_{ij} = \begin{pmatrix} - & a_{12} & \cdots & a_{1n} \\ a_{21} & - & \cdots & a_{2n} \\ \vdots & \vdots & - & \vdots \\ a_{n1} & a_{n2} & \cdots & - \end{pmatrix} \tag{3.7}$$

where a_{ij} denotes the observed number of times, stimulus i was judged to be of higher pitch than stimulus j. The total number of comparisons of the two stimuli is $t_{ij} = a_{ij} + a_{ji}$. Since a stimulus can not be compared to itself, the diagonal elements of the matrix are left vacant. From $\mathbf{A}$, a probability matrix is constructed with elements $p_{ij} = \frac{a_{ij}}{t_{ij}}$. It is the probability that the stimulus A_i is

preferred over stimulus A_j ($A_i \succ A_j$). The summation of the symmetric elements is $p_{ij} + p_{ji} = 1$. The elements of the basic Bradley-Terry form of the probability matrix $\mathbf{X}$ can be derived with:

$$x_{ij} = \frac{p_{ij}}{1 - p_{ij}} = \frac{p_{ij}}{p_{ji}} \tag{3.8}$$

and analogously

$$x_{ji} = \frac{p_{ji}}{1 - p_{ji}} = \frac{p_{ji}}{p_{ij}}. \tag{3.9}$$

Using the natural logarithm on the single elements of $\mathbf{X}$, the scale is transformed from a ratio scale to a difference scale. Finally, the values ($s(i)$) of the single stimuli can be obtained by averaging over the columns of the transformed matrix $\mathbf{X}$

$$s(i) = \frac{\sum_{j=1}^{n} \ln x_{ij}}{n}. \tag{3.10}$$

In the special case where all subjects judged the pitch of the stimulus A_i higher than the pitch of stimulus A_j ($a_{ij} = t_{ij}$), the values in equation 3.8 and in equation 3.9 equal infinity. According to (David, 1988), who suggested to assume that 'half a subject' rated the pitch vice versa, the value of a_{ij} was corrected to ($a_{ij} = t_{ij} - \frac{1}{2}$).

3.3 Magnetoencephalography

Information about brain activity can be obtained from recordings of the electric or magnetic field outside the skull. Both, electroencephalography (EEG) and magnetoencephalography (MEG) are reflections of the same neural phenomena: the synchronous activity of neurons (mostly postsynaptic currents). The advantage of both techniques is their temporal resolution, which is only limited by the sampling rate of the signal. Today, the possible resolution is in the order of tens of milliseconds and thus, allows to record the brain activity in real-time. The relative sensitivities of EEG and MEG of different events in the brain depend on the geometrical arrangement of the generator cells only, with respect to the sensors and on the conductivity distribution of the head. Compared to MEG, the interpretation of the EEG signals requires more precise knowledge of the thicknesses and conductivities of the tissues in the head since concentric inhomogenities do not affect the magnetic field at all but have to be taken into account in the analysis of EEG data.

3.3.1 Electromagnetic Fields of Biological Origin

Information is processed in the human cortex by a total of approximately 10^{10} neurons, which are connected by about 10^{14} synapses among each other. Electric pulses, called action potentials propagate along the axons and transmit the information from one neuron to another. As illustrated in Figure 3.4, chemical transmitters are liberated into the 50 nm-wide synaptic cleft that separates the two neurons, when a pulse arrives along the axon of a presynaptic cell. The transmitter molecules diffuse to the postsynaptic membrane where they activate selective ion channels. In the rest state of the cell, the resulting currents are balanced by diffusive and ohmic currents for each ion type (mainly Na^+, K^+, and Cl^-) leading to a concentration gradient between intra- and extracellular space and to a transmembrane potential of about –80 to –90 mV.
Thus, an increase of the permeability of the ion specific channels through the cell membrane causes either a depolarization (excitatory postsynaptic potential, EPSP) or a hyperpolarization (inhibitory postsynaptic potential, IPSP) of the second cell. A net current flows through the membrane and, as illustrated in Figure 3.4, the return current flows back through the surrounding tissue. The summed PSPs of all synapses of a neuron charge the membrane potential. When

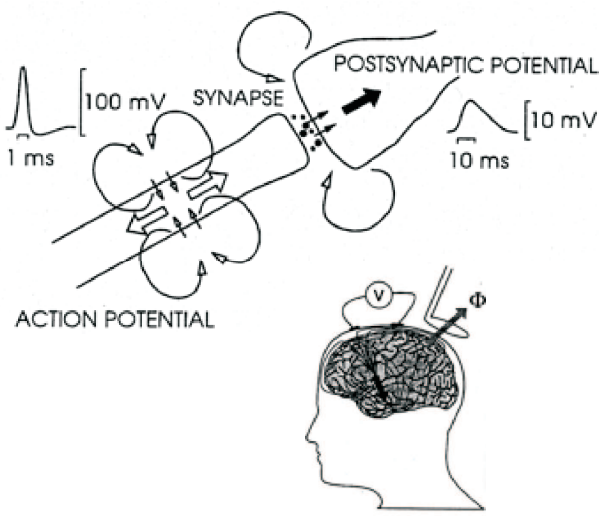

Figure 3.4: *Schematic view of a neuron: Generation of postsynaptic potentials from presynaptic activity. The presynaptic action potential propagates along the axon. At the synaptic cleft, the liberated chemical transmitters diffuse to the postsynaptic membrane and initiate a postsynaptic action potential. Note the amplitude and duration difference between the pre- and postsynaptic potential. Potential differences due to propagating action potentials, can be detected. Adopted from Ilmoniemi (1993).*

the transmembrane potential is increased (in the case of an EPSP) up to a threshold of about –40 mV, an action potential is initiated in the second cell's axon. Since the charge distribution of a single neuron is only of the order of approximately 20 fAm, it takes about 10^6 synchronized neurons to produce a magnetic field or an electric potential that can be recorded at the scalp. Thus, with the assumption of 20,000 pyramidal cells/mm^2 on the cortical surface, a generator of such an amplitude covers at least an area of about 0.5 cm^2 (Eggermont and Ponton, 2002) and the cortical activity measured with MEG and EEG always mirrors a sum activity.
Furthermore, two effects account for the measured signal. First, the resulting field of the PSP decays with the distance by only $1/r^2$ whereas the quadrupolar field of an action potential decreases with $1/r^3$. Second, a typical PSP lasts about 10 ms, whereas the duration of an action potential is only in the order of 1 ms. Thus, the temporal overlap, and in particular, the temporal synchronization of different neurons is remarkably reduced.
Altogether, the field distribution at the scalp is mainly an effect of the postsynaptic current flow with its approximate dipolar distribution.

3.3.2 Neuromag-122™ Gradiometer System

The auditory evoked magnetic field that is recorded from the human brain, typically has a peak amplitude of 100 fT. In contrast, the electromagnetic noise caused by MRI-scanners (3 T), the earth's static field (50 μT) or laboratory noise (100 nT) are many orders of magnitude larger compared to the far field outside the head evoked by the brain tissue. Therefore, the detection of this kind of biological activity requires highly sensitive instrumentation and, at the same time, attempt to eliminate extraneous magnetic fields.
The Neuromag-122™ whole head magnetoencephalograph provides the possibility to measure these auditory evoked magnetic fields. The cross-section of Figure 3.5 (left), illustrates the dewar, that consists of two cylinder shaped vessels, between which there is a thermally insulating vacuum. The subject's head is placed inside the helmet-shaped device (Figure 3.5 right). The construction is such that the 122 sensors cover the full cortex with an average distance of only 17 mm apart from the scalp. The sensors are planar gradiometers that consist of two figure figure-eight-shaped and perpendicular wristed pickup coils. The advantage of applying gradiometers instead of magnetometers is their higher sensitivity to sources near the coils, since the rather homogeneous field produced by distant noise sources is cancelled out. The time varying magnetic flux produced by the

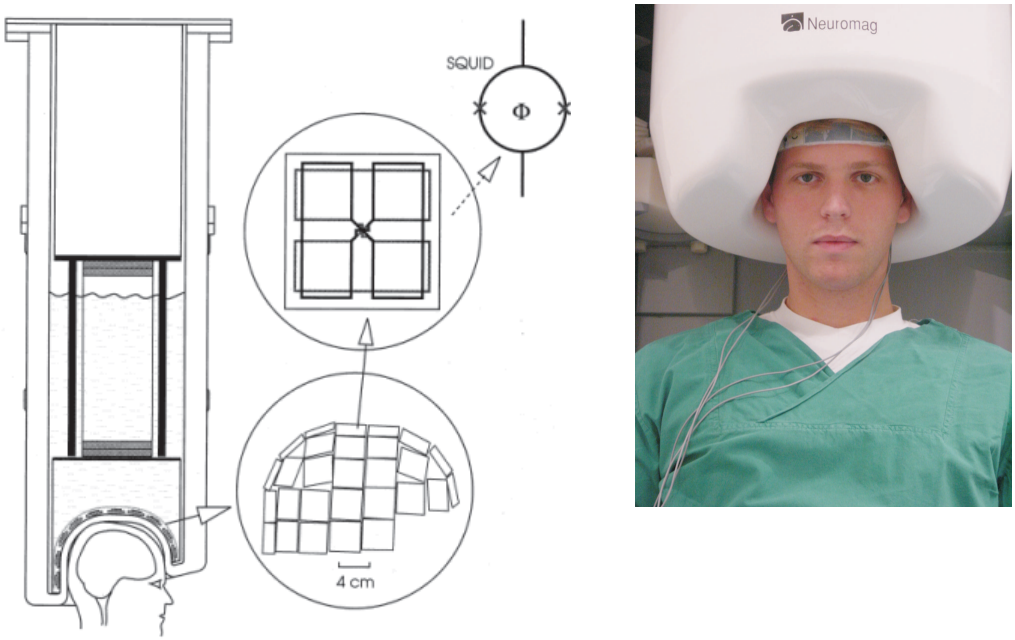

Figure 3.5: *Left: Cross section of the dewar, cooled with liquid helium as shown in Hämäläinen et al. (1993). The helmet shaped sensor elements (enlarged) illustrate the planar gradiometer array with its two 'figure-eight-shaped' pickup coils, which are coupled to the SQUIDs. Right: Front view of the dewar with subject SR.*

human brain passes through a pickup coil and induces a time-varying electrical current within the wire. In a typical coil, though, this current is quickly dissipated as heat by the electrical resistance of the wire. Thus, MEG measurements became possible only through the development of superconductive instrumentation, which ensures that even very small changes in the magnetic flux induce a certain amount of current within the coil. Each pickup coil is coupled to an input coil, which transforms the voltage back into a secondary magnetic field. It is detected by a so called SQUID (Super Conducting Quantum Interference Device), a superconducting ring, interrupted by two Josephson junctions (DC SQUID). The electrons of the induced current tunnel through the parallel junctions and interfere with each other. This is caused by a phase difference between the wave functions of the electrons, which depends upon the strength of the magnetic field through the loop. The phase difference of 2π is reached for a magnetic flux quant with the flux $\Phi_0 = h/2e = 2.07 \cdot 10^{-15}\,\mathrm{Vs}^2$. Hence, the two junctions in parallel can detect variations in the magnetic field very sensitively. To maintain the superconducting state, the pickup coils and the SQUIDs are immersed within a liquid helium bath at an operating temperature of 4.2 °K.
The amount of magnetic noise from non-neuronal sources is significantly reduced

by operating the whole device inside a magnetically shielded room made of alternating layers of high magnetic permeability (μ-metal with $\mu \geq 30,000$) and high electrical conductivity (aluminium). Depending on the frequency of the noise, the shielded room manufactured by Imedco (Switzerland), reduces the external magnetic noise by 40–100 dB/Hz.

3.3.3 Source Identification

In general, neural activity can be represented mathematically as a primary source with the current density $\vec{J}_p(\vec{r})$ in a closed volume G of finite conductivity $\sigma(\vec{r})$. The volume represents the head and thus, outside G, the conductivity and the current density is zero. The potential $\Phi(\vec{r})$ inside the head can be computed as the divergence of the source from Poisson's equation,

$$\nabla \vec{J}_p(\vec{r}) = \nabla(\sigma(\vec{r})\nabla\Phi(\vec{r})). \tag{3.11}$$

Using Maxwell's equations

$$\nabla \cdot \vec{E} = \frac{\rho}{\epsilon_0} \tag{3.12}$$

$$\nabla \cdot \vec{B} = 0 \tag{3.13}$$

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \tag{3.14}$$

$$\nabla \times \vec{B} = \mu_0 \left(\vec{J} + \epsilon_0 \frac{\partial \vec{E}}{\partial t} \right), \tag{3.15}$$

one can estimate $\vec{J}(\vec{r})$ from measurements of $\vec{E}$ and $\vec{B}$. Under quasistatic conditions, i.e., sufficiently small time derivatives, the associated field $\vec{E}(\vec{r})$ is

$$\vec{E}(\vec{r}) = -\nabla\Phi(\vec{r}). \tag{3.16}$$

The intracellular currents $\vec{J}_p(\vec{r})$ inside G give rise to an electric field in the extracellular space, which in turn, results in currents that flow passively through the conducting medium. The passive current density $\vec{J}_v(\vec{r})$ is given by Ohm's law,

$$\vec{J}_v(\vec{r}) = \sigma(\vec{r})\vec{E}(\vec{r}). \tag{3.17}$$

Thus, the total current density $\vec{J}(\vec{r})$ inside the head is the sum of the intracellular and the extracellular current density. Together with equations 3.16 and 3.17, $\vec{J}(\vec{r})$ can be written as

$$\vec{J}(\vec{r}) = \vec{J}_p(\vec{r}) + \vec{J}_v(\vec{r}) = \vec{J}_p(\vec{r}) - \sigma(\vec{r})\nabla\Phi(\vec{r}). \tag{3.18}$$

Using 3.18, then Maxwell's equation 3.15 becomes for the quasistatic approximation

$$\nabla \times \vec{B} = \mu_0 \vec{J}(\vec{r}). \tag{3.19}$$

A solution to 3.19 that obeys Maxwell's third equation (3.13), and under the condition that $\vec{B}$ vanishes at infinity, is given by the Ampère-Laplace law:

$$\vec{B}(\vec{r}) = \frac{\mu_0}{4\pi} \int \frac{\vec{J}(\vec{r}') \times \vec{R}}{R^3}\, dv' = \frac{\mu_0}{4\pi} \int \left(\vec{J}_p(\vec{r}) - \sigma \nabla' \Phi \right) \times \frac{\vec{R}}{R^3}\, dv', \tag{3.20}$$

where $\vec{R} = \vec{r} - \vec{r}'$ is the distance between the observation point $\vec{r}$ and the source point $\vec{r}'$. Using the mathematical identities, the magnetic field becomes for a current of finite spatial extent

$$\vec{B}(\vec{r}) = \frac{\mu_0}{4\pi} \int \left(\vec{J}_p(\vec{r}) + \Phi \nabla' \sigma \right) \times \frac{\vec{R}}{R^3}\, dv'. \tag{3.21}$$

In 3.21, both current densities contribute to the magnetic field. The passive return current density is represented by $\Phi \nabla' \sigma$. If the conductivity inside G is homogenous, then $\nabla \sigma = 0$. Thus, the extracellular current density does not contribute to the extracranial field.
The potential Φ of equation 3.21 can be replaced by its relation to 3.18 together with $\nabla \vec{J}(\vec{r}) = 0$, which is a consequence of 3.15. One obtains

$$\nabla \cdot (\sigma \nabla \Phi) = \nabla \cdot \vec{J}_p(\vec{r}), \tag{3.22}$$

Outside the head the conductivity is zero. Together with the continuity of the current components, which makes the current normal to ∇G inside G to zero, the boundary condition is

$$\hat{n}_{\nabla G} \nabla \phi = 0 \qquad \text{on } \nabla G. \tag{3.23}$$

The normal on the scalp ∇G is represented by $\hat{n}_{\nabla G}$.

The potential can be calculated from 3.22 together with the boundary condition given in 3.23. Knowing the potential, the calculation of the magnetic field based on 3.21 is straightforward.

If the volume element, where $\vec{J}_p(\vec{r}) \neq 0$, is small it can be replaced by

$$\vec{J}_p(\vec{r}) \approx I\ \vec{l}\ \delta(\vec{r} - \vec{r}_Q) = \vec{Q}\ \delta(\vec{r} - \vec{r}_Q). \tag{3.24}$$

In 3.24, $\vec{r}_Q$ is the position where the current density is different from zero, I is the current strength, $\vec{l}$ the direction vector of the current density, and $\vec{Q}$ is the current dipole moment,

$$\vec{Q} = \int dr'\ \vec{J}_p(\vec{r}'). \tag{3.25}$$

As illustrated in Figure 3.6, the contribution of a primary current described by the current dipole $\vec{Q}$ at position $\vec{r}_Q$ can be derived from equation 3.21 to

$$\vec{B}_p(\vec{r}) = \frac{\mu_0}{4\pi}\ \frac{\vec{Q} \times (\vec{r} - \vec{r}_Q)}{|\vec{r} - \vec{r}_Q|^3}. \tag{3.26}$$

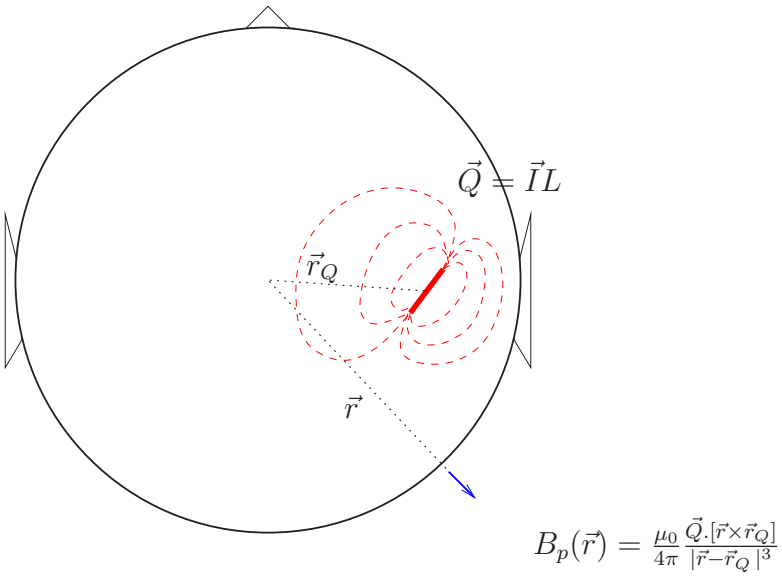

Figure 3.6: *Schematic view of the magnetic field $\vec{B}$ generated by a source Q inside the head.*

In summary, the potential and the magnetic field at any position outside and on the surface of the head can be calculated from the intracellular source density $\vec{J}_p(\vec{r})$ and depend linearly upon it:

$$\phi(\vec{r}) = \int d\vec{r}' \; \vec{\mathcal{L}}^{\text{pot}}(\vec{r}, \vec{r}')\vec{J}_p(\vec{r}'), \tag{3.27}$$

$$\vec{B}_\alpha(\vec{r}) = \int d\vec{r}' \; \vec{\mathcal{L}}_\alpha(\vec{r}, \vec{r}')\vec{J}_p(\vec{r}'), \tag{3.28}$$

where $\vec{\mathcal{L}}$ is the lead field vector depending on the geometry and conductivity of the head only.

3.3.4 The Inverse Problem

Calculating the potentials and magnetic field distribution that result from the intracellular currents is known as the forward problem. Conversely, the electromagnetic inverse problem is to estimate the cerebral current sources underlying an externally measured magnetic field (MEG) or potential (EEG). As shown by Helmholtz (1853), the inverse problem poses a very difficult mathematical problem without a unique solution. There can be primary current distributions that are either magnetically silent ($\vec{B} = 0$ outside G), electrically silent ($\vec{E} = 0$ outside G), or both. For example, a radial dipole in a spherically symmetric conductor produces an electric field, but is magnetically silent outside G. The case is illustrated in Figure 3.6, when $\vec{r} \times \vec{r}_Q$ is zero. Thus, a priori information needs to be incorporated into the analysis to make some assumptions regarding the location or the geometry of the source to derive a reliable solution.
The activated neurons are most commonly described in terms of a "single-dipole-in-a-sphere" model. That is a single equivalent dipole, located in the center of a spherically symmetric homogenous medium. The solution of the inverse problem can be obtained with different methods: The Principal Component Analysis (PCA) calculates the autocorrelation function of the sampled signal matrix. Eigenvectors and eigenvalues of this matrix define the topographies, that contribute to the measured signal. Another method is to use the MUSIC-algorithm (multiple signal classification). One assumes, that the sources and the noise are orthogonal to each other and that the number of the sources is known. For a comparison of the advantages and disadvantages of these methods, see Sieroka (2004).
The brain activity can also be modelled with a distributed source model, in which

several current dipoles are densely spaced along the brain's gray matter. A problem occurs when the number of dipoles exceeds the number of sensors that record the field distribution. One has to handle an underdetermined problem and needs to place constraints in order find a solution. One way to overcome the problem is given with the "minimum norm solution", a solution for which the norm of the activities of all sources exhibits its minimum (e.g. Hoechstetter (2001)).

The auditory evoked fields (AEFs), that are investigated in the Chapters 4 and 5 have the immense advantage, that the pyramidal cell dendrites of the auditory cortex (the putative generators of the electrical sources) are arranged in a columnar fashion with an orientation normal to the cortical surface. Assuming 20,000 pyramidal cells/mm^2 on the cortical surface and an evoked amplitude of about 20 nAm; the center of activity covers an area of approximately 0.5 cm^2 (Eggermont and Ponton, 2002).
fMRI studies conducted by Griffiths et al. (1998) and Warren et al. (2003) revealed, that pitch changes in a continuous sound (like melody) revealed increased activity in a very small area of the lateral Heschl's Gyrus only. Thus, the center of activity due to the change of pitch can be adequate modelled with one dipole in each auditory cortex of both hemispheres.

Chapter 4

Pitch and Neuromagnetic Representation of RIS

The aim of this Chapter is to correlate the neuromagnetic responses evoked by RIS with psychoacoustic results. The perceived pitch shift between RIS generated with positive and negative sign of the gain g finds particular attention, since the results on the perceived pitch of RIS generated with a negative gain are rather controversial. The pitch shift from ambiguous pitches in the region of $\frac{1}{0.9d}$ and $\frac{1}{1.1d}$ to the pitch of $\frac{1}{2d}$ is investigated in view of the controversial literature as described in section 2.1.3.
The psychoacoustic results, the neurophysiological responses and the perception of RIS are investigated not only in dependence of g, but also of the delay time d and the number of iterations n. We also included $n = 4,096$ iterations in the experiments, where the power spectrum of RIS approaches a line spectrum.
A continuous stimulation is used to extract the neuromagnetic pitch onset response (POR). In contrast to earlier studies (Pantev et al., 1988; Roberts and Poeppel, 1996), we concatenated segments of RIS with fixed n and d, but alternating sign of g to avoid responses evoked by the energy onset. In the psychoacoustic experiment, listeners performed pitch matching according to methods reported by Yost (1996a). Second, we circumvented their reported difficulty of pitch matching by using a simple two-alternative forced choice task. In the simulation part, the temporal pitch model introduced in section 2.3 is employed to simulate the perceived pitch of RIS. The outcome of the stabilized auditory images is used to predict the perceived octave shift and the transition in the perception to ambiguous pitches in the region of $\pm 10\%$ between RIS generated with positive and negative gain.

For convenience, we use the notation RIS(d, g, n) for RIS generated with a delay time d, a gain factor g of either plus or minus one and a number of iterations n.

4.1 Material and Methods

4.1.1 Subjects

Twenty adult listeners (ten male, ten female) with no reported history of peripheral or central hearing disorder participated in both experiments after giving informed consent. The mean age ($\pm$ standard deviation) was 33 ($\pm$ 9) years. During the MEG sessions, subjects watched a silent movie of their own choice. They were asked not to pay attention to the stimuli and concentrate on the movie.

4.1.2 Stimuli for MEG Recordings

Digitally generated white noise at a sampling rate of 48,000 Hz was used to produce RIS. The gain g was either set to plus or minus one. Pitch strength was varied by employing 2, 8 and 4,096 iterations (representing $n \rightarrow \infty$). According to Bilsen (1966) and Yost (1996a), delays of 2, 4, 8 and 16 ms (corresponding to 500, 250, 125, and 62.5 Hz for RIS(d,+1,n)) were used to cover a range of three octaves (oct). The single RIS segments had a length of 510 ms and were high-pass filtered at 500 Hz to compare our results to other studies (Krumbholz et al., 2003).
The RIS segments were balanced in energy and ramped with a 10-ms hanning window at the beginning and at the end. To avoid neuromagnetic responses to the energy onset, the segments were concatenated with alternating sign of g (+1/-1), but same d and n to produce a RIS sound with a length of 10.5 s. Thereby, the subsequent RIS segments of the continuous sounds overlapped in a way that there was no visible change of the sound pressure at the transition of two segments (Figure 4.1).

Sounds were presented diotically via a 16 bit soundcard at a sampling rate of 48,000 Hz. Etymotic Research (ER3) earphones equipped with 90 cm plastic tubes and foam ear pieces were used. The transfer function of the whole setup shows a band-pass characteristics with a bandwidth from approximately 500 to 5,000 Hz. A Brüel & Kjær sound level meter was used to set the overall level of the sounds to 63 dB SPL.
Neuromagnetic responses were recorded in four different sessions. In each session

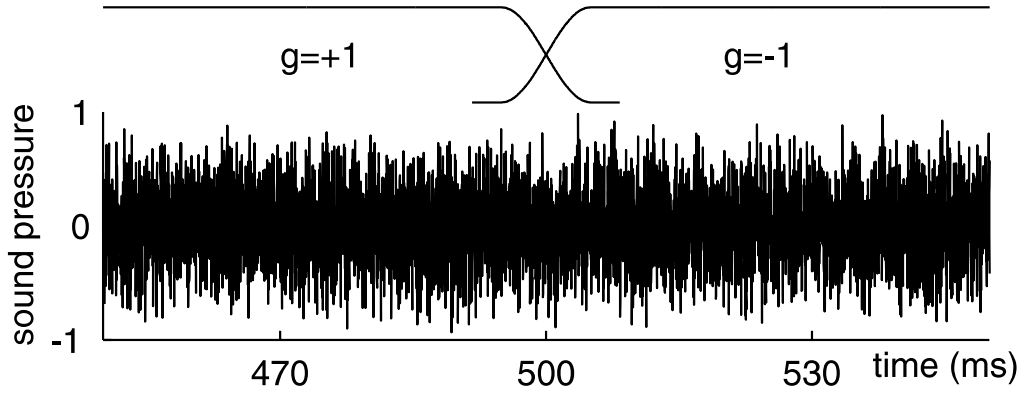

Figure 4.1: *The individual RIS-segments are concatenated with alternating gain. At the transition of the segments, energy is kept constant and there are no visible differences in the fine structure of the waveform.*

d was kept constant while n was varied randomly between the different stimuli. In this way, ten different RIS segments were generated for each condition. The stimuli were delivered in pseudo-random order. Altogether, 350 RIS segments of each condition were presented within one session.

4.1.3 Data Processing and Source Analysis

The neuromagnetic responses to the presented sounds were DC-recorded with a Neuromag-122™ whole head planar gradiometer system inside a magnetically shielded room (Imedco, Switzerland). Data were sampled with a rate of 1,000 Hz and lowpass filtered at 330 Hz. To co-registrate the 3D-MRI reconstructions, the landmarks of nasion, two pre-auricular- and additional 32 surface points were digitized with a Polhemus 3D Space Isotrack2 system.
The software package BESA2000 (MEGIS GmbH, Germany) was used to perform the source analysis offline. Before averaging the auditory evoked fields, data was inspected and noisy channels were excluded. Epochs with gradient changes greater than 800 fT per millisecond were rejected. In this way, approximately 330 sweeps per subject and condition were averaged over a period of -100 ms to 600 ms relative to the transition of the single RIS segments.
A spatio-temporal source model with one equivalent dipole in each hemisphere was used (Scherg et al., 1989). The averaged waveforms of RIS(2,+1,8) were 2–30 Hz band-pass filtered (zero-phase, 12 dB/oct), and the first prominent negative deflection (N100m') of this condition was fitted in a 30 ms-interval around that peak. The evoked responses of this condition produced a consistent fit for every

single subject and was used as a spatial filter to derive the equivalent source waveforms of all other conditions. No further constraints concerning dipole location, orientation or symmetry-conditions were applied. A principle component analysis (PCA) over an interval of 400–500 ms after the transition of the unfiltered auditory evoked responses was computed (Berg and Scherg, 1994) to compensate drifts and other low frequency artifacts due to the continuous stimulation. The PCA component that accounted for the largest variance in this interval was included in the spatio-temporal filter for the respective condition. The applied method of using fixed dipoles derived from one condition causes source strength differences. In order to determine these differences, we also fitted the source waveforms in response to the strong pitch of RIS generated with 4,096 iterations on every single condition. These averaged responses allowed to fit any condition for all subjects in the same way as described for the responses to RIS(2,+1,8). Furthermore, the source model defined with the responses to RIS(2,+1,4096) was used to derive the equivalent source waveforms in response to any other RIS condition generated with 4,096 iterations. For calculating the grand averaged and significances, all resulting equivalent source waveforms were low-pass filtered (zero-phase, 12 dB/oct) at 100 Hz. The period of 100 ms before the transition was used to define the baseline. Most individual source waveforms, especially those evoked by RIS conditions with 2 iterations and delay times of 8 and 16 ms, did not allow to derive latency and amplitude reliably. Thus, we applied a bootstrap resampling procedure for statistical evaluation. Latencies and amplitudes of the N100m' in response to all RIS conditions were determined for each of the 1,000 independent bootstrap resamples generated by sampling 20 waveforms with replacement from the original waveforms. Peak latency and amplitude was assessed by deriving the minimum and the corresponding value from the mean of the resamples. The critical t-intervals of latency and amplitude were computed using the resulting distribution of the single minima and the corresponding amplitude of each resample (Efron and Tibshirani, 1993). Latency differences of the N100m' peak were considered significant when the t-intervals of two conditions did not overlap. Additionally, we derived for the source waveforms evoked by RIS(d,g,4096) the latencies and amplitudes of N100m' component for each single subject.

T1-weighted magnetic resonance images (MRI) of the individual anatomical structure of 19 subjects were co-registrated to describe the location of the fitted dipoles. The voxel size was set to 1 mm. MR data was acquired using a Siemens Symphony 1.5 T-scanner. The resulting locations of the fitted dipoles were rescaled[1] accord-

[1] www.mrc.cbu.cam.ac.uk/Imaging/Common/mnispace.html

ing to the stereotactic space of Talairach and Tournoux (1988) and projected onto the probabalistic map provided by Schneider et al. (2004).

4.1.4 Psychoacoustics

The evoked pitch of RIS that is generated with a negative gain factor, was examined in different psychoacoustic studies, but with controversy results. Yost et al. (1978) reported for RIS generated with a negative gain factor ($g = -1$) and $n > 4$ a pitch shift of an octave below the corresponding RIS with $g = +1$. For RIS generated with less than four iterations, they reported the perceived pitch not in the expected region of $f = \frac{1}{2d}$, but around $\frac{1}{0.9d}$ and $\frac{1}{1.1d}$, independent of the delay time d.
Raatgever and Bakkum (1986) reported different results for their pitch matching experiment, using an infinite number of iterations. The change in the perception of RIS(d,-1,∞) depended on the delay d. Pitch matches at $\frac{1}{2d}$ only occurred for delay times of less than 6 ms, but with increasing delay, pitch matched around $\frac{1}{0.9d}$ and $\frac{1}{1.1d}$ compared to RIS(d,+1,∞) (see Figure 2.2). Yost (1996a) pointed out, that the task of pitch matching is very difficult and often subjects could not perform the task. For example, in his study published in 1996 four of six musicians were excluded. They were not able to rate the pitch of RIS generated with a delay time of 4 ms and a gain factor of $g = +1$ with that of a pulse train within $\pm 20\%$ accuracy. Thus, we also applied a paired comparison task to circumvent the reported problems.

Pitch Matching

In view of the existing controversy on the pitch of RIS generated with a negative gain factor, we first tried to reproduce the conditions of the study published by Yost (1996a). As in MEG, RIS were produced with digitally generated white noise at a sampling rate of 48,000 Hz. The stimuli had a length of 500 ms and were ramped at the beginning and at the end with a 10-ms hanning window. The unfiltered sounds were presented diotically via Sennheiser headphones at an overall level of 63–65 dB SPL. According to Bilsen (1966) and Yost (1996a), delays of 2, 4, 8 and 16 ms (corresponding to 500, 250, 125, and 62.5 Hz for RIS(d,+1,n)) were used. Pitch strength was varied using 1, 2, 3, 4, 6 and 8 iterations. To determine the pitch of RIS generated with negative gain, two sounds with an inter-stimulus interval of 250 ms were presented to the listener.

During one session, the delay time of the test RIS was fixed, while the delay of the matching RIS was changed randomly in an interval of an octave. The test sound was presented with fixed parameters, while the second sound was of variable pitch, adjusted by the subject. The test RIS was always generated with a gain factor of minus one, the second RIS was produced with a gain factor of plus one but with the same number of iterations. To find the matched pitch of RIS(d,-1,n), the listener could adjust the pitch of the second RIS generated with a positive gain factor by changing the delay either by small (1%), medium (3%) or large (9%) steps. Furthermore, the subject could listen to the two RIS sounds again, without any pitch change of the second sound. Once, the listener was satisfied with the matched pitch, octave changes of the matching RIS were also possible to make sure that the pitch match was in the same octave. The number of iterations (n) was changed randomly to a new value after each pitch match was determined. During one session in which the delay of the test RIS was fixed, listeners had to adjust each number of iteration ten times. The listeners were not told which conditions were being presented. Altogether, listeners had to adjust 60 pitch matches during one session. The subjects had to repeat the four sessions, leading to a total of 480 pitch matches.

Paired Comparison

Psychoacoustic measurements of the paired comparison task were carried out using the same audio-equipment and conditions as in the MEG experiment. Stimuli were generated with delay times of 2, 4, 8 and 16 ms, the gain g was plus or minus one and the number of iterations n was 2, 8 and 4,096. Sounds had a length of 500 ms and were also high-pass filtered at 500 Hz.
In each pitch-matching task, d and g were varied, while n was fixed. Therefore, comparisons across different numbers of iterations were not carried out. A two-alternative forced choice task for paired comparisons was applied. Two different RIS with fixed n, but varying d and g were presented with an inter-stimulus interval of 500 ms in random order. After repetition subjects had to judge which tone had the higher pitch before the next pair was presented. During one session, every combination of paired comparisons was presented twice with reversed order. Listeners judged all possible combinations of RIS, excluding the comparison of RIS to themselves, which led to a total of $2\binom{n}{2} = 56$ trials per session. Subjects had to repeat the session once. Thus, every pair of RIS was judged four times by each subject.
A scale for the relative 'height of pitch' was derived from the results of the paired comparison experiment, using the Bradley-Terry-Luce (BTL) method (David,

1988). According to the description of the algorithm (see section 3.2), the pitch of the RIS stimuli can be ordered according to a linear scale.
Additionally, the same experiment was conducted with completely unfiltered RIS, to compare our psychoacoustic data with the results published by Yost (1996a).

4.2 Results

4.2.1 Source Analysis

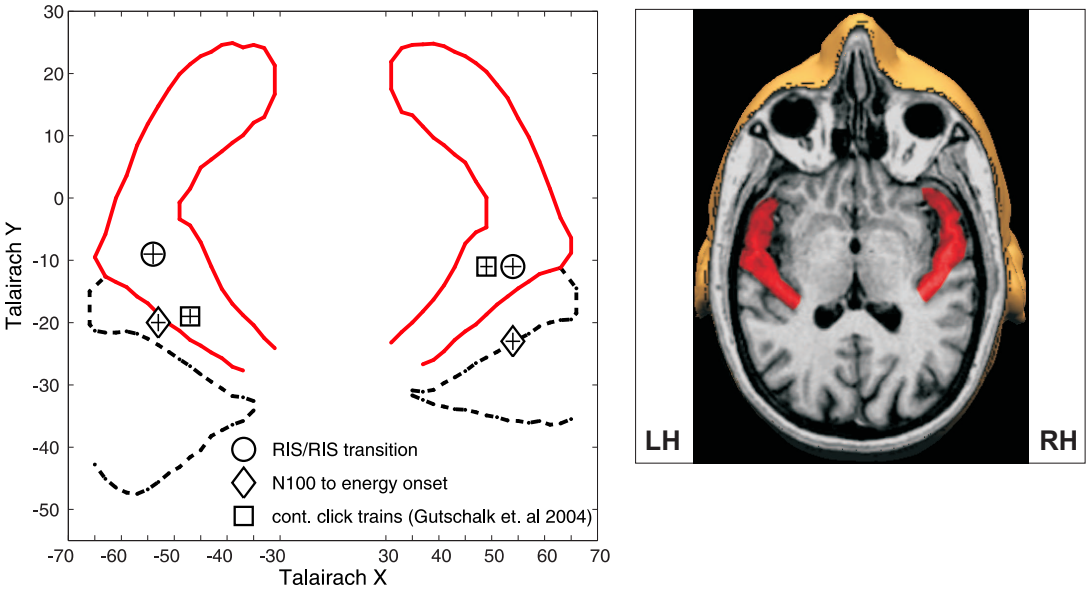

Figure 4.2: *Localization of the mean coordinates of the averaged equivalent dipoles projected onto an axial plain (left). The coordinates of the sulcal borders were provided by Schneider et al. (2004). Right: MRI of a single subject with highlighted Heschl's Gyrus. The fitted dipoles of the isolated POR (circles) are located in the left and right lateral part of Heschl's Gyrus. The diamonds show the onset of RIS from silence (N100). For comparison, the mean location of the fitted dipoles in response to continuous click trains with varying pitch strength (Gutschalk et al., 2004) are also depicted with squares. The vertical and horizontal bars represent the standard error of mean.*

The neuromagnetic responses evoked by RIS(2,+1,8) produced a good signal-to-noise ratio with a consistent fit for each subject. Therefore, this condition was always used for fitting a two-dipole model with one equivalent dipole in each hemisphere. In both hemispheres, the averaged coordinates of the equivalent source waveforms were located in the lateral aspect of the primary auditory cortex (left:

x=−54(±8), y=−9(±7), z=12(±11) and right: 54(±7), −11(±6), 14(±8), brackets indicate the standard deviation). The projection of the averaged location of 19 subjects onto an axial plane through Heschl's Gyrus and Planum temporale is shown with circles in Figure 4.2. Additionally, we derived the location of the N100m component evoked by the onset of the 10.5 s lasting sound by fitting the onset of the RIS(2,+1,8) condition. The diamonds of Figure 4.2 indicate the position of the fitted dipoles. The center of activity was found to be more posterior compared to the fit on the N100m' component (left: x=−53(±6), y=−20(±6), z=8(±9) and right: 54(±6), −23(±6), 10(±10))

4.2.2 Neuromagnetic Responses to the Change of Pitch

The grand-average of the source waveforms for all 20 subjects in response to the transition of concatenated RIS-segments with alternating gain g, but fixed d and n is presented in Figure 4.3. The comparison of the morphology of the left (black) and right (grey) hemisphere revealed no differences in latency and amplitude. Therefore, the responses of the right- and left-hemisphere were pooled to calculate significances with the bootstrap method. The morphology of the AEF is similar for all conditions. The responses to the transition of RIS with alternating g are characterized by a first peak at about 60 ms, followed by the N100m' deflection evoked by the pitch onset. The third evoked neuromagnetic response with a latency of 200 ms to 250 ms (P2) differed between conditions in both, amplitude and latency.

Pitch Onset Response

The mean amplitude of the POR is shown in Figure 4.4. A decrease of the absolute value of the evoked source waveforms was observed when the delay time d was increased while n was held constant, especially for RIS generated with two and eight iterations. For constant values of the delay time d, the amplitude of the POR increased with higher number of iterations n of the presented RIS. Especially the responses to RIS generated with 4,096 iterations were augmented compared to RIS consisting of two and eight iterations only. The latency of the N100m' in response to RIS with fixed n increased significantly when the delay time d was increased. A summary of all POR latencies and the corresponding bootstrap assessed t-intervals is given in Table 4.1. The mean N100m'-latencies of RIS generated with the same number of iterations n but opposite sign of g are presented in Figure 4.5 (left) for direct comparison.

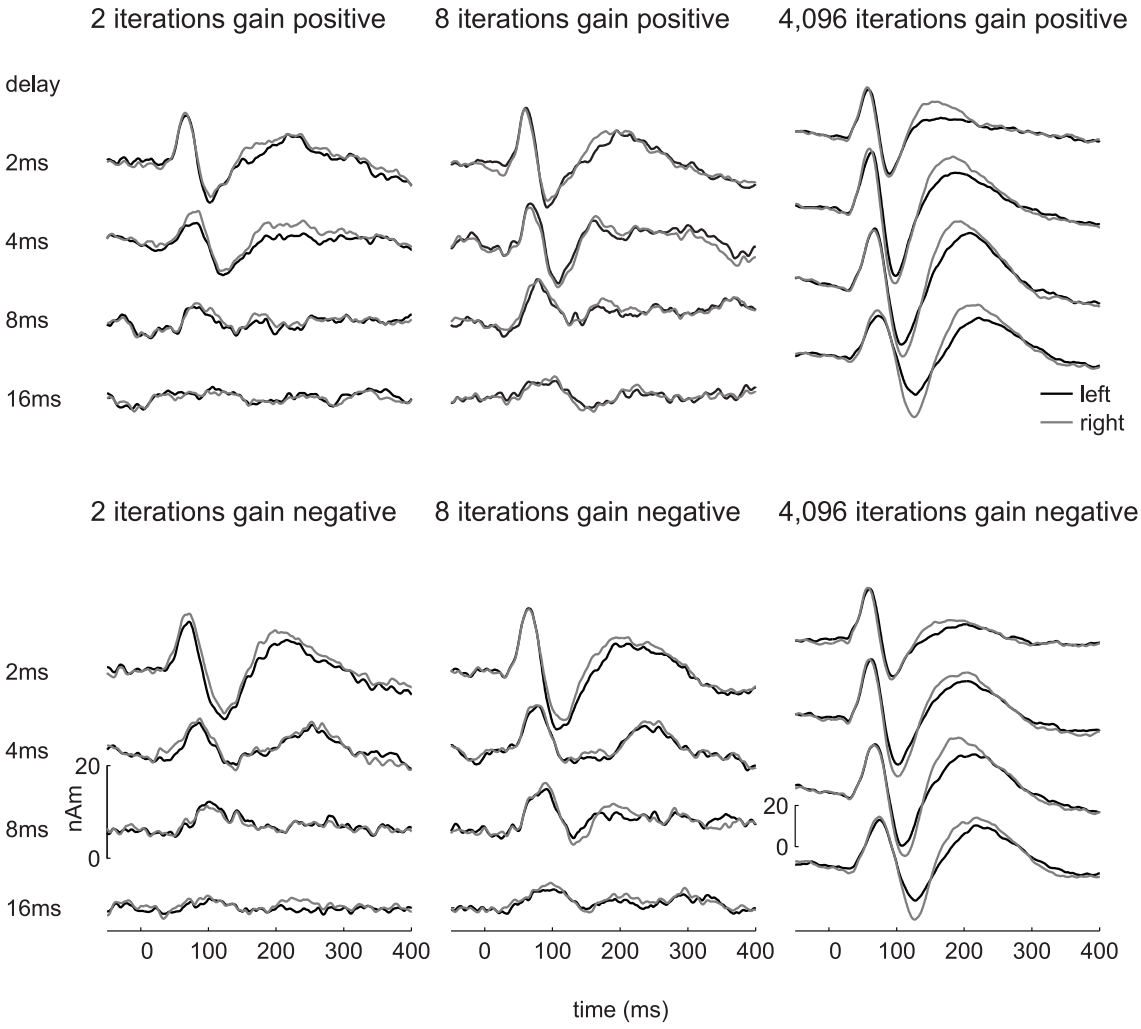

Figure 4.3: *Grand-average source waveforms of 20 subjects based on the fit of the POR evoked by RIS-segments. All waveforms of the left (black) and right hemisphere (grey) are zero-phase filtered from 1–100 Hz. Independent of gain (top/bottom) and the number of iterations (left to right: 2, 8 and 4,096), the N100m' latency was augmented significantly with increasing delay. The amplitude of N100m' decreases with increasing delay. Note the different scaling in the plot with 4,096 iterations.*

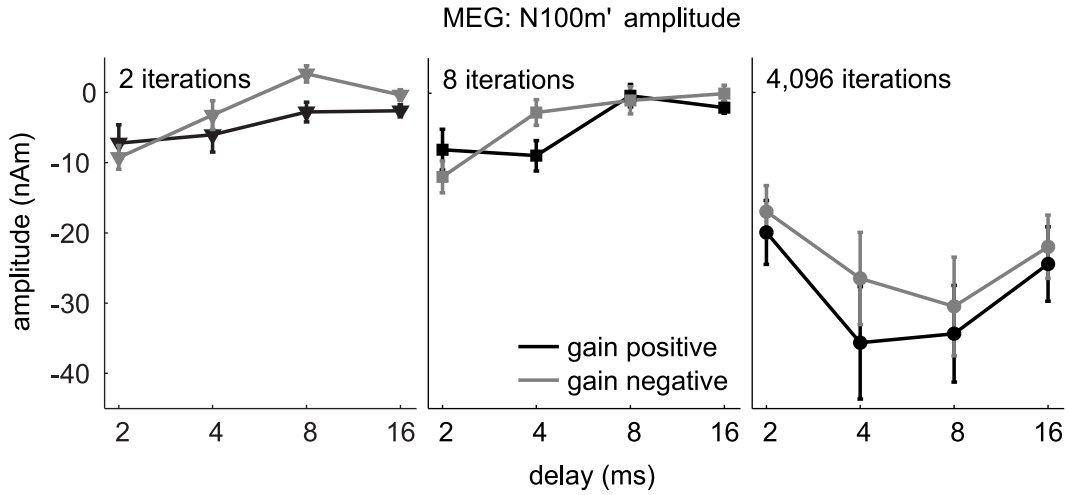

Figure 4.4: *Mean N100m' amplitude of the equivalent source waveforms in response to RIS generated with two, eight and 4,096 iterations (left to right). Error bars represent the bootstrap assessed standard error over the subjects. The amplitude evoked by RIS generated with 2 and 8 iterations did not change significantly, when either the sign of the gain factor (black/grey) was changed or the delay time d was varied. According to the bootstrap assessed* t*-intervals, the amplitude of the POR evoked by RIS generated with 4,096 iterations was increased significantly compared to the responses evoked by RIS generated with 2 and 8 iterations.*

Table 4.1: *Latency (in ms) of the N100m' component evoked by the transition of all applied RIS-segments. Critical t-intervals (20 subjects; t=1.7291) were bootstrap computed to assess significance.*

iterations	delay (ms)/gain							
	2/+1	2/–1	4/+1	4/–1	8/+1	8/–1	16/+1	16/–1
2 it	106	126	124	140	143	137	178	163
$\pm t$-int	±8	±4	±6	±8	±6	±11	±21	±17
8 it	97	115	109	141	134	136	155	150
$\pm t$-int	±5	±5	±3	±27	±15	±7	±16	±8
4,096 it	92	96	99	104	111	112	128	128
$\pm t$-int	±3	±3	±2	±3	±3	±3	±4	±4

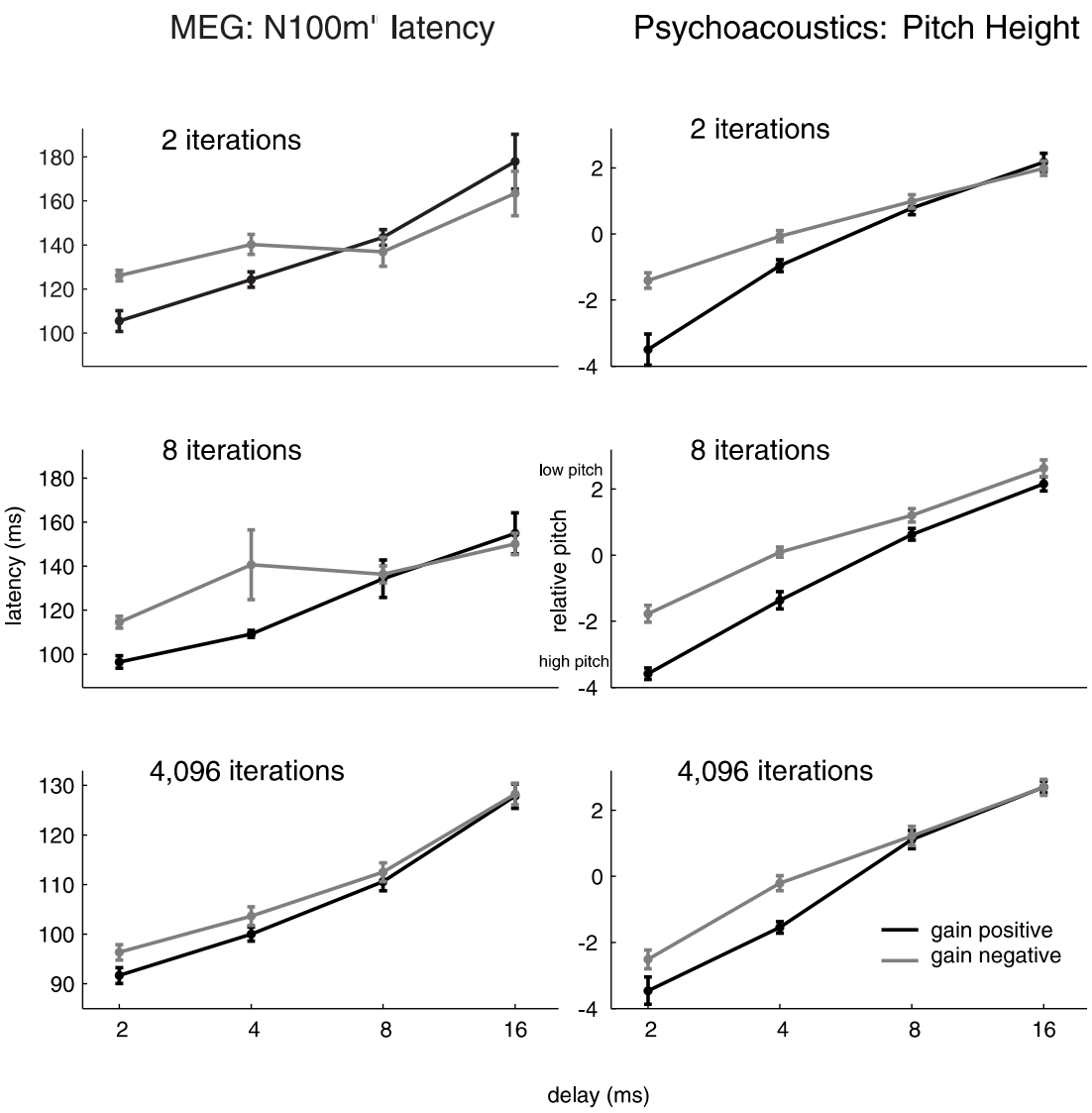

Figure 4.5: *Left: N100m' latency of the bandpass filtered source waveforms in response to RIS. Right: Relative pitch of the psychoacoustic judgements to the same RIS stimuli using the BTL method. Error bars indicate the bootstrap assessed standard error. The latencies of the N100m' component evoked by RIS are significantly increased for delay times of 2 and 4 ms when the gain factor is changed from positive to negative. In contrast, the peak latency of the POR did not differ significantly when the delay time of the presented RIS was set to 8 and 16 ms. Analogously, the relative pitch scale of the paired comparison task revealed significant differences between the pitch of RIS generated with 2 and 4 ms and opposite gain. For delay times of 8 and 16 ms, the pitch difference was found to be within errors. Note that the relative BTL-based pitch scales cannot be compared directly, since RIS-conditions with different number of iterations n were not tested within the trials of the same session.*

The POR evoked by RIS generated with a positive gain , revealed for all n a nearly linear increase in the latency of the d-logarithmic scaled diagrams. The comparison with the corresponding POR latencies evoked by RIS generated with negative gain exhibited a significantly later N100m' for delay times of 2 and 4 ms, independent of n. In contrast, the latency differences for delay times of 8 ms and 16 ms did not differ significantly. The good signal-to-noise ratio of RIS(d,g,4096) allowed to detect the N100m' peak in single subjects. The results were equivalent to the bootstrap calculated t-intervals. The latency difference between RIS generated with a positive and a negative gain factor was significant ($F_{1,19} = 11.09, p < 0.01$) for 2 and 4 ms delay times, whereas no increase in the latency of N100m' was observed for delay times of 8 and 16 ms ($F_{1,19} = 0.74, n.s.$). The latency of the POR in response to RIS(2,-1,n) has the same value compared to revealed latency of RIS(4,+1,n). This indicates also the perceived octave shift between RIS(2,+1,n) and RIS(2,-1,n). Similarily, the same latency of the N100m' component was evoked RIS(4,-1,n) and RIS(8,+1,n). However, the effect was not observed anymore for RIS(8,-1,n) compared to RIS(16,+1,n).

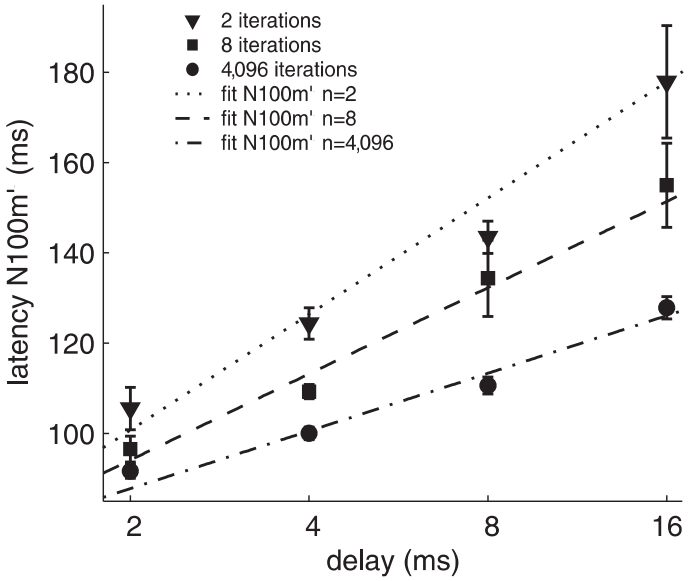

Figure 4.6: *Latency of pitch responses evoked by RIS with positive gain in dependence of the different delay times d. The regression lines represent the observed latencies evoked by RIS with different number of iterations n. The dashed lines correspond with the fit formula described in the text. The latency of the pitch-specific N100m' component depends on both, the delay time d and the number of iterations n of the presented RIS.*

The dependence of the evoked N100m' latency on the parameters of the presented RIS was investigated with the neuromagnetic responses evoked by RIS generated with positive gain. Figure 4.6 illustrates the latency of the POR in dependence of the delay time d. The slope of the POR-latency decreased with an increase of the number of iterations, respectively an increase of the pitch strength (top to bottom) from 24 ms/oct ($n = 2$) and 19 ms/oct ($n = 8$) to 12 ms/oct ($n = 4,096$).
Thus, the latency of the N100m' depended on both, pitch and pitch strength. The latencies of the POR in response to RIS generated with positive gain factor g can be fitted using the formula

$$lat_{N100m'}(d, n) = 18\,\text{ms} \cdot \frac{\sqrt{n} + \frac{3}{2}}{\sqrt{n}} \cdot \ln d + 75\,\text{ms}$$

where d is the delay in milliseconds and n the number of iterations. The N100m' latencies in response to RIS generated with positive gain are shown in Figure 4.6 together with the results of the fit-formula ($lat_{N100m'}$) for a fixed number of iterations (2, 8 and 4,096).

The applied spatio-temporal source model with one equivalent dipole in each hemisphere did not result in a stable fit for each subject and every single RIS condition. The neuromagnetic responses of RIS(2,+1,8) revealed a high SNR and produced a stable fit in each subject. Therefore, we modelled the equivalent dipoles for every subject to the responses of RIS(2,+1,8) and derived the equivalent source waveforms to the transition of all other RIS conditions with the fixed dipole model, since the exact location of the two dipoles does not affect the latency of the POR, but influences only the amplitude of the equivalent source waveforms. The error of the source strength is shown exemplarily for the RIS conditions generated with 4,096 iterations. The Talairach coordinates of the fitted dipoles in response to RIS(2,+1,4096) are also in the lateral part of Heschl's Gyrus (left: x=$-52(\pm1)$, y=$-12(\pm2)$, z=$5(\pm3)$ and right: $52(\pm1)$, $-11(\pm2)$, $8(\pm3)$). The difference of the source strength in dependence of the two analysis methods is shown in Figure 4.7 for the condition RIS(16,-1,4096). The mean of the equivalent waveforms over 20 subjects is derived either by modelling the responses of the condition RIS(2,+1,4096) (black line) or by fitting the source waveforms on the condition RIS(16,-1,4096) (dotted). The shadowed area represents the 90%-confidence interval over the subjects. The influence on the source strength of the POR is given for all RIS conditions generated with 4,096 iterations in Table 4.2. The values represent the ratio between the source strength of the POR assessed with the fit on the individual condition and the fit on RIS(2,+1,4096).

Table 4.2: *RIS generated with 4,096 iterations. Influence on the source strength of the POR when the dipole model is either fitted on the individual condition or held fixed on the fitted condition of RIS(2,+1,4096). The given values correspond to the ratio of the source strength assessed with the two different methods.*

RIS(2,+1,4096)	1	$\pm s.e.$
RIS(2,-1,4096)	0.99	±0.16
RIS(4,+1,4096)	0.80	±0.14
RIS(4,-1,4096)	0.88	±0.15
RIS(8,+1,4096)	0.83	±0.16
RIS(8,-1,4096)	0.88	±0.18
RIS(16,+1,4096)	0.90	±0.15
RIS(16,-1,4096)	0.90	±0.15

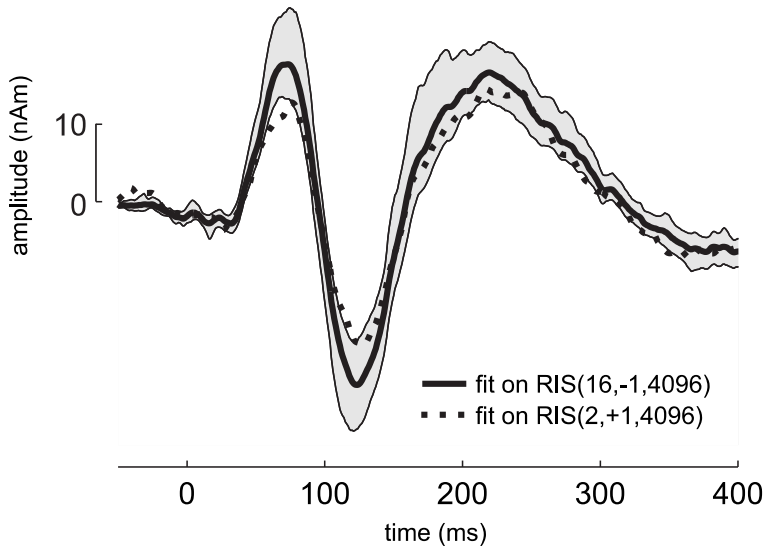

Figure 4.7: *Difference of the source strength waveforms due to the fitting method. Grand average of the equivalent source waveforms in response to RIS(16,-1,4096) when the model is fitted either on the individual condition (black line) or fixed (dotted line) on the fitted source waveforms in response to RIS(2,+1,4096). The area in grey defines the 90% confidence interval over 20 subjects. Note that the fixed spatio-temporal filter resulted in a smaller amplitude. However, the latencies are not affected by the different fitting methods.*

P50m–First Responses after the Transition

The first response that occurred after the transition between two RIS-segments varied in both, latency and amplitude. As shown in Figure 4.8, the latency of this component increased with an increase of the delay time of RIS for a fixed number of iterations. Furthermore, increasing the pitch strength of RIS caused a latency decrease. Compared to the N100m' component, there is no pitch-dependent latency difference between RIS generated with positive and negative gain: for RIS produced with two and eight iterations and a negative gain, the first positivity peaked significantly later for delay times of 2, 4, and 8 ms compared to RIS with positive gain. But for RIS(d,g,4096) there was no latency difference between RIS generated with opposite sign of the gain factor. And for delay times of 16 ms, the latency difference between RIS(16,+1,n) and RIS(16,-1,n) was within errors, independent of n. The peak-amplitude of this component (see Figure 4.9) decreased with increasing the delay time of the presented RIS, independent of the sign of the gain factor but only for two and eight iterations. The amplitude of the responses evoked by RIS generated with 4,096 iterations was completely independent on the delay time. The sign of the gain factor of RIS is totally irrelevant for the observed amplitude for all applied number of iterations.

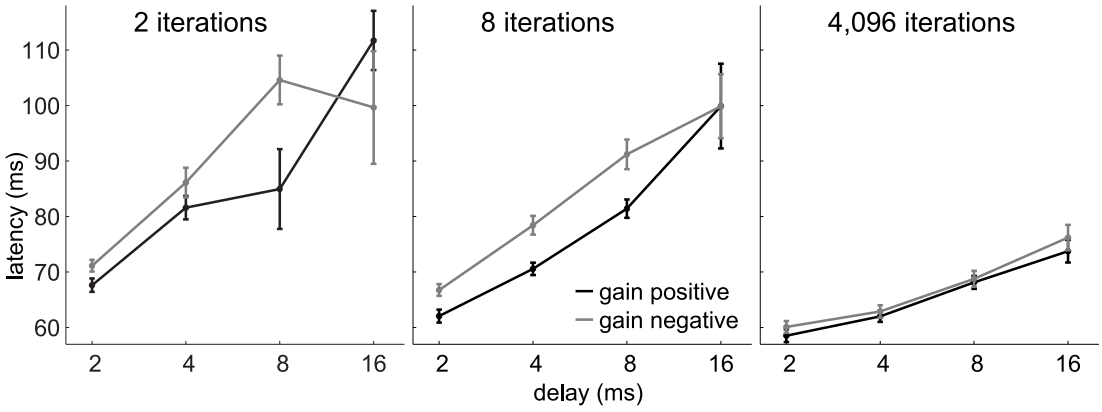

Figure 4.8: *P50m component evoked by the RIS-RIS transition: Latency (in ms) of the positivity for the different number of iterations (left to right). The latency of the P50m increased with the delay time d of the presented RIS.*

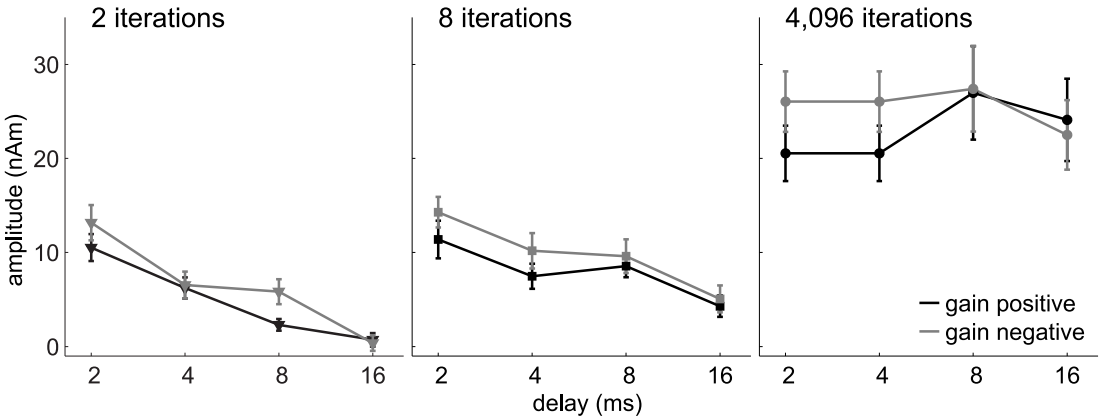

Figure 4.9: *P50m component evoked by the RIS-RIS transition. The amplitude of the positivity increased with an increase of the number of iterations n, independent of the gain factor of RIS. .*

4.2.3 Psychoacoustics

Pitch Matching

Altogether, five subjects which also took part in both, the paired comparison task and the MEG experiment, participated in the pitch matching experiment. The results of the five single subjects is illustrated in detail in Appendix B. As can be seen from the Appendix, two listeners were not able to match the pitch reliably and were excluded. Only the three musicians were able to perform the task and executed a total of 480 pitch matchings each. RIS(d,+1,n) was always used as the matching stimulus throughout the experiment. The test stimulus was generated with the same parameters but negative sign of the gain factor. In this way, the timbre of the matching RIS is similar to the test RIS. Figures 4.10 and 4.11 show the mean distribution of 1,920 pitch matches. From top to bottom, the number of iterations of the tested RIS increases from one to eight. The noisy timbre of RIS generated only with one or two iterations as compared to the more tonal quality of RIS(d,g,8) makes it more difficult to match the pitch reliably. As can be seen, the peaks of the distribution sharpen with increasing n.
The pitch of RIS(2,-1,n) was matched at 250 Hz, which is an octave below the pitch of RIS(2,+1,n), independent of the number of iterations. If n is set to one or two, octave like mistakes occurred at frequencies of 125 Hz and 62.5 Hz. Pitch matches at about ±10% (450 or 550 Hz) of the perceived pitch of RIS(d,+1,n) were not observed. RIS(4,-1,n) revealed the same results: octave like errors (at a frequency of 62.5 Hz and 31.25 Hz) were observed for the noisy test stimuli generated with up to three iterations. But again, no single subject rated the

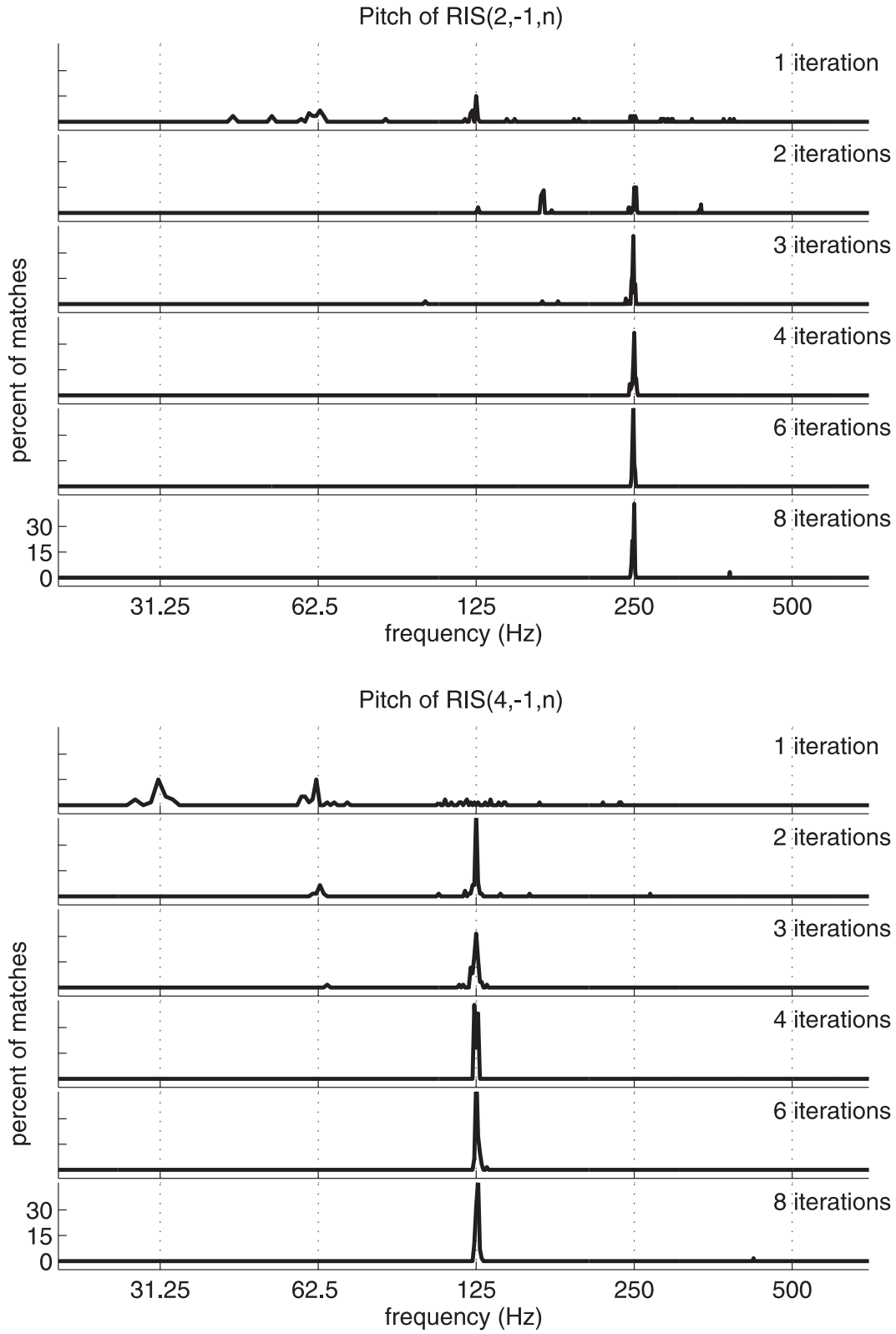

Figure 4.10: *Matched pitch of RIS generated with a delay of 2 ms (top) and 4 ms (bottom). The abscissa is log-scaled and represents the perceived pitch of RIS generated with a negative gain. Independent of the number of iterations, pitch was matched an octave below ($\frac{1}{2d}$) the corresponding pitch of RIS produced with the same parameters but opposite sign of g. Note that even octave like errors occurred for $n < 3$.*

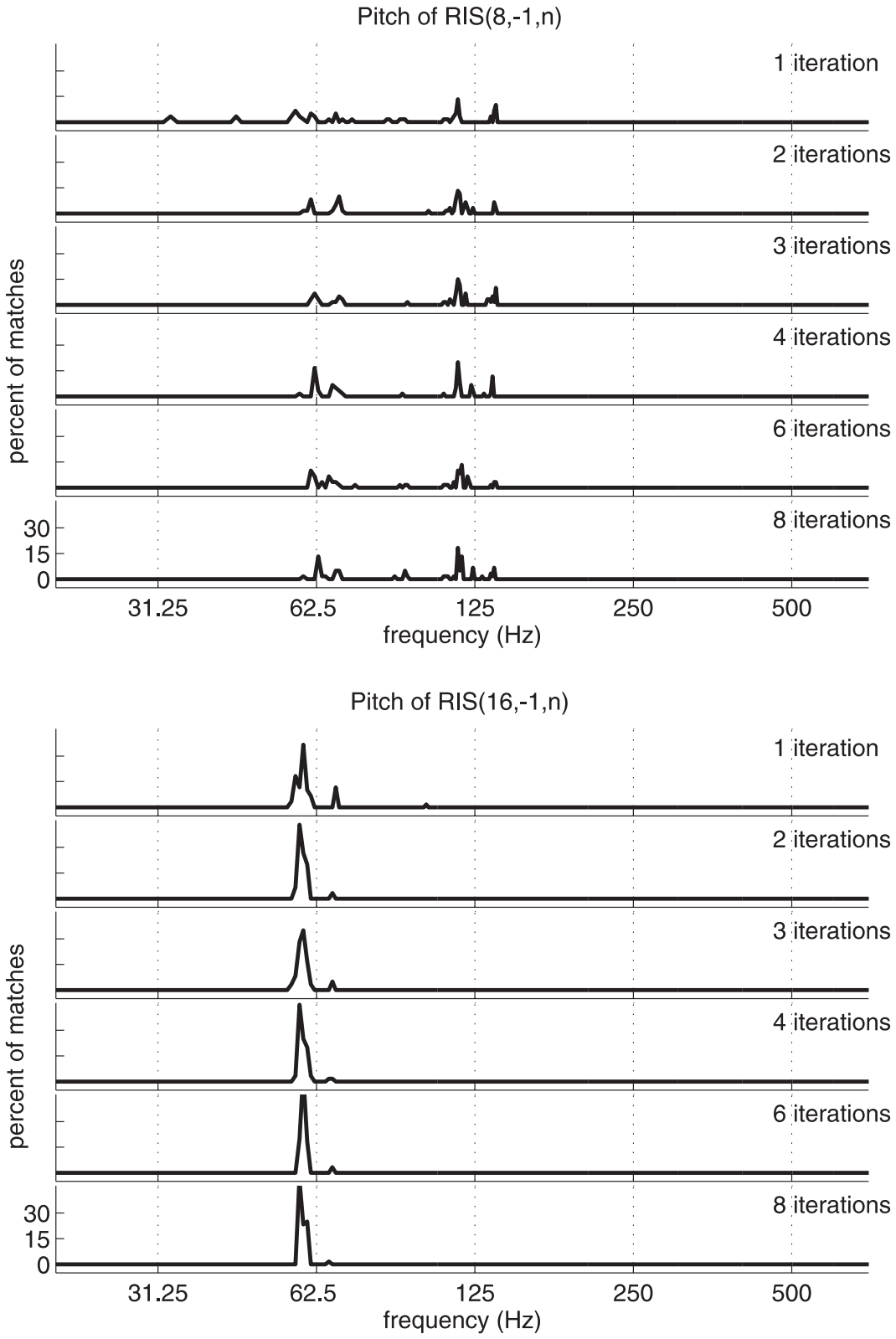

Figure 4.11: *Matched pitch of RIS generated with a delay of 8 ms (top) and 16 ms (bottom). With an increase of the delay time d, the perceived pitch of RIS generated with a negative gain was matched in the region of* $\frac{1}{0.9d}$ *and* $\frac{1}{1.1d}$*. The transition of the pitch shift to ambiguous pitch matchings was observed for delay times of 8 ms (upper), where still some octave matchings occurred.*

pitch of RIS(4,-1,n) around 225 Hz or 275 Hz, which would correspond to a 10% change of the sensation compared to RIS(4,+1,n). A change in the perception was found when the delay time was increased to 8 ms. Independent of n, subjects matched the pitch either an octave below (62.5 Hz) RIS(8,+1,n) or approximately at 137.5 Hz and at 112.5 Hz. A further increase of the delay time led to a main peak of the perceived pitch of RIS(16,-1,n) in a range of 55–60 Hz which differed to the pitch of RIS(16,+1,n) at about 10%. No octave shifts, which would correspond to 31.25 Hz were perceived.
In summary, an octave shift between RIS generated with opposite sign of the gain factor was perceived for delay times of 2 and 4 ms. The transition between the perception of an octave and an ambiguous pitch in the range of ±10% compared to RIS produced with $g = +1$ was found at a delay time of 8 ms. When the delay time was increased to 16 ms, the sensation of an octave shift vanished completely and the perception was found to be within the range of ±10%. These results were observed independent of the number of iterations, only the delay time of RIS was relevant.

Paired Comparison

The results of the paired comparison task performed by the 20 subjects were linear-scaled according to the BTL-method. The mean of the relative perceived pitch of the 500 Hz highpass filtered RIS is shown in Figure 4.12 (right) for the different sessions. In each block, the number of iterations n of the tested RIS was fixed (2, 8 or 4,096 iterations, top to bottom). Since RIS-conditions with different n were not tested within trials of the different blocks, the relative scales of the three diagrams cannot be compared directly. Only differences between RIS stimuli with the same number of iterations are relevant. The comparison of RIS(d,+1,n) with RIS(d,-1,n) (black versus grey) exhibited significant changes ($p < 0.05$) for the 2 ms and 4 ms delay conditions in the relative pitch according to the t-intervals (Table 5.2) for all tested conditions. A qualitative change of the relative pitch difference between RIS conditions with opposite sign of g was observed in all diagrams independent of the pitch strength: the pitch differences vanished with increasing the delay time to 8 ms and 16 ms.
The tested RIS covered the same frequency range and, although the diagrams only represent the relative pitch, all scales nearly had an identical range (5.67 ($n = 2$), 5.75 ($n = 8$), 6.16 ($n = 4,096$)), independent of the number of iterations.

Furthermore, the psychoacoustic experiment was repeated with unfiltered RIS stimuli in order to compare our results directly with other pitch-matching experiments (Yost, 1996a, e.g.). The values of the relative pitch between conditions of

Table 4.3: *Relative pitch of all tested RIS-conditions. Left: 500 Hz highpass filtered RIS. Right: unfiltered RIS conditions. Relative pitch values obtained by the BTL procedure for the same stimuli. Errors are based on bootstrap assessed t-intervals.*

RIS(d,g,n)	filtered RIS		unfiltered RIS	
(2,+1,2)	–3.50	±0.81	–2.62	±1.57
(2,–1,2)	–1.41	±0.41	–1.24	±0.43
(4,+1,2)	–0.96	±0.31	–0.67	±0.49
(4,–1,2)	–0.06	±0.29	–0.08	±0.23
(8,+1,2)	0.78	±0.33	0.64	±0.40
(8,–1,2)	0.99	±0.35	0.50	±0.85
(16,+1,2)	2.17	±0.47	1.71	±0.84
(16,–1,2)	1.99	±0.38	1.61	±0.69
(2,+1,8)	–3.59	±0.29	–3.26	±0.59
(2,–1,8)	–1.77	±0.45	–1.78	±0.56
(4,+1,8)	–1.36	±0.45	–1.22	±0.28
(4,–1,8)	–0.09	±0.26	–0.27	±0.62
(8,+1,8)	0.63	±0.31	0.99	±0.58
(8,–1,8)	1.21	±0.35	0.75	±0.80
(16,+1,8)	2.16	±0.36	2.32	±0.41
(16,–1,8)	2.63	±0.43	2.47	±0.52
(2,+1,4096)	–3.46	±0.67	–4.10	±0.54
(2,–1,4096)	–2.51	±0.50	–2.57	±0.86
(4,+1,4096)	–1.55	±0.52	–1.26	±0.47
(4,–1,4096)	–0.21	±0.45	–0.06	±0.47
(8,+1,4096)	1.11	±0.45	0.87	±0.29
(8,–1,4096)	1.22	±0.64	1.39	±0.27
(16,+1,4096)	2.70	±0.26	2.68	±0.24
(16,–1,4096)	2.69	±0.48	2.94	±0.44

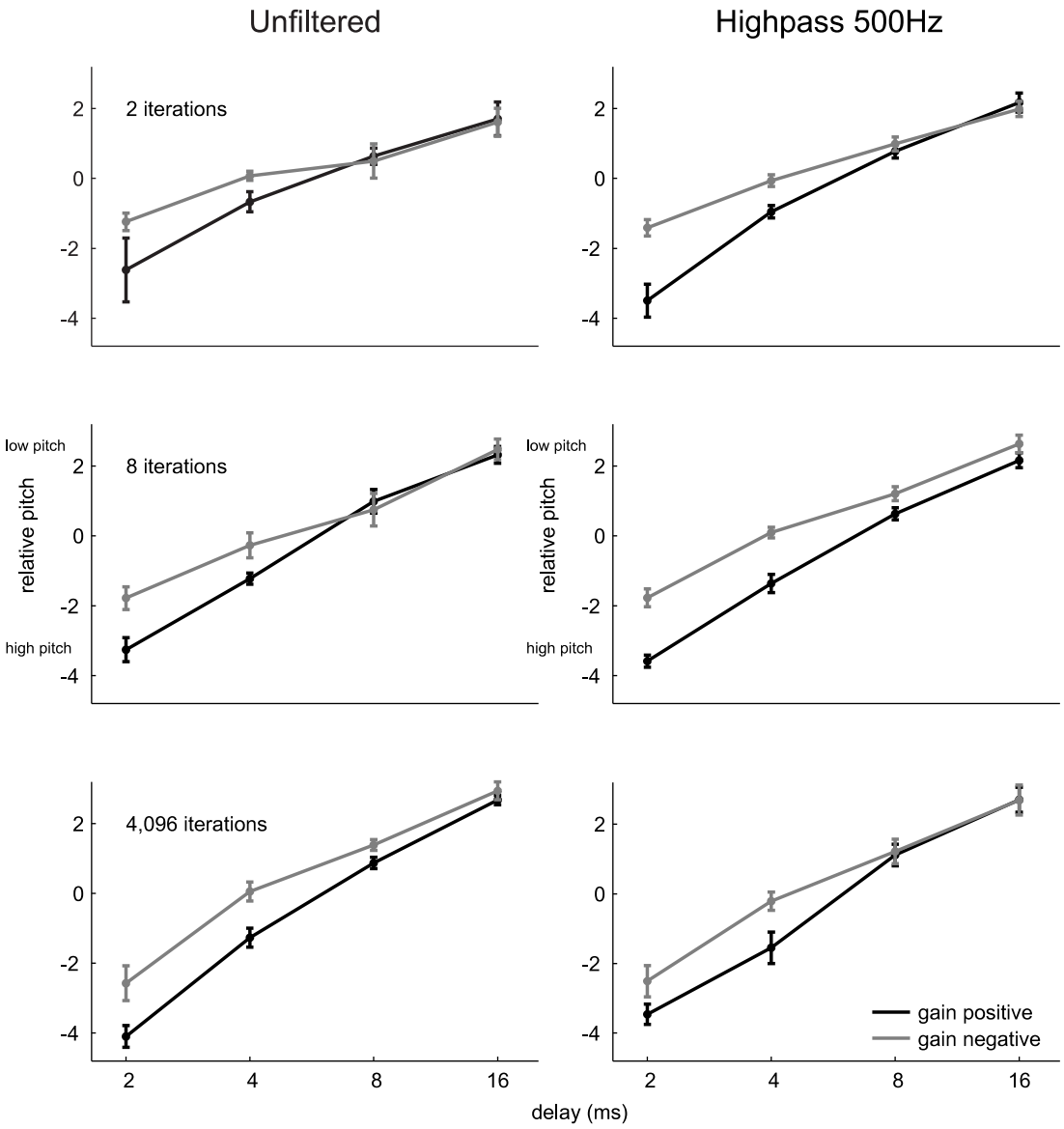

Figure 4.12: *Relative pitch scale obtained with the BTL-method of paired comparisons: The comparison demonstrates a high similarity between the perceived pitch based on unfiltered (left) and 500 Hz highpass-filtered RIS sounds (right).*

filtered and unfiltered RIS are given in Table 5.2, the diagrams are shown in Figure 4.12. The comparison shows, that independent of the filter condition and the number of iterations, the pitch differences are significant for delay times of 2 and 4 ms and within errors for delay times of 8 and 16 ms. Two exceptions are be found between the highpass filtered und the unfiltered RIS conditions. The filtered RIS(8,+1,8) condition differs significantly in the relative pitch compared to the RIS(8,-1,8) condition. This pitch difference is not observed in the unfiltered condition. The second difference between the filtered and unfiltered relative pitch is observed for RIS generated with 4,096 iterations and a delay time of 8 ms. However, the resulting differences of these two conditions seem to represent a statistical error.

4.2.4 Correlation of the Neuromagnetic POR to the Perceived Pitch of RIS

As shown in Figure 4.5, the psychoacoustic results agree very well with the findings of the POR in the MEG-experiment: differences in the latency of N100m' in response to RIS produced with opposite sign of g were significant for delay times of 2 and 4 ms and not significant for 8 and 16 ms, independent of the number of iterations n. The same observation was made for the paired comparison task of the psychoacoustic study. The BTL scale exhibited significant pitch differences for RIS generated with opposite gain factor g and delay times of 2 and 4 ms. The pitch differences disappeared when the delay was increased up to 8 and 16 ms. The correspondence between the neuromagnetic responses to the change of pitch and the perceived pitch of the psychoacoustic experiment is illustrated in Figure 4.13. The diagrams show the latency of N100m' plotted against the relative pitch scale assessed with BTL for the different number of iterations with high pitches at the bottom. Black markers represent RIS conditions generated with $g = +1$, whereas RIS(d,-1,n) are in grey. The delay of the POR is highly correlated ($\rho > 0.9$ for all tested n) with the inverse of the perceived pitch, independent of the gain factor and the number of iterations. That is, the delay gets longer as the pitch gets weaker, as would be expected. Although the BTL-method gives a relative scale for the perceived pitch, the range of the three pitch diagrams is nearly identical. Since subjects had to judge the pitch and not the pitch strength, the BTL-scale could be considered approximately as an absolute scale of pitch. Figure 4.14 shows the latencies of the N100m' plotted against the pitch scales to summarize the correlation between our neuromagnetic and psychoacoustic results. The latency of the N100m' evoked by the distinct pitch

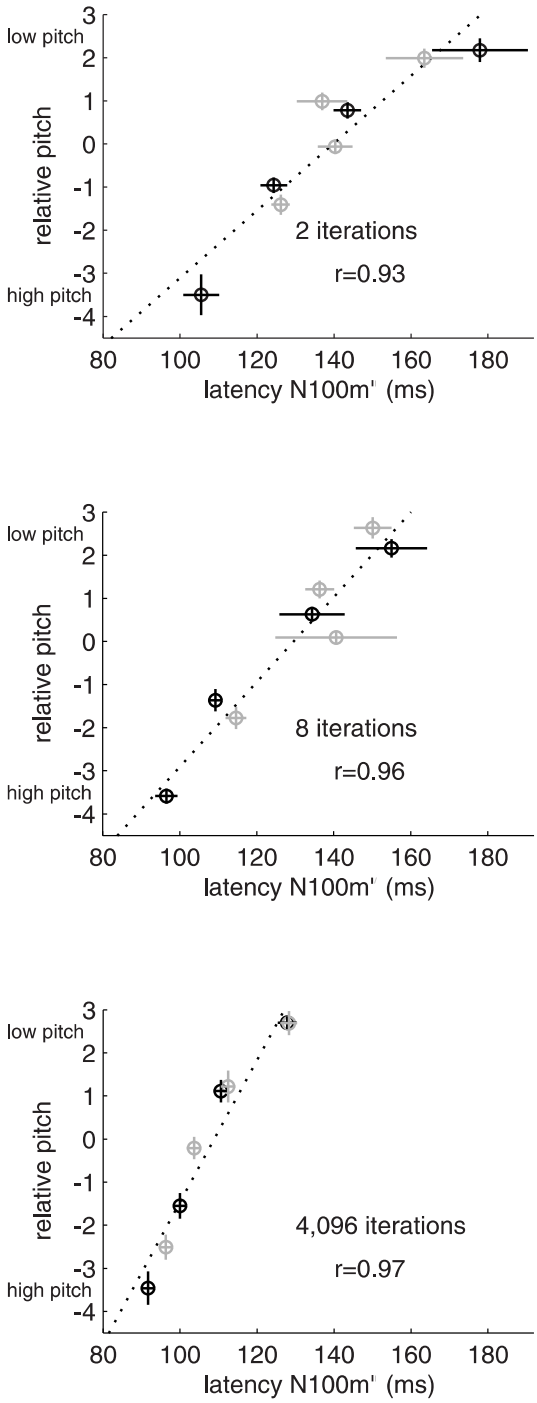

Figure 4.13: *Correlation between the latency of N100m' (POR) evoked by RIS and the relative perceived pitch of the corresponding sounds. RIS generated with positive gain are in black whereas RIS produced with a negative factor of g are in grey. Vertical and horizontal error bars represent the standard error of mean. For all tested numbers of iterations n, the pitch response and the perceived pitch are highly correlated.*

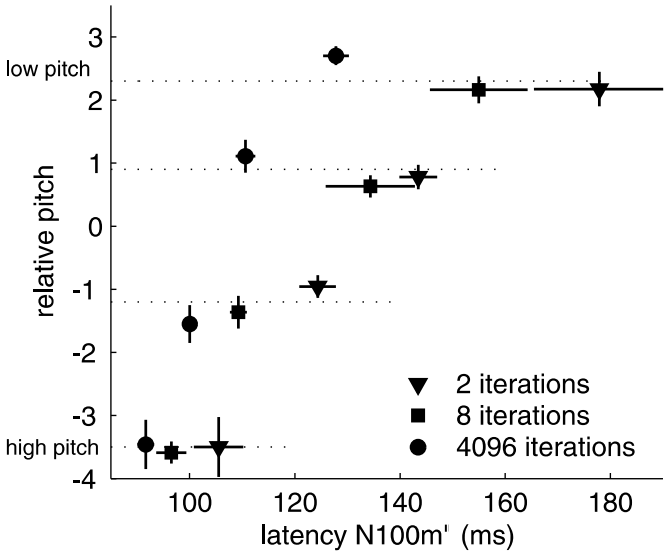

Figure 4.14: *Correspondence between latency of N100m' and the perceived pitch of RIS with positive gain. Error bars indicate the standard error of mean. The pitch response of the MEG-experiment depends on the pitch strength (n); the relative pitch is independent of n. The dashed lines show the mean of the pitch evoked by RIS with the different delay times (2, 4, 8 and 16 ms).*

of RIS generated with positive gain increases with the delay time d of RIS. The POR also depends on the number of iterations: increasing n with a fixed delay d revealed a significant decrease in the observed latency. In contrast to the POR of the MEG-experiment, the relative pitch height was not affected by varying n, giving evidence that the relative pitch of the BTL scale is independent from pitch strength.

4.3 Auditory Image Model: Simulation of RIS

In this section, a short description of how the Auditory Image Model (AIM) (Patterson et al., 1995) predicts the perceived pitch of RIS generated with positive or negative gain is given. The exact explanation of how the model extracts the pitch can be found in 2.3.1. AIM is based on the simulation of the spectral analysis performed along the basilar membrane using a bank of auditory filters: the neural transduction process of the inner hair cells and the primary auditory fibers is simulated in each of the frequency channels defined by the filters. In the next stage of processing, the time-interval information of each frequency channel is evaluated by strobed temporal integration. The array of these histograms,

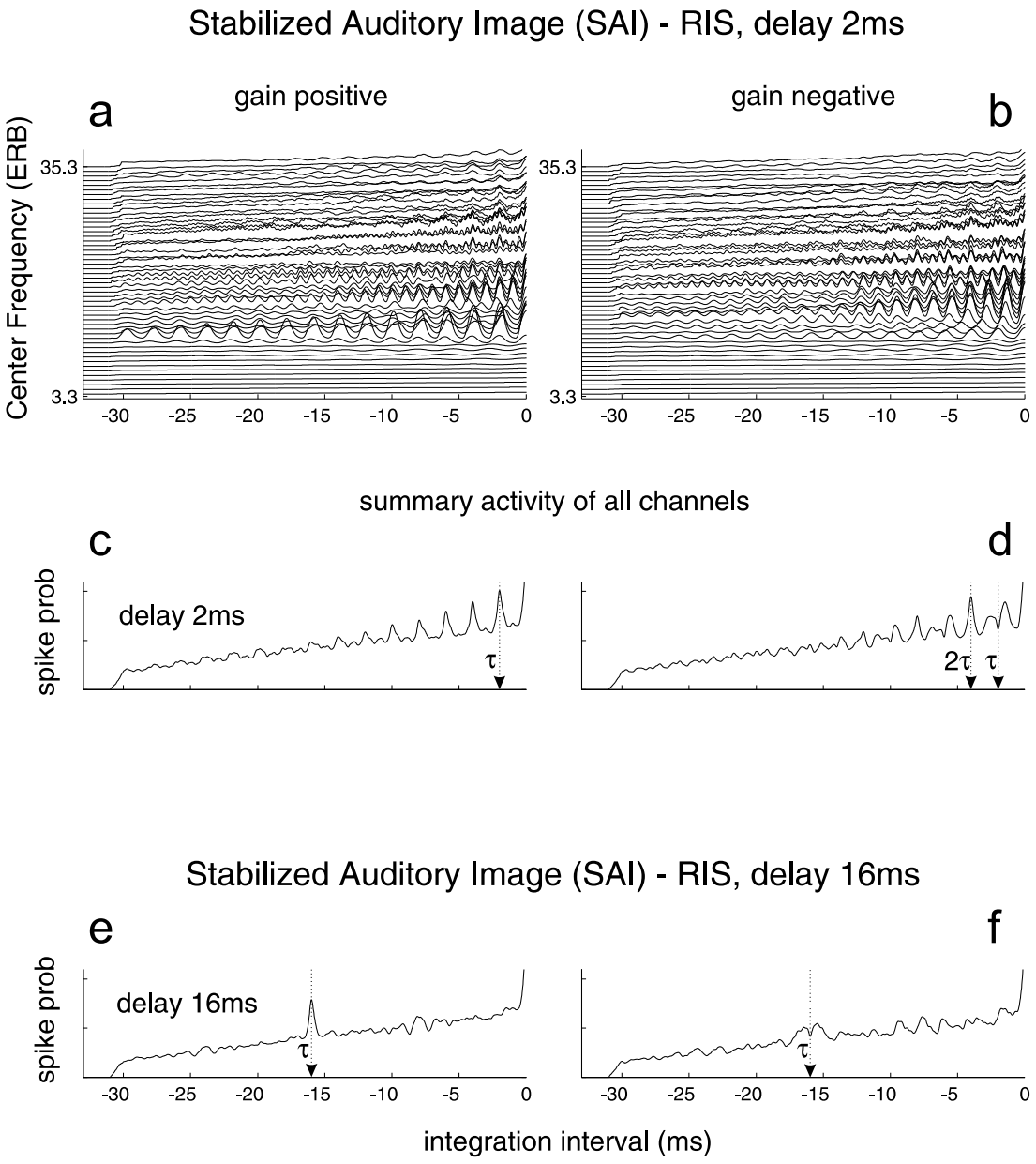

Figure 4.15: *Simulation of pitch extraction with RIS(d, g, n) with the Auditory Image Model (AIM; Patterson et al., 1995). Stabilized Auditory Images (SAIs) are created from the neural activity pattern by strobed temporal integration. The position and height of the first peak at time lag τ predicts the perceived pitch.*

the 'auditory image' (Figure 4.15a and 4.15b) is shown for RIS generated with a delay of 2 ms and the gain either set to plus or minus one. Respectively, the abscissa represents the time interval of the image buffer. It decays over time and is limited to 32 ms. The activity of each frequency channel is summed up and results in the summed stabilized auditory image (SAI).
The distance from 0 to the location τ of the first peak corresponds to the frequency of the perceived pitch $\frac{1}{\tau}$. Thus the predicted pitch of RIS(2,+1,8) (Figure 4.15c) is $\frac{1}{2\,ms} = 500\,\text{Hz}$. Changing the gain of RIS to $g = -1$ causes a significant change in the representation of the first peak (Figure 4.15d, RIS(2,-1,8)): now the SAI exhibits a minimum at lag τ and the first positive peak is located at 2τ. Thus, the predicted pitch is an octave below the pitch evoked by RIS with a positive gain factor.
Because of the decay in the buffer, the first positive peak at 2τ decays more and more with increasing the delay time d (Figure 4.15e). RIS stimuli with negative gain also reveal two side maxima at $\frac{1}{0.9d}$ and $\frac{1}{1.1d}$ which flank the observed minimum at lag τ (Figure 4.15d and 4.15f). For delay times of 2 and 4 ms, the first positive peak at lag 2τ is still prominent and determines the perceived pitch. But with an increasing delay time the height of the peak at 2τ decreases and the two side-peaks which predict the ambiguous pitch become more prominent (Figure 4.15f). In this way, the model explains the pitch shift of RIS(d,-1,n) from an octave below the pitch of RIS(d,+1,n) to the perceived ambiguous pitch in the region of ±10%.

4.4 Discussion

The aim of this experiment was to investigate the correlation between neuromagnetic responses and the perceived pitch of RIS. Especially, the perceived pitch shift of RIS generated with a negative gain compared to RIS produced with an opposite sign of g was investigated. The correlation between the perceived pitch shift and the neuromagnetic responses evoked by a RIS-RIS transition is built up.
Our results revealed a remarkable agreement between the neurophysiological and the psychoacoustic results (Figure 4.5). The POR evoked by RIS generated with opposite signs of the gain factor g and delay times of 2 and 4 ms showed significant differences in the latency of the N100m'.
The corresponding perception of the BTL scaled relative pitch of RIS also differed significantly for delay times of 2 and 4 ms. The latency difference of the POR and the pitch difference were consistent with each other: Both, the latency dif-

ference of the N100m' and the relative pitch difference did not differ significantly for delay times of 8 and 16 ms. This equivalence was independent of the number of iterations n. Both, the perceived pitch and the latency of the neuromagnetic pitch responses indicate the pitch shift. For delay times of 2 and 4 ms, the pitch of RIS generated with a gain factor of minus one was an octave lower than that of RIS produced with a positive sign of g, and rather similar for delay times of 8 and 16 ms. Our results showed that latency of the POR does not simply represent the spectral differences of the applied RIS conditions, as observed at the level of the basilar membrane. It is rather a neurophysiological correlate of the perceived pitch.

The simulation with AIM showed that the temporal pitch model introduced by Patterson et al. (1995) is able to predict the shift in the perception of RIS in dependence of the delay time d. The location of the first peak in the SAI (Figure 4.15c and 4.15e) at lag τ corresponds to the frequency of the perceived pitch ($\frac{1}{\tau}$). For RIS generated with a negative gain factor, the first positive peak at 2τ (Figure 4.15d) as well as the two side maxima at $\frac{1}{0.9d}$ and $\frac{1}{1.1d}$ (Figure 4.15f) can account for the perceived pitch.

Earlier studies conducted by Pantev et al. (1988) or Roberts and Poeppel (1996) used sinusoidal tones with a silent inter-stimulus interval to derive pitch-dependent properties of the N100. In contrast, the continuous stimulation allows to investigate the response to the onset of pitch in absence of the large proportion which is simply evoked by the energy onset. However, Ragot and Lepaul-Ercole (1996), who applied vowels like harmonic series with different spectral envelopes in their EEG study, reported latencies for the N100 which are comparable with our results of the POR evoked by RIS conditions generated with 4,096 iterations. This is interesting insofar as our neurophysiological responses were evoked by a continuous sound without an energy change at the transition of two RIS conditions (see Figure 4.1). The N100m' latency increased with increasing delay times (see also Figure 4.13), which is also in agreement with findings of Krumbholz et al. (2003) who reported comparable results of the POR latency in response to RIS(d,+1,n). We also observed decreasing amplitudes of N100m' with increasing delay times. Especially for RIS generated with two iterations and a delay time of 8 and 16 ms we observed smaller amplitudes, since our stimulation consisted of a continuous pitch and we recorded the responses to the transition of RIS. The N100m' evoked by RIS(d,+1,4096) exhibited latencies comparable to findings of Mäkelä et al. (1988) who measured the transition from noise to square waves. Krumbholz et al. (2003) proposed a fit formula for the latency of the POR with the delay time d as the only parameter determining the latency. However, the MEG data in our experiment revealed a strong dependence on the number of iterations. The slope

of the latency-increase per octave decreased with increasing n of RIS. Therefore, the fit formula from Figure 4.14 depended on d as well as on the number of iterations n. This behavior possibly represented the fact that the extraction of pitch takes longer when the signal-to-noise ratio is reduced and may be responsible for the latency differences of the POR in response to RIS generated with different n. The morphology of the recorded equivalent source waveforms exhibited a prominent positive peak at about 60–90 ms after the transition. The comparison with the data published by Rupp et al. (2005) suggests, that this peak is probably not only a response to the onset (P50m) of RIS, but also includes an off-response. Gutschalk et al. (2002) showed that the decay of the sustained field can account for this positive effect. The increased latency of this positivity evoked by RIS with increased delay times d can be explained with an overlapping effect of the positive P50m component and the pitch specific deflection (N100m').

The location of the POR evoked by RIS was in both hemispheres in the lateral aspect of Heschl's Gyrus. The applied fitting method with a fixed dipole model on the RIS(2,+1,8) condition can be justified, since the latency of the POR is not influenced by a change of the dipole location. The decrease of the amplitude strengths was only in the order of 10–20%. With regard to Figure 4.1, the fit of the dipoles in response to RIS/RIS transitions was more lateral compared to neuromagnetic source locations evoked by continuous click trains with varying pitch strength (Gutschalk et al., 2004). In another study, Griffiths et al. (1998) used PET to show that activation in Heschl's Gyrus increased with the temporal regularity of RIS. Further validation of the lateral activity is given by several neuroimaging studies based on functional MRI. Warren et al. (2003) reported that the medial Heschl's Gyrus is activated similarly when processing either pitch or noise. Only in a small area of the lateral Heschl's Gyrus, they reported an increased activity evoked by the change of pitch. The results of another functional MRI experiment conducted by Patterson et al. (2002) showed activity in the lateral aspect of Heschl's Gyrus for a melody of RIS versus fixed pitch condition. Interestingly, the equivalent dipoles evoked by our stimulation were located in the same area. (S. Uppenkamp, personal communication, 2004, see also Chapter 6).

The psychoacoustic experiment handled the difficult pitch matching with a simple two alternative forced choice task. The BTL-procedure was applied to derive the perceived pitch. RIS generated with opposite sign of g were connected with significant differences in the BTL scale for delay times of 2 and 4 ms, revealing the perceived octave shift. The pitch shift vanished when the delay time of RIS was set to 8 and 16 ms. Some subjects rated the ambiguous pitch of RIS generated with a negative gain factor higher, whereas others ranked it lower compared to

RIS consisting of positive gain. Even ratings within subjects were not unique, and therefore, the relative pitch differences in the BTL scale were within errors. Our results are in agreement with psychoacoustic experiments of Raatgever and Bilsen (1992). In their pitch matching study, subjects had to adjust the frequency of RIS(d,+1,∞) to fixed RIS(d,-1,∞). A pitch shift of an octave occurred for delay times of less than 3.5 ms respectively 6 ms (see also Figure 2.2). In the present experiments, we found the transition between the octave shift ($\frac{1}{2d}$) and pitch matchings in the region of $\frac{1}{0.9d}$ and $\frac{1}{1.1d}$ between delay times of 4 ms and 8 ms. These observations were made in the pitch matching experiment as well as in the paired comparison study. We also exhibited no differences in the perception between 500 Hz highpass filtered and completely unfiltered RIS. In contradiction to (Yost, 1996a), we did not observe in any psychoacoustic experiment an influence of the number of iterations n on the perceived pitch of RIS(d,-1,n). In a similar pitch matching experiment, Wiegrebe and Winter (2000) used highpass filtered RIS and found the pitch shift between RIS(d,+1,8) and RIS(d,-1,16) depending on the delay and on the cut-off frequency. At a cut-off frequency of 625 Hz, which is comparable to the highpass filtered sounds employed in our study, they observed an octave shift for stimuli with delay times of 2 and 4 ms, as illustrated in Figure 2.3. For delay times of 8 and 16 ms the deviation of RIS with negative g was found to differ by 10% compared to RIS with a positive sign of g. Their findings are consistent with our results. However, according to the results of Wiegrebe and Winter (2000) higher cutoff frequencies of the highpass filters applied to RIS, changes the perceived pitch shift. Thus, in the following Chapter, we apply bandpass conditions with different center frequencies to our RIS-like stimuli to determine the question if removing lower frequency components affects the perceived pitch of the sounds. Furthermore, we investigate the latency dependence of the pitch onset response evoked by the transition of bandpass filtered sounds and test if the auditory image model can account for pitch shifts due to different bandpass filter conditions.

Chapter 5

Perception and Representation of Filtered Complex Sounds

The results of Chapter 4 show, that the latency of the neuromagnetic N100m' component correlates highly with the perceived pitch of RIS. The spectrum of RIS generated with 4,096 iterations is nearly equal to the spectrum of a complex harmonic sound. Thus, the first question to be answered in this Chapter: Is the perceived pitch of RIS and a corresponding complex sound correlated?
Moreover, the simulation of RIS generated with a negative gain (see section 4.3) revealed two side maxima at $\frac{1}{0.9d}$ and $\frac{1}{1.1d}$ in the simulated activity of the summarized stabilized auditory images. These peaks seem to correspond to the ambiguous pitch. In addition, the stabilized auditory images of the single auditory channels (see Figure 4.15) exhibit, that the side maxima merge together with an increase of the center frequency of the single channels. To address the role of these two peaks, we applied harmonic complex sounds and filtered them with three different bandpasses. The neuromagnetic pitch onset response as well as the perceived pitch of the bandpass filtered complex sounds is investigated in this Chapter.

5.1 Material and Methods

The stimulus conditions were equal to Chapter 4. Also, recordings were made with the same hardware and acquisition settings described in the previous Chapter.

5.1.1 Subjects

Ten adult listeners (five male, five female) participated in the bandpass experiment. Eight of them took also part in the RIS experiment described in Chapter 4. The two additional subjects also had no reported history of peripheral or central hearing disorder. After giving their informed consent, all subjects participated in both experiments, the MEG measurements and the psychoacoustic study. The mean age of the ten subjects was 33.5 years with a standard deviation of ±8.2 years. In comparison to the RIS experiment, the same conditions were provided: subjects watched a silent movie of their own choice and were asked to concentrate on the movie and not to pay attention to the stimuli.

5.1.2 MEG stimuli

Harmonic series with a frequency spacing between adjacent components (Δf) of either 125 Hz or 250 Hz were applied. In the 'even'-condition, a complete harmonic series with a fundamental frequency ($f_0 = \Delta f$) of either 125 or 250 Hz was produced. The 'odd'-conditions were generated by shifting the spectrum to a fundamental frequency to 62.5 Hz and 125 Hz, leading to an incomplete series only with odd numbers of harmonics ($f_0 = \frac{\Delta f}{2}$). The spectra of all stimuli are shown in Figure 5.1. The spectra of the 'even' conditions are very similar to the applied RIS sounds generated with a delay time of 4 ms and 8 ms respectively, a positive gain factor, and 4,096 iterations (see also Chapter 4). RIS generated with a negative gain factor has a fundamental frequency of $\frac{f_0}{2}$ compared to RIS generated with a positive gain. The corresponding spectrum only consists of odd harmonics of the fundamental frequency (see Figure 3.1) and is the equivalent to the 'odd' conditions. As illustrated in Figure 5.1, all harmonic complexes were bandpass filtered. Three different bandpass conditions were used. A 'LOW'-bandpass with cutoff frequencies at 500 Hz and 1,000 Hz, a 'MID'-bandpass (1,000–2,000 Hz) and a 'HIGH'-bandpass (2,000–4,000 Hz) condition. The complex sounds had a total length of 510 ms, were balanced in energy and ramped with a 10-ms hanning window at the beginning and at the end. Neuromagnetic responses to the sound onset from silence were avoided by using a continuous sound. It consisted of complex stimuli with the same bandpass condition and the same frequency spacing between adjacent harmonics but alternated between the 'odd'- and 'even'-conditions. In analogy to the sounds applied in Chapter 4, the subsequent stimuli of the continuous sounds overlapped in a way that there was no visible change of the sound pressure at the transition.
The resulting sound had a total length of 10.5 s and was superimposed with a

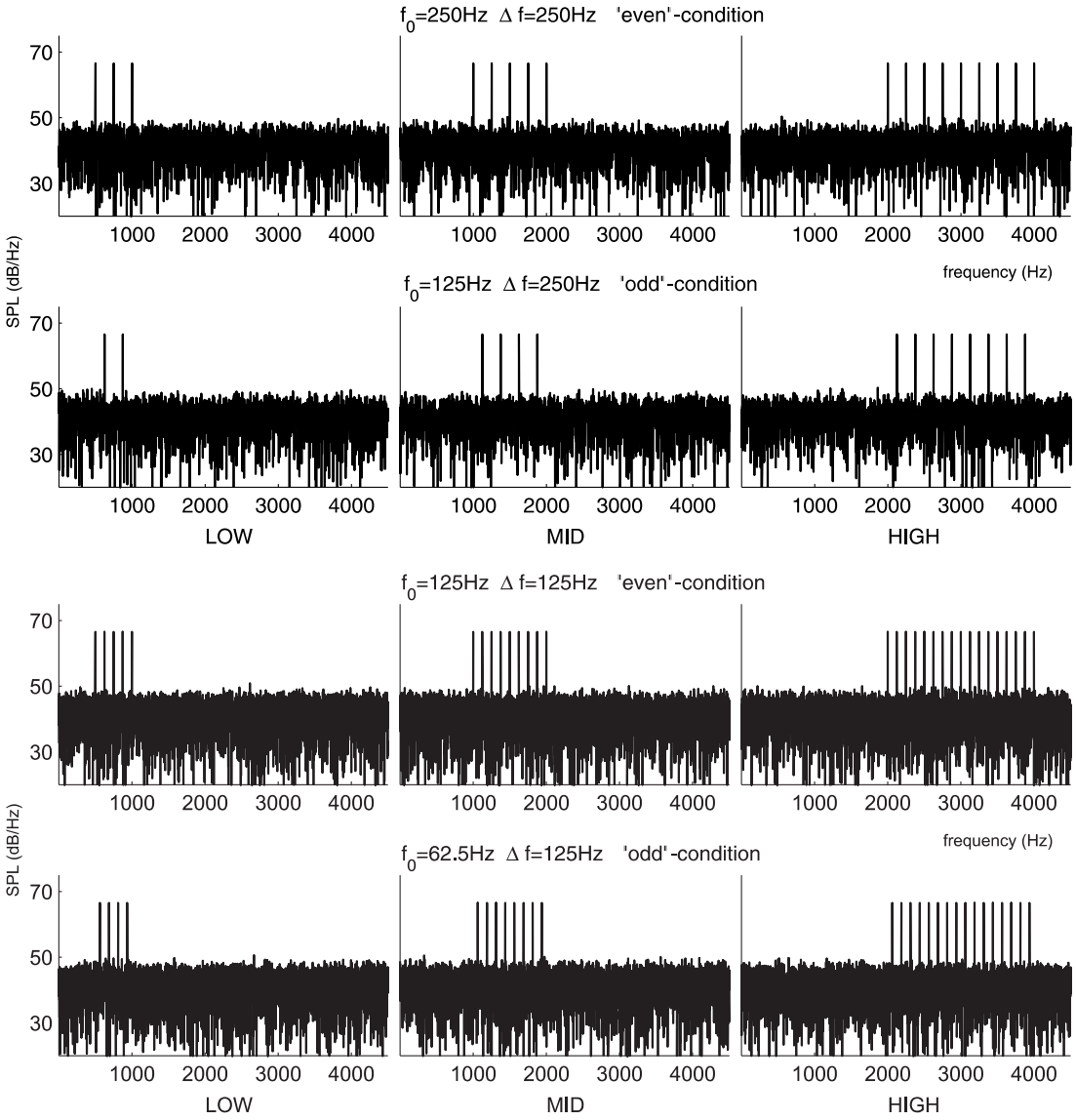

Figure 5.1: *The applied stimuli either had a spacing between adjacent frequency components of $\Delta f = 250\,Hz$ (upper two panels) or 125 Hz (lower two panels). The fundamental frequency was either f_0=250 Hz (first panel) or f_0=125 Hz (third panel) to generate complete harmonic complex sounds ('even'-condition). The 'odd'-condition consists of sounds that were frequency shifted by $f_0/2$ to produce harmonic complex sounds, but only with odd harmonics (second and fourth panel).*
As illustrated, all applied sounds were superimposed with white noise (DC–10,000 Hz), attenuated at 20 dB, relative to the signal to any suppress distortion effects.

white background noise (DC–10,000 Hz) attenuated at 20 dB relative to the overall level of the sound (see Figure 5.1) to avoid any distortion effects.

Neuromagnetic responses were recorded in two sessions. In each session the frequency spacing (Δf) of the concatenated sounds was fixed. Altogether 150 continuous sounds of 10.5 s-length were presented with a total of 500 odd/even transitions, leading to a length of 29 minutes per session.

5.1.3 Data Processing and Source Analysis

In analogy to Chapter 4, a spatio-temporal source model with one equivalent dipole in each hemisphere was used (Scherg et al., 1989) to estimate the moment and the space coordinates of the activated cortical area. The fit of the dipole sources was applied to the averaged waveforms of the transitions. In both sessions, the waveforms in response to the transition of the 'even-'condition of the lowest bandpass (500–1000 Hz) was 2–30 Hz bandpass filtered (zero-phase, 12 dB/oct). The first prominent negative deflection (N100m') of this condition was fitted in a 30 ms-interval around that peak. The responses of this condition produced a consistent fit for all ten subjects and was used in all cases as a spatial filter to derive the equivalent source waveforms of all other conditions. No further constraints concerning dipole location, orientation or symmetry-conditions were applied.
The drift and other low frequency artifacts of the continuous stimulation were compensated by computing a principle component analysis (PCA) over an interval of 400–500 ms after the transition of the unfiltered auditory evoked responses (Berg and Scherg, 1994). The PCA component that accounted for the largest variance in this interval was added to the spatial filter for the respective condition. For further calculations, the resulting equivalent source waveforms were zero-phase filtered from 1–100 Hz. The period of 100 ms before the transition was used to define the baseline.
In a second model, the averaged waveforms to the onset of the bandpass-filtered continuous sounds were also used to fit a single equivalent dipole in each auditory cortex of both hemispheres. Each onset of the three bandpass-conditions (LOW, MID, HIGH) were filtered and fitted in the same way as the averaged responses to the transitions.
The location of the fitted dipoles were rescaled[1] according to the stereotactic space of Talairach and Tournoux (1988) using the T1-weighted magnetic resonance images (MRI) of the individual anatomical structures of 9 subjects (see also section 4.1.3). The coordinates in space were projected onto the probabalistic map of the auditory cortex provided by Schneider et al. (2004).

[1] www.mrc.cbu.cam.ac.uk/Imaging/mnispace.html

5.1.4 Psychoacoustics

Psychoacoustic experiments were conducted in two different sessions: in the first session we used bandpass filtered complex sounds as illustrated in Figure 5.1. According to the MEG measurements, sounds were superimposed with a white background noise (DC–10,000 Hz), ramped with a 10 -ms hanning window and attenuated at 20 dB relative to the overall level of the sound complexes.
For a direct comparison between the psychoacoustic results of the complex sounds and RIS, listeners had to judge the pitch of RIS sounds with a similar spectrum: Delay times were set to 4 or 8 ms. The number of iterations was fixed at eight and the gain factor was either plus or minus one. The tested RIS stimuli were also generated with a sampling rate of 48,000 Hz; they had a total length of 500 ms and were ramped with a 10 ms hanning window at the beginning and at the end. In analogy to the complex sounds, RIS were also bandpass filtered (LOW, MID, and HIGH) and superimposed with white noise in the same way as the complex sounds, leading again to a total of 12 different conditions. The tested sounds of both sessions were presented using the ER3 earphones to ensure that the sounds were of the same quality as in the MEG experiment. The level of the sounds was set to a convenient level, adjusted by the subjects (60 dB $\pm$4 dB).
For both sessions, a two-alternative forced choice task for paired comparisons was applied: Two different sounds were presented randomly, with an inter-stimulus interval of 300 ms. Subjects had to indicate the tone with the higher pitch. During one session, every combination of paired comparisons was presented twice with reversed order. They judged all possible combinations of the sounds (12 $\times$ 12), excluding the comparison of the single sounds to themselves, which led to a total of 132 trials per session.
The scale for the relative 'height of pitch' was derived from the paired comparison experiment according to the algorithm described in section 3.2.

5.2 Results

5.2.1 Source Analysis

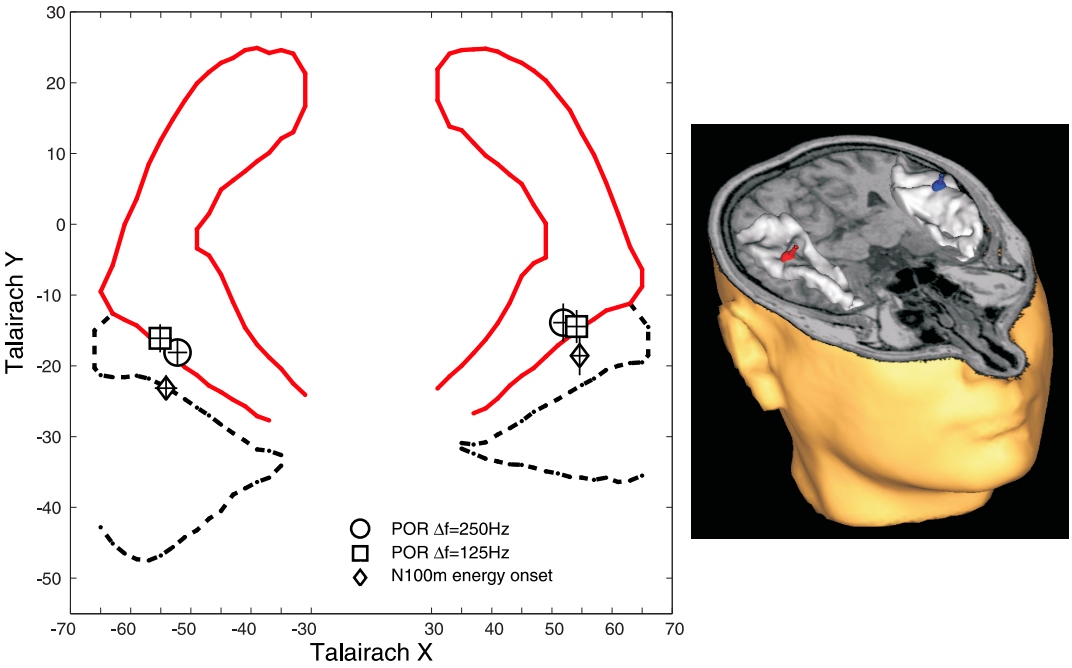

Figure 5.2: *Left: Mean coordinates of the averaged equivalent dipoles projected onto an axial plain. The sulcal borders were kindly provided by Schneider et al. (2004). Equivalent dipoles were evoked by the transition from complex sounds with a Δf of 250 Hz and 125 Hz. The fitted dipoles of the isolated POR (circles and squares) are located in the left and right lateral Heschl's Gyrus (solid borderline). The diamonds show the onset of the complex sounds from silence (N100m) averaged over all bandpass conditions. The standard errors of nine subjects are represented by vertical and horizontal bars. Right: MRI of a single subject (AR), with 3D-surface of Heschl's Gyrus.*

As for the experiment conducted with RIS, the neuromagnetic responses, evoked by the transition from the frequency-shifted ('odd') sound complexes to the harmonic complex ('even') sounds with the same Δf between adjacent frequency components were fitted on the N100m'-component with the two-dipole model. The coordinates of both equivalent dipoles were transformed in the space of Talairach and Tournoux (1988). In Figure 5.2, the mean coordinates of nine subjects are depicted on an axial plane through Heschl's Gyrus (solid borderline) and Planum temporale (dotted borderline). Both dipoles are located in the

lateral part of Heschl's Gyrus. As illustrated, the evoked equivalent dipoles of the Δf=125 Hz complex sound condition seem to be more laterally compared to the center of activity evoked by complex sounds of a Δf=250 Hz sound. The difference in the location of the POR evoked by both conditions is within errors. An overview of all derived dipole-coordinates are given in Table 5.1.

The table also shows the mean coordinates of the equivalent dipoles, fitted on the N100m component to the sound onset from silence. The three different bandpass conditions ('LOW','MID','HIGH') of the filtered complex sounds are

Table 5.1: *Coordinates of the equivalent dipoles located in the left and right auditory cortex in response to bandpass filtered complex sounds transformed in the space of Talairach and Tournoux (1988). The neurophysiological activity evoked by the transition (POR) of the complex sounds was modelled from the 'LOW'-condition. The coordinates of the equivalent dipoles evoked by the onset of the sounds from silence were derived for all bandpass conditions. All coordinates represent the averaged values of 9 subjects (± s.e.).*

		right audit. cortex			**left audit. cortex**		
Spacing	Bandpass	x	y	z	x	y	z
Δf=250 Hz	POR	52	-14	14	-52	-18	11
		±5	±8	±5	±5	±5	±7
	LOW	55	-19	14	-56	24	7
		±4	±9	±9	±5	±5	±7
	MID	52	-18	13	-51	-23	4
		±5	±8	±11	±5	±6	±9
	HIGH	56	-22	15	-56	26	11
		±4	±7	±11	±6	±6	±7
Δf=125 Hz	POR	54	-14	11	-55	-16	11
		±6	±7	±13	±6	±6	±13
	LOW	56	-16	13	-55	23	9
		±6	±8	±7	±5	±5	±10
	MID	55	-17	13	-54	-21	9
		±4	±8	±7	±5	±5	±9
	HIGH	53	-19	15	-52	23	9
		±5	±9	±8	±7	±4	±10

all located posterior to the POR-response. This is also illustrated in Figure 5.2 with diamonds, which represent the mean of all six conditions to the sound onset from silence.

5.2.2 Neuromagnetic Responses to the Change of Pitch

The evoked neuromagnetic responses fitted to the transition from the 'odd'- to the 'even'-condition of the 'LOW'-bandpass filtered sounds are shown in Figure 5.3 as grand-average of all 10 subjects. As can be seen from the Figure, the equivalent source waveforms of all tested complex sound conditions revealed no significant differences between the right (grey) and left (black) hemisphere. Thus, the mean of both source waveforms was used for further calculations. All values, either amplitude or latency of the N100m' component which are presented in this Chapter stand for mean values of both hemispheres.
The morphology of the AEFs is similar to the source waveforms of the RIS-RIS transitions described in section 4.2.2. All responses to the transitions of the bandpass filtered sounds presented in Figure 5.3 exhibit a first positivity at approximately 70–80 ms after the transition, followed by the prominent N100m' deflection. After the zero-crossing, a second positive peak, the P200 was observed at about 200–250 ms after the transition. Generally, the latency of this component was increased, when the Δf was decreased. The amplitude of the P200 depended strongly on the bandpass condition. The highest amplitudes were observed for the 'LOW'-condition, whereas this component nearly vanished after the 'HIGH'-bandpass filtered transition.

The latency of the pitch specific N100m' component was examined for all tested bandpass filtered sound transitions. The resulting mean latencies are presented in Figure 5.4. When Δf was fixed and the bandpass condition was increased, the latency of the POR increased significantly for tested harmonic sounds ($\Delta f = f_0 = 250\,\text{Hz}$: $F(2,18) = 6.93, p < 0.01$ and $\Delta f = f_0 = 125\,\text{Hz}$: $F(2,18) = 7.47, p < 0.01$). Between the 'even'- and 'odd'-conditions of a fixed Δf, the latency increased significantly for the harmonic sounds generated with $\Delta f = 250\,\text{Hz}$ ($F(1,9) = 12.26, p < 0.01$) and was found to be insignificant for the sounds with $\Delta f = 125\,\text{Hz}$ ($F(1,9) = 0.19, n.s.$). Between the 'even'- and 'odd'-conditions generated with a fixed Δf and filtered with the same bandpass, the latency difference reached the level of significance for the stimuli with $\Delta f = 250\,\text{Hz}$, filtered to the 'LOW'-frequency range ($t(9) = 2.07, p < 0.05$) and the sounds filtered with the 'MID'-bandpass filter ($t(9) = 3.92, p < 0.01$). The

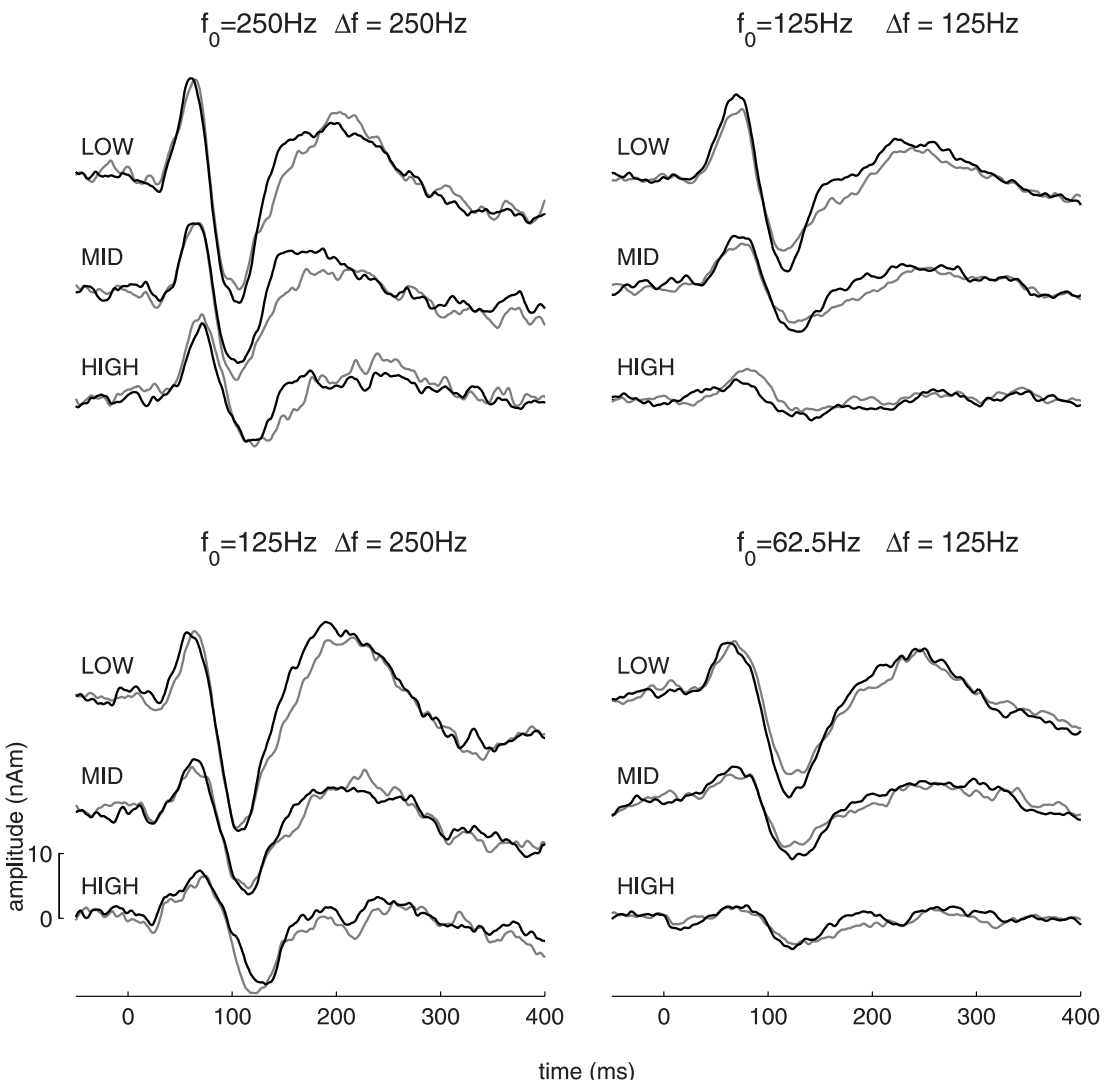

Figure 5.3: *Grand-average source waveforms of 10 subjects based on the fit of the POR evoked by the transition from the 'odd'- to the 'even'-condition of the 'LOW'-bandpass filtered sounds. All waveforms of the left (black) and right hemisphere (grey) are zero-phase filtered from 1–100 Hz. Independent of the 'even'- or 'odd'-condition and the fundamental frequency f_0, the N100m' latency increased when the filter condition of the bandpass filter was increased from 'LOW' (500–1,000 Hz) to 'MID'(1,000–2,000 Hz) and 'HIGH'(2,000–4,000 Hz). Increasing the bandpass filter from 'LOW' to 'MID' and 'HIGH' also caused a decrease of the N100m' amplitude in all tested conditions.*

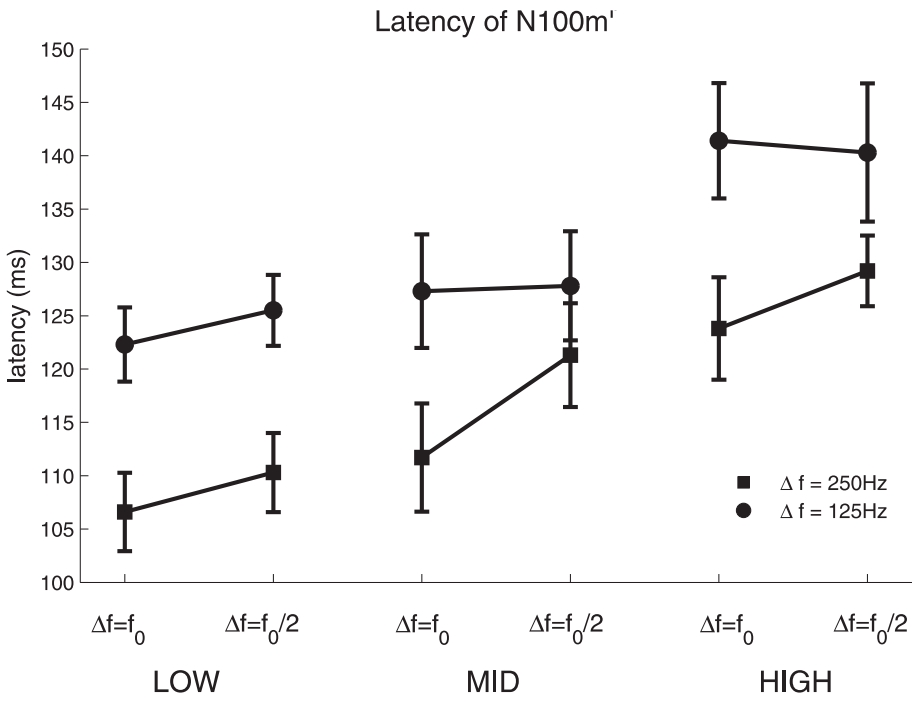

Figure 5.4: *Mean latency (in ms) of the N100m' component evoked by the transition of bandpass filtered complex sounds. The different bandpass conditions are along the* x*-axes from 'LOW' (500–1,000 Hz) to 'MID' (1,000–2,000 Hz) and 'High' (2,000–4,000 Hz). The latencies of the POR, evoked by harmonic sounds with the 'even'-conditions are represented with* $\Delta f = f_0$ *and the 'odd'-conditions with* $\Delta f = f_0/2$.
Independent of Δf*, the latency of the POR increased significantly, when the bandpass condition was increased. Between the 'even'- and 'odd'-conditions, the latency increased significantly for the harmonic sounds generated with* $\Delta f = 250\,Hz$ *(squares) and was found to be insignificant for the sounds with* $\Delta f = 125\,Hz$ *(circles). Between the 'even'- and 'odd'-conditions generated with a fixed* Δf *and filtered with the same bandpass, the latency difference reached the level of significance, only for the stimuli with* $\Delta f = 250\,Hz$*, filtered to the 'LOW'- and 'MID'-frequency range.*
Error bars represent the standard error of 10 subjects.

latency difference was insignificant, when the sounds were bandpass filtered with the 'HIGH'-bandpass filter ($t(9) = 0.93, n.s.$). When the sounds were generated with $\Delta f = 125\,\mathrm{Hz}$, the latency differences were insignificant for the three bandpass conditions ('LOW': $t(9) = 0.91, n.s.$, 'MID': $t(9) = 0.10, n.s.$ and 'HIGH': $t(9) = -0.37, n.s.$).

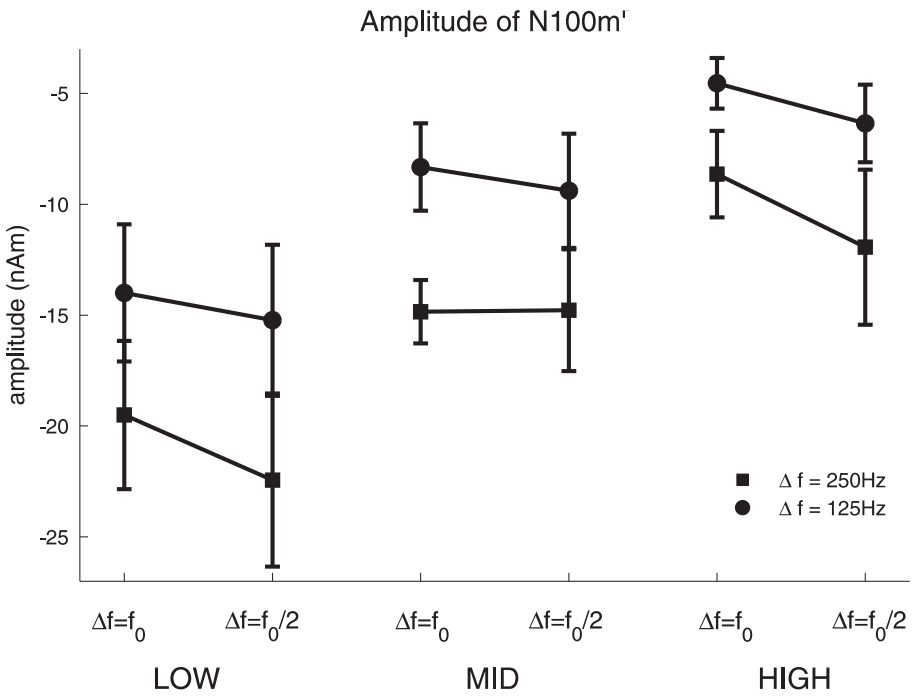

Figure 5.5: *Mean amplitude (in nAm) of the N100m' component evoked by the transition of bandpass filtered complex sounds. In analogy, to Figure 5.4, the different bandpass conditions are along the* x*-axes from 'LOW' (500–1,000 Hz) to 'MID' (1,000–2,000 Hz) and 'High' (2,000–4,000 Hz). The amplitudes of the 'even'-conditions are represented with* $\Delta f = f_0$ *and the 'odd'-conditions with* $\Delta f = f_0/2$.
Independent of Δf*, the latency of the POR decreased significantly, when the bandpass condition was increased. Between the 'even'- and 'odd'-conditions, there was no significant difference of the amplitude. Error bars represent the standard error of 10 subjects.*

Analogously, the amplitudes of the N100m' component evoked by the transitions of the complex sounds, were examined. Figure 5.5 shows the mean amplitudes of the POR over all subjects. Again, the amplitudes of the POR evoked by the 'even'-conditions of the complex sounds are represented with $\Delta f = f_0$, and the 'odd'-conditions with $\Delta f = f_0/2$ respectively.
When the harmonic sounds were generated with $\Delta f = 250\,\text{Hz}$, the latency of the N100m' component decreased significantly, when the bandpass condition was increased ($F(2, 18) = 17.45, p < 0.001$). Amplitude also decreased when the sounds consisted of harmonics with a $\Delta f = 125\,\text{Hz}$ and only the filter-passband was increased ($F(2, 18) = 15.28, p < 0.001$). Keeping the bandpass condition fixed, no significant effect was observed for amplitude differences between the 'even'- and 'odd'-conditions. Neither for $\Delta f = 250\,\text{Hz}$ ($F(1, 9) = 1.84, n.s$) nor for harmonic

sounds with a Δf of 125 Hz harmonic sounds ($F(1,9) = 2.52, n.s.$).

5.2.3 Psychoacoustics

In the first session of the paired comparison task, the 10 subjects rated the pitch of complex sounds and in the second part the corresponding RIS was judged. The results of both sessions were linear-scaled according to the BTL-algorithm described in section 3.2. All derived relative pitches are given together with the t-intervals of each value in Table 5.2. In Figure 5.6, the correlation between RIS and the complex sounds are illustrated for a direct comparison.

The upper part of the table shows the relative pitch values of RIS. With a fixed delay time d and fixed gain g, the relative perceived pitch increased with increasing the bandpass condition. This effect was found for both delay times (4 and 8 ms), and observed independent of a positive or negative sign of g. The comparison between RIS with opposite sign of the gain factor revealed, that the differences in relative pitch were not significant, except for the condition with a delay time of 4 ms and a passband of the filter of 2,000–4,000 Hz ($p < 0.05$, according to the t-interval).
In the lower part of the Table, the values of the bandpass filtered complex sounds are listed. Analogously, the perceived pitch of all tested conditions was rated higher for fixed Δf and f_0, when the passband of the filter was increased from 'LOW' to 'MID' and 'HIGH'. The comparison of the conditions with the same Δf between adjacent peaks in the spectra revealed, that the perceived pitch of the complex sound that consisted of odd harmonics ($\Delta f = f_0/2$) only, was judged higher than the pitch of the corresponding complete harmonic sound ($\Delta f = f_0$). The higher rating of the perceived pitch was consistent for all sound conditions tested, but only reached the level of significance ($p < 0.05$) in two cases: Δf=250 Hz, bandpass filtered between 2,000 and 4,000 Hz ('HIGH') and in the Δf=125 Hz sounds, that were bandpass filtered from 1,000–2,000 Hz.

The correlation between the pitch of the bandpass filtered complex sounds and the corresponding RIS generated with 8 iterations is illustrated in Figure 5.6. The filled symbols represent the sound conditions that consist of all harmonics and, respectively, the RIS conditions with a positive gain factor. The open symbols show the perceived pitch of complex sounds, with odd harmonics which correspond to RIS generated with a negative gain. The tested sounds with a Δf with 250 Hz (RIS: 4 ms delay time) are shown in squares and sounds with Δf=125 Hz (delay time of RIS 8 ms) are diamonds. In general, the relative pitch of the sounds with Δf=250 Hz was judged higher as the pitch of Δf=125 Hz.

Table 5.2: *Relative perceived pitch of all tested RIS and complex sound-conditions. The Relative pitch values were obtained by the paired comparison task by linear scaling according to the BTL procedure. All stimuli were bandpass filtered into three different passbands: 'LOW' (500–1,000 Hz), 'MID'(1,000–2,000 Hz) and 'HIGH'(2,000–4,000 Hz). Upper: Relative pitch of RIS(d, g,8). Lower: Relative pitch of complete harmonic sounds ($\Delta f = f_0$) and complex sounds consisting of odd harmonics ($\Delta f = f_0/2$) only. Error bars represent the bootstrap assessed t-intervals for 10 subjects (t=1.8331) of each value. Note that negative pitch values correspond to high pitches; positive values to low pitches.*

RIS	delay	gain	bp cond.	relative pitch	
	4	+	LOW	1.15	±0.80
		–		1.23	±1.24
		+	MID	-0.85	±0.61
		–		-0.68	±0.48
		+	HIGH	-1.70	±0.53
		–		-2.31	±0.37
	8	+	LOW	1.79	±0.30
		–		1.56	±1.14
		+	MID	0.66	±0.46
		–		0.53	±0.35
		+	HIGH	-0.83	±0.41
		–		-0.54	±0.57

Complex Sound	Δf	f_0	bp cond.	relative pitch	
	250	250	LOW	0.25	±0.52
		125		-0.04	±0.82
		250	MID	-0.55	±0.53
		125		-0.90	±0.37
		250	HIGH	-1.77	±0.54
		125		-2.42	±0.46
	125	125	LOW	2.33	±0.34
		62.5		1.70	±0.99
		125	MID	1.50	±0.41
		62.5		0.87	±0.45
		125	HIGH	-0.29	±0.52
		62.5		-0.67	±0.38

As expected, the correlation between the paired comparison task of the tested complex sounds and the paired comparison study conducted with the corresponding RIS exhibits a remarkable correlation coefficient of ρ =0.90.

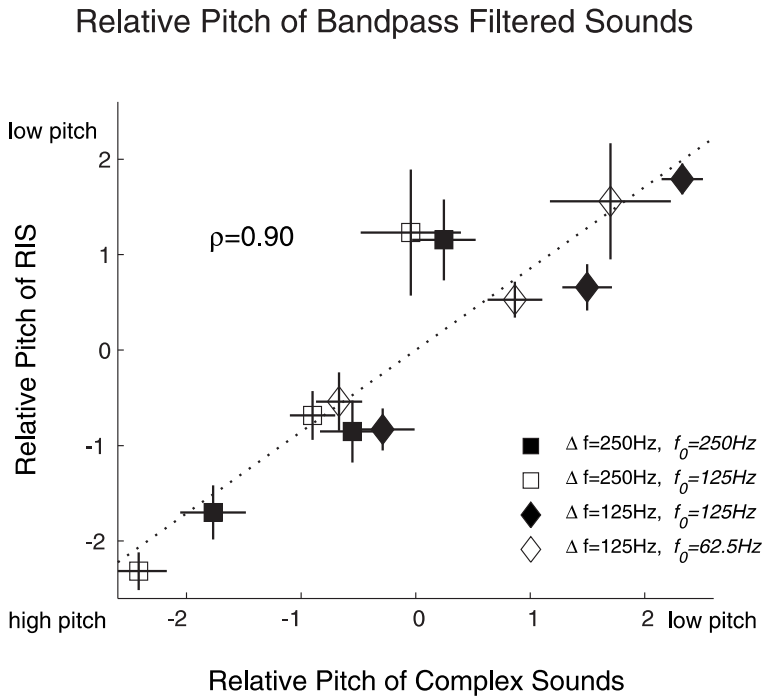

Figure 5.6: *Correlation between the relative perceived pitch of bandpass filtered complex sounds and the corresponding bandpass filtered RIS. The relative pitch was derived from the paired comparison task by linear scaling according to the BTL-method. Error bars represent the bootstrap assessed* t*-intervals for 10 subjects. Filled symbols represent the pitch of sound conditions that consist of all harmonics ($\Delta f = f_0$) and, respectively, the RIS conditions with a positive gain factor. The open symbols show the perceived pitch of complex sounds, with odd harmonics ($\Delta f = f_0/2$) which correspond to RIS generated with a negative gain. The sounds with $\Delta f = 250$ Hz (RIS: 4 ms delay time) are shown in squares; sounds with Δf=125 Hz (delay time of RIS equals 8 ms) are diamonds.*

5.2.4 Functional Dependence of the POR on the Perceived Pitch

The latencies of the POR evoked by the different bandpass filtered complex sounds were correlated to the corresponding relative perceived pitch of the harmonic sounds, as derived in the psychoacoustic study. Figure 5.7 illustrates the observed correlations for the tested harmonic sounds with a Δf of 250 Hz (left) and for Δf=125 Hz (right). In both diagrams, open symbols represent the evoked POR (and respectively the relative pitch) of harmonic complex sounds that consist of odd harmonics ($\Delta f = f_0/2$) only. Filled symbols show the sounds generated with harmonics of all integer multiples ($\Delta f = f_0$). Both diagrams comprise the applied three different bandpass conditions of the sounds in ascending order of the regression line. It can be seen, that the psychoacoustically derived relative pitch corresponds very well with the corresponding latency of the POR evoked

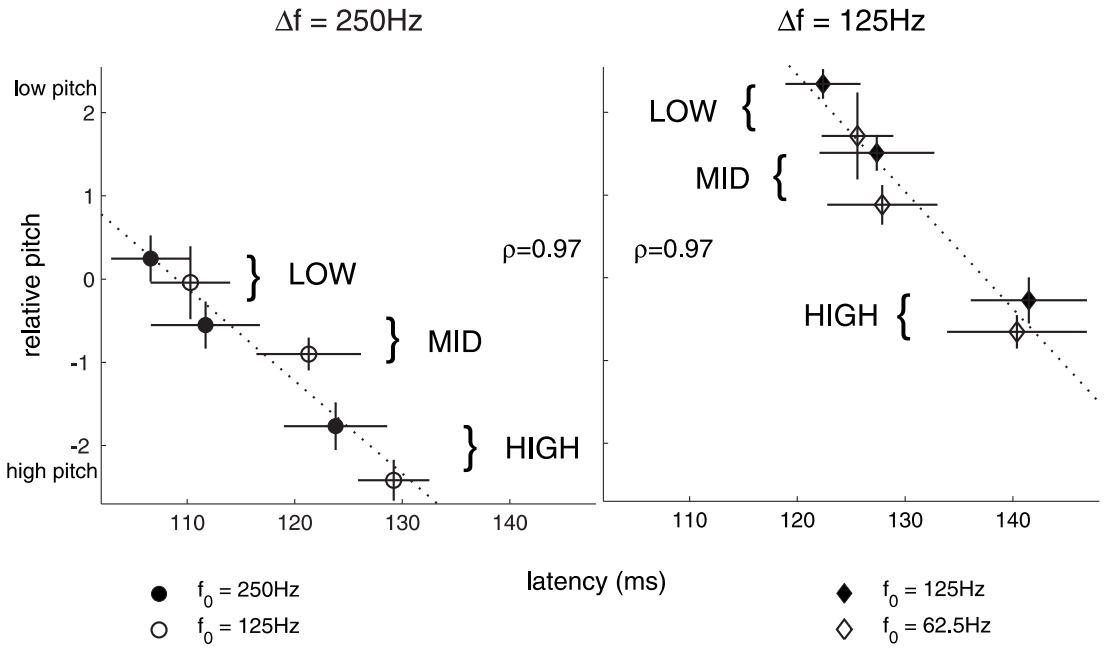

Figure 5.7: *Correlation between the latency of the N100m' component and the relative perceived pitch of the bandpass filtered complex sounds. Complex sounds produced with a Δf of 250 Hz (left) and Δf =125 Hz (right). Filled symbols represent the pitch of sound conditions that consist of all harmonics ($\Delta f = f_0$) and, respectively, the latency of the POR evoked by these sounds. The open symbols show the perceived pitch of complex sounds, with odd harmonics ($\Delta f = f_0/2$) and the corresponding latency of the N100m' component. Horizontal error bars represent the standard error of mean for the POR. The vertical error bars were derived using the bootstrap algorithm.*

by the same sound and the perceived pitch of the tested stimuli highly correlated with the latency of the N100m' component (ρ=0.97 for both correlations). The effect was found to be independent of the case that the complex sounds were produced either with the 'even'- or 'odd'-condition. The slope of the regression line also was nearly identical (Δf=250 Hz: $m = -0.0014/ms$, (Δf=125 Hz: $m = -0.0016/ms$). Thus, the mean of both regression lines suggests, that an increase of the relative perceived pitch of about 0.015 corresponds to a latency increase of 10 ms in the N100m' component.

5.3 Simulated Pitch of Complex Sounds

The Auditory Image Model (AIM), as introduced in section 2.3, is applied to predict the perceived pitch of bandpass filtered sounds. The simulation of the stabilized auditory images was computed for all harmonic complex sounds used for the MEG-measurements and for the psychoacoustic task. The summed activity of each auditory channel in the stabilized auditory image (SAI) results in the summed SAI. The simulation of the bandpass filtered sounds is illustrated in Figure 5.8 for a Δf of 250 Hz, and in Figure 5.9 for the sounds generated with Δf =125 Hz respectively. Since the sounds were bandpass filtered with three different conditions ('LOW', 'MID', and 'HIGH'), each of the Figures is divided into three subplots, showing the simulation of the filtered sounds. Each subplot contains the outcome of the model computed with the 'even'-condition ($\Delta f = f_0$) in black; the 'odd'-conditions ($\Delta f = f_0/2$) are in grey. The abscissa represents the time interval of the image buffer. The spike probability decays over time with a half life time of 20 ms. The distance from 0 to the location τ of the first prominent peak corresponds to the frequency of the predicted pitch with $\frac{1}{\tau}$. When the sounds, generated with 'even'-condition are simulated, the first peak in the summed SAI is independent of the bandpass filter condition at $\tau = \frac{1}{\Delta f}$. Sounds generated with 'odd'-harmonics reveal a minimum at time lag τ flanked by two side maxima. For these simulations, the first positive peak is at 2τ. As can be seen from the Figures, the distance ($\Delta\tau$) between both side maxima peaks and the minimum at τ is equal. The distance $\Delta\tau$ is independent of the spacing between adjacent frequencies (Δf) of the simulated sounds but depends on the bandpass filter condition only. The side maxima merge together when the bandpass condition is increased from 'LOW' ($\Delta\tau$=0.65 ms) to 'MID' ($\Delta\tau$=0.35 ms) and 'HIGH' ($\Delta\tau$ =0.15 ms).

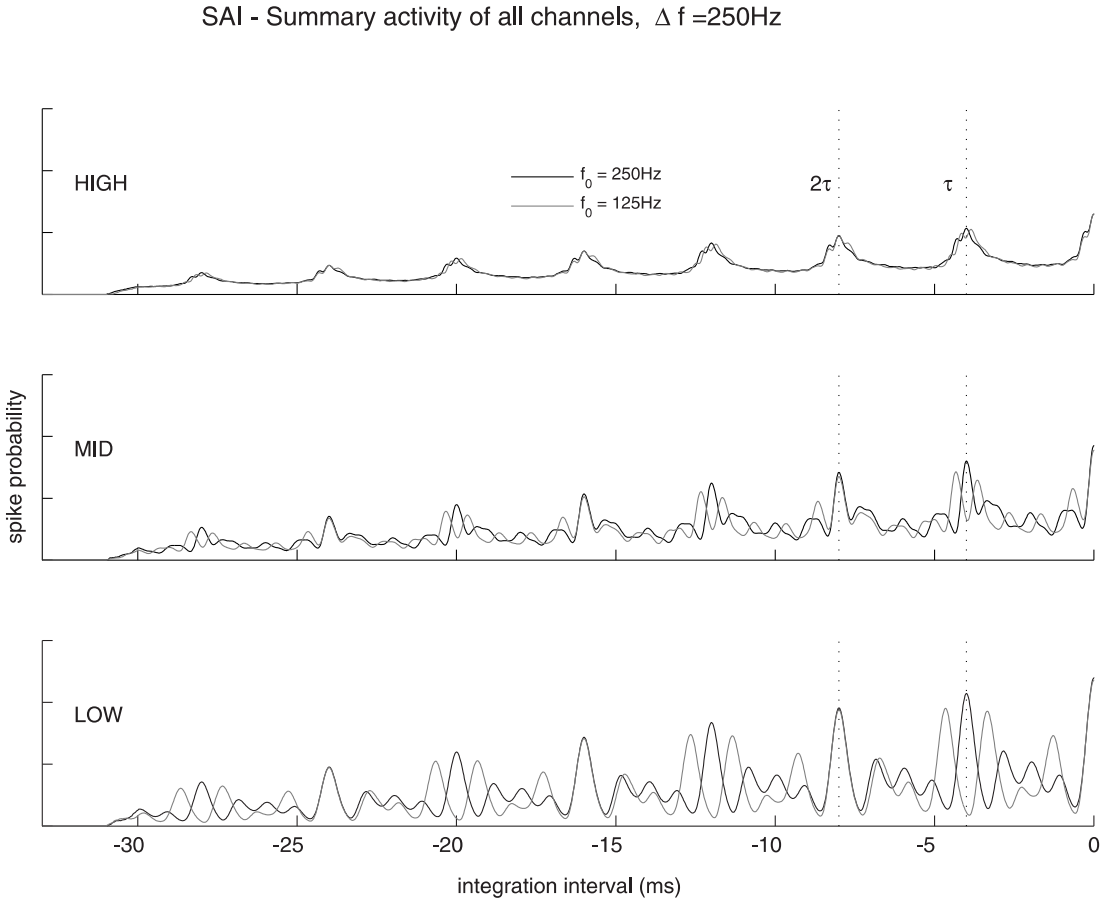

Figure 5.8: *Summed activity of the SAIs in response to simulated bandpass filtered complex sounds with a Δf of 250 Hz. From bottom to top, the simulation of the three different bandpass filtered sounds ('LOW' 500–1,000 Hz, 'MID' 1,000–2,000 Hz, and 'HIGH' 2,000–4,000 Hz) are shown. The 'even'-conditions ($\Delta f = f_0$) are in black and the 'odd'-conditions ($\Delta f = f_0/2$) in grey. The time interval of the spike probability decays over time and is limited to 32 ms. The SAI predicts the perceived pitch with distance from 0 to the location τ of the first prominent peak. The frequency of the perceived pitch is $\frac{1}{\tau}$. The sounds generated with 'even'-harmonics exhibit their first peak independent of the bandpass filter condition at $\tau = \frac{1}{\Delta f}$. In contrast, the sounds generated with 'odd'-harmonics reveal a minimum at time lag τ flanked by two side maxima. The first positive peak for these simulations is at 2τ. The predicted pitch is ambiguous and corresponds either to one of the side maxima or to the first positive peak at time lag 2τ.*

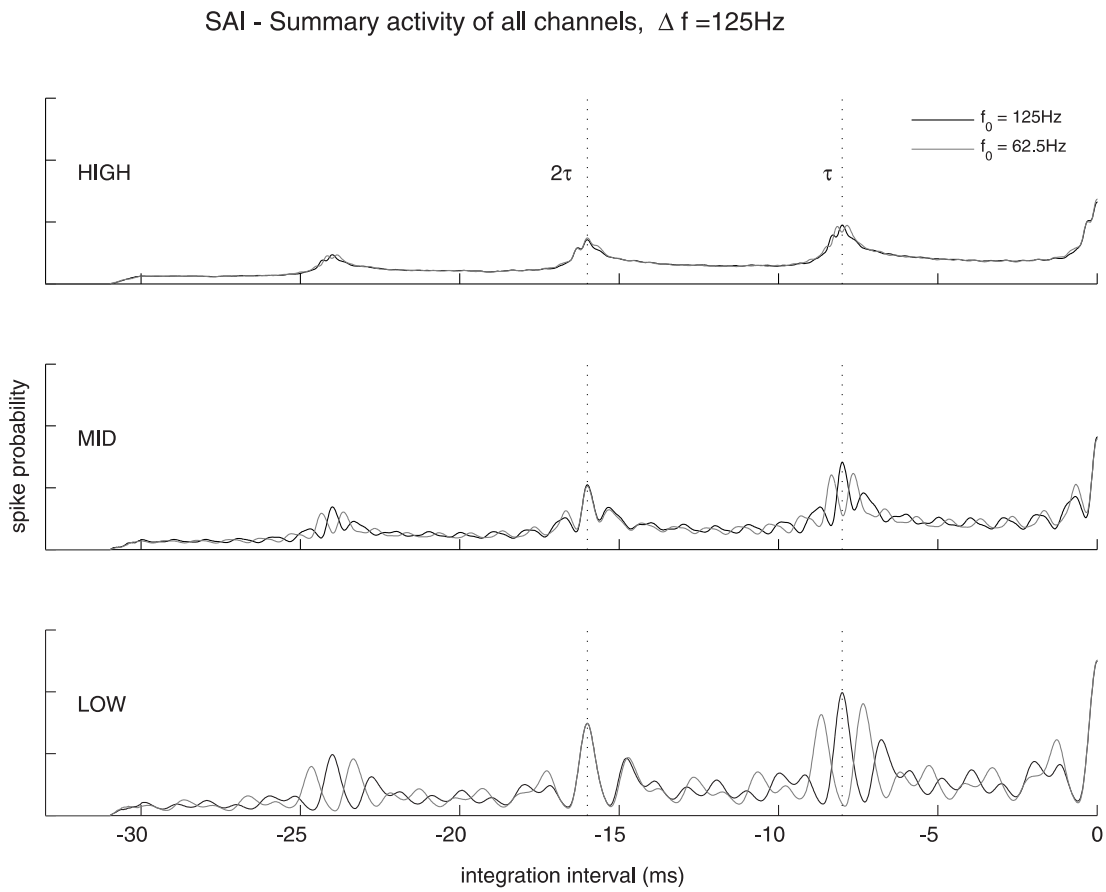

Figure 5.9: *In analogy to Figure 5.8, the summed activity of the SAIs in response to simulated bandpass filtered sounds, here with a Δf of 125 Hz, is shown. Bottom to top, the simulation of the three different bandpass filtered sounds 'LOW', 'MID', and 'HIGH'. 'Even'-conditions ($\Delta f = f_0$) are in black, 'odd'-conditions ($\Delta f = f_0/2$) in grey. Note that in comparison to Figure 5.8 the distance from 0 to the first positive peak at $\tau = \frac{1}{\Delta f}$ =8 ms is doubled.*

5.4 Discussion

In this experiment, the pitch of bandpass filtered sounds was investigated. A high correlation between the perception and the pitch-specific N100m' component evoked by these sounds was found. It was tested, whether the auditory image model can predict the pitch of bandpass filtered sounds.
The paired comparison task of the psychoacoustic experiment demonstrated, that complex sounds can be perceptually regarded as a correlate to RIS. Subjects rated the pitch of complex sounds in the first session, and pitch of RIS generated with eight iterations in the second session. Both, the complex sounds and RIS were

filtered, using three different bandpass frequencies ('LOW' 500–1,000 Hz, 'MID' 1,000–2,000 Hz, and 'HIGH' 2,000–4,000 Hz). The pitch scaling according to the BTL-algorithm revealed a correlation of ρ =0.90 between the relative pitch of complex sounds and the perceived pitch of RIS (see Figure 5.6). The results showed, that for sounds with the same fundamental frequency (f_0), the perceived pitch increases when the bandpass filter condition is changed from 'LOW' to 'MID', and 'HIGH'. This is in line with Zwicker and Fastl (1999), who reported that not all complex sounds from which the lower harmonics have been removed, elicit a pitch that corresponds to f_0. They demonstrated that rather a specific combination between f_0 and the lowest component must occur to produce the sensation of *virtual pitch*, that is a pitch of the fundamental frequency. They suggested an existence region in which a pitch of f_0 can be evoked. According to the definition of this region, a pitch of f_0 is only perceived when the sounds are filtered with the 'LOW' and 'MID'- bandpass filter.
The results of our psychoacoustic experiment showed that pitch also increased between the 'LOW'- and 'MID'- bandpass conditions, independent of the applied Δf and f_0. Additionally, an increase of pitch due to changes of the applied bandpass characteristics was observed This could also be caused by the fact that timbre influenced the ratings of listeners. Within a fixed bandpass filter condition, and in absence of timbre changes, subjects rated the pitch of complex sounds generated with the 'even'-condition ($\Delta f = f_0$) lower than pitch of sounds produced only with 'odd' harmonics ($\Delta f = f_0/2$).

The equivalent dipoles fitted on the N100m, evoked by the onset of the sound from silence revealed a center of activity in Planum temporale. Näätänen and Picton (1987) reported that this deflection, occurring 100 ms after the onset of the stimulus consists of at least three different components. Thus, the two-dipole model with one equivalent dipole in each hemisphere represents the center of activity built from a sum of different N100m deflections. The use of a continuous stimulation allowed to investigate the POR (N100m') evoked by the onset of pitch in absence of activity due to the energy onset. The pitch-specific N100m' was located more posterior in the lateral aspect of Heschl's Gyrus.
The latency of the POR depended on the fundamental frequency of the presented sound. When f_0 was decreased from 250 Hz to 125 Hz (see Figure 5.4), the POR latency increased significantly, independent of the applied filter condition. Furthermore, the latency also increased, when the bandpass condition was increased from 'LOW' to 'MID' and 'HIGH'. These results are in line with Bernstein and Oxenham (2003), who reported that low and high order harmonics make different contributions to the perceived pitch. The single components of the

presented sounds are equally spaced on a linear frequency scale, but the absolute bandwidths of the auditory filters increase with increasing the center frequency. Thus, the spacing between the individual components within a complex sound becomes smaller and smaller in the cochlear representation. The number of harmonics of each auditory filter increases with increasing harmonic number. As a conclusion, low order harmonics are resolved from one another, whereas higher-order harmonics begin to interact within single auditory filters and become unresolved. Thus, the latency increase might also be evoked by the presentation of less resolved harmonics. Carlyon and Shackleton (1994) demonstrated, that the ability to distinguish between two sounds drops steadily with an increase of the lowest harmonic number present. Thus, it is suggested that extraction of pitch takes longer when lower harmonics are removed. Furthermore, the prolongation of the extraction time might also have caused the observed significant drop of the N100m'-amplitude, since the activity over time of the single neurons is blurred.

The latency increase of the POR was highly correlated with the perceived pitch of the presented complex sounds. Figure 5.7 reveals that subjects did not perceive the missing fundamental (virtual) pitch anymore, when the bandpass condition was changed to higher center frequencies which is physiologically represented in the latency increase of the N100m' component. However, the loss of the virtual pitch alone cannot explain the increase of the latency. Roberts et al. (2000) reported that the N100 latency, evoked from the onset of sounds from silence was further modulated according to a secondary spectral analysis. They employed triangle and square wave stimuli of the same fundamental frequency and similar spectral center of gravity and revealed latency differences of the N100 component. Interestingly, they also reported that the spectral substructure analysis of their stimuli corresponded to a subjective perception of the sounds. The conclusion was that the latency of the N100 component is determined by both, an acoustic analysis (allowing contributions of both center of gravity as well as substructure or periodicity) and by a perceptual processing stage which recognizes a sound‘s timbre or the presence of the missing fundamental, respectively. When removing the lower harmonics of our applied stimuli, the perception of the complex sounds changes in that way that for a constant Δf of adjacent harmonics, the chroma shifted. Probably, listeners used both, timbre and pitch cues in the psychoacoustic task. In an EEG study, Ragot and Lepaul-Ercole (1996) presented harmonic complex sounds, filtered in three different formant regions, whose center frequencies corresponded to 350 Hz, 700 Hz and 2,300 Hz. Harmonic complex sounds with a different f_0 were applied. Their experiment showed, that the latency of the N100 peak rather depended on f_0 than on the frequency spec-

trum of the sound. However, all sounds presented in their experiment evoked a clear pitch at f_0 and timbre effects were suppressed. But their results also reveal a increase of the N100 latency when f_0 was fixed and the center frequency of the formant filter was changed from 350 Hz to 700 Hz.

Figure 5.8 illustrates the simulation of the summarized stabilized auditory images for complex sounds generated with a Δf of 250 Hz. Respectively, the simulations of the sounds produced with Δf =125 Hz are shown in Figure 5.9. The predicted pitch by the model corresponds to the distance from 0 to the first prominent peak of the neural autocorrelation function. With regard to the simulated outcome of the different filter conditions applied to the complex sounds generated with all harmonics ($\Delta f = f_0$), the pitch is predicted. Independent of the bandpass condition, the first prominent peak was observed at time lag $\tau = \frac{1}{\Delta f}$. Thus, the predicted pitch is the same for all tested filter conditions. In contrast, the psychoacoustic results showed that the pitch increased with an increase of the applied bandpass filter from 'LOW' to 'MID' and 'HIGH'. Removing the lower harmonics of the complex sounds weakens the pitch of the missing fundamental until no pitch around f_0 is perceived anymore.
The simulation of the complex sounds generated with the 'odd'-condition ($\Delta f = f_0/2$) revealed a minimum at time lag τ. Thus, the first prominent peak, that predicts the pitch of the sound can be either one of the side maxima that flank the minimum at τ, or the peak at time lag 2τ predicts the pitch. In the first case, the perceived pitch either would be slightly higher (or lower) between a tone and a semitone compared to the 'even'-condition. In the second case, the predicted pitch is an octave.
However, our psychoacoustic study revealed, that the pitch of the sounds generated with odd harmonics only, is higher for all tested conditions, compared to the corresponding sound with the same f_0 but generated with even harmonics and filtered into the same frequency passband. The result was consistent for complex sounds either generated with Δf =250 Hz or Δf =125 Hz.
Thus, the outcome of the model cannot account for the change of the perceived pitch of the applied sounds when influences of timbre increased.

Chapter 6

General Discussion and Conclusion

The present work was conducted to elucidate special aspects of the mechanisms on pitch processing in human. Particularly, the relation between psychoacoustic measurements and their auditory evoked neuromagnetic responses were examined. MEG was used to record the pitch-specific responses evoked by pitch changes. To determine the location and temporal course of the neuromagnetic activity in the auditory cortex, source analysis was applied. Both, neuromagnetic responses as well as psychoacoustic results were compared with simulations of the pitch processing based on the auditory image model provided by Patterson et al. (1995).

In the first experiment, regular interval sounds (RIS–Huygens noise) were applied to investigate the perception and the representation in the auditory cortex. The spectral envelope of RIS is a rippled continuum and thus, activity across the tonotopic dimension of neural activity in the central auditory pathway is rather uniform. The fact, that even RIS generated with one or two iterations is perceived as a tone with a distinct pitch suggests that this pitch is extracted using temporal regularity. In the second experiment, the study was systematically extended, to investigate the influence of bandpass filtered harmonic sounds on pitch processing, i.e. to determine the role of resolved harmonics.

The perceived pitch of RIS generated with a delay-and-add process (gain $g = +1$) is proportional to the inverse of the delay time d. Pitch salience is increasing with the number of iterations n. When RIS is generated with a negative gain g (delay-and-subtract), the perceived pitch differs considerably. However, the resulted pitch change is discussed controversially in the literature. The present work was carried out with special emphasis on this debate. Yost (1996a) reported

that the pitch of RIS generated with a negative gain depended on the number of iterations n. When RIS was produced with less than four iterations, pitch was found in the region of $\frac{1}{0.9d}$ and $\frac{1}{1.1d}$. The perceived pitch was shifted to $\frac{1}{2d}$ when n was increased.

The results of our pitch matching experiment are in contrast to these findings. The musically trained listeners who performed the task reliably, adjusted the pitch of RIS generated with a negative gain g an octave below ($\frac{1}{2d}$) the pitch of RIS produced with an opposite sign of g when the delay time was 2 and 4 ms. For delay times of 8 and 16 ms, pitch matches were observed at about $\frac{1}{0.9d}$ and $\frac{1}{1.1d}$. Pitch was found to be independent of the number of iterations, since n was varied in the experiment between one and eight. These results are in line with Raatgever and Bilsen (1992), who reported that the pitch depended on the delay time d. In analogy to our results, they found the transition between ambiguous pitches and the pitch shift of an octave below the pitch of RIS produced with a positive gain within delay times of 4–6 ms.

However, pitch matching tasks are very difficult and Yost (1996a) had to exclude four out of six subjects in his study. The pitch matching experiment of the present work also showed that the task could only be performed reliably by musically trained listeners. The results of two listeners, who were excluded from the test can be found in Appendix B.

To circumvent the observed difficulties, a simple two alternative forced choice task was applied, together with the BTL-algorithm that allows to derive a linear pitch scale. It was found that RIS generated with opposite sign of g exhibited significant differences in the relative pitch scale for delay times of 2 and 4 ms, which mirrors the perceived octave shift. The pitch difference vanished when the delay time of RIS was increased to 8 and 16 ms. Some subjects rated the ambiguous pitch of RIS generated with a negative gain factor higher, whereas others ranked it lower compared to RIS based on positive gain. Even ratings within subjects were not unique, and therefore, the relative pitch differences in the BTL scale were within errors between RIS generated with an opposite sign of g.

Wiegrebe and Winter (2000) also investigated the pitch shift of RIS but applied different highpass filters on the sounds. They reported, that the perceived pitch shift depended on the delay time d but also on the cutoff frequency of the filter. To determine the influence of the filter on the perceived pitch, bandpass filtered complex harmonic sounds were presented in the second experiment.

In order to receive similar spectra compared to RIS generated with positive and negative gain, complex sounds were generated with even and odd harmonics. Complex harmonic sounds with a frequency spacing (Δf) of 125 and 250 Hz between adjacent harmonics were bandpass filtered with three different center

frequencies, to investigate, whether different bandpass conditions affect the perceived pitch. For all sounds tested, listeners judged the pitch of the complex sounds that were only produced with odd harmonics, higher compared to complex sounds generated with all harmonics and filtered with the same bandpass condition. Furthermore, the following effect was observed: The perceived pitch also changed, when Δf was fixed, but the center frequency of the bandpass filter was increased. Two arguments are relevant for the perceived pitch shift. First, the pitch of the missing fundamental disappeared, with increasing the center frequency of the filter passband. According to Zwicker and Fastl (1999), the perceived pitch at f_0 vanishes if lower harmonics of a complex sound are removed. They claimed that rather a specific combination between f_0 and the lowest component present must occur to elicit a pitch of the missing fundamental f_0. Second, it might be possible that the increasing timbre influence on the presented sounds provoked listeners to rate octave mistakes.

The neuromagnetic representation of pitch changes were identified using MEG. Earlier studies demonstrated that the latency of the N100m component evoked by the onset of a sinusoidal sound from silence depended on the frequency of the presented tone (Pantev et al., 1988). In an EEG study, Roberts and Poeppel (1996) used pure tones within a frequency range of 100 to 5,000 Hz and showed, that the latency of the N100m component increased as a function of the frequency of the stimulus tone. Ragot and Lepaul-Ercole (1996) applied complex tones with a fundamental frequency (f_0) between 100 and 330 Hz and showed that the latency of the N100 increased from about 100 ms to 120 ms with decreasing f_0. They concluded, that the N100 latency might be used as a physiological, non-subjective index of the perceived pitch.
However, Stufflebeam et al. (1998) reported, that the latency of the N100m component is increased systematically by decreasing intensity and Roberts et al. (1998) demonstrated that the latency of the N100m is also sensitive to perceptual changes of the stimulus.
All above mentioned experiments have the common aspect, that the N100(m) component was evoked by the stimulus onset from silence. Lütkenhöner (2001) reported that the N100m deflection is composed of multiple, partially temporally overlapping, independent components and Näätänen and Picton (1987) identified at least three different sources of the N100, which were probably generated over a wide region of the supratemporal plane.
In an early investigation, Clynes (1969) recorded EEG using a continuous stimulation with changing frequency to isolate the specific response to the onset of pitch

and to avoid overlapping responses evoked by the simple energy onset. Later, Mäkelä et al. (1988) used a continuous stimulation with transitions from noise to square waves. At about 100 ms after the transition, a prominent deflection (N100m') was found to be sensitive to both, pitch height and pitch salience.
Thus, a continuous stimulation with concatenated RIS segments of alternating gain was used to study the pitch-specific neuromagnetic pitch onset response (POR).
The observed latency of the N100m' component was highly correlated with the perceived pitch of the applied sounds. The neuromagnetic responses evoked by RIS generated with opposite signs of the gain factor and delay times of 2 or 4 ms produced significant latency differences of the N100m' component. However, when the delay time of the presented RIS was increased to 8 and 16 ms, the latency differences stayed equal.
The results of this experiment showed clearly that the latency of the POR does not simply represent the spectral differences of the applied RIS conditions observed at the level of the basilar membrane but is rather a neurophysiological correlate of the perceived pitch of a sound.
To test, whether perception changes evoked by shifting the chroma are also reflected in the latency of the POR, MEG measurements were extended by presenting complex sounds with a fixed Δf, analogously to the psychoacoustic experiment. In the 'even'-condition, sounds were produced with all harmonics ($\Delta f = f_0$), and in the 'odd'-condition with odd harmonics ($\Delta f = f_0/2$) only. The stimuli were concatenated with an alternation between 'even' and 'odd' and were bandpass filtered to the same frequency interval.
The perception of bandpass filtered sounds were also highly correlated with the observed latency of the N100m' component. However, the results showed, that the latency of the POR is prolonged when the center frequency of the bandpass filter was increased, but stimuli were presented at a fixed fundamental frequency. Thus, the results support strongly the hypothesis, that the N100 latency reflects the perceived pitch, rather than the applied spectrum of the presented stimuli as suggested by Ragot and Lepaul-Ercole (1996). The POR revealed significant latency differences between the presentation of 'even'- and 'odd'-conditions for a Δf of 250 Hz and for bandpass conditions of 'LOW' (500-1,000 Hz) and 'MID' (1,000–2,000 Hz). The latency difference of the 'HIGH' (2,000–4,000 Hz) filtered sounds were insignificnat. This corresponds to reports of Wiegrebe and Winter (2000). They found that an octave shift is perceived when RIS was generated with an opposite sign of g and a delay time of 4 ms (Δf =250 Hz), but they also demonstrated that removing lower harmonics, or increasing the cutoff frequency of the highpass filter caused a pitch shift of RIS produced with a negative gain

to the region of ±10% compared to RIS produced with $g = +1$ (see also Figure 2.3). The transition was found between cutoff frequencies of 1,250 Hz and 2,500 Hz. They also reported that the pitch of RIS generated with a delay time of 8 ms revealed pitch differences of about 10%, independent of the applied high-pass cutoff frequency. This is also conform with our experiment. The latency differences of the bandpass filtered sounds generated with a Δf of 125 Hz were found to be within errors, independent of the tested filter condition.

The location of the POR was fitted with two equivalent dipoles in each auditory cortex. As illustrated in Figure 6.1 (circles), the source dipoles were located in the lateral part of Heschl's Gyrus. Recently, Krumbholz et al. (2003) applied transitions from noise to RIS and reported, that the specific N100m' response arose in the medial Heschl's Gyrus. Rupp et al. (2005) found the activity of the

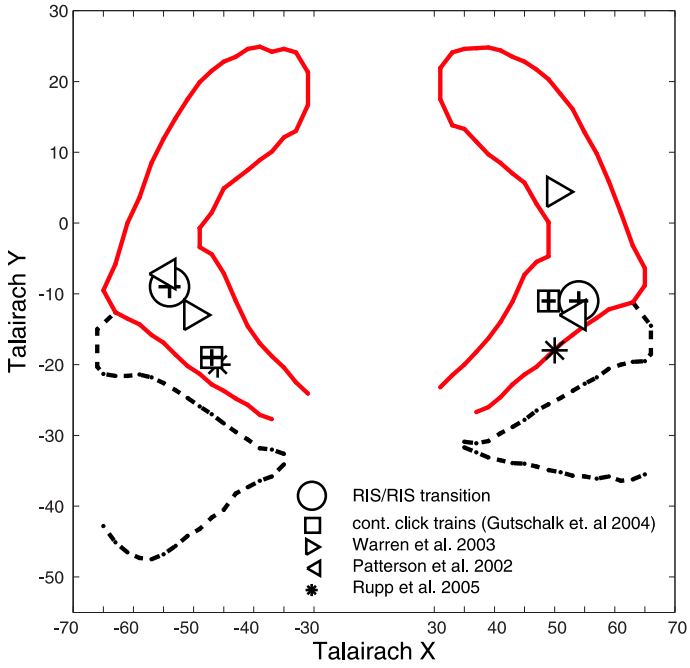

Figure 6.1: *Coordinates of activity from different experiments, projected onto an axial plain with sulcal borders provided by Schneider et al. (2004).*
The fitted dipoles of the isolated pitch specific response evoked by RIS (circles) are located in the left and right lateral Heschl's Gyrus. Mean coordinates of the averaged equivalent dipoles from MEG studies with a continuous stimulation (Gutschalk et al., 2004; Rupp et al., 2005) are also illustrated Additionally, the plot shows results of fMRI experiments: The center of activity evoked by pitch changes from Warren et al. (2003) and Patterson et al. (2002).

N100m' in the region of medial Heschl's Gyrus (depicted in the Figure with stars) for an alternating noise-RIS transition. Similarly, Gutschalk et al. (2004) applied a continuous click train stimulation with varying pitch strength. As can be seen from Figure 6.1 (squares), the dipole pair, fitted to the N100m' accounted for the pitch strength and was located in the lateral part of Heschl's Gyrus. Further validation of the observed activity in the lateral Heschl's Gyrus is given by several other neuroimaging studies. Based on functional MRI, Warren et al. (2003) reported that the medial Heschl's Gyrus is activated similarly when processing either pitch or noise. Only a small area of the lateral part of Heschl exhibited an increased activity evoked by pitch changes. The functional MRI data of Patterson et al. (2002) also revealed activity in the lateral half of Heschl's Gyrus, when pitch of RIS was varied. The mean coordinates of the activated area (S. Uppenkamp, personal communication, 2004) are also depicted in Figure 6.1. As can be seen from the Figure, the location of the equivalent source dipoles of the applied continuous stimulation with changing pitch matched exactly the activity of the MRI data, confirming (1) the applied model with two dipoles and (2) the fitted location of the equivalent source dipoles. Furthermore, Patterson et al. (2002) concluded, that the lateral Heschl's Gyrus is involved in detecting pitch changes and represents the center of melody processing in the auditory cortex. The present results support this hypothesis and beyond, demonstrated that the N100m' component is involved in pitch processing.

It has been shown in the present work that the POR represents a neurophysiological correlate of the perceived pitch. To simulate how temporal pitch information can be extracted in the auditory pathway, the auditory image model (AIM) was applied. A strong correspondence between the perceived pitch shift (due to the changing sign of the gain) of RIS and the simulation with AIM was found. The location of the first positive peak in the stabilized auditory image at time lag τ mirrored the frequency of the perceived pitch ($\frac{1}{\tau}$) for RIS generated with a positive gain. The results support the hypothesis of Yost et al. (1998), that the first prominent positive peak observed in the stabilized auditory image seems to play a major role in pitch extraction.
For RIS generated with a negative gain, the model was also able to predict the shift in the perception of RIS in dependence of the delay time d. The stabilized auditory images revealed a first positive peak at 2τ which predicts a pitch shift of an octave. Since a decaying buffer with a half life time of 20 ms is integrated in the model, increasing the delay time of the simulated RIS results in a smaller amplitude of the peak at 2τ. At the same time, the observed two side maxima

at $\tau = \frac{1}{0.9d}$ and $\tau = \frac{1}{1.1d}$ become more prominent in comparison to the peak at 2τ and can account for the ambiguity of the pitch. The location of the peaks represent exactly the perceived pitch shift of 10% between RIS with opposite sign of g. Thus, the temporal pitch model is able to predict the octave shift as well as the ambiguous pitch of RIS generated with a negative gain.
The bandpass filtered sounds were also simulated with AIM to test, whether the model can also extract the timbre-induced shift in the perception and the corresponding latency change of the POR. The outcome of the simulated sounds generated with all harmonics ($\Delta f = f_0$) revealed a peak at time lag $\tau = \frac{1}{\Delta f}$. Independent of the applied bandpass condition, the stabilized auditory images exhibited the first prominent peak at time lag $\tau = f_0$. However, as revealed by the psychoacoustic experiments of the present work and according to the above mentioned existence region of the missing fundamental (Zwicker and Fastl, 1999), pitch at f_0 was not perceived anymore, when sufficient lower harmonics were removed from the sounds.
As expected from the simulation of RIS generated with a negative gain, the complex sounds generated with the 'odd'-condition ($\Delta f = f_0/2$) revealed a minimum at time lag τ. Analogously, the first prominent peak, that predicts the pitch of the sound can be either one of the side maxima that flank the minimum at τ, or the peak at time lag 2τ. In the first case, the predicted pitch varies between a tone and a semitone (up or down) compared to the 'even'-condition. In the second case, the predicted pitch shift is an octave. However, the psychoacoustic measurements revealed, that the perceived pitch of the 'odd'-condition was always rated higher than the pitch of the corresponding 'even'-condition.
AIM failed to predict the pitch shift, when the influence of spectral envelope increased.

In the present work, two effects were observed that influenced the POR latency in opposite direction. When RIS was applied, the latency of the N100m' deflection was highly correlated to the perceived pitch of RIS. The latency of the N100m' increased when the perceived pitch of RIS decreased. Furthermore, the latency of the POR depended on the pitch strength of RIS. With increasing number of iterations n, the latency also decreased. In the experiment where bandpass filtered sounds were applied, the fundamental frequency was fixed, but the center frequency of the bandpass filtered sound was increased. Removing lower harmonics of the presented sounds resulted in an increase of the N100m'-latency. The latency increase was also highly correlated to the perceived pitch of the psychoacoustic experiment conducted with bandpass filtered sounds.
Two effects can account for this result. First, it could be that listeners were

mainly influenced by timbre changes and rated octave mistakes. However, the stronger argument is given secondly by the harmonic resolvability within the auditory filters. The number of harmonics of each filter increases with increasing harmonic number and begin to interact. The latency increase of the POR might be due to the presentation of less resolved harmonics which prolong the pitch processing, i.e. the extraction time of pitch in the auditory system. This effect was also observed when RIS was presented. With a fixed delay time d and fixed gain g, the latency of the POR was also increased with decreasing number of iterations. For RIS generated with two iterations, the noise component of the sound is still prominent but the tone/noise ratio increases with increasing n. Thus, the extraction of a prominent pitch is easier, which resulted in a shorter latency of the N100m'. In summary, the POR rather reflects the evoked sensation of a sound, than the spectrum.

The present work settled the controversial debate on the pitch of RIS generated with a delay-and-subtract process. The results of the psychoacoustic studies are in line with the reported findings of Raatgever and Bilsen (1992) and Wiegrebe and Winter (2000). Furthermore, the POR was found to represent a neurophysiological correlate of the perceived pitch.
It is important to note that the observed perceptual pitch shift between RIS produced with a positive and negative gain factor cannot be explained by spectral auditory mechanisms. Temporal properties are able to explain the perceived pitch shift of RIS. However, the temporal pitch model failed to predict timbre induced influences on the perceived pitch. Thus, pitch shifts, effected by changing the spectrum but fixed fundamental frequency, should be taken into account in temporal models to increase their ability of correct pitch processing simulations. The neuromagnetic POR can probably provide valuable physiological parameters which could be integrated in further models, since the latency of the pitch specific response was found to integrate both, timbre and pitch.
In summary, the results of the present work support the view of Carlyon and Shackleton (1994) who proposed that two different extraction mechanisms account for the perceived pitch. Spectral cues are involved when the stimuli contain resolved harmonics, but the auditory system also relies on temporal cues for pitch extraction, regardless of resolvability, yielding some ambiguous pitch percepts.

Appendix A

Huygens (1693): Pitch of Rippled Noise

The original report by Huygens is taken from the paper by Bilsen and Ritsma (1969/70):

"Je veux ajouter ici au sujet de la réflexion du son une observation assez singulière, que j'ai fait autrefois étant à la belle maison de Chantilly de la Cour où est la statue Equestre on descend avec un degré large de [. . .] marches dans le parterre où il y a une fontaine de celles qu'on appelle gerbe d'eau, qui fait un bruit continuel. Quand on est descendu en bas et qu'on se tient entre le degré et la fontaine, on entend du coté du degré une résonance qui a un certain ton de musique qui dure continuellement, tant que la gerbe jette de l'eau. On ne savait pas d'où venait ce son ou en découvrait des causes peu vraisemblables ce qui me donna envie d'en chercher une meilleure. Je trouvai bientôt qu'il procédait de la réflexion du bruit de la fontaine contre les pierres du degré. Car, comme tout son, ou plutôt bruit, réitéré à des intervalles égaux et très petits, fait un ton de musique et que la longueur d'un tuyau d'orgue détermine le ton qu'il a par sa longueur, parce que les battements de l'air arrivent également dans les petits intervalles de temps que ses ondoiements emploient à savoir quand il est fermé par le bout, ainsi je concevois que chaque bruit tant soit peu distingué que venait de la fontaine, étant réfléchi contre les marches du degré, devait arriver à l'oreille de chacune d'autant plus tard qu'elle soit plus éloignée, et cela par des différences de temps justement égales à celle que les ondoiements de l'air emploient à aller et venir autant qu'était la largeur d'une marche. Ayant mesuré cette largeur qui est de 17 pouces, je fis un rouleau de papier qui avait cette longueur, et je trouvai qu'il avait le même ton qu'on entendait au bas du degré. Je trouvai, comme j'ai dit, que la gerbe n'allant point l'on cessait d'entendre ce ton. Et ayant eu l'occasion

d'aller à Chantilly pendant l'hiver, qu'il était tombé beaucoup de neige qui ôtait la forme aux marches, je remarquai qu'on entandait rien quoique la gerbe allât et fît du bruit à l'ordinaire."

Appendix B

Pitch Matching Results of RIS

The pitch matching results of RIS according to the described method in Chapter 4. In the results section of Chapter 4, only the mean of the subjects who matched all RIS sounds is shown. Here, the results of all single subjects is illustrated. Yost (1996a) reported that some subjects had difficulties to perform the task, which was also observed in the present task.

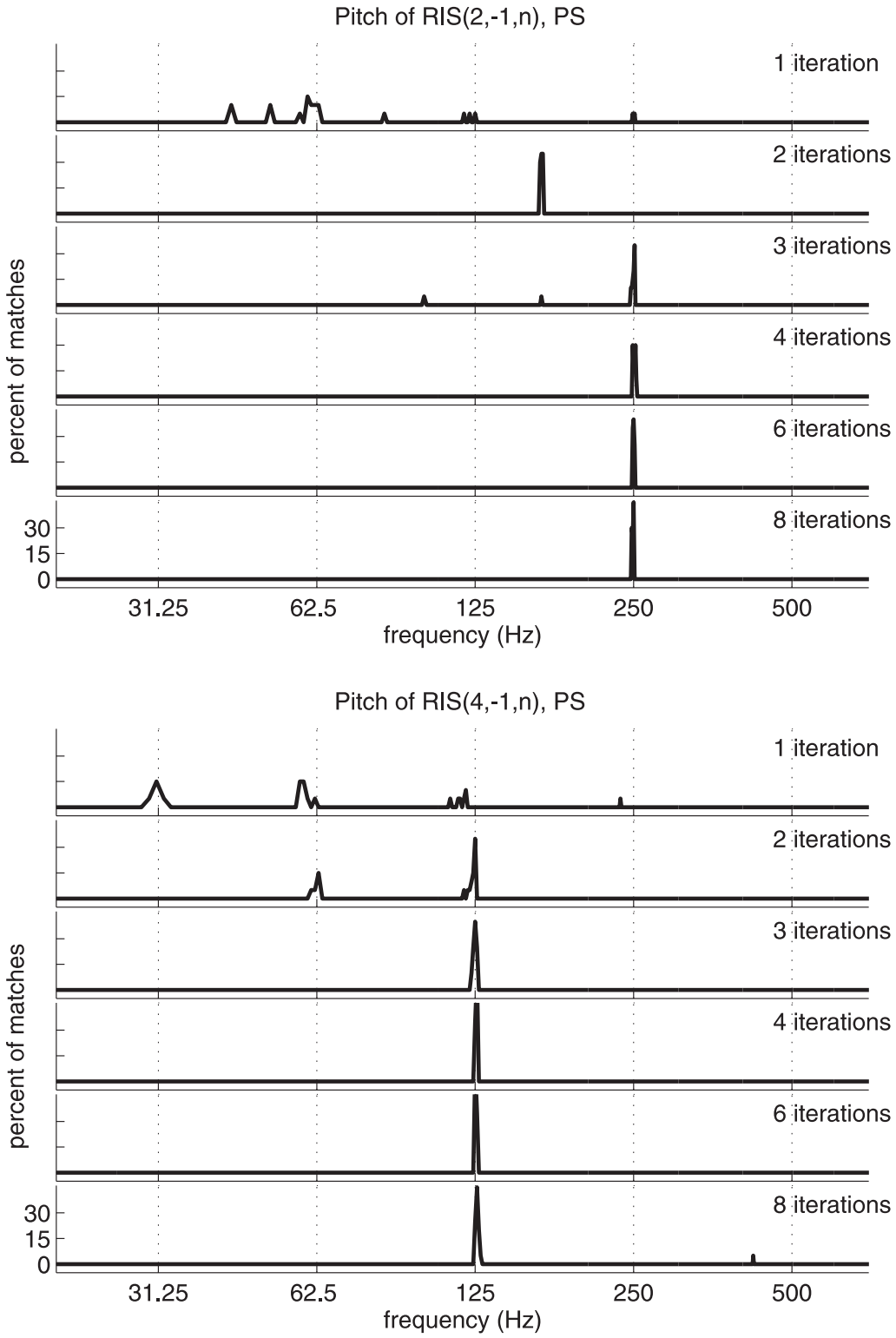

Figure B.1: *Subject PS: Matched pitch of RIS generated with a delay of 2 ms and 4 ms.*

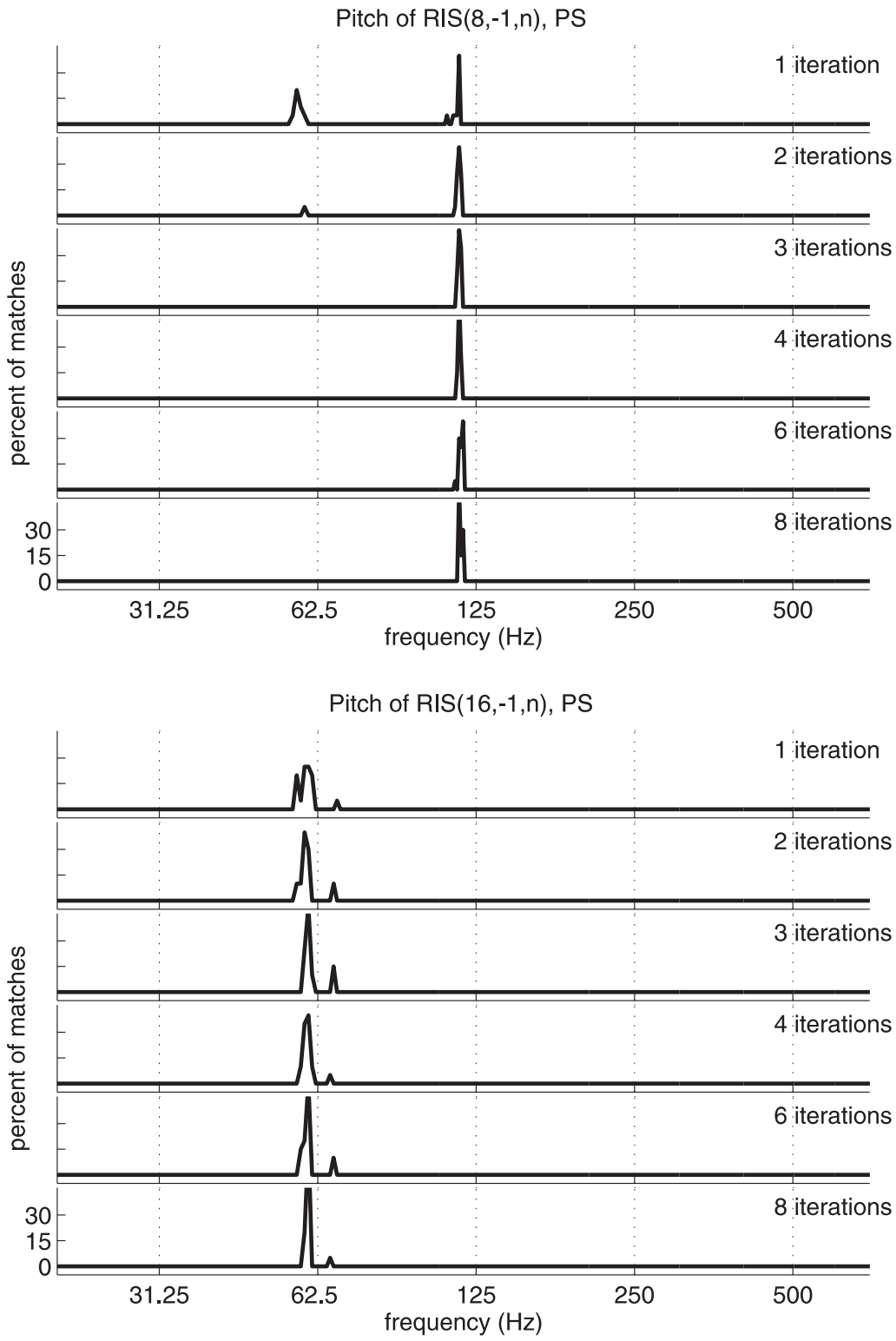

Figure B.2: *Subject PS: Matched pitch of RIS generated with a delay of 8 ms and 16 ms.*

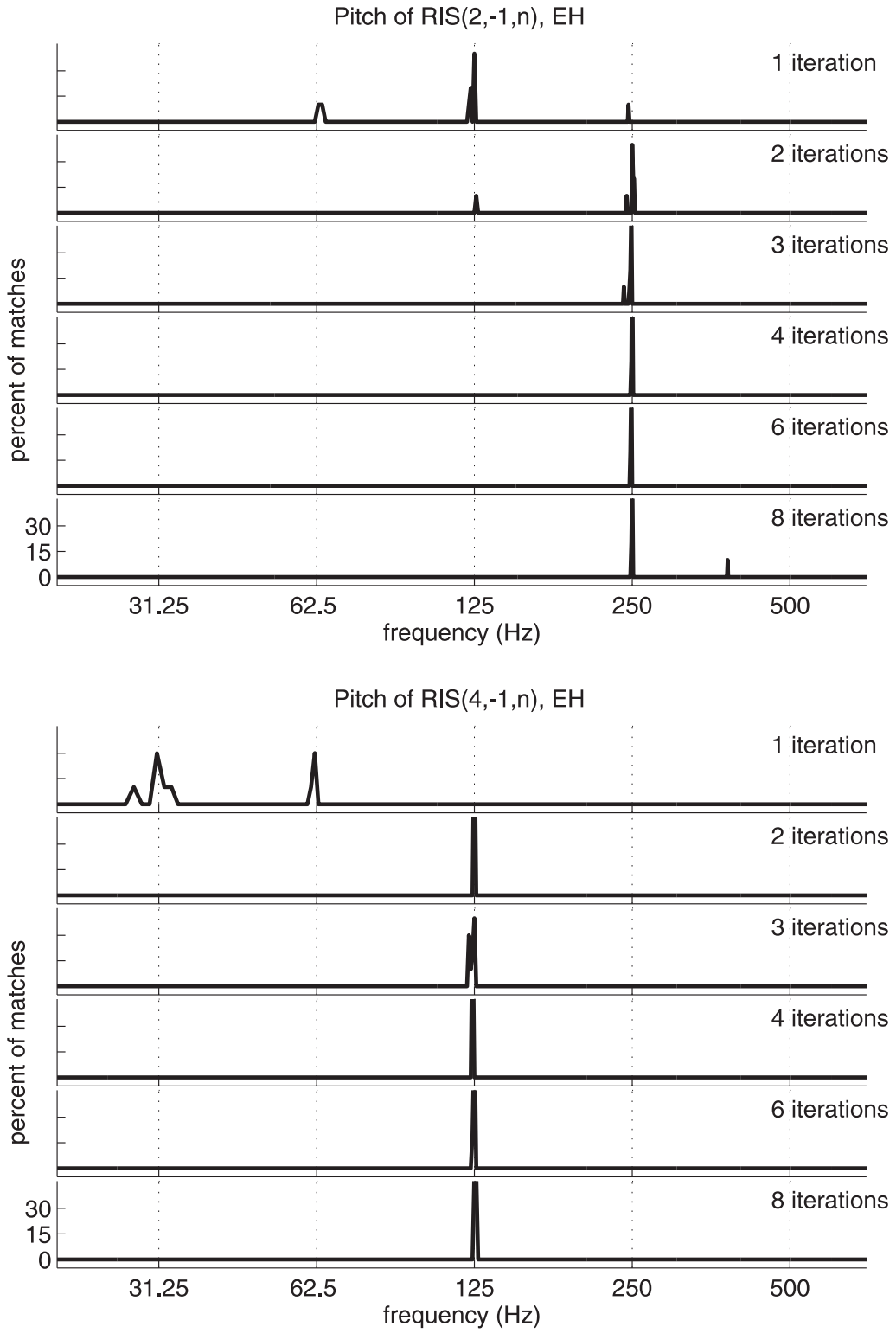

Figure B.3: *Subject EH: Matched pitch of RIS generated with a delay of 2 ms and 4 ms.*

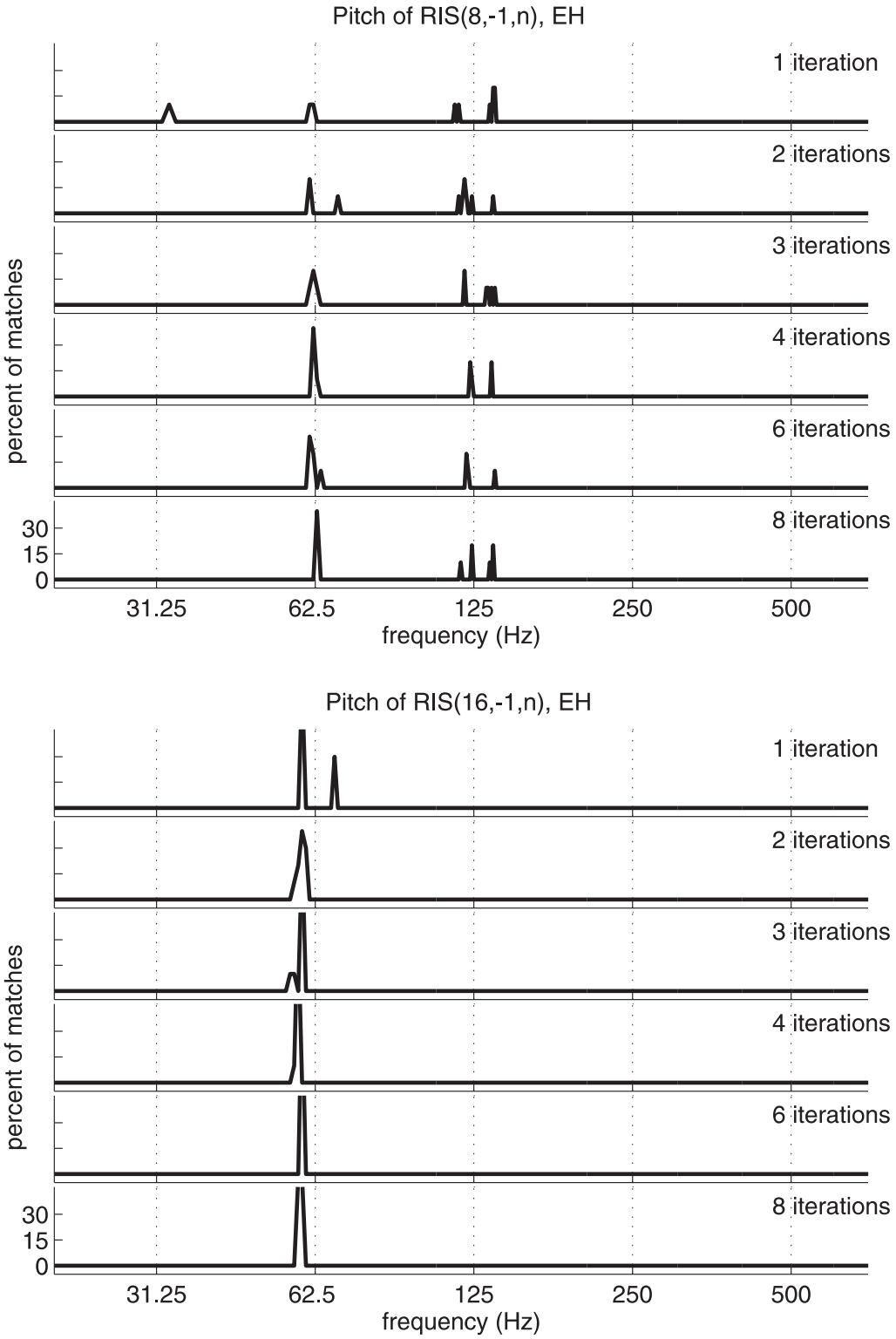

Figure B.4: *Subject EH: Matched pitch of RIS generated with a delay of 8 ms and 16 ms.*

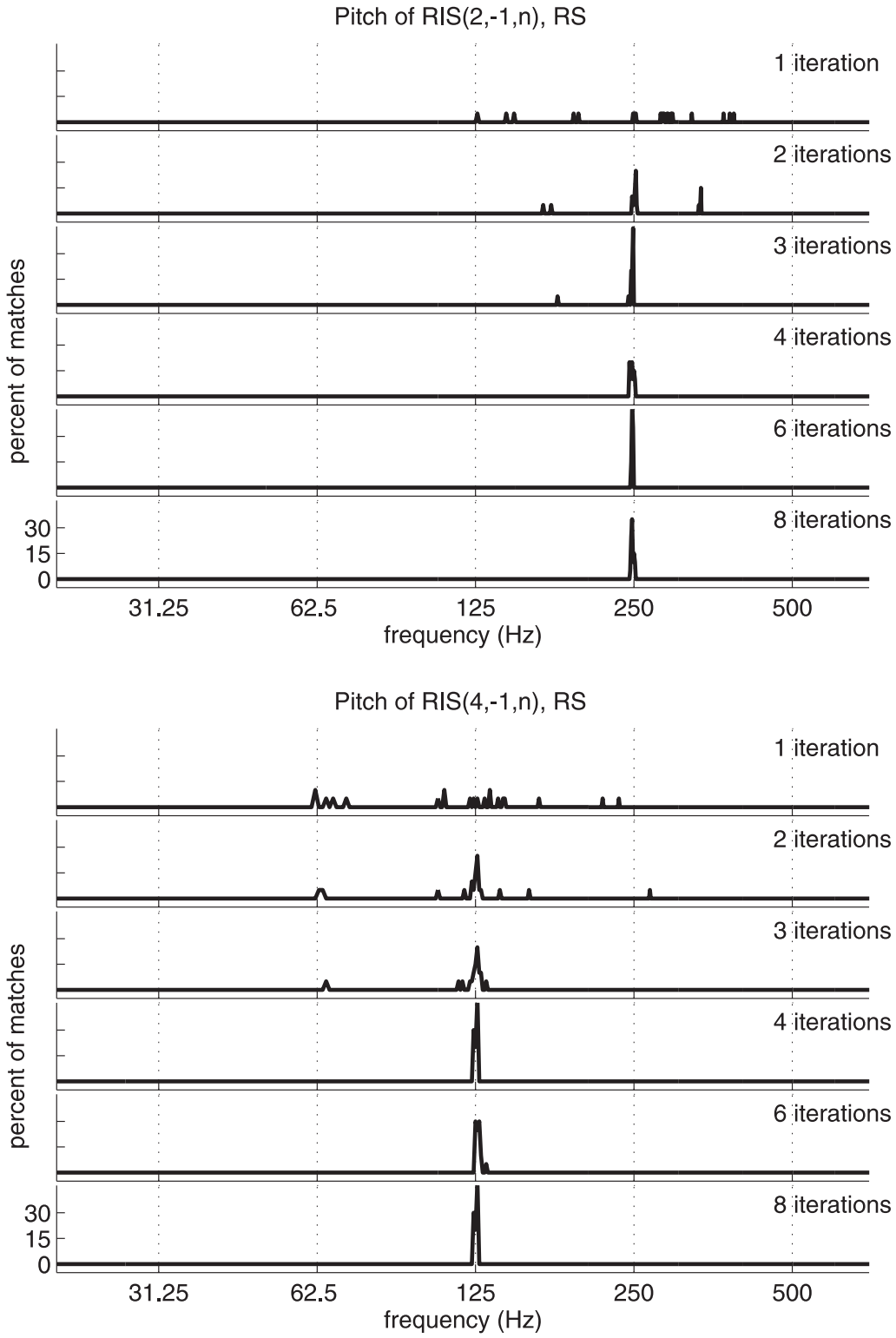

Figure B.5: *Subject RS: Matched pitch of RIS generated with a delay of 2 ms and 4 ms.*

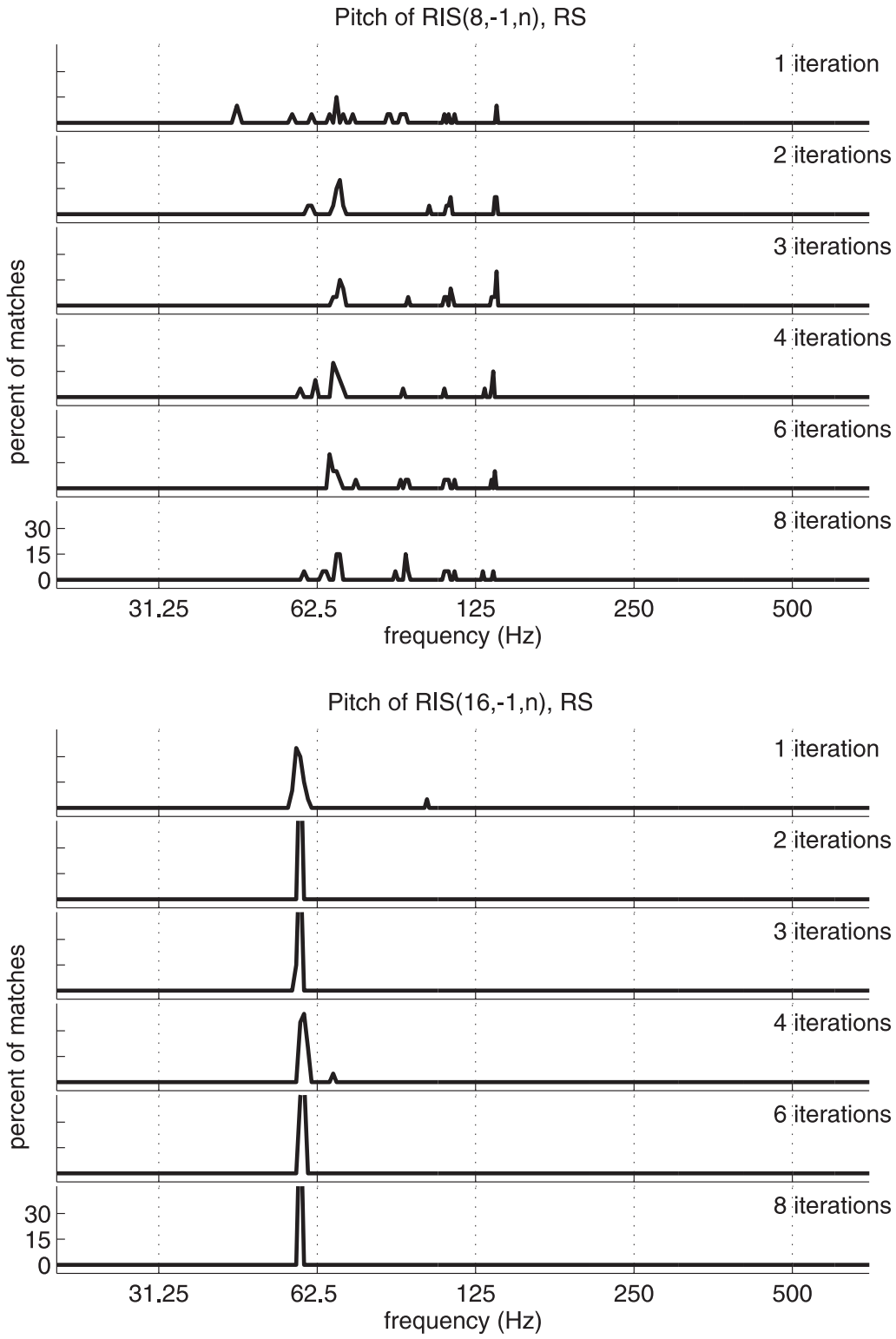

Figure B.6: *Subject RS: Matched pitch of RIS generated with a delay of 8 ms and 16 ms.*

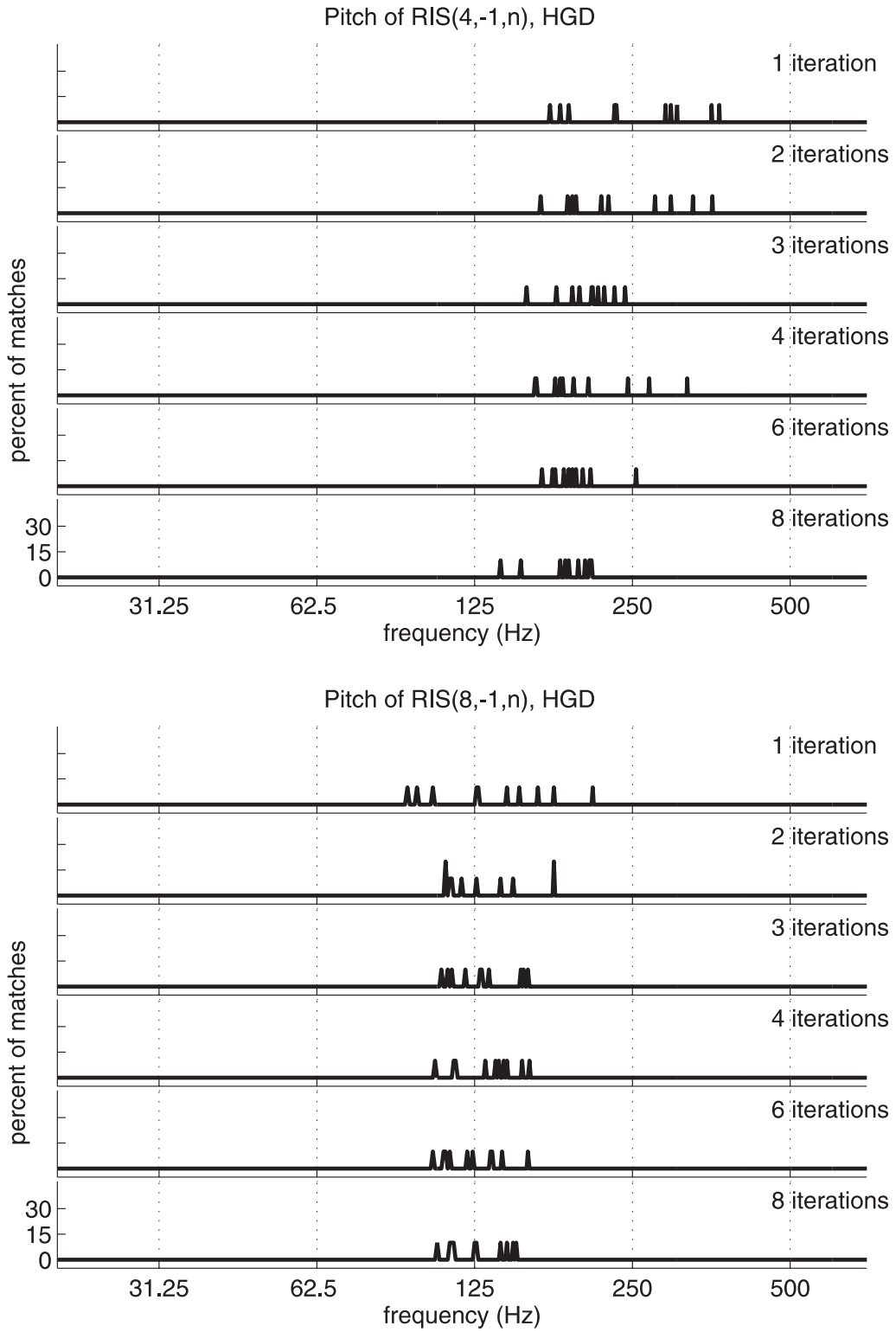

Figure B.7: *Subject HGD: Matched pitch of RIS generated with a delay of 4 ms and 8 ms.*

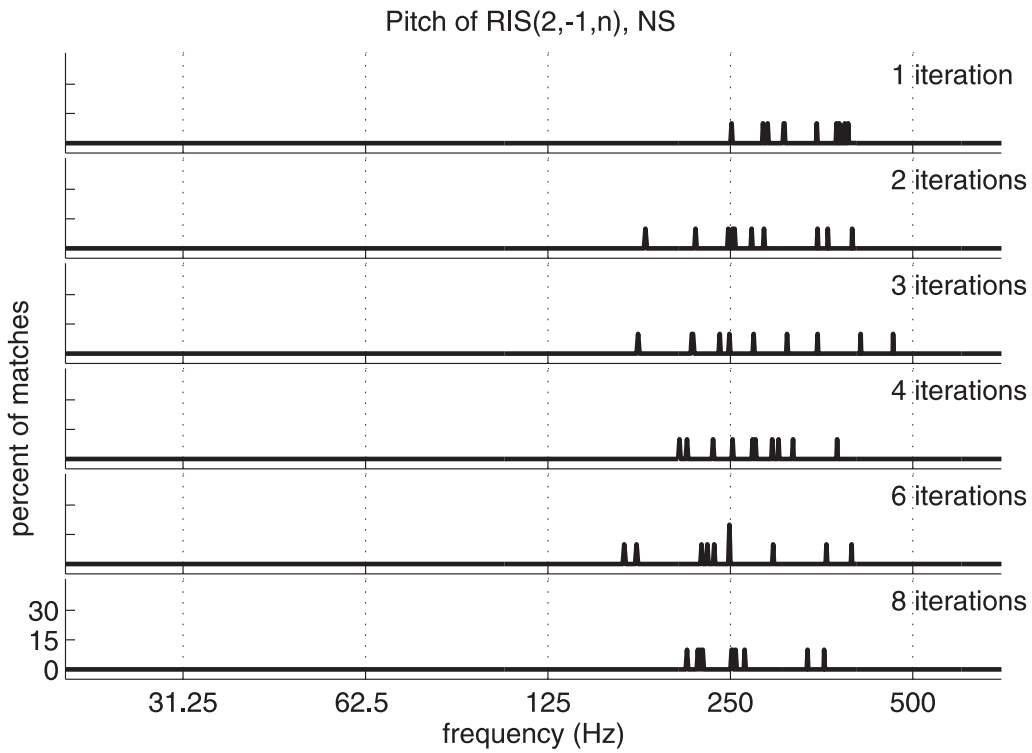

Figure B.8: *Subject NS: Matched pitch of RIS generated with a delay of 2 ms.*

Bibliography

Berg, P., Scherg, M., 1994. A multiple source approach to the correction of eye artifacts. *Electroencephalogr. Clin. Neurophysiol.* 90: 229-241.

Bernstein, J. G. and Oxenham, A. J., 2003. Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number? *J. Acoust. Soc. Am.* 113: 3323-3334.

Bilsen, F. A., 1966. Repetition Pitch: Monaural interaction of a Sound with the Repetition of the Same, but Phase Shifted Sound. *Acustica* 17: 295-300.

Bilsen, F. A. and Ritsma, R. J., 1969/70. Repetition Pitch and Its Implication for Hearing Theory. *Acustica* 22: 63-73.

Bilsen, F. A., ten Kate, J. H., Buunen, T.J.F., and Raatgever, J., 1975. Responses of single units in the cochlear nucleus of the cat to cosine noise. *Acustica* 58: 858-866.

Bradley, R. A. and Terry, M. E., 1952. Rank analysis of incomplete block designs. I. the method of paired comparisons. *Biometrika* 39: 324-345.

Brugge, J. F., 1985. Patterns of organization in auditory cortex. *J. Acoust. Soc. Am.* 78: 353-359.

Campain, R. and Minckler, J., 1976. A note on the gross configurations of the human auditory cortex. *Brain Lang.* 3: 318-323.

Cariani, P., 1999. Temporal Coding of Periodicity Pitch in the Auditory System: An Overview. *Neural Plasticity* 6: 147-172.

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

Clynes, M., 1969. Dynamics of vertex evoked potentials: The R-M brain function. In: *Average evoked potentials: Methods, results and evaluations edited by* E.Donchin & D.B. Lindsley, US Government Printing Office, Washington, (NASA SP-191), 363-374.

David, H. A., 1988. *The method of paired comparisons.* 2nd edn., Oxford University Press, New York.

Efron, B. and Tibshirani, R. J., 1993. *An Introduction to the Bootstrap.* Chapman and Hall, New York.

Eggermont, J. J. and Ponton, C. W., 2002. The Neurophysiology of Auditory Perception: From Single Units to Evoked Potentials. *Audiol. Neurootol.* 7: 71-99.

Fastl, H., 1988. Pitch and pitch strength of peaked rippled noise. In: *Basic Issues in Hearing, edited by* Duifhuis, H., Horst, J. W., and Wit, H. P., Academic Press, San Diego, 1988, 370-379.

Fay, R. R., Yost, W. A., and Coombs, S., 1983. Psychophysics and neurophysiology of repetition noise processing in a vertebrate auditory system. *Hearing Res.* 12: 31-55.

Fechner, G. T., 1860/1965. *Elements of Pschophysics.* (Adler, H. E., Trans., Vol. I).) Holt, Rinehart and Winston, New York.

Galaburda, A., and Sanides, F., 1980. Cytoarchitectonic organization of the human auditory cortex. *J. Comp. Neurol.* 190: 597-610.

Giguère, C. and Woodland, P. C., 1994. A computational model of the auditory periphery for speech and hearing research. I. Ascending paths. *J. Acoust. Soc. Am.* 95: 331-342.

Giguère, C. and Woodland, P. C., 1994. A computational model of the auditory periphery for speech and hearing research. I. Descending paths. *J. Acoust. Soc. Am.* 95: 343-349.

Glasberg, B. R. and Moore, B. C. J., 1990. Derivation of auditory filter shapes from notched-noise data. *Hearing Research* 47: 103-138.

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.

Griffiths, T. D., Büchel, C., Frackowiak, R. S. J., and Patterson, R. D., 1998. Analysis of the temporal structure in sound by the human brain. *Nat. Neurosci.* 1: 422-427.

Griffiths, T. D., Uppenkamp, S., Johnsrude, I., Josephs, O., and Patterson, R. D., 2001. Encoding of the temporal regularity of sound in the human brainstem. *Nat. Neurosci.* 4: 633-637.

Gutschalk, A., Patterson, R. D., Rupp, A., Uppenkamp, S., and Scherg, M., 2002. Sustained Magnetic Fields Reveal Separate Sites for Sound Level and Temporal Regularity in Human Auditory Cortex. *NeuroImage* 15: 207-216.

Gutschalk, A., Patterson, R. D., Scherg, M., Uppenkamp, S., and Rupp, A., 2004. Temporal dynamics of pitch human auditory cortex. *NeuroImage* 22: 755-766.

Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., and Lounasmaa, O. V., 1993. Magnetoencephalography – theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics* 65(2): 413-497.

Helmholtz, H. L. F. von, 1853. Über einige Gesetze der Vertheilung elektrischer Ströme in körperlichen Leitern, mit Anwendung auf die thierisch-elektrischen Versuche. *Annalen der Physik und Chemie,* 89: 211-233, 353-377.

Helmholtz, H. L. F. von, 1863. *Die Lehre von Tonempfindungen als physiologische Grundlage für die Theorie der Musik.* F. Vieweg & Sohn, Braunschweig.

Hoechstetter, K., 2001. *Magnetic Source Imaging of Tactile Evoked Activity in the Human Secondary Somatosensory Cortex.* PhD Thesis, Faculty of Physics, Heidelberg.

Hoekstra, A., 1979. *Frequency discrimination and frequency analysis in hearing.* PhD thesis, University of Groningen.

Holdsworth, J. W. and Patterson, R. D., 1993. Analysis of waveforms. UK Patent No. GB 2-234-078-B, UK Patent Office, London.

Houtsma, A. J. M. and Goldstein, J. L., 1972. The central origin of the pitch of complex tones: Evidence from musical interval recognition. *J. Acoust. Soc. Am.* 51: 520-529.

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

Houtsma, A. J. M. and Fleuren, J. F. M., 1991. Analytic and synthetic pitch of two-tone complexes. *J. Acoust. Soc. Am.* 90: 1674-1676.

Houtsma, A. J. M., 1995. In: *Hearing,* edited by Moore, B. C. J., Academic Press, London, 1995, 267-295.

Huygens, C., 1693. *Oeuvres Completes.* Vol. 10: 570.

Ilmoniemi, R. J., 1993. Models of source currents in the brain. *Brain Topogr.* 5: 331-336.

Kandel, E. R. and Schwartz, J. H., 1985. *Principles of neural sciences.*, Elsevier, New York.

Krumbholz, K., Patterson, R. D., Seither-Preisler, C., Lammertmann, C., and Lütkenhöner, B., 2003. Neuromagnetic Evidence for a Pitch Processing Center in Heschl's Gyrus. *Cereb. Cortex* 13: 765-772.

Kulynych, J. J., Vladar, K., Jones, D. W., and Weinberger, D. R., 1994. Gender differences in the normal lateralization of the supratemporal cortex: MRI surface-rendering morphometry of Heschl's gyrus and the planum temporale. *Audiol. Neurootol.* 7: 71-99.

Leonard, C. M., Puranik, C., Kuldau, J. M., and Lombardino, L. J., 1998. Normal variation in the frequency and location of human auditory cortex landmarks. Heschl's gyrus: Where is it? *Cereb. Cortex* 8: 397-406.

Licklider, J. C. R., 1951. The duplex theory of pitch perception. *Experientia* 7: 128-133.

Liberman, M. C., 1978. Auditory nerve response from cat raised in a low-noise chamber. *J. Acoust. Soc. Am.* 63: 442-455.

Liégeois-Chauvel, C., Musolino, and Chauvel, P., 1991. Localization of the primary auditory area in man. *Brain.* 114: 139-153.

Liégeois-Chauvel, C., Musolino, A., Badier, J. M., Marquis, P., and Chauvel, P., 1994. Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. *Electroencephalogr. clin. Neurophysiol.* 92: 204-214.

Luce, D., 1959. *Individual choice behavior.* John Wiley & Sons, New York.

Lütkenhöner, B., and Steinsträter, O., 1998. High-precision neuromagnetic study of the functional organization of the human auditory cortex. *Audiol. Neurootol.* 3: 191-213.

Lütkenhöner, B., Lammertmann, C., and Knecht, S., 2001. Latency of Auditory Evoked Field Deflection N100m Ruled by Pitch or Spectrum? *Audiol. Neurootol.* 6: 263-278.

Mäkelä, J. P., Hari, R., and Leinonen, L., 1988. Magnetic responses of the human auditory cortex to noise/square wave transitions. *Electroencephalogr. clin. Neurophysiol.* 69: 423-430.

Mäkelä, J. P., Hämäläinen, M., Hari, R., and McEvoy, L., 1994. Whole-head mapping of middle-latency auditory evoked magnetic fields. *Electroencephalogr. clin. Neurophysiol.* 92: 414-421.

Meddis, R., 1988. Simulation of auditory-neural transduction: Further studies *J. Acoust. Soc. Am.* 83: 1056-1063.

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

Merzenich, M. M. and Brugge, J. F., 1973. Representation of the cochlear partition on the superior temporal plane of the macaque monkey. *Brain Res.* 50: 275-296.

Moore, B. C. J. and Rosen, S. M., 1979. Tune recognition with reduced pitch and interval information. *Quaterly Journal of Experimental Psychology* 31: 229-240.

Moore, B. C. J. and Peters, R. W., 1992. Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity. *J. Acoust. Soc. Am.* 91: 2881-2893.

Morel, A., Garraghty, P. E., and Kaas, J. H., 1993. Tonotopic organization, architectonic fields and connections of auditory cortex in macaque monkeys. *J. Comp. Neurol.* 335: 437-459.

Morgan, C. T., Garner, W. R., and Galambos, R., 1951. Pitch and intensity. *J. Acoust. Soc. Am.* 23: 658-663.

Näätänen, R. and Picton, T., 1987. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* 24: 375-425.

Ohm, G. W., 1843. Über die Definition des Tones nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen. *Annalen für Physik und Chemie* 59: 513-565.

Pantev, C., Hoke, M., Lehnertz, K., Lütkenhöner, B., Anogianakis, G., and Wittkowski, W., 1988. Tonotopic organization of the human auditory cortex revealed by transient auditory evoked magnetic fields. *Electroencephalogr. clin. Neurophysiol.* 69: 160-170.

Patterson, R. D. and Moore, B. C. J., 1986. Auditory filters and excitation patterns as representations of frequency resolution.
In: *Frequency selectivity in Hearing edited by* B.C.J. Moore, Academic PressLimited, London, 123-177.

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

Patterson, R. D., Allerhand, M., 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.

Patterson, R. D., Handel, S., Yost, W. A., and Datta, A. J., 1996. The relative strength of tone and noise components in iterated rippled noise. *J. Acoust. Soc. Am.* 100: 3286-3294.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., and Griffith, T. J., 2002. The processing of Temporal Pitch and Melody Information in Auditory Cortex. *Neuron* 36: 767-776.

Pick, G. F., 1980. Level dependence of psychophysical frequency resolution and auditory filter shape. *J. Acoust. Soc. Am.* 68: 1085-1095.

Picton, T. W., Hillyard, S. A., Krausz, H. I., and Galambos, R., 1974. Human auditory evoked potentials. I: evaluatiuon of components. *Electroencephalogr. Clin. Neurophysiol.* 36: 179-190.

Plomp, R., 1967. Pitch of complex tones. *J. Acoust. Soc. Am.* 41: 1526-1533.

Popper, A. N. and Fay, R. R., 1992. *The mammalian auditory pathway: Neurophysiology.*, Springer, New York.

Pressnitzer, D., Patterson, R. D., and Krumbholz, K., 2001. The lower limit of melodic pitch. *J. Acoust. Soc. Am.* 109: 2074-2084.

Raatgever, J. and Bakkum, M. J., 1986. 'Spectral dominance for noise signals with monaural and dichotic comb spectra'. *Proc. 12th Int. Congr. Acoust.,* Toronto, B2-4.

Raatgever, J. and Bilsen, F. A., 1992. The pitch of anharmonic comb filtered noise reconsidered. In *Auditory Physiology and Perception, edited by* Y. Cazals, L. Demany, and K. Horner, Pergamon, Oxford.

Ragot, R. and Lepaul-Ercole, R., 1996. Brain potentials as objective indexes of auditory pitch extraction from harmonics. *Neuroreport* 7: 905-909.

Roberts, T. P. and Poeppel, D., 1996. Latency of auditory evoked M100 as a function of tone frequency. *Neuroreport* 7: 1138-1140.

Roberts, T. P., Ferrari, P., and Poeppel, D., 1998. Latency of evoked neuromagnetic M100 reflects perceptual and acoustic stimulus attributes. *Neuroreport* 9: 3265-3269.

Roberts, T. P., Ferrari, P., Stufflebeam, S., and Poeppel, D., 2000. Latency of the auditory evoked neuromagnetic field components: Stimulus dependence and insights toward perception. *J. Clin. Neurophysiol.* 17: 114-129.

Reite, M., Edrich, J., Zimmermann, J. T., and Zimmermann, J. E., 1978. Human magnetic auditory evoked fields. *Electroencephalogr. clin. Neurophysiol.* 45: 114-117.

Reite, M., Teale, P., Zimmermann, J. T., Davis, K., and Whalen, J., 1988. Source location of a 50 ms latency auditory evoked field component. *Electroencephalogr. clin. Neurophysiol.* 92:149-160.

Ritsma, R. J., 1967. The existence region of the tonal residue. *J. Acoust. Soc. Am.* 34: 1224-1229.

Ritsma, R. J., 1967. Frequencies dominant in the perception of the pitch of complex sounds. *J. Acoust. Soc. Am.* 42: 191-198.

Romani, G. L., Williamson, S. J., and Kaufman, L., 1982. Tonotopic organization of the human auditory cortex. *Science* 216: 1339-1340.

Rouiller, E. M., and de Ribaupierrre, F., 1989. Note on the tonotopic organization in the cat medial geniculate body: influence of sampling of units. *Exp. Brain Res.* 74: 220-226.

Rouiller, E. M., and Welker, E., 1991. Morpholgy of corticothalamic terminals arising from the auditory cortex of the rat: a phaseolus vulgaris-leucoagglutinin (PHA-L) tracing study. *Hearing Research* 56: 179-190.

Rupp, A., Uppenkamp, S., Bailes, J., Gutschalk, A., and Patterson, R., 2005. Time constants in temporal pitch extraction: a comparison of psychophysical and neuromagnetic data. In: *Auditory signal processing: physiology, psychoacoustics, and models, edited by* D. Pressnitzer, A. de Cheveigne, S. McAdams, L. Collet, Springer, New York.

Rutherford, E., 1886. A new theory of hearing. *J. Anat. Physiol.* 21: 166-168.

Scheich, H., Heil, P., and Langner, G., 1993. Functional organization of auditory cortex in the mongolian gerbil (Meriones unguiculatus). II. Tonotopic 2-deoxyglucose. *Europ. J. Neurosci.* 5: 898-914.

Scherg, M., and von Cramon, D., 1986. Evoked dipole source potentials of the human auditory cortex. *Electroencephalogr. clin. Neurophysiol.* 65: 344-360.

Scherg, M., Vajsar, J., and Picton, T. W., 1989. A source analysis of the human auditory evoked potentials. *J. Cognit. Neurosci.* 1: 336-354.

Scherg, M., 1991. *Akustisch evozierte Potentiale: Grundlagen - Entstehungsmechanismen - Quellenmodell.* Kohlhammer, Stuttgart.

Schneider, P., Sluming, V., Roberts, N., Bleeck, S., and Rupp, A., 2003. Gray matter volume asymmetry in antero-lateral Heschl Gyrus is related to instrumental preference in musicians. *Proceedings of the International Conference on auditory cortex, Magdeburg.*

Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., and Rupp, A., 2004. Tonotopy, Tonochrony and Pitch Mapping in Heschl's Gyrus of Musicians. *Proceedings of the 14th International Conference on Biomagnetism (BIOMAG),* 570.

Schouten, J. F., 1940. The residue and the mechanism of hearing. *Proceedings of the Koninklijke Akademie van Wetenschap* 43: 991-999.

Schouten, J. F., Ritsma, R.J., and Cardozo, B.L., 1962. Pitch of the residue. *J. Acoust. Soc. Am.* 34: 1418-1424.

Seebeck, A., 1841. Beobachtungen über einige Bedingungen der Entstehung von Tönen. *Annalen für Physik und Chemie* 53: 417-436.

Shackleton, T. M. and Carlyon, R. C., 1994. The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *J. Acoust. Soc. Am.* 95: 3529-3540.

Shofner, W. P., 1991. Temporal representation of ripple noise in the anteroventral cochlear nucleus of the chinchilla. *J. Acoust. Soc. Am.* 90: 2450-2466.

Shofner, W. P., 1999. Responses of cochlear nucleus units in the chinchilla to iterated rippled noises: quantitative analysis of neural autocorrelograms. *J. Neurophysiol.* 81: 2662-2674.

Sieroka, N., 2004. *Neurophysiological Aspects of Time Perception.* PhD Thesis, Faculty of Physics, Heidelberg.

Smoorenburg, G. F., 1970. Pitch perception of two-frequency stimuli. *J. Acoust. Soc. Am.* 48: 924-941.

Stevens, S. S., Volkmann, J., and Newman, E. B., 1937. A scale for the measurement of the psychological magnitude of pitch. *J. Acoust. Soc. Am.* 8: 185-190.

Stufflebeam, S. M., Poeppel, D., Rowley, H. A., and Roberts, T. P., 1998. Peri-threshold encoding of stimulus frequency and intensity in the M100 latency. *Neuroreport* 9: 91-94.

Talairach, P., Tournoux, J., 1988. *A Stereotactic Coplanar Atlas of the Human Brain.* Thieme, Stuttgart.

Terhardt, E. and Fastl, H., 1971. Zum Einfluss von Störtönen und Störgeräuschen auf die Tonhöhe von Sinustönen. *Acustica* 25: 53-61.

Terhardt, E., 1979. Calculating virtual pitch. *Hearing Research* 1: 155-182.

Warren, J. D., Uppenkamp, S., Patterson, R. D., and Griffiths, T. D., 2003. Separating pitch chroma and pitch height in the human brain. *PNAS* 100: 10038-10042.

Wiegrebe, L. and Winter, I., 2000. Psychophysics and physiology of regular-interval noise: critical experiments for current pitch models and evidence for a first-order temporal pitch code. In *Physiological and Psychophysical Bases of Auditory Function - 12th International Symposium on Hearing, edited by* DJ Breebart, AJM Houtsma, A Kohlrausch, VF Prijs, R Schoonhoven. Shaker, Maastricht, 121-128.

Wightman, F. L., 1973. The pattern-transfomation model of pitch. *J. Acoust. Soc. Am.* 54: 407-416.

Yost, W. A. and Hill, R., 1978. Strength of the pitches associated with ripple noise. *J. Acoust. Soc. Am.* 64: 485-492.

Yost, W. A., Hill, R., and Perez-Falcon, T., 1978. Pitch and pitch discrimination of broadband signals with rippled power spectra. *J. Acoust. Soc. Am.* 63: 1166-1173.

Yost, W. A., Allerhand, M., Robinson, K., and Patterson, R., 1993. The Pitch and Pitch Strength of Iterated Rippled Noise. *Assoc. Res. Otol.* Abstr., 47.

Yost, W. A., Sheft, S., Shofner, B., and Patterson, R., 1995. A temporal account of complex pitch. In: *Proc. ATR Workshop on a Biological Framework for Speech Perception and Production, edited by* H. Kawahara, Kyoto, Japan.

Yost, W. A., Patterson, R. D., and Sheft, S., 1996. A time domain description for the pitch strength of iterated rippled noise. *J. Acoust. Soc. Am.* 99: 1066-1078.

Yost, W. A., 1996a. Pitch of iterated rippled noise. *J. Acoust. Soc. Am.* 100: 511-518.

Yost, W. A., 1996b. Pitch strength of iterated rippled noise. *J. Acoust. Soc. Am.* 100: 3329-3335.

Yost, W. A., 1997. Pitch strength of iterated rippled noise when the pitch is ambiguous. *J. Acoust. Soc. Am.* 101: 1644-1650.

Yost, W. A., Patterson, R. D., and Sheft, S., 1998. The role of the envelope in processing of iterated rippled noise. *J. Acoust. Soc. Am.* 104: 2349-2361.

Zwicker, E. and Fastl, H., 1999.
Psychoacousstics–Facts and Models. Springer, Heidelberg.

Acknowledgements

It is a pleasure to thank those people who contributed in various ways to the success of this thesis:

Prof. Dr. Hans Günter Dosch and Prof. Dr. Hans-Joachim Specht for their interest in the present research and for supervising this thesis. With their open minded, straight forward working style, they understood to encourage my interest in research. Nevertheless, their guidance left me the freedom to include my own ideas.
Thanks to Prof. Dr. Michael Scherg for being the second referee of this work.
In particular I would like to thank Dr. André Rupp who contributed in a special way to this work by creating the workplace and by initiating the possibility to conduct the research of this thesis, by spending hours on discussing, reading, re-discussing and re-reading. He and all members of the Section of Biomagnetism deserve my thanks for the friendly atmosphere that made work enjoyable. Special thanks to Barbara Burghardt, Esther Tauberschmidt and Heide Rogatzki who did all the work behind the scenes. Without their help, this thesis would not have been possible. Dr. Alexander Gutschalk, Anita Kult, Dr. Peter Schneider, and Dr. Norman Sieroka for numerous helpful discussions. Sebastian Hack, who always helped to solve hardware and software problems (on different platforms) promptly.
I owe many thanks to all who provided their brain activity to this work i.e. who dared to serve as test persons.
Thanks to Marc for reading the manuscript and for his most helpful comments.
Finally, I would like to thank my family for their constant support; my mother for always having an open ear for my problems and a good advise; my sister for her quick, spontaneous assistance whenever needed; Caroline for being there and always sharing my problems for her great patience and tolerance during the last time. I hope we can do a lot more important things together in the next years, despite working on a thesis.