Pitch discrimination in optimal and suboptimal acoustic environments: electroencephalographic, magnetoencephalographic, and behavioral evidence

Nikolai Novitski

Academic dissertation to be publicly discussed, by due permission of the Faculty of Behavioral Sciences at the university of Helsinki in auditorium XII on the 8th of November, at 12 o'clock

Cognitive Brain Research Unit, Department of Psychology,
University of Helsinki, Finland

Neuroscience Unit, Institute of Biomedicine/Physiology,
University of Helsinki, Finland

Helsinki Brain Research Center, Finland

Helsinki 2006

Supervisors: Docent Mari Tervaniemi,
Cognitive Brain Research Unit,
Department of Psychology,
University of Helsinki,
Finland

Prof. Synnöve Carlson,
Faculty of Medicine,
University of Tampere
Finland

Reviewers: Docent Elina Pihko,
BioMag laboratory,
Helsinki University Central Hospital
Finland

Dr. Bertram Opitz, Assistant Professor
Experimental Neuropsychology Unit,
Department of Psychology,
Faculty of Behavioral Science
Saarland University,
Germany

Opponent: Prof. Josef Rauschecker,
Department of Physiology and Biophysics
Georgetown Institute for Cognitive and Computational Sciences
Georgetown University Medical Center
USA

ISBN 952-10-3455-6 (paperback)
ISBN 952-10-3456-4 (PDF) (http://ethesis.helsinki.fi)
ISSN: 0781-8254 (no.39)

To my father Yuri Valentinovich Novitski, PhD

ABSTRACT

Pitch discrimination is a fundamental property of the human auditory system. Our understanding of pitch-discrimination mechanisms is important from both theoretical and clinical perspectives. The discrimination of spectrally complex sounds is crucial in the processing of music and speech. Current methods of cognitive neuroscience can track the brain processes underlying sound processing either with precise temporal (EEG and MEG) or spatial resolution (PET and fMRI). A combination of different techniques is therefore required in contemporary auditory research. One of the problems in comparing the EEG/MEG and fMRI methods, however, is the fMRI acoustic noise.

In the present thesis, EEG and MEG in combination with behavioral techniques were used, first, to define the ERP correlates of automatic pitch discrimination across a wide frequency range in adults and neonates and, second, they were used to determine the effect of recorded acoustic fMRI noise on those adult ERP and ERF correlates during passive and active pitch discrimination. Pure tones and complex 3-harmonic sounds served as stimuli in the oddball and matching-to-sample paradigms.

The results suggest that pitch discrimination in adults, as reflected by MMN latency, is most accurate in the 1000-2000 Hz frequency range, and that pitch discrimination is facilitated further by adding harmonics to the fundamental frequency. Newborn infants are able to discriminate a 20% frequency change in the 250-4000 Hz frequency range, whereas the discrimination of a 5% frequency change was unconfirmed. Furthermore, the effect of the fMRI gradient noise on the automatic processing of pitch change was more prominent for tones with frequencies exceeding 500 Hz, overlapping with the spectral maximum of the noise. When the fundamental frequency of the tones was lower than the spectral maximum of the noise, fMRI noise had no effect on MMN and P3a, whereas the noise delayed and suppressed N1 and exogenous N2. Noise also suppressed the N1 amplitude in a matching-to-sample working memory task. However, the task-related difference observed in the N1 component, suggesting a functional dissociation between the processing of spatial and non-spatial auditory information, was partially preserved in the noise condition.

Noise hampered feature coding mechanisms more than it hampered the mechanisms of change detection, involuntary attention, and the segregation of the spatial and non-spatial domains of working-memory. The data presented in the thesis can be used to develop clinical ERP-based frequency-discrimination protocols and combined EEG and fMRI experimental paradigms.

AKNOWLEDGEMENTS

This work was performed in the Cognitive Brain Research Unit (CBRU), Department of Psychology and in the Neuroscience Unit, Institute of Biomedicine/Physiology, University of Helsinki. The work on the thesis was financially supported by Center for International Mobility (CIMO, Finland), Academy of Finland (project numbers 53294 and 78265) and Centre for Advanced Study, University of Leipzig (ZHS). The experimental part was performed in CBRU, in Max Planck Institute for Human Cognitive and Brain Sciences (Leipzig, Germany), and in the Hospital for Children and Adolescents of the Helsinki University Central Hospital.

I owe a great deal to my supervisors, Doc. Mari Tervaniemi and Prof. Synnöve Carlson for their support, guidance and encouragement at all stages of thesis preparation. I also express my gratitude towards the head of CBRU, Academy Prof. Risto Näätäänen, for granting me the opportunity to work in the CBRU and for his enormous patience and understanding during my progress as a graduate student. I reserve my warmest thanks to Dr. Irina Anurova, Dr. Burkhard Maess and Prof. Kimmo Alho for fruitful collaboration and educating experience. I duly appreciate the contribution of my coauthors Doc. Minna Huotilainen, Dr. Oleg Korzyukov, Dr. Sami Martinkauppi, Doc. Teemu Rinne, Prof. Hannu Aronen and Prof. Vineta Fellman. This thesis would be impossible without assistance of Marja Junnonaho, Piiu Lehmus, Sampo Antila, Marcus Kalske, Teemu Peltonen, Yvonne Wolff and Tarja Ilkka, to whom I give my thanks.

I would like also to thank all my present and former colleagues in CBRU, Neuroscience Unit and MPI for Human Cognitive and Brain Sciences for maintaining supportive atmosphere and corporative spirit. The teachers and students of the Pythagoras graduate school provided interesting discussion that promoted this thesis for which I am also thankful.

I am grateful to Prof. Josef Rauschecker for agreeing to be my opponent and Doc. Elina Pihko and Dr. Bertram Opitz for pre-examining the thesis.

Finally, I would like to thank my parents Yuri Novitski and Irina Novitskaya for their unconditional love and for emotional support that they have provided me during the preparation of the current thesis.

TABLE OF CONTENTS

LIST OF THE ORIGINAL PUBLICATIONS

The thesis is based on five publications referred to in the text by Roman numerals I-V:

I. Novitski N, Tervaniemi M, Huotilainen M, Näätänen R. (2004) Frequency discrimination at different frequency levels as indexed by electrophysiological and behavioral measures. *Cogn Brain Res*, **20**: 26-36.

II. Novitski N, Huotilainen M, Tervaniemi M, Näätänen R, Fellman V. (in press) Neonatal frequency discrimination in 250-4000 Hz range: electrophysiological evidence. *Clin Neurophysiol.*

III. Novitski N, Maess B, Tervaniemi M. (2006) Frequency specific impairment of automatic pitch change detection by fMRI acoustic noise: An MEG study. *J Neurosci Methods*, **155**: 149-159.

IV. Novitski N, Alho K, Korzyukov O, Carlson S, Martinkauppi, S, Escera C, Rinne T, Aronen HJ, Näätänen R. (2001) Effects of acoustic gradient noise from functional magnetic resonance imaging on auditory processing as reflected by event-related brain potentials. *Neuroimage*, **14**: 244-251.

V. Novitski N, Anourova I, Martinkauppi S, Aronen HJ, Näätänen R, Carlson S. (2003) Effects of noise from functional magnetic resonance imaging on auditory event-related potentials in working memory task. *Neuroimage*, **20**:1320-1328.

ABBREVIATIONS

(f)MRI – (functional) magnetic resonance imaging
ABR - auditory brainstem responses
ANC – active noise cancellation
BOLD – blood oxygenation level-dependent
BW - bandwidth
DFL – difference limens for frequency
ECD – equivalent current dipole
EEG – electroencephalography, electroencephalogram
EPI – echo-planar imaging
ERF - event-related field
ERP – event-related potential
ES – echo spacing
FA - false alarm
FDML – frequency-modulation difference limens
HG –Heschl's gyrus
HR – hit rate
IRN – iterated rippled noise
MEG – magnetoencephalography
MMN – mismatch negativity
MNE – minimum norm estimate
NIHL – noise-induced hearing loss
PET – positron-emission tomography
PFC - prefrontal cortex
PP – planum polare
PT – planum temporale
RF – radio frequency
RT – reaction time
SCD - source current distribution
STG – superior temporal gyrus
TR – time to repeat

1. INTRODUCTION

Frequency is one of the most crucial features of sound. Analysis of incoming sounds according to their frequency is performed in the auditory system at all levels of organization, from cochlea to the primary auditory cortex, and may also take place in the other cortical areas (Moore, 1997). In addition, frequency analysis gives rise to the perception of pitch that is defined as "an attribute of auditory sensation in terms of which sounds may be ordered on a musical scale" (American Standard Association, 1960; as cited in Moore, 1997, p. 177). Pitch is one of the major defining features of an auditory object (Bregman, 1990). Pitch- and frequency-discrimination studies are important from both theoretical and clinical perspectives. Although frequency discrimination has been extensively studied from the earliest days of psychophysics, there are still areas where the data is insufficient.

First, the behavioral methods that have been used in psychophysics require the active cooperation of the subject; these methods are able to estimate only the output of the whole system that is the human brain. To overcome these constraints, electroencephalography (EEG), magnetoencephalography (MEG), and, most recently, functional magnetic resonance imaging (fMRI) have also been applied to the frequency-discrimination research. Second, pure tones have been used as stimuli in the vast majority of frequency-discrimination studies. However, the majority of natural sounds have complex spectral time-varying composition, and most tonal sounds consist of several harmonics. Thus, the use of harmonic tones instead of pure tones represents a move towards a more natural acoustic stimulation. Third, the study of sound perception against a silent background is another deviation from natural conditions where the acoustical background is often polluted with irrelevant noises. In laboratory environments, the fMRI is a method in which a silent condition is not achievable. Therefore, data on frequency discrimination in suboptimal, noisy conditions are valuable for both theoretical and methodological purposes.

This thesis is a collection of studies of frequency discrimination in a wide frequency range by means of electroencephalography (EEG), magnetoencephalography (MEG), and behavioral methods. Both pure tones and spectrally complex sounds were used. Special emphasis was placed on frequency discrimination under conditions in which pre-recorded acoustic noise from fMRI experiments was presented as background sound.

1.1 The organization of the auditory system

Compared with the other sensory systems, auditory pathways are characterized by a larger number of synaptic transductions on the way from the sensory periphery to the cortex. The receptors - hair cells - are located in the cochlea, a specialized organ within the temporal bone. Hair cells are distributed along the spiral of the cochlea between the basal and tectorial membranes and are collectively known as the organ of Corti. The hair cells are activated by the displacement of the tectorial membrane. The design of the cochlea provides different maxima of displacement of the tectorial membrane for the different frequencies of sounds that come through the cochlea. As a result, the hair cells at different loci along the organ of Corti are excited by different frequencies, from the high frequencies at the base of the cochlea to the low ones at its top. The hair cells are synaptically connected to the auditory nerve fibers and transmit information through the spiral ganglion into the brain stem (Moore, 1997). The tonotopical representation of sounds continues into the auditory nerve, the fibers sensitive to higher frequencies being located closer to the outer surface of the nerve, while the fibers responding to the lower frequencies are concentrated on the inner part of the nerve. The other structures of the auditory pathway are also tonotopically organized.

Several stages of auditory processing occur within the medulla oblongata and the pons, specifically in the cochlear nuclei, the olivary complex, and the lateral lemniscus (Moore, 1997). The inferior olives are the first place where information from both ears is integrated. The crucial role of the inferior olives in sound localization has been shown in several studies (for an overview, see Eggermont, 2001). It is also the point where most of auditory information crosses the midline and is further transmitted contralaterally to the ear of stimulation. Among the brainstem structures, the inferior colliculus in the midbrain is believed to contain the representation of the different auditory features (for discussions, see Langner, 1992; Eggermont, 2001). From the inferior colliculus, the auditory pathway continues to the medial geniculate body of the thalamus, which in turn is projected into the primary auditory cortex (Moore, 1997). In addition to the feedforward pathways from the cochlea to the cortex, the auditory system contains a number of top-down projections, e.g., the olivo-cochlear tract.

The auditory cortex of the primates comprises several tonotopically organized areas (Rauschecker *et al.*, 1995). They form the core of the auditory cortex and are surrounded by the belt and parabelt areas. The core area, receiving direct afferentation from the thalamus, is responsive to simple sounds, especially pure tones, whereas the belt and parabelt receive their input from the core area and are more responsive to the complex, wider-frequency range sounds, such as narrow-band noises and, in monkeys, vocalizations. The anterior part of the belt and parabelt are responsive to the frequency characteristics of sounds, whereas the posterior part is responsive to the location of the sounds. This suggests a segregation of the “what” and the “where” perceptual streams in the auditory system by analogy with a similar division in the visual system (Rauschecker and Tian, 2000).

The human homologues of the belt and parabelt regions have not been precisely localized, and, in general, the division of the functional areas within the human auditory cortex is not as clear as that of lower primates. The human primary auditory cortex is located in the medial two-thirds of Heschl’s gyrus (HG) and is homologous to the core areas of the macaque. The presence of the tonotopical representation of sounds in the human primary auditory cortex has been shown in the MEG (Romani *et al.*, 1982; Tiitinen *et al.*, 1993; Pantev *et al.*, 1996; but see Lütkenhöner *et al.*, 2003 for criticism) and the fMRI (Wessinger *et al.*, 1997; Talavage *et al.*, 2000; Schönwiesner *et al.*, 2002). The primary auditory cortex is more sensitive to the loudness of the sound than are the surrounding areas (Hart *et al.*, 2002). The areas anterior and posterior to the primary auditory cortex at the superior temporal plane and in the superior temporal gyrus are usually described as auditory associative areas. Among them, the planum temporale, the posterior part of the superior temporal plane, is larger in the left than in the right hemisphere, thus constituting the largest anatomical asymmetry between the cerebral hemispheres (Galaburda *et al.*, 1978). Only non-primary auditory areas react to spectrally complex sounds, such as narrow-band noises (Wessinger *et al.*, 2001), harmonic tones (Hall *et al.*, 2002), and amplitude- and frequency-modulated tones (Hart *et al.*, 2003).

1.2 Frequency analysis and pitch perception

The analysis of sound according to its frequency includes frequency selectivity and frequency discrimination. The frequency selectivity is a property of the auditory system to separate simultaneous tones of different frequency from each other. Its clearest manifestation is auditory masking. Masking data suggest that two simultaneous sounds at the same level can only be separated if their difference in frequency exceeds a certain value or, in other words, if they fall in different critical bands (Moore, 1997; Zwicker and Fastl, 1999). Frequency discrimination is the detection of frequency change in time. Initially, it was assumed that the accuracy of frequency discrimination is defined by frequency selectivity, as also suggested by frequency-modulation difference limens (FMDL; Moore, 1997). However, the difference limens for frequency

(DFL) for tone pulses at frequencies lower than 5 kHz are smaller than predicted on the basis of the critical-band theory (Wier *et al.*, 1977; Sek and Moore, 1995). These data led to the conclusion that frequency discrimination is based on a more complex mechanism than the interaction of the excitatory patterns on the basilar membrane of the cochlea.

The frequency of a periodic sound is closely connected with the perception of pitch. A periodic sound is perceived as containing a pitch if it has 10 or more cycles; otherwise it is perceived as a click (Pierce, 1999). The pitch of the pure tone is defined by its frequency. In complex tones, the fundamental frequency defines the pitch; however, it is not necessary for the fundamental frequency to be physically present. Three successive harmonics of a tone are enough to create the sensation of the pitch, which is called the virtual or residual pitch. The same is true for pitches lower than 800 Hz; no virtual pitch exceeding 1000 Hz can be achieved (Pierce, 1999). Virtual pitch emerges even with very high-order harmonics that could not be resolved at the basilar membrane (Houtsma and Smurzynski, 1990). However, if lower- and higher-order harmonics contain different pitch information, the former “dominate”, and the pitch is perceived on the basis of the lower-order harmonics only (Ritsma, 1967). Pitch sensation can also be created by sounds with a continuous spectrum, such as iterated rippled noise (IRN; Patterson *et al.*, 2002).

Historically, there are two competing theories explaining pitch perception: the “place” and the “time” theories. According to the “place” or pattern-recognition theory, pitch perception is preceded by the extraction of individual frequency components. Thereafter, the pitch is calculated on the basis of the frequency intervals between the components. Time theories, in contrast, take advantage of the phenomenon of the phase-locking of the spike activity in the auditory nerve with the frequency of stimulation. The pitch, according to this theory, is extracted from the inter-spike interval distribution over the auditory nerve.

Neither of the two theories can explain all the phenomena related to pitch perception, however. The “place” theory has difficulty with the virtual pitch of the non-resolved harmonics and the pitch of non-tonal sounds (such as IRN). The “time” theory cannot explain the dominance of low-order harmonics and the pitch from the two harmonics presented separately to different ears. Also, phase-locking was found for frequencies lower than 5 kHz only. The most modern theories combine the “place” and “time” mechanisms in their explanation of pitch (e.g., Moore, 1997, p. 204). The interaction of the two mechanisms happens between 300 and 1600 Hz (Pierce, 1999), with the time mechanism (periodic pitch) prevailing at lower frequencies and the place mechanism (spectral pitch) at higher frequencies.

Spectral and periodical pitch are both represented at the level of the auditory cortex (Hall, 2005). There are neurons in the auditory cortex that are activated by both the pure-tone and the harmonic complex with the corresponding virtual pitch (Bendor and Wang, 2005). Less salient virtual pitch caused by the substitution of the resolved harmonics for the unresolved harmonics led to the decreased activation in anterior Heschl's gyrus (HG), but not in the primary auditory cortex or brainstem structures (Penagos *et al.*, 2004). In another study (Schneider *et al.*, 2005), subjects were divided into two groups on the basis of their pitch-extraction strategies. One group relied on periodical pitch, whereas the other relied on spectral pitch. The difference between the groups was evident in the volume of the gray matter in the antero-lateral Heschl's gyrus and in the MEG responses. In the periodic-pitch subjects the gray-matter volume and the P50m response were larger in the right hemisphere, whereas the left-hemisphere dominance was observed in the spectral-pitch group (Schneider *et al.*, 2005).

Right-hemisphere dominance in complex spectral processing has been shown in several studies. In positron-emission tomography (PET), the automatic detection of change in chord sequence activated the right temporal cortex, whereas phoneme discrimination activated the left temporal cortex (Tervaniemi *et al.*, 2000b). Yet in another PET study (Zatorre and Belin, 2001), the increase of tone-pitch diversity in the tone sequence led to a pronounced activation of the anterior part of the right superior temporal gyrus (STG). In contrast, an increase in the tone-duration diversity led to the activation of the auditory core region in the left STG.

According to the model developed by Griffiths (2003), the analysis of the temporal regularity in sounds starts at the level of the cochlear nuclei, where the increase in the temporal regularity increases neuronal activation due to synchronized activity in the neuronal population. In the inferior colliculi, the regularity is already presented in the form of a local rate code. At the cortical level, the salient stable pitch activates the lateral part of Heschl's gyrus. When pitch is changing within a sequence of sounds, the anterior (planum polare, PP) and posterior (planum temporale, PT) parts of STG are additionally activated. The PP activation is related to the pitch chroma, while PT is responsible for the pitch height (Griffiths, 2003).

1.2.1 Pure tones and complex sounds in perception

In all natural-pitched sounds, the sound spectrum consists of a time-varying pattern of multiple harmonic partials across a large frequency spectrum. However, to date, the overwhelming majority of the experiments on pitch perception in psychoacoustics and auditory neuroscience used sinusoidal tones consisting of one harmonic partial (fundamental) only. Yet the frequency-change discrimination of tones with a rich spectral structure, as reflected by the mismatch negativity (MMN, see p. 23), is facilitated when compared with that of pure sinusoidal tones (Tervaniemi *et al.*, 2000a). The MMN amplitude was larger and the latency shorter in spectrally rich tones consisting of 3 harmonic partials than in pure tones, both with a 500 Hz fundamental. Consistent with this, the subject's behavioral responses in detecting the frequency change in a separate condition were more accurate for spectrally rich tones than for sinusoidal tones. Further, 3 harmonic partials were sufficient for this facilitation to occur (Tervaniemi *et al.*, 2000c). This sufficiency was established by comparing the MMN elicited by a frequency change of 2.5% in 1-, 3-, and 5-partial harmonic sounds (in separate blocks). The frequency-MMN amplitude was larger with 3- and 5-partial tones when compared with a 1-partial tone, with no significant difference between the 3- and 5-partial tones. This suggests that, at least with sounds having a fundamental frequency of 500 Hz, relatively few harmonic partials result in the maximal pitch discrimination accuracy.

1.2.2. Frequency discrimination in child development

The development of the cochlea is complete by full-term birth (Abdala and Sininger, 1996). However, the maturation of the brainstem continues further into infancy and that of the auditory cortex continues into late childhood (Moore, 2002). As determined by otoacoustic-emission studies (Abdala and Sininger, 1996), cochlear frequency resolution is fully mature by term birth; behavioral studies report sound-frequency discrimination in infants as young as 3 months (Olsho *et al.* 1987; reviewed in Werner and Gray, 1998). However, early attempts to find evidence for tonal frequency discrimination in newborns by behavioral methods have failed (Leventhal and Lipsitt, 1964; Trehub, 1973). Yet the sucking-rate measurement and other methods indicate that newborns can discriminate differences in the fundamental frequency of the human voice (DeCasper and Fifer, 1980; Mehler *et al.*, 1988; for survey see Gerken and Aslin, 2005). At 3 months, the DFLs for low-frequency tones are smaller than those for high-frequency tones (Olsho *et al.*, 1987), but the frequency discrimination at high frequencies develops faster, reaching the adult level by the age of 6 months. The low-frequency DFLs still remain immature until late childhood (Maxon and Hochberg, 1982).

However, the behavioral data in infants younger than 5 months may provide incomplete and unreliable information (Stapells and Kurtzberg, 1991). This uncertainty emphasizesthe importance of the electrophysiological measures of frequency discrimination, which are attention-free and can provide direct access to infants' auditory discriminative abilities (Kurtzberg and Vaughan, 1985; Stapells and Kurtzberg, 1991). In adults, frequency discrimination can be objectively assessed with the mismatch negativity (MMN) recording (Näätänen *et al.*, 1978; Hari *et al.*, 1984; Sams *et al.*, 1985). The MMN correlates with behavioral performance (Tervaniemi *et al.*, 1993; Tiitinen *et al.*, 1994; Amenedo and Escera, 2000; Novitski *et al.*, 2004). Furthermore, it does not require the subject's attention or task performance and can therefore be used in subject groups that are not able to cooperate with the experimenter. The MMN to frequency change was found in newborns (Alho *et al.*, 1990), and its MEG analog (Hari *et al.*, 1984) was recently discovered even in fetuses (Draganova *et al.*, 2005; Huotilainen *et al.*, 2005). Unlike adults, in young infants, the mismatch-like response was reported in different studies to be of either negative (Alho *et al.*, 1990; Cheour *et al.*, 1998; Ceponiene *et al.*, 2002) or positive polarity (Dehaene-Lambertz and Dehaene, 1994; Leppänen *et al.*, 1997; Morr *et al.*, 2002; Sambeth *et al.*, 2006). Despite the discrepancies in the polarity of the response, it is clear that the discriminative ability of newborns can be measured with ERP recordings. This ability can be emeasured not only for physical features of a stimulus, but also for abstract rules within an auditory stream (Ruusuvirta *et al.*, 2003, 2004; Winkler *et al.*, 2003; Carral *et al.*, 2005).

In the majority of the afore-mentioned ERP studies, the frequency difference between the standard and the deviant was large: the minimum was 10% while the most often used difference was 20%. Also in non-speech studies, the frequency range of predominantly 1000-2000 Hz was exploited while the discriminative response in the other frequency ranges was examined. In adults, a 5% frequency difference elicits a significant MMN in the frequency range of 250-4000 Hz. To compare the automatic frequency discrimination in adults and newborns, the present thesis includes a study using the MMN paradigm in the 250–4000 Hz frequency range with two levels of difference, 5% and 20%.

1.3 The effects of noise on humans

Even if the majority of studies in auditory neuroscience are performed in a silent laboratory environment, such a situation is rare in the daily life. We are continuously surrounded by interfering sounds. Noise can be broadly defined as unwanted sound (Miller, 1974). Noise is always acting on the human organism through the auditory system, since the sensitivity of the ear to air-pressure changes by far exceeds the corresponding sensitivity of the rest of the body (Miller, 1974). The effect of noise can be divided into three major groups. First, loud noise can cause damage to the ear and hearing loss. Second, noise can interfere with auditory perception and reduce the capability of signal processing, such as masking and distraction. Finally, noise as a part of the ecological environment for certain human populations can cause a variety of negative health effects (Stansfeld and Matheson, 2003).

1.3.1. Noise-induced hearing loss

Noise-induced hearing loss (NIHL) is the second largest cause of the sensorineural hearing deficit after age-induced hearing loss (presbyacusis; Rabinowitz, 2000). The hearing loss measured as an increase in the hearing threshold can be either temporary or permanent. The effect of noise on hearing depends on both the amplitude of noise and the exposure time. Even a short exposure to noise louder than 140 dB causes acoustical trauma, and it unavoidably leads to a permanent hearing loss (Miller, 1974). Current USA work regulations mandate that exposure to noise levels greater than 85 dB during the 8-hour working day is unsafe (Department of labor occupational noise exposure standard 29 CFR §48FR9776, 1983; as cited in Clark and Bohne, 1999). Hearing loss not resulting from direct acoustic trauma is caused by metabolic changes in the organ of Corti that eventually lead to the death of the hair cells (Miller, 1974). The detailed mechanism of this process is still under investigation. It has been found (Hu *et al.*, 2002) that both necrosis (passive death) and apoptosis (cell suicide) are taking place. Beside damage to hair cells, loud noise causes reduction in cell density in auditory brainstem structures as well as in the thalamus and the auditory cortex (Basta *et al.*, 2005). Research on the cellular mechanism of the noise-induced damage opened a perspective for the drug therapy of the NIHL, which is based on the blockade of the different stages of apoptosis (Harris *et al.*, 2005; Kopke *et al.*, 2005).

The combination of two features makes NIHL especially dangerous. The hearing loss is not obvious immediately after exposure to the noise. Since the hair cells do not regenerate, it is usually too late for any therapeutic intervention by the time the deficit manifests itself (Miller, 1974; Clark and Bohne, 1999; Rabinowitz, 2000). The levels of MRI acoustic noise raise concerns about NIHL. The peak level of noise in high-field scanners comes very close to the level of acoustic trauma (140 dB) and the average levels exceed by far the limit allowed for industrial workers in the USA (85 dB; Ravicz *et al.*, 2000). Indeed, 43% of patients have demonstrated temporal hearing threshold increase by more than 15 dB after 20-40 minutes of anatomical MRI scanning with a 0.35-T magnet (Brummett *et al.*, 1988). Therefore, whatever the purpose and design of the fMRI experiment, ear protection is a necessary prerequisite for those studies. If protection is provided, the next concern is about the masking effect of noise.

1.3.2 Auditory masking

Masking is normally defined as “the amount (or the process) by which the threshold of audibility of a sound is raised by the presence of another (masking) sound” (American Standards Association; 1960; Acoustical Terminology SI, 1-1960; ASA, NY., cited in Moore, 1997, p. 89). If the test sound is not audible in the presence of a masker, then the masking is complete and the test sound is totally masked. Otherwise, the masker can simply reduce the loudness of the test sound, and in this case, the masking is partial (Zwicker and Fastl, 1999). The most important factors in masking are the level and frequency composition of the masker and the masked sound. If the masked sound is a pure tone and the masker has the same spectral density at all audible frequencies (white noise), then the increase of masking produced by the masker is equal to the increase in the level of the masker. However, masking by white noise depends on the frequency of the masked sound, so that at frequencies higher than 500 Hz, a ten-fold increase in frequency leads to a 10-dB stronger masking (Zwicker and Fastl, 1999). A special type of noise with attenuated high frequencies (the so-called pink noise) is used if a uniform masking across all masked frequencies is required.

The masking of the pure tone by narrow-band noise increases as the bandwidth of the noise increases until the bandwidth reaches the critical band for this particular frequency, and with further widening of the band upwards, the noise has the masking properties of a wide-band noise (Zwicker and Fastl, 1999). The critical band is about 100 Hz wide at frequencies lower than 500 Hz and about 0.2*f at higher frequencies (Zwicker and Fastl, 1999). The narrow-band noise at levels higher than 40 dB masks tones with frequencies higher than its central frequency more than lower-frequency tones (Scharf and Buus, 1986; Moore, 1997; Zwicker and Fastl, 1999). Pure tone as a masker acts in a very similar way to the narrow-band noise, and a combination of 5or more pure tones within a critical band completely mimics the masking by narrow-band noise at those frequencies (Zwicker and Fastl, 1999).

The spectrum of the fMRI acoustic noise suggests that it should have frequency-selective masking properties. The most susceptible frequencies are those near the peak value of the fMRI noise spectrum (around 700-1000 Hz). For the rest of the spectrum, the theory predicts stronger masking for higher frequencies than for lower ones. Indeed, when the hearing thresholds were measured at 125-8000-Hz frequency range in the presence of fMRI scanner noise, then the maximal threshold elevation was found at 750 and 1000 Hz (Ulmer *et al.*, 1998). The threshold was less elevated at frequencies higher than

1000 Hz and at lower than 750 Hz, but asymmetrically, so that the low-frequency slope of the curve was steeper, and at 125 Hz the threshold was no different from that in quiet. At a shallower high-frequency slope, the threshold elevation present was as high as at 8000 Hz (Ulmer *et al.*, 1998).

One important special case is the masking of speech by environmental noise. In real-life situations, the levels of noise and of voices as well as the distance between the speaker and listener should be considered. With a usual distance for public communication of 5 to 6 meters, the level of noise should not exceed 50 dB for normal conversation without raising the voice (Miller, 1974). Even at intimate distances of less than 30 cm, noise of more than 75 dB makes communication problematic (Miller, 1974). It was found (Moelker *et al.*, 2004) that in MR-scanner conditions normal communication during imaging requires shouting. The masking effect of the fMRI acoustic noise is aggravated by the fact that its dominant spectral peaks overlap with the dominant spectral region of speech (500-2000 Hz; Ulmer *et al.*, 1998).

Although measured in behavioral terms, masking is largely a peripheral phenomenon. The “line-busy hypothesis” explains that the auditory nerve fibers are too preoccupied with the transmission of information about the masker to allow the information about the test tone to come through (Scharf and Buus, 1986). The hypothesis is consistent with the profile of masking by a tone of fixed frequency. Indeed, a function of the test tone frequency mirrors the shape of the travelling wave on the basilar membrane: shallow towards high frequencies and steep towards the low ones (Scharf and Buus, 1986). In addition, phenomena such as binaural unmasking and interaural masking suggest the existence of some central masking mechanism (Scharf and Buus, 1986). However, the effect of noise on human performance is not limited to masking. Noise can also cause distraction and other related phenomena.

1.3.3 Distraction by noise

Distraction is a lapse of attention from the task at hand (Miller, 1974). It can be measured as a disruption of the task performance. Disruption has been extensively studied in a serial recall paradigm, in which a list of items is visually presented and the subject, after a certain delay, has to write down the items in the same order they were presented. The presentation of irrelevant speech during the task performance causes disruption as manifested by an increase in the number of errors (Colle and Welsh, 1976). The acoustical level of the distracter does not affect the disruption if the level is within the dynamic range of the human voice (Colle, 1980; Tremblay and Jones, 1999), nor do the semantic contents of the irrelevant speech. In fact, similar disruptive effects are found in normal speech, inverted speech, and foreign speech, and also for the sequence of changing tones (Hughes and Jones, 2001). Broadband continuous noise does not affect performance (Salame and Baddeley, 1987). According to one explanation, the disruption of a serial task by irrelevant sound depends on how much change is present in a stream of irrelevant sounds (Jones *et al.*, 2000). To act as an effective distracter, the background sound sequence should, on the one hand, change in time along some acoustical parameter (e.g., pitch) but, on the other hand, contain sounds similar enough to be perceived as one acoustic stream (e.g., with pitch differences less than 6-10 semitones).

The data on the disruptive properties of fMRI acoustic noise are contradictory. Most reviews (Amaro *et al.*, 2002; Moelker and Pattynama, 2003) mention a possible distractive effect of fMRI noise as a self-evident assumption. On the other hand, the sound of the scanner is regularly repeating and thus should not produce disruption. Repeated syllables with or without pauses between them produced no disruption in a serial recall task (Bridges and Jones, 1996). Such stimulation parallels the acoustical conditions of the continuous and sparse sampling fMRI. In a number of studies the noise produced change in the ERP or fMRI brain activation during a working memory task; however, no significant changes in the task performance were revealed (Novitski *et al.*, 2003; Haller *et al.*, 2005; Tomasi *et al.*, 2005). fMRI acoustic noise impaired the ability to retrieve the relative position of a geometric shape in a visual scene as well as changed PET activations during that task (Mazard *et al.*, 2002). Interestingly, in that study, the behavioral accuracy correlated with hemodynamic response in calcarine cortex in silence, but no such activations were found when fMRI noise was played.

Two conclusions could be drawn. First, the disruptive effect of the fMRI noise on task performance has been poorly investigated, and assumptions about it may be overly pessimistic. Second, the effects of noise on brain activity are not necessarily expressed in behavioral changes.

1.3.4 Noise effect on ERPs and ERFs

Noise effects on event-related potentials (ERPs) and event-related fields (ERFs) can be highly selective. For instance, broadband noise affects the magnetic counterparts of ERPs to speech sounds differently in the right and left hemispheres (Shtyrov *et al.*, 1999). The effects of noise may also be different for the different ERP components (see pp. 23-30 for a summary of the auditory ERP components), as was shown by Martin and colleagues (1999) who recorded ERPs of speech sounds masked by high-pass filtered noise. The spectral power of the noise was increased by lowering the cutoff frequency of the filter. The N1 amplitude gradually decreased with the lowering of the high-pass filter cutoff of the noise. In contrast, the MMN and P3 amplitudes did not change until the cutoff reached 1000 Hz, after which the amplitudes dropped dramatically. In another auditory investigation (Herrmann *et al.*, 2000), fMRI noise suppressed MEG responses to sentences in a time window of 120-200 ms (reflecting early linguistic processes), but not in an earlier time window of 20-100 ms (reflecting early auditory processes). The MMN for speech sounds was suppressed by 5 different kinds of noises more than the MMN for non-speech sounds (Kozou *et al.*, 2005). In addition, long-term exposure to industrial noise shifted hemispheric lateralization of the MMN (Brattico *et al.*, 2005). The MMN to speech contrasts had a larger source amplitude in the left hemisphere in the control subjects, but in the right hemisphere in the noise-exposed subjects. Thus, it could be concluded that the effect of noise may be selective vis-à-vis to the different stages of auditory processing.

1.4. Working memory

The analysis of incoming auditory stimulation in the brain is performed with the goal of providing information for task execution. Working memory is a brain system that keeps task-relevant information in an active state during task execution (Baddeley, 1992). According to the most influential model (for the most recent modification of the model, see Baddeley, 2000), working memory uses separate mechanisms for processing verbal and spatial information. The capacity of the working-memory store is limited. Working memory is often considered a counterpart of long-term memory - a passive store of retrievable information with almost unlimited capacity (Baddeley, 1992). Working-memory task execution activates a vast network of the cortical areas, where the prefrontal cortex (PFC) plays the integrative role (Carlson *et al.*, 1998; Martinkauppi *et al.*, 2000; Prabhakaran *et al.*, 2000). The primary and secondary sensory cortices were also shown to be activated during working-memory tasks (Ruchkin *et al.*, 2003; Winkler and Cowan, 2005).

Working-memory mechanisms could be classified, first, by the type of cognitive operation (e.g., maintaining, manipulating) and, second, by the type of information that is processed in working memory (e.g., spatial, verbal). Domain-specificity hypothesis of Goldman-Rakic, supported by a large body of lesion and single-unit recording data in non-human primates, suggests that working memory representation in the brain is physiologically segregated on the basis of the type of information (for discussion, see Levy and Goldman-Rakic, 2000; Romanski, 2004). According to this model, dorsolateral PFC is responsible for spatial processing, while the ventrolateral PFC is responsible for feature and object identity processing (Levy and Goldman-Rakic, 2000; Romanski, 2004). Domain specificity does not imply modality specificity; visual, auditory and somatosensory information can be integrated within spatial and non-spatial domains.

The majority of studies on the subdivision of working memory have been performed with visual stimuli on nonhuman primates. Recently, it was shown that there are task-related differences between the ERPs recorded during auditory spatial and nonspatial working memory tasks (Alain *et al.*, 2001; Anourova *et al.*, 2001; Anurova *et al.*, 2003). Both the electric and magnetic counterparts of the N1 component and the amplitudes of later event-related potentials were affected by the type of task. A recent fMRI study has demonstrated that the processing of location and the identity of human voices engages dorsal and ventral cerebral pathways correspondingly (Rämä *et al.*, 2004). These findings suggest that the processing of spatial and pitch information in audition is at least partially segregated in the human brain. When working memory processing is studied using fMRI, the noise inherent to this technique might have an influence on the results (Martinkauppi *et al.*, 2000). A combination of the fMRI with EEG and MEG is required to investigate this problem further.

1.5 Cognitive neuroscience methods

The neuronal activity in the brain can be detected from outside the head in several ways. One way is a direct measurement of the change in the electric or magnetic field caused by the neuronal activity as is done with the EEG and MEG, respectively. The other way is to detect metabolic changes that necessarily accompany the change in neuronal activation, as is done in the PET and the fMRI.

1.5.1. Electroencephalography and magnetoencephalography

The EEG is the oldest and most straightforward method of investigating brain functions. The neuronal excitation is manifested as a spreading change of permeability of the neuronal membrane to certain ions, resulting in a sequence of ion currents through the membrane (the so called primary current). The currents also create a difference in the potential between the surface of the activated and non-activated neurons. These differences are reflected as a change in the pattern of the electric field of the brain. Although distorted by the skull and the scalp, the pattern of the electric field can also be detected from the electrodes on the outer surface of the head (Malmivuo and Plonsey, 1995). The EEG recorded from the scalp is more likely to be caused by the summation of the postsynaptic potentials, and not the action potentials (spikes). Therefore, the EEG and the fMRI have common neuronal basis (Logothetis *et al.*, 2001; Mukamel *et al.*, 2005; Shmuel *et al.*, 2006).

The currents created within the brain are accompanied by a change in the magnetic field. Those changes can also be detected from the outside of the head by the magnetoencephalogram (for a full description of the method, see: Hämäläinen *et al.*, 1993). The magnetic field detection is not compromised by the low conductive properties of the bones, the skin, and the air. But in a spherical conductor, only the dipole sources tangential to the surface can be detected by the MEG.

Changes in the primary current cause practically simultaneous changes in the electric and magnetic field distributions, and therefore, the time course of the neuronal activity can be traced with the EEG and MEG with almost infinite precision. However, the spatial resolution of both methods is limited, especially that of the EEG. The inverse problem in the case of the EEG and MEG has no unique solution, since the same two-dimensional distribution of the electric or magnetic field at the surface of the head can be created by several different sources in the head's three-dimensional space. The actual locus can be modelled by a variety of methods, including the equivalent current dipole (ECD) and the minimum norm estimate (MNE; for a review, see: Hämäläinen *et al.*, 1993). For more precise spatial information, the EEG and the MEG should be combined with other imaging methods, such as the fMRI.

The event-related potential (ERP) and event-related field (ERF) methods are based on the analysis of EEG or MEG activity, which is time-locked to an event. The ERPs are related to the postsynaptic activity at the membrane of the pyramidal cells in the cortex. This interpretation suggests that the excitatory synaptic input from the extracellular perspective is seen as a sink or negative change in the potential. If the sink is situated at the surface of the cortex, as in the case of the apical dendrites, it is seen as negativity in the ERP (Kandell *et al.*, 2000; Luck, 2005). In the case of a deep synaptic input, what is visible at the surface of the head would not be a primary input of the current (sink), but rather its secondary output (source), and the resulting ERP would be positive. Since the thalamic inputs innervate the deep layers of the cortex and the cortico-cortical connections innervate apical dendrites, the negativities in the auditory ERPs may reflect the neuronal processing within the cortex, while a positivity may reflect thalamo-cortical activation. This scheme may be too simplified, and the role of inhibition in the circulation may be underestimated. An important role is also played by neuronal activation via top-down connections from the upper cortical structures (Friston *et al.*, 2005) usually not considered in the ERP literature.

The ERPs can be divided into obligatory and endogenous components (Donchin *et al.*, 1978; Näätänen, 1992). The obligatory potentials can be predicted from the properties of the stimulus. Those with a latency shorter than

10-12 ms are generated in the brainstem, whereas the later potentials have their sources in the cortex. The most prominent cortical obligatory ERP is the N1, a negative deflection peaking around 100 ms after sound onset (Näätänen and Picton, 1987; Woods, 1995). Endogenous ERPs are generated in the cortex, and their characteristics depend more on the condition of the subject than on the properties of the stimulus. The most prominent are the MMN, P3a, and P3b.

The auditory input is dissected in the subcortical structures into auditory features, such as frequency, intensity, duration, pitch, and direction of the frequency sweep (Nelken *et al.*, 2003). The auditory cortex solves the problem of modelling the world and predicting the future on the basis of these features (Näätänen and Winkler, 1999; Näätänen *et al.*, 2005). Models that are created in the cortex include a memory component, which provides access to the information perceived in the preceding time moments. It has been suggested that the N1 is a manifestation of auditory-feature processing in the auditory cortex, whereas the mismatch negativity (MMN) is a reflection of representational (memory-dependent) processing (Näätänen and Winkler, 1999). The MMN is related to the representation of the events in the current auditory environment in the sensory memory and is elicited as a negative peak at around 200 ms whenever a new sound violates that representation (Näätänen and Winkler, 1999). An existing body of data supports the view that the N1 and MMN indeed are two distinct processes in the brain (Näätänen *et al.*, 2005). If the new, incoming sound is salient enough, it may trigger involuntary attention shift, which is reflected in the ERP as a positive peak at about 300 ms after a stimulus onset called the P3a (Escera *et al.*, 2000). Studies of the MMN elicited by complex rule violations led to the conclusion that primitive unconscious intelligence may operate in the auditory cortex (Näätänen *et al.*, 2001).

1.5.2. Functional magnetic resonance imaging

Currently, the most popular and rapidly developing method of brain imaging is the fMRI. It has been developed on the basis of the conventional or anatomical magnetic resonance imaging (MRI). Anatomical MRI makes use of the inherent differences in the passive magnetic properties of the white matter, gray matter, and cerebrospinal fluid. The magnetic properties of the tissue also depend on the level of blood oxygenation. Neuronal activity increases the cerebral blood flow in that area (Roy and Sherrington, 1890) and thus increases the amount of oxygenated blood flowing through this area. It is possible to detect activation as a local distortion of the magnetization in what is called the blood oxygenation-level dependent (BOLD) contrast.

The localization of the signal in the MRI and fMRI is achieved by applying of the gradients of the magnetic field in 3 dimensions. Thus, each voxel of the image is uniquely defined, and there is a one-to-one correspondence between the image and the brain. In fMRI, unlike the EEG and MEG, the inverse problem has a unique solution. The spatial resolution of the fMRI (3-5 mm) is highest among the imaging methods, such as PET and SPECT (Nair, 2005; Kim *et al.*, 2000). However, the method has some drawbacks.

First, the fMRI is a metabolic, indirect measure of neuronal activity. The activation-related change in the local blood flow is coupled with neuronal activation through a complex set of interactions between the neurons, glial cells, and blood-vessel walls (Magistretti and Pellerin, 1999). Therefore, the relationship between the BOLD response and the underlying neuronal activation can be non-linear. The local field potentials in the monkey brain predict the BOLD response fairly well (Logothetis *et al.*, 2001), and the decrease in neuronal activity is coupled with the decrease in the BOLD response (Shmuel *et al.*, 2006). Second, the fMRI has a relatively low time resolution. The hemodynamic response lags behind the external stimulation, due to the sluggishness of the neurovascular coupling processes, reaching its peak several seconds from stimulus onset (Menon *et al.*, 1995; Belin *et al.*, 1999). The time course of the BOLD response differs between the different brain areas, a difference that has to be taken into account when comparing the timing of events in the different parts of the brain. The fMRI can give a plausible picture of the participants in the brain network during a certain task, but it does not provide accurate timing of activation within a network (Nair, 2005). To obtain the temporal information, the fMRI has to be combined with other imaging methods, such as EEG and MEG.

The simultaneous acquisition of the EEG and the fMRI has also been used in the oddball paradigm (Liebenthal *et al.*, 2003). Still it is not a standardized technique, as the gradient switching in the MRI pulse sequence creates artefacts in EEG recording. In addition, there are difficulties in the synchronization of EEG and fMRI devices (Garreffa *et al.*, 2004). In the case of the MEG, the simultaneous acquisition with the fMRI is not possible at the current level of technical development. Therefore, separate sessions for fMRI and EEG/MEG recording are conducted to compare the results of the two methods (Opitz *et al.*, 2002; Doeller *et al.*, 2003). In auditory studies one should take into account the third major drawback of the fMRI i.e., the EEG and MEG studies are normally conducted in a silent acoustical background, whereas fMRI acquisition is accompanied by loud acoustic noise.

1.6 fMRI acoustic noise

The fMRI acoustic noise is a necessary by-product of the scanning procedure. The structure of the gradient noise is tightly connected to the experimental setup and thus is highly predictable. Its acoustic properties are determined by user-defined parameters of the pulse sequence as well as by the resonance properties of the gradient coil assembly.

1.6.1 Origins of the fMRI acoustic noise

The magnetic resonance imaging device mainly consists of 3 parts: the main magnet, the gradient magnets (coils), and the radio frequency (RF) coils. The main magnet is always on. Its functioning requires superconductivity, which in turn requires low temperature provided by liquid helium. The maintenance of the main magnet includes in-room air conditioning, the MR ventilation system, and cryogen pumping, all of which produce noise during their functioning. This ambient noise has an amplitude of 45-75 dB (Ravicz *et al.*, 2000). However, the major source of the noise during the scanning is in the gradient coils.

The gradient coils are the main appliance providing the precise spatial resolution of magnetic resonance imaging. The principle of imaging is roughly as follows: a signal is generated when the precession of hydrogen molecules resonate with the incoming radio frequency impulse. The precession frequency is directly proportional to the strength of the external magnetic field. A predictably changing magnetic field across the object of interest is needed to obtain a signal from a limited area. This predictable pattern of the magnetic field is produced by the gradient coils. They create a gradient in the field in the orthogonal directions and, by manipulation with the phase of the signal, finally limit the origin of the signal in 3 dimensions, constructing a voxel as a part of the resulting image. Thus, there are 3 sets of gradient coils along the axes X (the readout, or the frequency-encoding gradient), Y (the phase-encoding gradient), and Z (the slice selection gradient), correspondingly.

The gradient coil basically is an electromagnet. It induces a magnetic field when the electric current travels through the loop of the conducting wire. The coil functions within an external magnetic field. Any movement of the charge in the magnetic field creates a Lorentz force, which acts upon the charge or, in the case of the MRI machine, on the wire that conducts the electric current. The wire is driven by the force in a perpendicular direction both to the direction of the magnetic field lines and to that of the current. A closed wire loop or a coil placed orthogonally to the direction of the magnetic field lines either expands or shrinks, depending on the direction of the current in the coil. The expansion or shrinking of the coil creates changes in the air pressure, which are perceived as sound. In a gradient coil of girth l and electric current I placed in a magnetic field B will be affected by the Lorentz force, equal to:

$$F = I * l \times B,$$

Since the sound amplitude will be proportional to the Lorentz force, the stronger gradients and the stronger magnets will produce louder sounds. The sound will also be louder with larger gradient coils.

The peak loudness during scanning was reported to be 123 dB and 138 dB for 1.5 and 3 T magnets, respectively. Average loudness for a 2-second time to repeat (TR) and 15-slice volume was 109 and 126 dB for the same scanners (Ravicz *et al.*, 2000).

In conventional anatomical MRI scanning, the gradients switch once per TR. Therefore, the sounds created by the gradient switch are rare, and they are perceived as single clicks, with one click per each of the slice selection, readout, and phase-encoding gradients. In the fMRI, where an echo-planar sequence is used, the auditory scene is completely different. Echo-planar imaging (EPI) encounters the same slice-selection principle as that in conventional imaging, but EPI allows the acquisition of the data of the whole slice in “one-shot” and within one TR. The resolution of the image depends not only on the strength of the gradient, but also on the time when the gradient is on. Thus with a stronger gradient, an image with a satisfactory resolution can be obtained in a shorter time than with weaker gradients. In echo-planar imaging, stronger gradients are used, and therefore the data can be acquired in a shorter time (although with lower space resolution than in conventional anatomical MRIs). Yet stronger gradients produce louder acoustic noise. Moreover, the fast switching of the gradients turns a series of clicks into a periodic sound.

The fundamental frequency of that periodic sound depends on the echo spacing (ES) in the EPI pulse sequence. To achieve a more effective use of the readout gradients’ switching time, the adjacent readout gradients are directed opposite to each other. This in turn creates a deflection of the coil in different directions, with an extension followed by a shrinking. The full period of the resulting sound is equal to 2 ES, and the fundamental frequency therefore is 1/(2ES). The frequency can also be calculated from more readily available imaging parameters, matrix size (Mx by My) and the bandwidth of the receiver (BW). If the signal is recorded during the whole time the gradient is on (otherwise, the rise and fall time of the gradients should be considered), then ES=Mx/BW, and the fundamental frequency of noise is Fr=BW/ (2Mx). In typical fMRI studies with 64 X 64 pixel matrix size and 100-125 kHz bandwidth, the resulting noise has the fundamental frequency of 781-980 Hz. In addition, the sound, as it is elicited by the coil, comprises the harmonic multiples of the fundamental, due to the trapezoidal form of the gradient switch.

In a typical MRI experiment, the multi-slice method is used. It takes advantage of the fact that the signal acquisition *per se* takes only a fraction of the TR, with the rest being needed for getting the magnetization to the baseline level. Since the magnetization in separate slices proceeds independently, the next slice can be activated while the previous one is still restoring its original magnetization. The number of slices in that procedure is limited by the duration of each slice-data acquisition. In terms of noise, each slice produces a sound burst and the sequence of several slices produces an amplitude-modulated sound. The frequency of the modulation is proportional to the inverse of the product of the acquisition time and the number of slices. In the case of continuous acquisition, the acquisition time is the whole TR. The usual 10-20 slice per 2 s TR will produce a 2.5-5 Hz modulation.

The sound from the coil will normally be a complex harmonic tone with a fundamental frequency of about 1 kHz, amplitude-modulated at about 3 Hz. The sound is further modified by the acoustic resonance properties of the coils, defined by their size and the elastic properties of the sound (Tomasi and Ernst, 2003).

1.6.2 Methods of noise abatement in the MRI

There are 3 possibilities in noise reduction. First, one can attempt to modify the hardware of the scanner so that the resulting noise amplitude is reduced. Second, the pulse sequence can be modified so that the resulting noise is either lower in amplitude or specially distributed in time. Finally, the subject can be isolated from the action of noise either passively or actively.

1.6.2.1 Low-noise hardware modifications

The most profound effort to eliminate the source of gradient acoustic noise has been made by Mansfield and colleagues (Mansfield *et al.*, 1995, 2001; Mansfield and Haywood, 2000; Mansfield, 2004). Their suggestion is to couple mechanically the coils that are generating Lorentz forces of different directions so that the resulting force of the whole assembly would be nullified. While theoretically elegant, this method requires elaborate calculations and in practice does not eliminate the noise, although it dramatically reduces its amplitude. With the latest modification of the assembly, the reduction was as much as 35 dB (Mansfield *et al.*, 2001; Mansfield, 2004).

If the vibration of the coils is unavoidable, then the device can still be designed so that those vibrations do not transform into sounds that reach the subjects' ears. There are two main pathways are responsible for this transformation: an airborne pathway and the vibration of the scanner. The airborne pathway can be blocked by encapsulating the gradient coils in a vacuum chamber, while the solid vibrations can be mitigated by decoupling the gradient coils from the main magnet with some elastic material and by independent mounting of the gradient coils on the floor (Katsunuma *et al.*, 2002;

Edelstein *et al.*, 2002). These measures together with the additional prevention of eddy currents in the system can reduce the intensity of the noise by 34 dB (Katsunuma *et al.*, 2002).

There are also other approaches to reduce noise generation in the hardware (Moelker and Pattynama, 2003). Most of the above-mentioned methods are at the stage of experimental design and are only slowly being incorporated into commercially available scanners. However, the user of a typical commercial scanner can manipulate the noise by modifying the pulse sequence or/and by isolating the subject from the noisy environment.

1.6.2.2 The manipulation of pulse sequences

The frequency response of the gradient coil is lower at low frequencies (up to 200-400 Hz) than at high frequencies (Hedeen and Edelstein, 1997). Therefore, for a pulse sequence with a low fundamental frequency (like conventional spin-echo sequence), only the higher harmonics create the unwanted effects. These effects could effectively be removed from the spectrum by substitution of the usual trapezoidal gradient switching for sinusoidal ones and by making the slopes as shallow as possible (Hennel *et al.*, 1999). However, this substitution is not possible for the EPI sequence where the fundamental frequency itself is well above the “safety region” of the coil frequency response. Therefore, the low-noise modifications of the EPI, such as sparse sampling and cluster acquisition, are gradually emerging as a standard of auditory fMRI research.

Instead of reducing the noise, one can try to avoid especially hazardous effects of noise in a particular experiment (as long as safety requirements are fulfilled). In an auditory fMRI study, the noise may, at first, prevent the subject from hearing the sounds that are the test material and, second, may generate a BOLD response in the auditory cortex (Bandettini *et al.*, 1998; Hall *et al.*, 2000). This response can obscure the BOLD response to the test stimulation by shifting the activation baseline and reducing the dynamic range of the response. The interaction of the noise and target responses can happen within the same volume acquisition (intra-acquisition interaction) or between the adjacent acquisitions (inter-acquisition interaction). The fact that the BOLD response is delayed in relation to stimulus onset could be used to dissociate the stimulus from noise as well as to dissociate the BOLD response to the stimulus and that from the noise (Eden *et al.*, 1999; Edmister *et al.*, 1999; Hall *et al.*, 1999). This idea was first implemented in the so-called clustered volume acquisition (Edmister *et al.*, 1999), where slices of the same volume were shifted as close together as possible, leaving the rest of the TR suitable for test-sound presentation. If volume acquisition is shorter than latency of the hemodynamic response (about 2 s), then the design also avoids the intra-acquisition interaction of noise and the test-stimulus responses. To prevent inter-acquisition interaction, the TR must be increased to create the so-called sparse imaging paradigm (Hall *et al.*, 1999). In the sparse paradigm the TR should be 7-9 s to allow the hemodynamic response to the noise to settle down to the baseline level. Simultaneously, the signal-to-noise ratio and therefore, the image brightness, increase. The clustered and sparse imaging produce better results than usual continuous-distributed imaging paradigms for both block- and event-related fMRI designs (Edmister *et al.*, 1999; Hall *et al.*, 1999; Amaro *et al.*, 2002). However, sparse imaging takes more time than does the conventional imaging due to longer TRs.

1.6.2.3 Passive noise reduction

The most straightforward way to protect the subject from the undesired effects of noise is mechanical isolation. Sound reaches a subject's cochlea mainly through the ear canal, while bone conduction in the head and the whole body constitute an additional channel. Protection of the subject's ears with earplugs prevents the propagation of the sound waves through the ear canal by 28-31 dB. Foam earplugs are reported to provide better isolation than other types. In addition, the ear-canal sound propagation can be checked by special earmuffs, which by themselves provide approximately 35 dB attenuation (Ravicz and Melcher, 2001). The earmuffs in combination with earplugs produce about 40-dB attenuation. The higher frequencies are attenuated more than the lower ones. The combined attenuation is less than the summed attenuation, especially at frequencies higher than 500 Hz, due to the greater role of bone conduction at those frequencies. To overcome this limitation, it has been suggested that the subject's head be covered with a special helmet. The combination of earplugs, earmuffs, and helmet resulted in approximately 60-dB noise attenuation (Ravicz and Melcher, 2001). This is the largest attenuation reported for any noise-abating technique so far. However, the subject's comfort is significantly compromised by the introduction of an isolating helmet. Earplugs and earmuffs have become *de rigueur* even in studies with no auditory stimulation, mostly because of their reliability in terms of safety concerns. Of all the methods, passive noise reduction is also the easiest to implement.

1.6.2.4 Active noise cancellation

Active noise cancellation (ANC) is based on the phenomenon of destructive interference. When the sound containing the same frequency components with the same amplitudes but with a 180° phase-shift relative to the original sound is presented in the same point of space, then the resulting waveform is zero. Even though theoretically elegant, this method requires the precise time and amplitude matching of the cancellation sound with the noise to be cancelled. Furthermore, if two signals happen to be in phase, then the noise will be doubled (a 3 dB increase) instead of being cancelled.

There are two main approaches in ANC, i.e., the feedback and feed-forward methods. In the feedback method, the noise is recorded at the spot where it is desirable to cancel it, normally near the subject's ear. Then its phase is reverted and the resulting sound is presented through the loudspeaker at the same place. However, this method can be applied only to frequencies less than 200-500 Hz, (Chambers *et al.*, 2001). Besides, the microphone located in the vicinity of the ear not only catches the noise, but also the stimulation and the speech messages from the operator, and those will be cancelled as well. In contrast, in the feed-forward technique, the microphone is located close to the source of the sound; it catches the sound that is still to reach the subject's ear. The phase of the incoming sound is thereby predicted and reversed, and the noise is cancelled (Chambers *et al.*, 2001). When the feed-forward technique was applied for the fMRI acoustic-noise cancellation (Chambers *et al.*, 2001; Jung *et al.*, 2005), the objective reduction of noise was substantial (30-40 dB in Chambers *et al.*, 2001, and 20-25 dB in Jung *et al.*, 2005), but the subjective reduction was more modest (5-12 dB in Chambers *et al.*, 2001). The ANC can only affect airborne noise, and therefore, its limit is defined by the bone-conduction level, that is, 40 dB less than air-conduction. In the experiments in which the passive isolation reduces noise in the ear-canal to the level of bone-conducted sound, the role of the ANC is minimal. It can still be complementary to the passive isolation at low frequencies (e.g., ambient noise), where earplugs and muffs are less effective.

To conclude, a great variety of methods of fMRI noise abatement has been suggested in recent years. None of them completely eliminate noise if used alone. The combination of methods is promising, but not always possible. Often the mitigating noise comes with a price in subject comfort, image quality, or the scanning time. Therefore, noise should be considered an inherent part of the fMRI experimental set-up with all of it consequences for data interpretation.

2. GOALS OF THE STUDY

The present thesis has the following goals:

Study I: To describe automatic pitch discrimination as indicated by an electroencephalographic measure (the MMN) in different frequency bands within the 250-4000-Hz frequency range by varying contrasts and comparing the results with those in a behavioral discrimination task;

Study II: To determine the limits of pitch discrimination within the 250-4000-Hz frequency range in newborns by applying the same experimental set-up that was used with adults in **Study I**;

Study III: To determine the effects of fMRI acoustic noise on magnetoencephalographic measures (the MMNm and P3am) of automatic pitch discrimination of simple and complex tones with fundamental frequencies overlapping and not overlapping the spectral maximum of noise and to compare those effects with the effects of noise on behavioral measures in the same task;

Study IV: To determine the effects of fMRI acoustic noise on electroencephalographic measures (the MMN and P3a) of automatic pitch discrimination and involuntary attention with simple tones, musical chords, and unique natural sounds;

Study V: To determine the effects of fMRI acoustic noise on electroencephalographic correlates of perception (N1, N2, and P3) during pitch and location delayed-matching to sample working-memory tasks;

Study	**N total (female)**	**Age (mean)**	**Hardware**	**Stimulation**	**fMRI noise**	**Task**
I	10 (4)	19-40 (24.2) Years	EEG	Sinusoidal and harmonical tones 250-4000 Hz	No	Watching film, ignore stimuli
II	11 (8)	1-4 (2) Days	EEG	Sinusoidal tones 250-4000 Hz	No	No task
III	20(10)	21-30 (25.2) Years	MEG	Sinusoidal and harmonical Tones 240-1240 Hz	64 dB, continuous	Watching film, ignore stimuli
IV	10(6)	18-30 (24) Years	EEG	Sinusoidal tones 262, 275, chords C/major, C#major	57 dB, continuous	Watching film, ignore stimuli
V	11 (6)	18-32 (25) Years	EEG	Sinusoidal tones 1000, 1500	72 dB, sparse	Matching-to sample task

TABLE 1. Main characteristics of the studies

3. METHODS

The main characteristics of study methods are summarized in table 1.

3.1. Subjects

Healthy adult volunteers participated in **Study I** and **Studies III-V**. Their ages ranged from 18 to 40 years with a mean of 24-25 in **Studies I, III-IV**. All groups were approximately balanced according to the gender. Eleven full-term sleeping newborns participated in **Study II** (the mean age was 2 days, APGAR scores 8-10, 8 females), predominantly in the active sleep phase (67 %).

3.2. Stimuli and presentation paradigm

Pure sinusoidal tones were used in all 5 studies. The frequencies of the pure tones were in the 200-4800-Hz range in **Studies I** and **II**, in the 240-1240 Hz range in **Study III**, in the 262 to 275 Hz range in **Study IV**, and in the 1000 to 1500 Hz range in **Study V**. In addition, complex harmonic tones (harmonic sounds) were used in **Studies I** and **III** and musical chords and novel sounds were used in **Study IV**. The complex harmonic tones in **Studies I** and **III** consisted of a fundamental and its 2 lowest overtones, with all these 3 partials having equal intensity. The fundamental of harmonic sounds mirrored the frequencies in pure tones of the same Study (**Studies I** and **III**). The chords in **Study IV** were C major and C# major, covering the 262-6640 Hz range in frequency. Novel sounds were randomly taken from the same pool of 180 natural sounds. The duration of the stimuli (including 10 ms rise and fall times) was 100 ms in **Studies I and II**, 200 ms in **Studies IV** and **V,** and 250 ms in **Study III**. The sounds were delivered to the subjects through headphones in **Studies I, IV,** and **V**; through loudspeakers in **Study II;** and through the plastic tubes in **Study III**.

The stimuli in **Studies I, II,** and **IV** were presented in the oddball paradigm with frequently repeating standard sounds and rare, deviant sounds. The probability of the standard was 0.76 in **Studies I and II**, and 0.8 in **Study IV**. The order of the sounds in the experimental block was random except that each deviant tone was preceded by at least one standard tone. The stimulus onset asynchrony (SOA) was 400 ms in **Study I**, 700 ms in **Study IV,** and 800 ms in **Study II**. In **Study I** the standard tone (and the fundamental of the harmonic tones) was, in separate blocks, 250, 500, 1000, 2000, or 4000 Hz in frequency. The deviant sounds were decrements and increments of 2.5, 5, 10, and 20% in frequency (p=0.03 each) relative to the standard tone. In **Study II**, the standard tones had a frequency of 250, 1000, or 4000 Hz in different blocks. The deviants were decrements and increments of 5 and 20% (p=0.06 each) in frequency relative to the standard tone. In the subsequent analysis, the responses to the lower- and higher-frequency deviants at each deviance level were pooled together. In **Study IV**, the standard sounds were pure tones of 262 Hz or C major chords in different blocks. Correspondingly, the deviant sounds (p=0.01) were pure tones of 275 Hz or C# major musical chords. In addition, novel sounds (p=0.1) were presented in all blocks, none of which appeared twice in a block of a given condition.

In **Study III** the stimuli were presented in the roving-standard paradigm (Cowan *et al.*, 1993; Shestakova *et al.*, 2002). Here, the stimulus sequence was

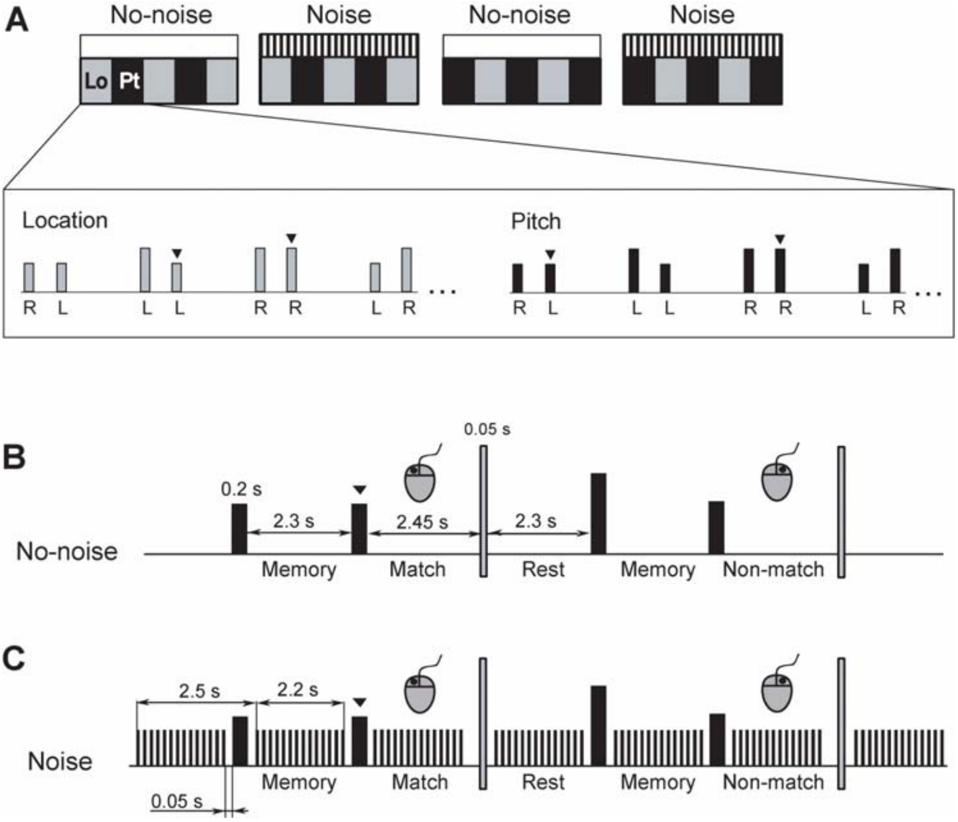

Figure 1. Study V: Experimental paradigm. (**A**) Location (Lo) and pitch (Pt) tasks were presented in an alternating manner. Half of the blocks were accompanied by fMRI noise. L = left, R = right; high and low bars correspond to high and low pitch; the match trial is marked with a triangle. (**B**) Timing of the tones in match and non-match trials in the no-noise condition. The gray bar indicates the time when the screen flashed once. (**C**) The timing of the tones and the fMRI noise bursts in a match and a non-match trial in the noise condition. The noise burst is illustrated by a set of thin vertical lines.

comprised of the mini-sequences of 8-12 sounds of the same pitch. The mini-sequences were presented at different pitch levels, and the first stimulus of each mini-sequence was treated as a deviant stimulus. The second to seventh stimuli were treated as “fillers,” and the responses to them were not analyzed. The rest of the stimuli (1-5 stimuli after the 7th) were treated as standards. The SOA between the subsequent sounds was on average 400 ms, randomly jittering from 370 to 460 ms. In the subsequent analysis, pitches from 240 to 498 Hz were classified as “low frequencies” and those from 598 to 1240 Hz as “high frequencies.” Harmonic and sinusoidal stimuli were presented in separate blocks.

In **Study V,** tones of either 1000 or 1500 Hz were binaurally presented. Left (L) and right (R) locations were simulated by using both an interaural intensity and time difference. The intensity in the ipsilateral channel was 77 dB SPL, and it was attenuated in the opposite channel by 15 dB, while the interaural time difference was 500 μs. According to the subjects' reports, the two locations were clearly perceived as right or left, and the two different frequencies, as high or low. Each trial began with a cue that was followed by a probe after a delay of 2.3 s. The interval between the probe and the cue of the next trial was 4.8 seconds. The experimental paradigm is presented in **Fig. 1.**

3.3. Experimental conditions

In **Studies I-IV,** subjects were passively listened to sounds during an EEG or MEG recording. The adult subjects in **Studies I, III,** and **IV** were instructed to remain still and focus on a silent, sub-titled film presented on a screen. The infants in **Study II** were asleep during the recordings.

In addition, a behavioral session with the same stimuli was held, after an EEG or MEG session in **Studies I** and **III**. In the behavioral session of **Study I,** subjects were instructed to indicate whether the tones delivered in pairs differed in pitch from each other. The pairs had a 300 ms silent within-pair interval and were presented with a 1900 ms between-pair interval. The session also included a 10-minute break in the middle. The hit rate (HR) was separately calculated for each deviant pair as a proportion of hits to the total number of pairs containing this deviant type within a block. The d' sensitivity measure was calculated according to the formula $d'=Z(HR_p)-Z(FA)$, in which the HR_p is the pooled average hit rate of 4 deviance levels within a block and the FA is the ratio of false alarms. In the behavioral session of **Study III**, subjects were instructed to press a button after detecting a deviant stimulus, while the presentation paradigm was the same as in the ignore session (see Stimuli). In both **Studies III** and **IV** subjects were not informed, either of the nature of the sounds or of their proportion. No practice was given before the actual experiment.

In **Study V,** subjects performed location and pitch delayed matching-to-sample tasks (Anourova *et al.*, 2001). In all blocks, the pitches and locations were randomly mixed. Before each block, subjects were instructed as to the type of task they were to perform. In the pitch task, subjects were instructed to indicate by pressing a button whether the pitch of the probe matched the pitch of the cue irrespective of the locations of the probe and the cue. In the location task, the location was to be matched irrespective of the pitch (**Fig. 1**). Match and nonmatch trials were presented in a pseudo-random order in equal proportion. All experiments started with a short training session.

3.4. fMRI noise (Studies III-V)

In **Studies III-IV,** one half of the experimental blocks included prerecorded fMRI acoustic noise. The noise consisted of the sound of an EPI

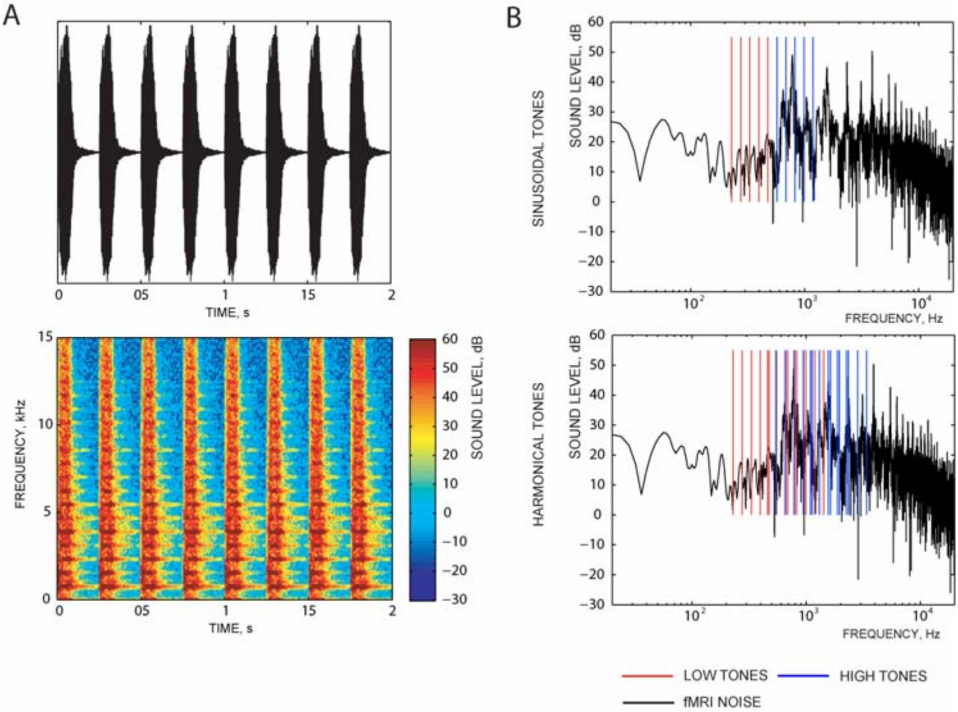

Figure 2. Study III: A. The time-course (upper panel) and spectrogram (lower panel) of fMRI noise. **B.** Schematics of the spectra of the sinusoidal (upper panel) and harmonic (lower panel) tones (red lines - low tones; blue lines - high tones) overlaid with the Fourier transform of the fMRI acoustic noise (black line).

sequence with 12-25 slices acquired in 2-2.5 s TR in 1.5 T (**Studies IV** and **V**) or a 3 T magnet (**Study III**). The time course, the spectrogram of the noise, and its spectrum in relation to the tones in **Study III** are presented in **Fig. 2**. In **Studies III** and **IV** the noise was presented via loudspeakers independent of the onset of tones, mimicking the fMRI block design. In **Study V** the noise was embedded in the experimental paradigm and presented through headphones between tone stimulations, mimicking sparse imaging. The resulting noise level at the subject’s head was 63, 54, or 72 dB SPL in **Studies III, IV,** and **V,** respectively.

3.5. EEG Recording and analysis (Studies I, II, IV, V)

In all studies the EEG was recorded with Ag-AgCl electrodes in an acoustically dampened and electrically shielded room. NeuroScan hardware and software were used in the data recording and analysis. The analogous passband was 0.1-40 Hz frequency range in **Study I**, 0.1-30 Hz in **Study II**, 0.1–100 Hz in **Study IV,** and 0.03–100 Hz in **Study V**. The sampling frequency was 500 Hz in **Studies I** and **IV** and 250 Hz in **Studies II** and **V**.

The EEG was recorded from 10 scalp electrodes in **Study I** (Fpz, Fz, Cz, Pz, LM, RM, L1, L2, R1 and R2), 9 scalp electrodes in **Study II** (F3, F4, FCz, C3, C4, P3, P4, T3 and T4), 10 electrodes in **Study IV** (Fz, Cz, Pz, Oz, C3, C4, T3, T4, LM and RM), and with a 32-electrode cap in **Study V**. The EEG was referenced against the nose electrode in **Studies I, IV,** and **V** and against the average of the mastoid leads in **Study II**. Eye-related activity was monitored by additional electrodes positioned at the outer canthus and below the right eye (**Study II**), by a single electrode attached in the vicinity of the left canthus (**Study IV**), or by two pairs of bipolar electrodes that recorded vertical and horizontal electro-oculograms (EOG; **Studies I, III,** and **V**).

The EEG was divided into epochs locked to stimulus onset. 500-ms epochs (-100...400 ms) were defined in **Study I**, 900 ms (-100…800 ms) in **Study II**, 800 ms epochs (-100 … 700) in **Study IV**, and 700 ms (-100…600 ms) in **Study V**. Epochs containing voltage variation exceeding a certain threshold (±100 μV in **Studies I, II, V** and ±80 μV in **Study IV**) were considered artefacts and were rejected from further analysis. The responses were off-line filtered with a digital filter having a passband of 1-12 Hz in **Study I**, 1-15 Hz in **Study II**, 0.5–30 Hz in **Study IV**, and 0.5–20 Hz in **Study V**. The baseline level for measuring the amplitudes of the ERPs and ERP difference waveforms was defined as the mean voltage of the 100-ms prestimulus period. The epochs were averaged for each stimulus type and for each condition. Difference waveforms (deviant-tone ERP minus standard-tone ERP) were calculated in **Studies I, II,** and **IV**. The standards following deviants were excluded from the analysis in **Studies I** and **II**.

In **Study I**, the MMN was quantified by determining the MMN peak amplitude and latency from the Fz grand-average difference waves separately for each deviant as the most negative peak between 100-300 ms in the individual difference waves. The statistical significance of the MMN was tested with one-tailed t-tests by comparing the MMN amplitude at Fz and that at the right mastoid lead with zero. Previous studies (see, e.g., Alho, 1995) showed that, when a nose reference is used, the MMN has its maximum amplitudes at Fz (the most negative value) and at the mastoid leads (the most positive value). In the further analyses, the Fz values were used after rereferencing the data against the average of the mastoid-lead amplitudes. This procedure was used to project the strength and timing of the MMN from several electrodes with opposite polarities (fronto-centrally negative values, mastoidally positive values) to one topographical point. The P3a was quantified by determining the P3a peak latency from the Cz grand-average difference waves separately for each deviant as the most positive peak between 200-400ms. The statistical significance of the P3a was tested with one-tailed t-tests by comparing the P3a amplitude at Cz with zero.

In **Study II**, the difference ERP was further divided into nine 100-ms intervals starting from the pre-stimulus interval (-100-0 ms) and averaged within each interval for statistical analysis.

In **Study IV**, the N1 was defined as the first negative peak of the ERP to standard sounds at around 100 ms from stimulus onset, with the P1 and P2 being taken as the positive peaks preceding and following, respectively, the N1. Further, the N2 was determined as the negative peak following the P2. The MMN in turn was determined as the maximal negative peak within the latency interval 120–250 ms of the difference waveforms obtained by subtracting the ERP to the standard sound from those to the deviant and novel sounds. However, since novel sounds also enhance the N1 amplitude (Alho *et al.*, 1998a), the difference-wave negativity for novel sounds will be called the N1/MMN in the following discourse. The P3a was measured as the maximal positive peak of the difference waveform between the ERPs to the novel and standard sounds within the latency interval of 200–300 ms. The amplitudes and latencies of the ERP peaks were separately measured for each subject, sequence type, and condition. The P1, N1, P2, N2, N1/MMN, and P3a were measured at Cz, while the MMN was measured at Fz, because MMN is more frontally distributed on the scalp than the other ERP components studied (Näätänen, 1992). The mean ERP and difference-wave amplitudes were also measured over the consecutive 50-ms windows centered at the individual peak latencies across all channels.

In **Study V**, the amplitudes of the ERP components were determined within a set of 9 electrodes chosen according to the spatial distribution of each component. A fronto-central set was used for the P1, N1, and P2, a frontal set for the N2, and a centro-parietal set for the P3 (**Fig. 8A**).

The P1 component was determined as the most positive peak within the 30–100 ms time window, the N1 as the most negative peak within the 60–160-ms time window, the P2 as the most positive peak within the 130–250-ms time window, the N2 as the most negative peak following the P2 within the 200–350-ms window, and the P3 as the most prominent positive peak within the 300–450-ms window. The latency of the peak amplitude of the component analyzed was first determined at the representative electrode site individually for each subject. Then the amplitudes of the ERP components were determined at this latency for each electrode site of a given set. The amplitudes were calculated as a mean value within a 30-ms interval around the peak latency of the P1, N1, and P2 components and within a 50-ms interval for the N2 and P3. Statistical analysis of the latencies and amplitudes of the ERP components was performed for the sets of 9 electrode sites.

3.6. MEG measurement and analysis (Study III)

The continuous MEG was recorded using the BTI Magnes WHS 2500 whole-head system; 148 magnetometer channels, 11 magnetic reference channels and 2 bipolar EOG-channels were used. The signals were digitized using a bandwidth of 0.1-50 Hz and a sampling rate of 254.31 Hz. At the beginning of the MEG measurement, the subject's head fiducial coordinate system was estimated, using a 3D-Digitizer. Whenever the movement of the subject's head position was larger than 8 mm, the data of this block were discarded.

The continuous MEG-data were filtered off-line with a 1.1-12-Hz bandpass filter. Epochs were rejected from further analysis if in any EOG channel, the standard deviation exceeded 100 µV or in any MEG channel the standard deviation exceeded 1100 fT in a gliding 200-ms window. The bad channels were determined on the basis of their low correlation with the neighboring channels, and they were rejected from the analysis.

For each subject, a realistically-shaped volume conductor was constructed on the basis of the Montreal brain phantom (Collins *et al.*, 1998). The size of the volume conductor was adjusted to the subject's real head size and shape. The source current distribution (SCD) maps with 1222 dipole locations were modelled with a linear minimum norm estimate procedure (L2-norm/LMNE; Hämäläinen *et al.*, 1993) using SimBio-software (http://www.simbio.de/, Knösche, Dümpelmann, Anwander, and Wolters, unpublished software). Further, the SCD maps of the responses to the standard tones were subtracted from the corresponding maps for the deviants. Six ROIs were defined on the basis of the visual inspection of the maximum activation on a grand-average SCD map for harmonic sounds against a background of silence (**Fig. 5**). The dipole strength was averaged within each ROI. The peak amplitudes and latencies of the dipole strength were measured within 100-200-ms (MMNm) and 200-300-ms (P3am) time windows by automatically determining the maximum peak in the window.

3.7. Statistical analysis

The EEG, MEG, and behavioral results were analyzed with the multifactor repeated-measures analysis of variance (ANOVA). The Greenhouse-Geisser correction was applied whenever the factor contained more than two levels.

In **Study I**, 4-way ANOVAs were used to determine whether the component type (MMN or P3a), sound structure (levels: Sinusoidal/Harmonic), magnitude of deviance (levels: 2.5, 5, 10, and 20%), or the standard-stimulus frequency (250, 500, 1000, 2000, 4000 Hz) affected the ERP amplitude and latency at the Fz electrode (for the MMN) or Cz (for P3a). If so, then the Tukey HSD test was applied as a post-hoc comparison. In the case of a significant interaction, the data were divided accordingly to determine the main effects of each factor on each level of the other factor (Roberts and Russo, 1999). Bonferroni correction was applied to the main effects to compensate for the repetitive measures.

In **Study II**, each averaged post-stimulus interval value (pooled across the frequencies, deviances, and electrodes) was compared with the pre-stimulus averaged interval value in a t-test with Bonferroni correction for multiple comparisons. Thereafter, a 4-factor repeated-measures ANOVA with factors Interval (-100-0, 0-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800 ms), Frequency (250, 1000, and 4000 Hz), Deviance (5%, 20%), and Electrode (F3, F4, C3, C4, FCz, P3, P4, T3, T4) was performed. Since in one interval (200-300 ms) the difference wave was marginally significantly different from the pre-stimulus interval, this interval and the pre-stimulus interval were further analyzed with a 4-factor repeated-measures ANOVA with factors Interval (-100-0, 200-300), Frequency (250, 1000, 4000 Hz), Deviance (5%, 20%), and Electrode (F3, F4, C3, C4, FCz, P3, P4, T3, T4).

In **Study III**, the amplitudes and latencies were analyzed with a 5-way ANOVA with factors Window (100-200 vs. 200-300), Noise (noise vs. silence), Sound (harmonic vs. sinusoidal), Frequency (high vs. low), and Locus (Right Inferior Frontal vs. Right Centro-Parietal vs. Right Supratemporal vs. Left Inferior Frontal vs. Left Centro-Parietal vs. Left Supratemporal). The significant interactions of the factors were post-hoc analyzed with the Tukey Honest Significant Difference (HSD) test.

In **Study IV,** 2-way ANOVAs with factors Noise Condition (Noise vs. No Noise) and Sequence Type (Tones vs. Chords) were used to compare the amplitudes and latencies of all ERP peaks with each other. Four-way ANOVAs with factors Noise Condition, Sequence Type, Hemisphere (electrodes C3/T3/LM vs. C4/T4/RM), and Laterality (C3/C4 vs. T3/T4 vs. LM/RM) were performed for the mean ERP amplitudes over the 50-ms windows to examine their lateral distributions.

In **Study V,** for all statistical comparisons of the EEG data, a 4-way ANOVA for repeated measures was applied, with factors Noise (no noise vs. fMRI noise), Task (location vs. pitch), Condition (memory, match, and nonmatch), and Electrode (9 sites). If significant main effects or interactions were observed ($p < 0.05$), then post-hoc analyses were performed using the Newman-Keuls test for the main effects and the Tukey HSD test for the interactions. The effect of the electrode was caused by minor differences in the

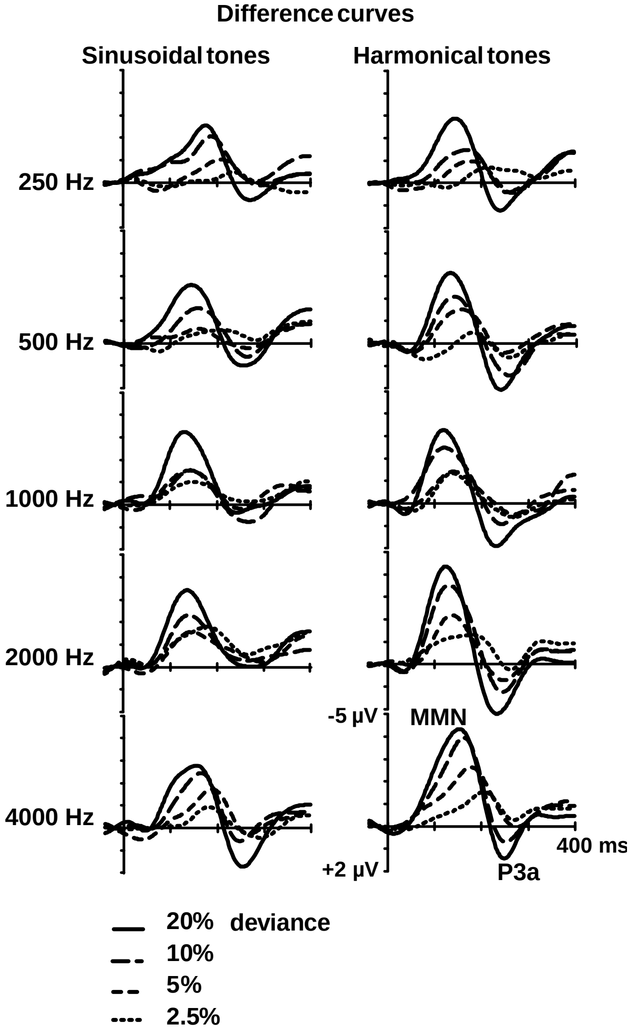

Figure 3. Study I: The difference curves (deviant- minus standard-tone ERPs) at the Fz electrode with the sinusoidal (left panel) and harmonic tones (right panel) at different standard-stimulus frequency levels and with the different magnitudes of deviation as indicated below the figure. These data, used in statistical analysis, were referenced to the mastoid-lead data.

ERP scalp distribution within the electrode matrix and was therefore not analyzed further. With the results of a previous Study (Anourova *et al.*, 2001) taken into account, planned comparisons between the N1 latencies in the location and pitch tasks were performed in the match condition in the noise and no-noise blocks.

4. RESULTS

4.1. Study I: Pitch discrimination in the 250-4000 Hz frequency range

In this study, ERPs were recorded in an oddball paradigm with 5 different standard-tone frequencies and 4 degrees of frequency change between the deviant and standard. Both pure tones and 3-harmonic complex tones were used. In a separate session, subjects performed a behavioral same-different task with the same stimuli. We made a systematic investigation of the effect of standard-sound frequency, the magnitude of change, and the sound type on the amplitude and latencies of change-related ERPs (the MMN and P3a) and behavioral measures (HR, RT, and d') as well as on the correlation between the ERP and behavioral parameters.

Frequency changes elicited a significant MMN as indexed by the negative displacement of the ERPs for deviant tones at the Fz electrode and by the accompanying positivity at the corresponding latency at the right mastoid for all standard-stimulus frequencies (250 … 4000 Hz; **Fig. 3**). The MMN amplitude was significant for all pitch-change levels (2.5 … 20%) used. At larger pitch changes, the MMN was followed by the P3a.

The standard-stimulus frequency affected both the ERP latency and amplitude as well as the HR. The MMN and P3a latencies for the 2.5% and 5% pitch changes were shorter at 1000 Hz than at the 250 Hz and 4000 Hz standard-stimulus frequencies. On the other hand, the MMN was smaller in amplitude at the 250 and 500 Hz than at the 2000 and 4000 Hz standard-stimulus frequencies. The HR for the 2.5% and 5% deviance discrimination was less accurate at the 250 Hz standard-stimulus frequency than at the higher standard-stimulus frequencies. The MMN amplitude was larger than that of the P3a for the higher standard-stimulus frequencies of 1000, 2000, and 4000 Hz with the 10% deviance.

The magnitude of the frequency change (deviance level) affected the MMN and P3a latencies and amplitudes as well as the HR and RT. The MMN and P3a latencies were longer for the 2.5% than those for the 10% and the 20% deviances, and those for the 5% deviance were longer than those for the 20% deviances at the 250, 500, and 4000 Hz standard-stimulus frequencies. The MMN amplitude steadily increased with the deviance increase from the 5% to 20% deviance, whereas the P3a amplitude increased only with the increase of deviation from 10% to 20%. In addition, the HR was less accurate at the 2.5% deviance level than at the 10% and 20% levels; it was also less accurate at the 5% level than at the 20% deviance level with the 250, 500, and 1000 Hz standard-stimulus frequencies. The RT was longer for the 2.5% deviance level

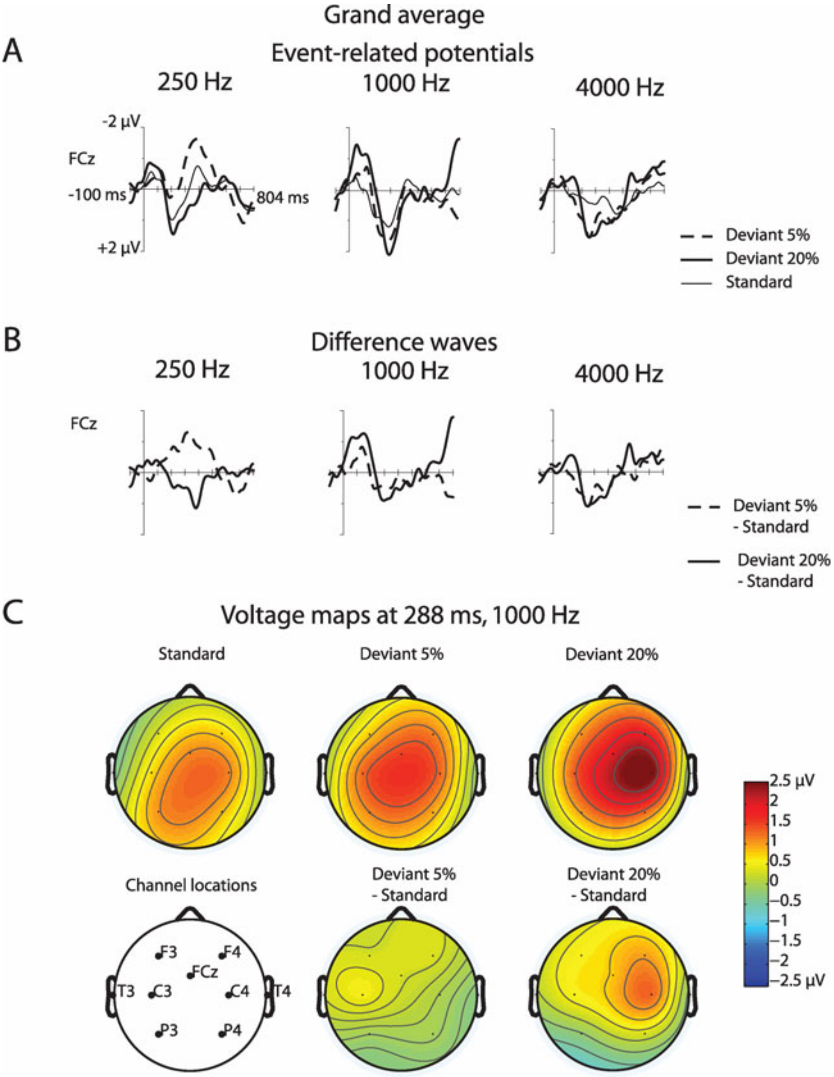

Figure 4. Study II: The ERPs to standards and deviants (A) and the grand-average deviant-standard difference waves (B) at FCz for the 5% and 20% frequency deviances at the 250, 1000, and 4000 Hz frequencies of the standard. The grand-averaged ERPs and difference waves were plotted with the standard Matlab interpolation function at the 288-ms time point that corresponded to the peak in the interval of interest yielded by statistical analysis. Note that voltage distribution outside the electrode locations is hypothetical and should be interpreted with caution (C).

than that for the 10% and 20% deviances. In addition, the RT was longer for the 5% than for the 20% deviance.

The effect of the sound structure was present in the MMN and P3a amplitudes and latencies as well as in the RT and d' results. At the 20% deviation, the harmonic sounds elicited an MMN and P3a with a shorter latency and a larger amplitude than did the sinusoidal sounds, resulting in a significant main effect of the sound structure for the whole data set. The RT was longer for the sinusoidal than for the harmonic sounds with the 2000 Hz standard-stimulus frequency. The discrimination of the harmonic tones resulted in a higher d' sensitivity than the discrimination of the sinusoidal tones with the 250 and 500 Hz standard-stimulus frequencies.

Finally, the MMN and P3a amplitudes and latencies significantly correlated with the HR and RT measures.

To summarize, the MMN amplitude was significant for the 250-4000 Hz standard-frequency levels for a change as low as 2.5% and the amplitude increased with the deviation level from 5%. The latencies of the MMN and P3 had their minima at the 1000 Hz standard-stimulus frequency and they were prolonged with higher and lower frequencies, whereas the MMN and P3a amplitudes were higher at the high frequencies (>500 Hz) than at the low standard-sound frequencies. At the 20% frequency change the harmonic tones elicited higher amplitude and shorter latency MMNs and P3as than those elicited by the pure tones. If judged on the basis of ERP latency, automatic frequency discrimination was the most accurate at the 1000 Hz standard-sound frequency.

4.2 Study II: Frequency discrimination in the 250-4000 Hz frequency range in newborns

In this study, a simplified version of the oddball paradigm relative to that in **Study I** was applied to newborn babies. Pure tones with 3 standard frequencies and 2 degrees of frequency change between deviant and standard were used. The objective was to establish the limits of frequency discrimination within the 250-4000 frequency range in newborns and compare the results with the corresponding results of adults.

There was substantial individual variation in the newborns' waveform shape, with both negative and positive deflections being present. The most prominent deflection in their grand-average ERPs was a positivity with a maximum around 200-300 ms and centro-frontal distribution. This positivity was more prominent at high frequencies than at low standard frequencies (**Fig. 4**).

Comparison of the post-stimulus intervals versus pre-stimulus interval indicates that only the averaged response in the 200-300-ms post-stimulus interval was marginally significantly larger than that in the pre-stimulus baseline interval. The response to the 20% deviant was more positive than the pre-stimulus interval and was larger in amplitude than the response to the 5% deviation.

The data suggest that newborns can discriminate a 20% frequency change in the 250-4000 Hz frequency range, whereas the discrimination of a 5% change was not confirmed. The threshold of automatic frequency discrimination in newborns was thus higher than in adults (cf. **Study I**).

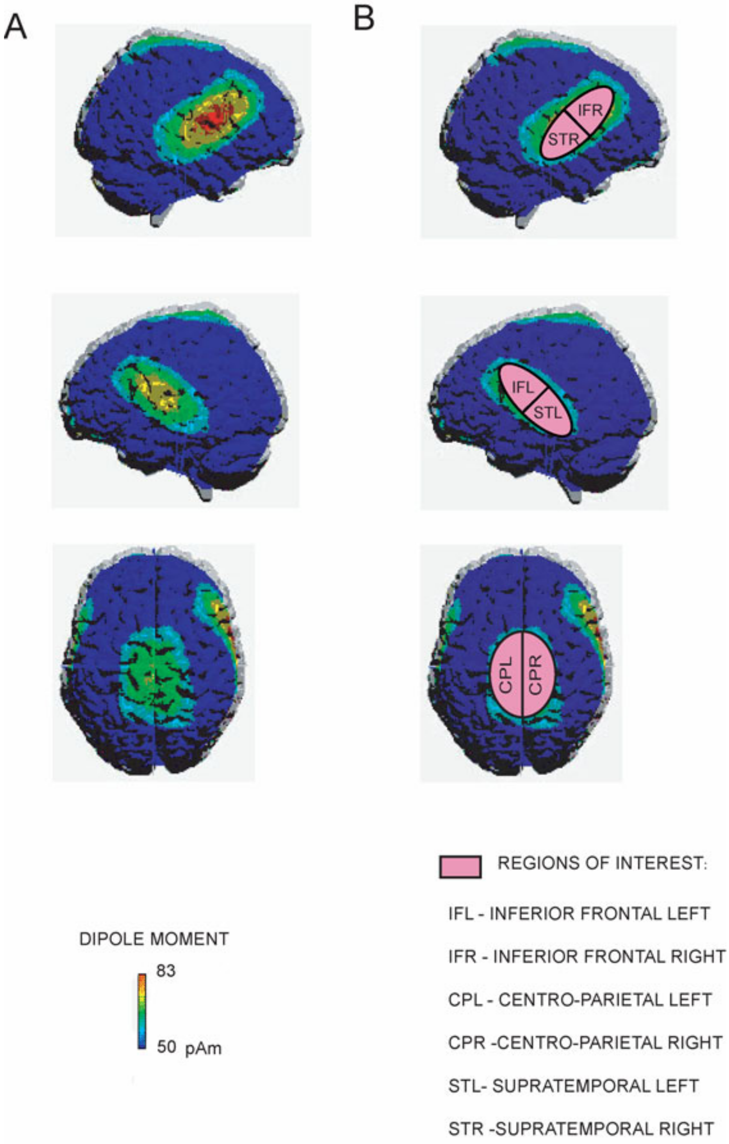

Figure 5. Study III: The grand-average (17 subjects) source-current distribution (SCD) map for harmonic tones in silence (the left column) and the regions of interest derived from this map (the right column): right supratemporal (STR), right inferior frontal (IFR), right centro-parietal (SPR), left supratemporal (STL), left inferior frontal (IFR), and left centro-parietal (SPL) regions. The SCD map is superimposed upon the Montreal brain phantom.

4.3. Study III: Frequency-specific effects of fMRI acoustic noise on pitch discrimination

In this study pure tones and 3-harmonic complex tones of 240-1240 Hz fundamental frequencies were presented in a roving-standard paradigm with a silent background and with the background of the fMRI acoustic noise. The MEG was recorded in an ignore condition, and behavioral discrimination was measured in a separate session. The neuronal activity underlying the MEG

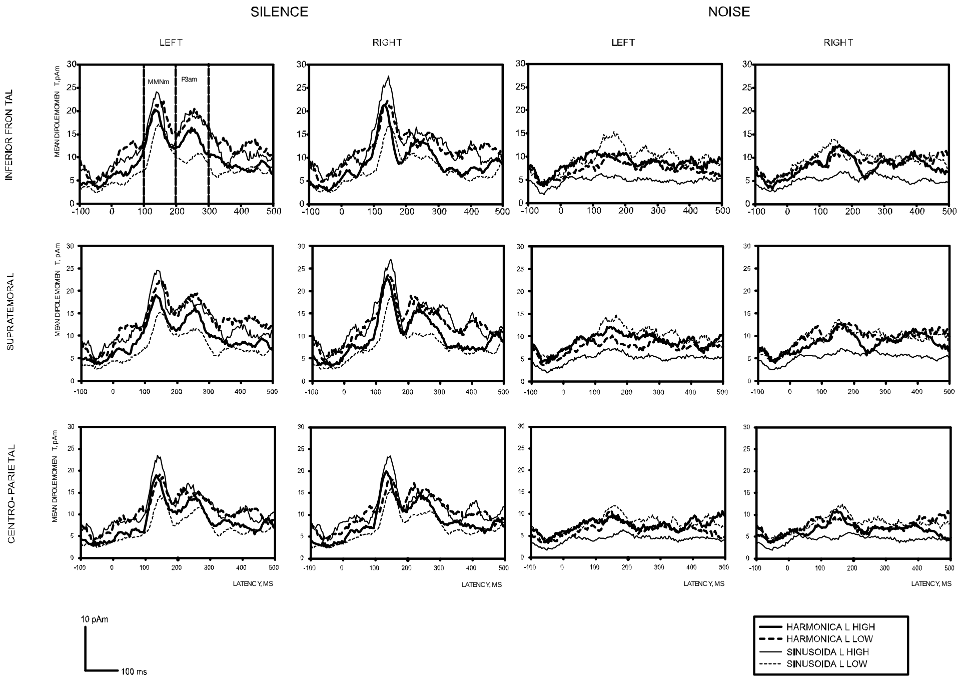

Figure 5. Study III: The mean time-course of the dipole moment amplitude for the SCD map in the regions of interest (grand-averaged across 17 subjects).

responses was modelled with the L2 norm estimate and measured in the regions of interest (**Fig. 5**). The aim was to compare the effect of noise on the MMN response to the pitch change that overlapped the spectral maximum of the noise (>500 Hz) with MMN to the pitch change that did not overlap the noise spectrum. Since the amplitude of the MMN in silence was larger for a pitch range higher than 500 Hz in **Study I**, the interaction of this effect and the possible effect of noise was of particular interest.

The time course of activation showed two prominent peaks, one within the 100-200-ms time window and the other within the 200-300-ms time window (**Fig. 6**). Based on their latency, the peaks are interpreted as the magnetic counterparts of the MMN and P3a (Alho *et al.*, 1998b).

The SCD peak amplitude was higher in the 100-200-ms window (MMNm) than in the 200-300-ms window (P3am). The SCD averaged amplitude in both time windows was lower in the centro-parietal loci than in the supratemporal and inferior frontal loci. The MMNm was statistically significant for all tones in silence and for the low sinusoidal, high harmonic, and low harmonic tones in noise. The P3am amplitude was significant for the high sinusoidal, low sinusoidal, high harmonic, and low harmonic tones in silence, shorter for high tones than for low tones and shorter in the right supratemporal region than in the left supratemporal region.

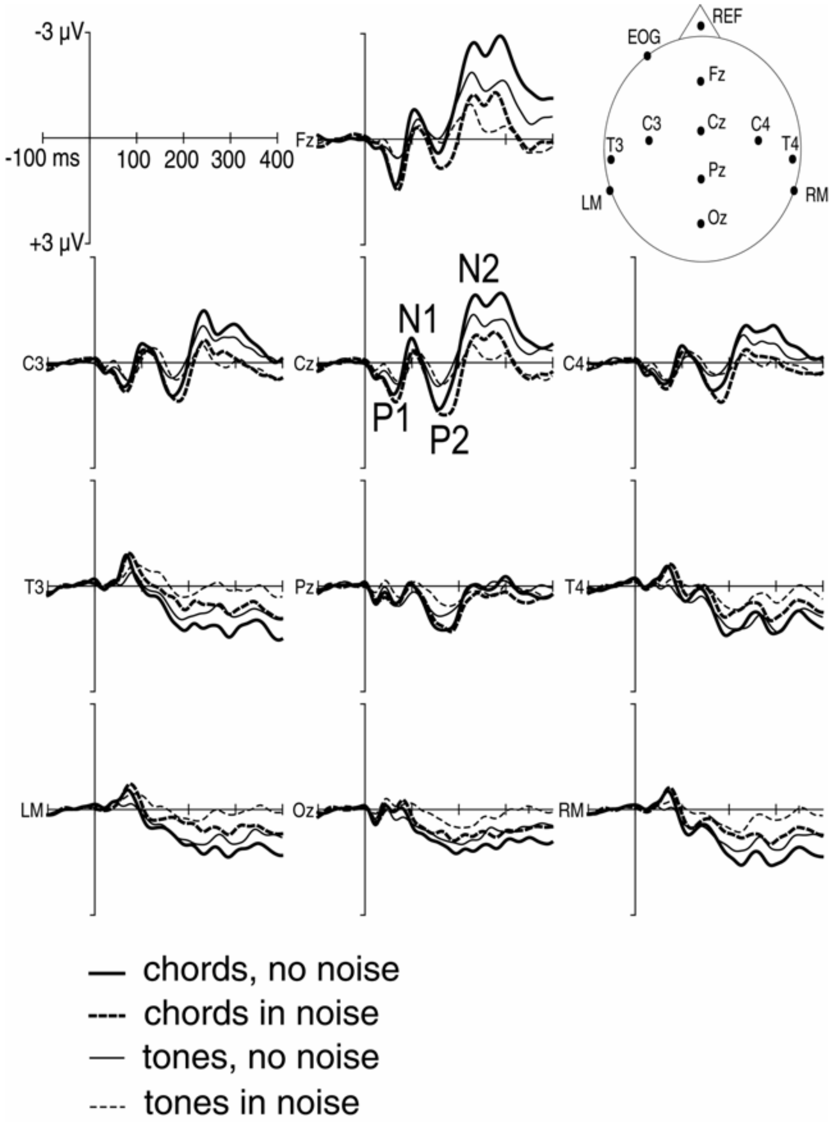

Figure 7. **Study IV**: The difference waves obtained by subtracting the grand–average ERPs to standard sounds from those to deviant sounds as recorded in two stimulus sequence types and in two noise conditions: chord sequence with background noise (dashed thick line), chord sequence without noise (solid thick line), tone sequence with noise (dashed thin line), and tone sequence without noise (solid thin line). Stimulus onset at 0 ms. The schematic representation of the position of 10 electrodes is indicated in the upper right corner together with the location of the reference electrode (REF) and electro-oculogram (EOG) electrode. Time and amplitude scales are given in the upper left corner.

The fMRI acoustic noise suppressed the MMNm and P3am amplitudes for the high sinusoidal, high harmonic, and low harmonic tones. In addition, the acoustic noise decreased the HR for the high tones and prolonged the RT for the harmonic and high tones. In silence, high sinusoidal tones elicited larger-amplitude MMNm and P3am than those elicited by low sinusoidal tones, whereas in noise, low sinusoids elicited larger-amplitude responses than those elicited by high sinusoids. In silence, low harmonic tones elicited larger-amplitude MMNm and P3am responses than those elicited by low sinusoidal tones, and in noise, the responses to high harmonic tones were larger in amplitude than those to high sinusoids. In noise, the HR was lower for high harmonic tones than for high sinusoidal tones.

To summarize, the fMRI acoustic noise suppressed amplitudes of the MMNm and P3am to pitch change within the frequency range that at least partly overlapped the spectral maximum of the noise. The behavioral responses were also deteriorated for these sounds. The MMN and P3a had larger amplitudes for the high tones by contrast with the low tones in silence (as in **Study I**), whereas their amplitudes for the low tones were larger than those for the high tones in the fMRI acoustic noise background.

4.4. Study IV: Effects of fMRI acoustic noise on ERPs in the oddball paradigm

In this study, an oddball paradigm with pure tones and musical chords (fundamental frequencies 262-415 Hz) was used in different blocks and natural novel sounds as additional deviants. The sounds were presented either against silent background or with the fMRI noise background. The objective was to determine the effect of the noise on the obligatory ERPs to the sounds as well as on the MMN and P3a to sound change. The sounds were of different spectral complexity and attention-catching salience.

Standard chords and tones elicited ERPs with P1, N1, P2, and exogenous N2 waves (**Fig. 7**). The mean ERP amplitudes at the lateral electrodes over the 50-ms periods centered at the N1, and N2 peak latencies were prolonged over the right as compared to the left hemisphere. The deviant-standard ERP difference waves showed MMN for deviant tones and chords at 120–250 ms from stimulus onset, and the novel-standard ERP difference waves revealed a negativity, presumably caused by an enhanced N1 response and the MMN, followed by a large P3a response. The mean peak amplitudes and latencies of these waves are given in **Table 2**.

The P1, N1, and P2 peak latencies at Cz were prolonged, and the N2 peak amplitude was reduced by the fMRI acoustic noise. There were no significant effects of the noise on the MMN for the deviant tones or chords or on the N1/MMN or P3a for the novel sounds.

The peak latency of the P1 was longer and the peak amplitudes of the P1, P2, and N2 were smaller for tones than for chords. The MMN peak latency at Fz was longer and the MMN peak amplitude smaller for the deviant tones than for deviant chords (**Table 2**). The N1/MMN peak latency at Cz was longer

Type of ERP response	Component of ERP	Type of stimulus sequence	Latency of the peak, Ms		Amplitude of the peak, μV	
			No noise	MRI noise	No noise	MRI noise
Standard sounds	P1	Chord	63±3.3	69±3.3	1.2±0.15	1.4±0.11
		Tone	73±4.9	80±3.2	0.7±0.18	0.8±0.17
	N1	Chord	102±3.8	109±3.7	-1.1±0.32	-0.7±0.26
		Tone	112±3.5	114±2.9	-0.7±0.20	-0.5±0.19
	P2	Chord	157±3.8	165±4.6	1.4±0.35	1.6±0.26
		Tone	158±4.7	163±6.5	0.8±0.25	0.7±0.14
	N2	Chord	232±3.0	233±2.4	-2.9±0.41	-1.3±0.25
		Tone	227±1.8	220±4.2	-2.1±0.36	-1.1±0.17
Deviant sounds	MMN	Chord	144±7.4	153±8.4	-3.8±0.76	-4.2±0.62
		Tone	251±24.1	250±27.8	-3.2±0.47	-2.4±0.33
Novel sounds	N1/MMN	Chord	149±5.2	141±4.2	-5.1±0.52	-4.4±0.72
		Tone	118±4.5	119±6.3	-4.7±0.76	-3.8±0.67
	P3a	Chord	240±6.6	240±6.6	7.2±1.23	6.9±0.96
		Tone	228±11.8	226±10.3	9.2±1.16	8.8±1.07

TABLE 2: Study IV. Mean (±SE) latencies and peak amplitudes of ERP components. *Note.* Mean (±SE) latencies and peak amplitudes of the P1, N1, P2 and N2 deflections in the ERP to standard sounds at Cz, the MMN to deviant sounds (measured at Fz from deviant-standard ERP difference waves) and of N1/MMN and P3a responses to novel sounds (measured from the novel-standard ERP difference waves at Cz) in two types of stimuli sequence and two noise conditions

for novel sounds among chords than for those among tones. The P3a peak amplitude at Cz, in turn, was larger for novel sounds among tones than that for those among chords.

To summarize, the noise had a suppressing effect on the obligatory responses to both spectrally simple and complex sounds. However, the MMN and P3a responses to sound change were not significantly affected by noise. Since the fundamental frequency of the tones and chords in this study was lower than 500 Hz and did not overlap with the major spectral maxima of the noise, the results are compatible with those of **Study III.** The ERPs to the spectrally rich sounds were in general shorter in latency and larger in amplitude than those to the pure tones, which is consistent with the results of **Study I**.

4.5. Study V: Effects of fMRI acoustic noise on ERPs in the working-memory task

In this study, a paired matching-to-sample paradigm with pure tones that varied in both pitch (1000 vs. 1500 Hz) and location (left vs. right) was applied while subjects had to match the pairs in either pitch or location in different

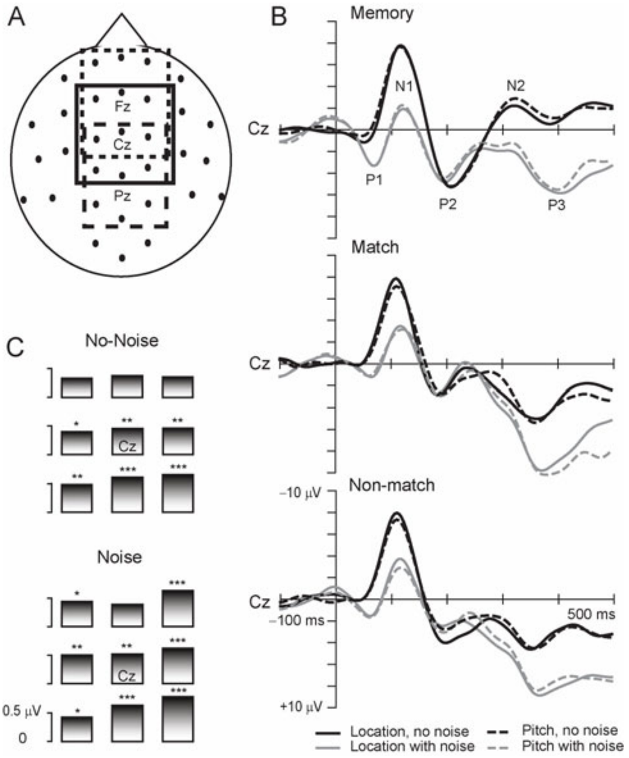

Figure 8. Study V: (**A**) The schematic representation of the positions of the electrodes in a 32-channel cap. The 9-electrode matrices used for the analysis of the different components are marked with rectangles. (**B**) Grand-averaged ERPs in the memory, match, and non-match conditions. (**C**) The difference in the N1 amplitude between the pitch and location tasks plotted as bars at each electrode of the matrix separately for the no-noise and the noise conditions. The significant task-related differences are marked with asterisks (* = p<0.05, ** = p<0.01, *** = p<0.001).

blocks. The pauses between the tones were either silent or filled with bursts of fMRI acoustic noise. The objective was to determine the effect of noise on the ERPs of the tonal cues (memory) and probes (match and non-match conditions) and on the difference in the ERPs between the location and pitch tasks. Thus the effect of noise on the ERPs in the active pitch-discrimination task was compared with the noise effect on the ERPs in automatic pitch discrimination in **Studies III** and **IV**.

Both location and pitch tasks were subjectively rated as more difficult in the noise blocks than in the no-noise blocks, although there was no difference in the behavioral performance between the noise and silent conditions. In addition, the N1 amplitude was larger in the location than the pitch task. This task-related difference was significant in 6 out of the 9 electrode sites of the matrix analyzed in the no-noise condition and in 8 sites during the presentation of noise (**Fig. 8C**). The planned comparisons demonstrated that the N1 latency in the match condition was shorter in the location than the pitch task in the no-noise blocks, but not in the noise blocks.

The fMRI noise increased the P1 amplitude and prolonged its latency (**Fig. 8B**). In addition, the N1 amplitude was reduced and the N1 latency in the match and non-match conditions was prolonged by the fMRI noise. Also, the N2 amplitude was reduced by the fMRI noise in the memory condition. The N2 latency was shortened, the P3 amplitude was increased and the P3 latency was prolonged by the fMRI noise in all conditions.

In the responses to the probes, the P1 and P2 amplitudes were decreased, while the N1 amplitude in the noise condition and the P3 amplitude were increased as compared with the response to the cues. The latencies of the P1, N1, and N2 were shortened in the responses to the probes relative to those in the responses to the cues. The N2 amplitude in the location task was larger in the match than in the non-match condition.

To summarize, the fMRI acoustic noise increased the amplitude, but delayed the latency of P1, diminished the amplitude, and delayed the latency of the N1, diminished the amplitude but shortened the latency of the N2, and increased the amplitude but prolonged the latency of the P3 response. The prolongation of the P1 and N1 latencies and the increase in the N2 amplitude are consistent with the results of **Study IV**. The noise spared the task-related amplitude difference in the N1 amplitude, whereas the amplitude itself was substantially diminished. This suggests that the noise suppressed more the early stages of analysis reflected in the afferent part of the N1 response, while the cognitive aspect was less affected. The latter can even be enhanced in the presence of noise as suggested by the increase in the P3 amplitude.

5. GENERAL DISCUSSION

5.1 The MMN as an objective measure of pitch discrimination

The MMN is sensitive to very small pitch changes and correlates with behavioral accuracy (Sams *et al.*, 1985; Amenedo and Escera, 2000; Tiitinen *et al.*, 1994; Jaramillo *et al.*, 2000). **Study I** of the present thesis demonstrated that pitch discrimination as indexed by the latency of the MMN is the most accurate in the 1000-2000 Hz frequency range, which corresponds to known behavioral data (Wier *et al.*, 1977; Sek and Moore, 1995). The MMN amplitude increases with pitch contrast and correlates with behavioral performance. The addition of higher harmonics improves pitch discrimination as indexed by the MMN in all frequency ranges. Therefore, the MMN can be used as a measure of pitch perception. The application of the same paradigm in newborns in **Study II** produced a mismatch response (MMR) of positive polarity, which can only be obtained for larger pitch changes than the MMN in adult subjects. Taken together with the results from the previous studies, these results demonstrate that it is possible to determine pitch-discrimination accuracy in populations where behavioral measurement is difficult or impossible.

Study I found that the MMN and P3a amplitudes as well as latencies correlated highly significantly with the subject's behavioral discrimination accuracy as indicated by the HR and RT at each standard-stimulus frequency level. In spite of the fact that the accuracy of pitch discrimination was lower than in the psychophysical studies (Wier *et al.*, 1977), this is an encouraging result when educational and clinical applications of the ERP recordings are considered. For instance, in exploring the limits of expertise of a population with specific, e.g., musical, talents (e.g., Tervaniemi, 2001), using stimulation with acoustic properties relevant to the ability under interest is of essential importance. Thus, if active paradigms are not suitable for practical or methodological reasons (e.g., in infants), then the MMN recordings might offer a probe into the integrity and limits of neural auditory discriminative functions (Tervaniemi and Huotilainen, 2003). It is clear by now that the neonatal auditory ERPs can serve as a predictor for future risks of language and learning problems (Molfese, 2000; Molfese *et al.*, 2001; Benasich and Tallal, 2002; Guttorm *et al.*, 2005). The role of infantile change-related responses may in the future be even more important than the mere responses to sound onset, since these MMN-type responses require the functioning of a higher-level cognitive system than the mere onset detection. Adequate functioning of short-term memory and automatic sound comparison are prerequisites for normal auditory perception.

The predominant polarity of the mismatch response in newborn data in **Study II** was positive. On comparing the newborn MMR of **Study II** with the corresponding adult MMN in **Study I,** it is evident that the threshold of MMR elicitation is substantially higher in newborns than in adults. In adults, a 5% deviance elicited a significant MMN in the 250 to 4000 Hz range in **Study I,** and the threshold value for MMN elicitation at 1000 Hz is known to be about 0.8% (Sams *et al.*, 1985). In contrast, in the newborns the 5% difference produced no MMR. These data indicate that the behavioral data suggesting the immaturity of frequency discrimination in older infants (Olsho *et al.*, 1982, 1987) reflects genuine auditory-system immaturity rather than undeveloped attention. It can thus be concluded that newborn infants discriminate a frequency change of 20% in the 250-4000 Hz frequency range, whereas the discrimination of 5% difference was not confirmed.

In **Study I** pitch deviances in the harmonic tones elicited an ERP with a shorter latency and larger amplitude than those elicited in sinusoidal tones. No sound type –frequency interaction was obtained in statistical analysis, thus, the earlier finding of pitch discrimination among harmonic sounds being facilitated when compared to sinusoidal sounds at only 500 Hz (Tervaniemi *et al.*, 2000a, 2000c) was confirmed and extended to a wide frequency band. Interestingly, the difference between the 1-harmonic (sinusoidal) and the 3-harmonic tones was found for the standard-stimulus fundamental frequency to be as high as 4000 Hz, even though its highest harmonic (12000 Hz) is already close to the limits of human hearing. The MMN for the deviant chords, too, in **Study IV** was larger in amplitude and shorter in latency than the MMN for deviant tones.

Furthermore, the results of **Study I** show that the P3a latency is shortened with spectrally rich sounds when compared with that for pure sinusoidal tones. This might be due to the larger frequency deviance employed in the present (2.5-20%) than in the previous study (2.5-10%; Tervaniemi *et al.*, 2000a). In previous studies (Tervaniemi *et al.*, 2000a; Tervaniemi *et al.*, 2000c) comparing pitch discrimination for sinusoidal vs. spectrally rich sounds, the 10% frequency was not large enough for the P3a elicitation; furthermore the P3a elicited by novel sounds among tones in **Study IV** was larger in amplitude than that elicited by novel sounds among chords. This finding could be explained by comparing the spectra of the stimulation, namely, tones, chords, and novel sounds. The difference between a spectrally rich natural sound and a chord is smaller than that between a natural sound and a pure tone. Therefore, a natural sound among tones is more salient than the same sound among chords, probably causing a stronger involuntary attention shift, which might explain a large P3a (cf. Escera *et al.*, 1998; 2000).

The present data might also have implications for the pitch perception theory. The change-related ERPs in **Study I** recorded at lower frequencies (250

Hz and 500 Hz) were clearly different from those recorded at higher frequencies (2000 and 4000 Hz). The most prominent is the discrepancy in the way the MMN amplitudes and latencies change as a function of frequency. From 250 to 1000 Hz, the MMN amplitude rises and its latency shortens with the increase of frequency. However, on further increase of frequency, the amplitude continues to grow, whereas the latency starts to be prolonged. In **Study III,** high-frequency tones were associated with a larger MMNm and P3am amplitude and with a shorter MMNm latency than with low-frequency tones. The frequency range of 500-1000 Hz might indicate a transition between two different neuronal mechanisms of pitch discrimination, presumably the rate and place mechanisms. Thus, these data indirectly support the view of a border between the rate and time mechanisms being at 500-1000 Hz (cf. Pierce, 1999).

5.2. The ERP and ERF as correlates of different stages in pitch processing

Studies I-V addressed the modification of the ERP/ERF components that corresponded to pitch change. The conspicuous components in different studies are the N1, MMN, P3a, and P3b. Those components are connected to the different stages of sound processing (Näätänen and Winkler, 1999). The N1 corresponds to the feature tracing, the MMN to stimulus representation, the P3a to the change in the attention resource allocation, and the P3b to the update in working memory (for specifications, see below). Of particular interest was comparison of the behavior of those components in the silent and noisy environments with each other.

N1 is elicited whenever a change in the sound pressure occurs, e.g., when sound onset happens against a silent background (Näätänen and Picton, 1987). The N1 amplitude is reduced with the repetition of the sound due to refractoriness, but it recovers whenever a sound is different from the previous one in some acoustical feature, such as frequency. As a response to the contrast in sound pressure, the N1 is diminished when sounds are presented against a noisy background, especially if noise includes the frequencies of the sounds presented (Martin *et al.*, 1999). In the present data, the fMRI acoustic noise prolonged the latency (**Studies IV** and **V**) and decreased the amplitude of the N1 (**Study V**).

The MMN is elicited by any violation of the regularity in the auditory stream (Näätänen, 1992; Näätänen *et al.*, 2004). The MMN represents an ERP probe of the early stages of memory formation (Winkler and Cowan, 2005). In our data, the MMN for pitch change was affected by noise only when the spectrum maximum of the noise overlapped with the frequency of the sounds

(**Studies III** and **IV**). This suggests that the memory-formation mechanisms had a stronger resistance to noise than did the feature-tracing mechanisms.

The P3a and P3b reflect higher-level cognitive processing than did the MMN and N1, involving both attention and memory (Knight and Scabini, 1998; Linden, 2005). The P3a is especially larger for novel sounds that catch the subject's attention (Escera *et al.*, 2000). The P3b during active discrimination is increased with stimulus relevance and decreased with its probability. In our studies, the P3a for novel sounds was not affected by noise (**Study IV**), and the P3b in a matching-to-sample task was increased in amplitude (although delayed in latency) by noise (**Study V**). These data suggest that involuntary attention to ecologically salient sounds is a very robust mechanism in terms of noise resistance, whereas the mechanisms of working-memory update may be enhanced against a noisy background.

In **Study I**, the MMN and P3a for pitch change were directly compared with each other. While the MMN was proved to be a reflection of automatic memory-based change detection (Escera *et al.*, 1998; Näätänen and Winkler, 1999), the P3a correlates with an involuntary attention shift in response to change (Escera *et al.*, 1998; Näätänen and Winkler, 1999). The MMN was elicited at all deviance levels, and its amplitude started to increase from the 5% deviance level. In contrast, the P3a was not always significant, even at the lowest deviance levels, and its amplitude increased only from the 10% to 20% deviance level. The frequency-dependent course of the P3a latency is similar to that of the MMN, whereas the frequency-dependent courses of their amplitudes were different. The MMN amplitude increased with frequency, whereas the P3a amplitude remained stable. Finally, the correlation coefficients between the P3a parameters and the behavioral indices (HR and RT) were smaller than those for the corresponding parameters of the MMN. This supports the view of a threshold-like mechanism behind the P3a elicitation; in other words, for P3a elicitation, a relatively large deviance is necessary.

The N1 amplitude in **Study V** was larger in the location than the pitch task, especially in the nonmatch condition. This finding is in line with earlier studies suggesting a dissociation between the spatial and nonspatial auditory information processing in the human brain (Anourova *et al.*, 2001; Rauschecker and Tian, 2000; Alain *et al.*, 2001). The paradigm in **Study V** was similar to that used in an earlier study (Anourova *et al.*, 2001), demonstrating that the N1 component to the second stimulus of a pair had a shorter latency and larger amplitude in the location than the pitch task when the stimuli of the pair were matched. The finding that the N1 amplitude was larger in the location than in the pitch task demonstrates that despite the general noise-induced decrease of the N1 amplitude, the task-related difference was not abolished by noise. However, the latency difference between the N1 components of the pitch and location tasks occurred in the no-noise condition only. This suggests that the subtle latency differences between the tasks are more vulnerable to the suppressing effect of noise than are the corresponding amplitude differences. Interestingly, this relation between the two tasks was preserved in noise, although the N1 in general was suppressed by noise. This is one more indication that higher cognitive processes are more resistant to noise than is the processing of physical features.

Altogether the observations on the discrepancies between the ERP components against a noisy background imply that fMRI acoustic noise affected the processes of pitch processing in a non-linear way (cf. Talavage and Edmister, 2004). Auditory feature-tracing mechanisms were deteriorated by noise more than the mechanisms of change detection, involuntary attention, and the segregation of the spatial and non-spatial domains.

5.3 Application in the combined EEG/MEG and fMRI studies

According to the present findings (**Studies III-IV**), fMRI acoustic noise has a non-linear effect on the ERPs and ERFs. The MMN response to pitch change is suppressed when the fundamental frequency of the tones overlaps with the spectral maximum of the gradient noise. The spectral maximum of noise lies in the 1000-2000 Hz range coinciding with the best frequency-discrimination region. Therefore, despite the importance of that frequency range in human audition, this spectral area should be avoided in fMRI studies. Although the MMN is not affected when the fundamental frequency is outside the spectral maximum of noise, the N1 is diminished and/or delayed, even when target stimulation is not overlapping the spectrum of the noise or is separated in time from the noise bursts (**Studies IV** and **V**). On the other hand, the P3a to the novel sound was not affected, even though the novel sounds had a broadband spectrum. Finally, the dissociation between the pitch- and location-discrimination processing was preserved in the fMRI noise background. Thus, the noise mostly affected early sensory processes, whereas cognitive processing was not affected by noise. Therefore, in future fMRI studies, the stimulation designs should carefully check for pure sensory differences between the stimulations. Furthermore, when fMRI and EEG/MEG are conducted in separate sessions, fMRI acoustic noise should be played as a background in the EEG/MEG study to make the results of the two methods as compatible as possible.

Although both ERP and fMRI reflect postsynaptic activity (Logothetis *et al.*, 2001; Shmuel *et al.*, 2006), the exact relation of ERP or fMRI to neuronal activity has not been established. Yet the combined fMRI/ERP studies suggest that there is a correlation between the results obtained with these two methods

(Opitz *et al.*, 2002). Further investigation is needed to determine the relationship between the ERP-amplitude and -latency changes in detail and the change in the fMRI signal. It is reasonable, however, to think that the effect of noise on the ERPs reflects its effect on neuronal processing. Therefore, the noise accompanying fMRI most probably also has an influence on the results obtained with this method.

The findings of **Study IV**, namely, that the P3a for novel sounds is more pronounced among tones than among complex sounds and, further, that P3a is not affected by scanner noise, suggest that it advisable to use pure tones instead of complex sounds as standard stimuli when studying involuntary attention to novel sounds with the fMRI.

Study V showed that fMRI noise strongly affects the ERPs recorded during auditory delayed matching-to-sample tasks. However, despite the modulatory effect of noise on the ERPs, the task-related difference observed in the N1 component, suggesting a functional dissociation between the processing of the spatial and nonspatial auditory information, was partially preserved in the noise condition. Therefore, the effects of the fMRI noise, the characteristics of the task-related stimuli, and the difficulty level of the task, which modulates the fMRI signal (Carlson *et al.*, 1998; Martinkauppi *et al.*, 2000), should be taken into consideration when designing auditory working-memory experiments for the fMRI.

The current tendency in auditory fMRI studies is to use so-called sparse sampling (Eden *et al.*, 1999; Edmister *et al.*, 1999; Hall *et al.*, 1999). In sparse sampling, data collection (and, thus, noise-provoking gradient switching) is compressed in time, and the remaining silent part of the TR is prolonged. The auditory stimulation is presented during the silent intervals between the fMRI-data acquisitions. The TR in the sparse-sampling paradigms is often as long as 9-18 s (Hart *et al.*, 2002; Gaab *et al.*, 2003; Scarff *et al.*, 2004), and thus the duration of the experiment is substantially increased. Despite this methodological constraint, our current results support the usefulness of the sparse sampling technique in the fMRI studies of audition.

6. CONCLUSIONS

The results of Studies I-V allow the following conclusions to be drawn:

1. In adults, the accuracy of pitch discrimination as reflected by the MMN latency is best in the 1000-2000 Hz frequency range and is facilitated by adding harmonics to the fundamental frequency;
2. Newborn infants are able to discriminate a 20% frequency change in the 250-4000 Hz frequency range, whereas the discrimination of a 5% frequency change was not confirmed;
3. The effect of the fMRI gradient noise on the automatic processing of tones was more prominent for tones with frequencies higher than 500 Hz, overlapping with the spectral maximum of the noise;
4. When the fundamental frequency of the tones was outside of the spectral maximum of the noise, then the MMN and P3a were not affected by fMRI noise, whereas the N1 and the exogenous N2 were delayed and suppressed by the noise;
5. Noise suppressed the N1 response in a matching-to-sample task, but the task-related difference observed in the N1 component, suggesting a functional dissociation between the processing of spatial and non-spatial auditory information, was partially preserved in the noise condition;
6. Automatic pitch discrimination in the optimal acoustic background was more accurate in adults than in newborns. The fMRI acoustic noise affected pitch processing in a non-linear way. Auditory feature-extraction mechanisms were hampered by noise more than were the mechanisms of change detection, involuntary attention, and the segregation between the spatial and non-spatial domains. Therefore, pitch discrimination can be effectively studied with the MMN paradigm in adults and newborns; however, the factor of acoustic noise should be taken into account when those studies are combined with the fMRI.

REFERENCES

Abdala C, Sininger YS, (1996) The development of cochlear frequency resolution in the human auditory system. *Ear Hear*; **17**: 374-385.

Alain C, Arnott SR, Hevenor S, Graham S, Grady CL, (2001) "What" and "where" in the human auditory system. *Proc Natl Acad Sci USA*; **98**: 12301-12306.

Alho K, (1995) Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear Hear*; **16**: 38-51.

Alho K, Connolly JF, Cheour M, Lehtokoski A, Huotilainen M, Virtanen J, Aulanko R, Ilmoniemi RJ, (1998a) Hemispheric lateralization in preattentive processing of speech sounds. *Neurosci Lett*; **258**: 9-12.

Alho K, Sainio K, Sajaniemi N, Reinikainen K, Näätänen R, (1990) Event-related brain potential of human newborns to pitch change of an acoustic stimulus. *Electroencephalogr Clin Neurophysiol*; **77**: 151-155.

Alho K, Winkler I, Escera C, Huotilainen M, Virtanen J, Jääskeläinen IP, Pekkonen E, Ilmoniemi RJ, (1998b) Processing of novel sounds and frequency changes in the human auditory cortex: magnetoencephalographic recordings. *Psychophysiology*; **35**: 211-224.

Amaro E, Jr., Williams SC, Shergill SS, Fu CH, MacSweeney M, Picchioni MM, Brammer MJ, McGuire PK, (2002) Acoustic noise and functional magnetic resonance imaging: Current strategies and future prospects. *J Magn Reson Imaging*; **16**: 497-510.

Amenedo E, Escera C, (2000) The accuracy of sound duration representation in the human brain determines the accuracy of behavioural perception. *Eur J Neurosci*; **12**: 2570-2574.

Anourova I, Nikouline VV, Ilmoniemi RJ, Hotta J, Aronen HJ, Carlson S, (2001) Evidence for dissociation of spatial and nonspatial auditory information processing. *Neuroimage*; **14**: 1268-1277.

Anurova I, Artchakov D, Korvenoja A, Ilmoniemi RJ, Aronen HJ, Carlson S, (2003) Differences between auditory evoked responses recorded during spatial and nonspatial working memory tasks. *Neuroimage*; **20**: 1181-1192.

Baddeley A, (1992) Working memory. *Science*; **255**: 556-559.

Baddeley A, (2000) The episodic buffer: a new component of working memory? *Trends Cogn Sci;* **4**: 417-423.

Bandettini PA, Jesmanowicz A, Van Kylen J, Birn RM, Hyde JS, (1998) Functional MRI of brain activation induced by scanner acoustic noise. *Magn Res Med*; **39**: 410-416.

Basta D, Tzschentke B, Ernst A, (2005) Noise-induced cell death in the mouse medial geniculate body and primary auditory cortex. *Neurosci Lett*; **381**: 199-204.

Belin P, Zatorre RJ, Hoge R, Evans AC, Pike B, (1999) Event-related fMRI of the auditory cortex. *Neuroimage*; **10**: 417-429.

Benasich AA, Tallal P, (2002) Infant discrimination of rapid auditory cues predicts later language impairment. *Behav Brain Res*; **136**: 31-49.

Bendor D, Wang X, (2005) The neuronal representation of pitch in primate auditory cortex. *Nature;* **436**: 1161-1165.

Brattico E, Kujala T, Tervaniemi M, Alku P, Ambrosi L, Monitillo V, (2005) Long-term exposure to occupational noise alters the cortical organization of sound processing. *Clin Neurophysiol*; **116**: 190-203.

Bregman AS, (1990) Auditory scene analysis: the perceptual organization of the sound, The MIT press, Cambridge, Massachusets, London, England.

Bridges AM, Jones DM, (1996) Word dose in the disruption of serial recall by irrelevant speech: phonological confusions or changing state? *Quat J Exp Psychology*; **49A**: 919-939.

Brummett RE, Talbot JM, Charuhas P, (1988) Potential hearing loss resulting from MR imaging. *Radiology*; **169**: 539-540.

Carlson S, Martinkauppi S, Rämä P, Salli E, Korvenoja A, Aronen HJ, (1998) Distribution of cortical activation during visuospatial n-back tasks as revealed by functional magnetic resonance imaging. *Cereb Cortex*; **8**: 743-752.

Carral V, Huotilainen M, Ruusuvirta T, Fellman V, Näätänen R, Escera C, (2005) A kind of auditory 'primitive intelligence' already present at birth. *Eur J Neurosci*; **21**: 3201-3204.

Ceponiene R, Kushnerenko E, Fellman V, Renlund M, Suominen K, Näätänen R, (2002) Event-related potential features indexing central auditory discrimination by newborns. *Brain Res Cogn Brain Res*; **13**: 101-113.

Chambers J, Akeroyd MA, Summerfield AQ, Palmer AR, (2001) Active control of the volume acquisition noise in functional magnetic resonance imaging: method and psychoacoustical evaluation. *J Acoust Soc Am*; **110**: 3041-3054.

Cheour M, Alho K, Ceponiene R, Reinikainen K, Sainio K, Pohjavuori M, Aaltonen O, Näätänen R, (1998) Maturation of mismatch negativity in infants. *Int J Psychophysiol*; **29**: 217-226.

Clark WW, Bohne BA, (1999) Effects of noise on hearing. *JAMA*; **281**: 1658-1659.

Colle HA, (1980) Auditory encoding in visual short-term recall: effects of noise intensity and spatial location. *J Verb Learn Verb Behav*; **19**: 722-735.

Colle HA, Welsh A, (1976) Acoustic masking in primary memory. *J Verb Learn Verb Behav*; **15**: 17-31.

Collins DL, Zijdenbos AP, Kollokian V, Sled JG, Kabani NJ, Holmes CJ, Evans AC, (1998) Design and construction of a realistic digital brain phantom. *IEEE Trans Med Imaging*; **17**: 463-468.

Cowan N, Winkler I, Teder W, Näätänen R, (1993) Memory prerequisites of mismatch negativity in the auditory event-related potential (ERP). *Journal of Experimental Psychology. Learning, Memory, and Cognition*; **19**: 909-921.

DeCasper AJ, Fifer WP, (1980) Of human bonding: newborns prefer their mothers' voices. *Science* **208**: 1174-1176.

Dehaene-Lambertz G, Dehaene S, (1994) Speed and cerebral correlates of syllable discrimination in infants. *Nature*; **370**: 292-295.

Doeller CF, Opitz B, Mecklinger A, Krick C, Reith W, Schröger E, (2003) Prefrontal cortex involvement in preattentive auditory deviance detection: neuroimaging and electrophysiological evidence. *Neuroimage*; **20**: 1270-1282.

Donchin E, Ritter W, McCallum WC, 1978. Cognitive psychophysiology: The endogenous components of the ERPs, in, E. Callaway, P. Tueting, and S. H. Koslow, eds. Event-related brain potentials in man. Academic Press, New York, 349-441.

Draganova R, Eswaran H, Murphy P, Huotilainen M, Lowery C, Preissl H, (2005) Sound frequency change detection in fetuses and newborns, a magnetoencephalographic study. *Neuroimage*; **28**: 354-361.

Edelstein WA, Hedeen RA, Mallozzi RP, El-Hamamsy SA, Ackermann RA, Havens TJ, (2002) Making MRI quieter. *Magn Reson Imaging*; **20**: 155-163.

Eden GF, Joseph JE, Brown HE, Brown CP, Zeffiro TA, (1999) Utilizing hemodynamic delay and dispersion to detect fMRI signal change without auditory interference: the behavior interleaved gradients technique. *Magn Res Med*; **41**: 13-20.

Edmister WB, Talavage TM, Ledden PJ, Weisskoff RM, (1999) Improved auditory cortex imaging using clustered volume acquisitions. *Hum Brain Mapp*; **7**: 89-97.

Eggermont JJ, (2001) Between sound and perception: reviewing the search for a neural code. *Hear Res;* **157:** 1-42

Escera C, Alho K, Schröger E, Winkler I, (2000) Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiol Neurootol*; **5**: 151-166.

Escera C, Alho K, Winkler I, Näätänen R, (1998) Neural mechanisms of involuntary attention to acoustic novelty and change. *J Cogn Neurosci*; **10**: 590-604.

Friston KJ, Penny W, David O, (2005) Modeling brain responses. *Int Rev Neurobiol;* **66**: 89-124.

Gaab N, Gaser C, Zaehle T, Jancke L, Schlaug G, (2003) Functional anatomy of pitch memory-an fMRI study with sparse temporal sampling. *Neuroimage*; **19**: 1417-1426.

Galaburda AM, LeMay M, Kemper TL, Geschwind N, (1978) Right-left asymmetrics in the brain. *Science*; **199**: 852-856.

Garreffa G, Bianciardi M, Hagberg GE, Macaluso E, Marciani MG, Maraviglia B, Abbafati M, Carni M, Bruni I, Bianchi L, (2004) Simultaneous EEG-fMRI acquisition: how far is it from being a standardized technique? *Magn Reson Imaging*; **22**: 1445-1455.

Gerken L, Aslin RN, (2005) Thirty years of research on infant speech perception: the legacy of Peter W. Jusczyk. *Lang Learn Dev*; **1**: 5-21.

Griffiths TD, (2003) Functional imaging of pitch analysis. *Ann N Y Acad Sci;* **999**: 40-49.

Guttorm TK, Leppänen PH, Poikkeus AM, Eklund KM, Lyytinen P, Lyytinen H, (2005) Brain event-related potentials (ERPs) measured at birth predict later language development in children with and without familial risk for dyslexia. *Cortex*; **41**: 291-303.

Hall DA, (2005) Representations of spectral coding in the human brain. *Int Rev Neurobiol*; **70**: 331-369.

Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW, (1999) "Sparse" temporal sampling in auditory fMRI. *Hum Brain Mapp*; **7**: 213-223.

Hall DA, Johnsrude IS, Haggard MP, Palmer AR, Akeroyd MA, Summerfield AQ, (2002) Spectral and temporal processing in human auditory cortex. *Cereb Cortex*; **12**: 140-149.

Hall DA, Summerfield AQ, Goncalves MS, Foster JR, Palmer AR, Bowtell RW, (2000) Time-course of the auditory BOLD response to scanner noise. *Magn Res Med*; **43**: 601-606.

Haller S, Bartsch AJ, Radue EW, Klarhofer M, Seifritz E, Scheffler K, (2005) Effect of fMRI acoustic noise on non-auditory working memory task: comparison between continuous and pulsed sound emitting EPI. *MAGMA*; **18**: 263-271.

Hämäläinen M, Hari R, Ilmoniemi J, Knuutila J, Lounasmaa OV, (1993) MEG: theory, instrumentation and applications to noninvasive studies of the working human brain. *Rev Mod Phys*; **65**: 413-497.

Hari R, Hämäläinen M, Ilmoniemi R, Kaukoranta E, Reinikainen K, Salminen J, Alho K, Näätänen R, Sams M, (1984) Responses of the primary auditory cortex to pitch changes in a sequence of tone pips: neuromagnetic recordings in man. *Neurosci Lett*; **50**: 127-132.

Harris KC, Hu B, Hangauer D, Henderson D, (2005) Prevention of noise-induced hearing loss with Src-PTK inhibitors. *Hear Res*; **8**: 8.

Hart HC, Palmer AR, Hall DA, (2002) Heschl's gyrus is more sensitive to tone level than non-primary auditory cortex. *Hear Res*; **171**: 177-190.

Hart HC, Palmer AR, Hall DA, (2003) Amplitude and frequency-modulated stimuli activate common regions of human auditory cortex. *Cereb Cortex*; **13**: 773-781.Hedeen RA, Edelstein WA, (1997) Characterization and prediction of gradient acoustic noise in MR imagers. *Magn Reson Med*; **37**:7-10.

Hennel F, Girard F, Loenneker T, (1999) "Silent" MRI with soft gradient pulses. *Magn Res Med*; **42**: 6-10.

Herrmann CS, Oertel U, Wang Y, Maess B, Friederici AD, (2000) Noise affects auditory and linguistic processing differently: an MEG study. *Neuroreport*; **11**: 227-229.

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

Hu BH, Henderson D, Nicotera TM, (2002) Involvement of apoptosis in progression of cochlear lesion following exposure to intense noise. *Hear Res*; **166**: 62-71.

Hughes R, Jones DM, (2001) The intrusiveness of sound: Laboratory findings and their implications for noise abatement. *Noise Health*; **4**: 51-70.

Huotilainen M, Kujala A, Hotakainen M, Parkkonen L, Taulu S, Simola J, Nenonen J, Karjalainen M, Näätänen R, (2005) Short-term memory functions of the human fetus recorded with magnetoencephalography. *Neuroreport*; **16**: 81-84.

Jaramillo M, Paavilainen P, Näätänen R, (2000) Mismatch negativity and behavioural discrimination in humans as a function of the magnitude of change in sound duration. *Neurosci Lett*; **290**: 101-104.

Jones DM, Alford D, Macken WJ, Banbury SP, Tremblay S, (2000) Interference from degraded auditory stimuli: linear effects of changing-state in the irrelevant sequence. *J Acoust Soc Am*; **108**: 1082-1088.

Jung KJ, Prasad P, Qin Y, Anderson JR, (2005) Extraction of overt verbal response from the acoustic noise in a functional magnetic resonance imaging scan by use of segmented active noise cancellation. *Magn Reson Med*; **53**: 739-744.

Kandell ER, Schwartz JH, Jessell TM, (2000) Principles of neural science, McGraw-Hill Medical.

Katsunuma A, Takamori H, Sakakura Y, Hamamura Y, Ogo Y, Katayama R, (2002) Quiet MRI with novel acoustic noise reduction. *MAGMA*; **13**: 139-144.

Kim DS, Duong TQ, Kim SG, (2000) High-resolution mapping of iso-orientation columns by fMRI. *Nat Neurosci*; **3**: 164-169.

Knight RT, Scabini D, (1998) Anatomic bases of event-related potentials and their relationship to novelty detection in humans. *J Clin Neurophysiol*; **15**: 3-13.

Kopke R, Bielefeld E, Liu J, Zheng J, Jackson R, Henderson D, Coleman JK, (2005) Prevention of impulse noise-induced hearing loss with antioxidants. *Acta Otolaryngol*; **125**: 235-243.

Kozou H, Kujala T, Shtyrov Y, Toppila E, Starck J, Alku P, Näätänen R, (2005) The effect of different noise types on the speech and non-speech elicited mismatch negativity. *Hear Res*; **199**: 31-39.

Kurtzberg D, Vaughan HG, Jr., (1985) Electrophysiologic assessment of auditory and visual function in the newborn. *Clin Perinatol*; **12**: 277-299.

Langner G, (1992) Periodicity coding in the auditory system. *Hear Res*; **60**: 115-142.Leppänen PH, Eklund KM, Lyytinen H, (1997) Event-related brain potentials to change in rapidly presented acoustic stimuli in newborns. *Developmental Neuropsychology*; **13**: 175-204.

Leventhal AS, Lipsitt LP, (1964) Adaptation, pitch discrimination, and sound localization in the neonate. *Child Dev*; **35**: 759-767.

Levy R, Goldman-Rakic PS, (2000) Segregation of working memory functions within the dorsolateral prefrontal cortex. *Exp Brain Res;* **133**: 23-32.

Liebenthal E, Ellingson ML, Spanaki MV, Prieto TE, Ropella KM, Binder JR, (2003) Simultaneous ERP and fMRI of the auditory cortex in a passive oddball paradigm. *Neuroimage*; **19**: 1395-1404.

Linden DE, (2005) The P300: where in the brain is it produced and what does it tell us? *Neuroscientist*; **11**: 563-576.

Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A, (2001) Neurophysiological investigation of the basis of the fMRI signal. *Nature*; **412**: 150-157.

Luck S, (2005) An introduction to the event-related potential technique, MIT Press, Cambridge, MA.

Lütkenhöner B, Krumbholz K, Seither-Preisler A, (2003) Studies of tonotopy based on wave N100 of the auditory evoked field are problematic. *Neuroimage*; **19**: 935-949.

Magistretti PJ, Pellerin L, (1999) Cellular mechanisms of brain energy metabolism and their relevance to functional brain imaging. *Philos Trans R Soc Lond B Biol Sci*; **354**: 1155-1163.

Malmivuo J, Plonsey R, (1995) Bioelectromagnetism. Principles and applications of bioelectric and biomagnetic fields, Oxford University Press.

Mansfield P, (2004) Snapshot magnetic resonance imaging (Nobel lecture). *Angew Chem Int Ed Engl*; **43**: 5456-5464.

Mansfield P, Chapman BL, Bowtell R, Glover P, Coxon R, Harvey PR, (1995) Active acoustic screening: reduction of noise in gradient coils by Lorentz force balancing. *Magn Res Med*; **33**: 276-281.

Mansfield P, Haywood B, (2000) Principles of active acoustic control in gradient coil design. *MAGMA*; **10**: 147-151.

Mansfield P, Haywood B, Coxon R, (2001) Active acoustic control in gradient coils for MRI. *Magn Res Med*; **46**: 807-818.

Martin BA, Kurtzberg D, Stapells DR, (1999) The effects of decreased audibility produced by high-pass noise masking on N1 and the mismatch negativity to speech sounds /ba/ and /da/. *J Speech Lang Hear Res*; **42**: 271-286.

Martinkauppi S, Rämä P, Aronen HJ, Korvenoja A, Carlson S, (2000) Working memory of auditory localization. *Cereb Cortex*; **10**: 889-898.

Maxon AB, Hochberg I, (1982) Development of psychoacoustic behavior: sensitivity and discrimination. *Ear Hear*; **3**: 301-308.

Mazard A, Mazoyer B, Etard O, Tzourio-Mazoyer N, Kosslyn SM, Mellet E, (2002) Impact of fMRI acoustic noise on the functional anatomy of visual mental imagery. *J Cogn Neurosci*; **14**: 172-186.

Mehler J, Jusczyk P, Lambertz G, Halsted N, Bertoncini J, Amiel-Tison C, (1988) A precursor of language acquisition in young infants. *Cognition*; **29**: 143-178.

Menon RS, Ogawa S, Hu X, Strupp JP, Anderson P, Ugurbil K, (1995) BOLD based functional MRI at 4 Tesla includes a capillary bed contribution: echo-planar imaging correlates with previous optical imaging using intrinsic signals. *Magn Reson Med;* **33**: 453-459.

Miller JD, (1974) Effects of noise on people. *J Acoust Soc Am*; **56**: 729-764.

Moelker A, Maas RA, Pattynama PM, (2004) Verbal communication in MR environments: effect of MR system acoustic noise on speech understanding. *Radiology*; **232**: 107-113.

Moelker A, Pattynama PM, (2003) Acoustic noise concerns in functional magnetic resonance imaging. *Hum Brain Mapp*; **20**: 123-141.

Molfese DL, (2000) Predicting dyslexia at 8 years of age using neonatal brain responses. *Brain Lang*; **72**: 238-245.

Molfese VJ, Molfese DL, Modgline AA, (2001) Newborn and preschool predictors of second-grade reading scores: an evaluation of categorical and continuous scores. *J Learn Disabil*; **34**: 545-554.

Moore BCJ, (1997) An introduction to the psychology of hearing, Academic Press, London.

Moore JK, (2002) Maturation of human auditory cortex: implications for speech perception. *Ann Otol Rhinol Laryngol Suppl*; **189**: 7-10.

Morr ML, Shafer VL, Kreuzer JA, Kurtzberg D, (2002) Maturation of mismatch negativity in typically developing infants and preschool children. *Ear Hear*; **23**: 118-136.

Mukamel R, Gelbard H, Arieli A, Hasson U, Fried I, Malach R, (2005) Coupling between neuronal firing, field potentials, and FMRI in human auditory cortex. *Science*; **309**: 951-954.

Nair DG, (2005) About being BOLD. *Brain Res Brain Res Rev;* **50**: 229-243.

Näätänen R, (1992) Attention and Brain Function, Erlbaum, Hillsdale, NJ.

Näätänen R, Gaillard AW, Mantysalo S, (1978) Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol (Amst)*; **42**: 313-329.

Näätänen R, Jacobsen T, Winkler I, (2005) Memory-based or afferent processes in mismatch negativity (MMN): A review of the evidence. *Psychophysiology*; **42**: 25-32.

Näätänen R, Pakarinen S, Rinne T, Takegata R, (2004) The mismatch negativity (MMN): towards the optimal paradigm. *Clin Neurophysiol*; **115**: 140-144.

Näätänen R, 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.

Näätänen R, Tervaniemi M, Sussman E, Paavilainen P, Winkler I, (2001) "Primitive intelligence" in the auditory cortex. *Trends Neurosci*; **24**: 283-288.

Näätänen R, Winkler I, (1999) The concept of auditory stimulus representation in cognitive neuroscience. *Psychol Bull*; **125**: 826-859.

Nelken I, Fishbach A, Las L, Ulanovsky N, Farkas D, (2003) Primary auditory cortex of cats: feature detection or something else? *Biol Cybern;* **89**: 397-406.

Novitski N, Anourova I, Martinkauppi S, Aronen HJ, Näätänen R, Carlson S, (2003) Effects of noise from functional magnetic resonance imaging on auditory event-related potentials in working memory task. *Neuroimage*; **20**: 1320-1328.

Novitski N, Tervaniemi M, Huotilainen M, Näätänen N, (2004) Frequency discrimination at different frequency levels as indexed by electrophysiological and behavioral measures. *Cogn Brain Res*; **20**: 26-36.

Olsho LW, Koch EG, Halpin CF, (1987) Level and age effects in infant frequency discrimination. *J Acoust Soc Am*; **82**: 454-464.

Olsho LW, Schoon C, Sakai R, Turpin R, Sperduto V, (1982) Preliminary data on frequency discrimination in infancy. *J Acoust Soc Am*; **71**: 509-511.

Opitz B, Rinne T, Mecklinger A, von Cramon DY, Schröger E, (2002) Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. *Neuroimage*; **15**: 167-174.

Pantev C, Roberts LE, Elbert T, Ross B, Wienbruch C, (1996) Tonotopic organization of the sources of human auditory steady-state responses. *Hear Res*; **101**: 62-74.

Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD, (2002) The processing of temporal pitch and melody information in auditory cortex. *Neuron*; **36**: 767-776.

Penagos H, Melcher JR, Oxenham AJ, (2004) A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *J Neurosci*; **24**: 6810-6815.

Pierce J, 1999. Introduction to pitch perception, in, P. R. Cook, ed. Music, cognition, and computerized sounds. MIT Press, Cambridge, MA, 57-70.

Prabhakaran V, Narayanan K, Zhao Z, Gabrieli JD, (2000) Integration of diverse information in working memory within the frontal lobe. *Nat Neurosci*; **3**: 85-90.

Rabinowitz PM, (2000) Noise-induced hearing loss. *Am Fam Physician*; **61**: 2749-2756, 2759-2760.

Rauschecker JP, Tian B, (2000) Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc Natl Acad Sci USA*; **97**: 11800-11806.

Rauschecker JP, Tian B, Hauser M, (1995) Processing of complex sounds in the macaque nonprimary auditory cortex. *Science*; **268**: 111-114.

Ravicz ME, Melcher JR, (2001) Isolating the auditory system from acoustic noise during functional magnetic resonance imaging: examination of noise conduction through the ear canal, head, and body. *J Acoust Soc Am*; **109**: 216-231.

Ravicz ME, Melcher JR, Kiang NY, (2000) Acoustic noise during functional magnetic resonance imaging. *J Acoust Soc Am*; **108**: 1683-1696.

Ritsma RJ, (1967) Frequencies dominant in perception of pitch of complex sounds. *J Acoust Soc Am*; **42**: 191-198.

Roberts MJ, Russo R, (1999) A student’s guide to analysis of variance, Routledge, London.Romani GL, Williamson SJ, Kaufman L, Brenner D, (1982) Characterization of the human auditory cortex by the neuromagnetic method. *Exp Brain Res*; **47**: 381-93.

Romanski LM, (2004) Domain specificity in the primate prefrontal cortex. *Cogn Affect Behav Neurosci*; **4**: 421-429.

Roy C, Sherrington CS, (1890) On the Regulation of the blood supply of the brain. *J. Physiol;* **11**: 85-108.

Ruchkin DS, Grafman J, Cameron K, Berndt RS, (2003) Working memory retention systems: a state of activated long-term memory. *Behav Brain Sci*; **26**: 709-728; discussion 728-777.

Ruusuvirta T, Huotilainen M, Fellman V, Näätänen R, (2003) The newborn human brain binds sound features together. *Neuroreport*; **14**: 2117-2119.

Ruusuvirta T, Huotilainen M, Fellman V, Näätänen R, (2004) Newborn human brain identifies repeated auditory feature conjunctions of low sequential probability. *Eur J Neurosci*; **20**: 2819-2821.

Rämä P, Poremba A, Sala JB, Yee L, Malloy M, Mishkin M, Courtney SM, (2004) Dissociable functional cortical topographies for working memory maintenance of voice identity and location. *Cereb Cortex;* **14**: 768-780. Epub 2004 Apr 2014.

Salame P, Baddeley A, (1987) Noise, unattended speech and short-term memory. *Ergonomics*; **30**: 1185-1194.

Sambeth A, Huotilainen M, Kushnerenko E, Fellman V, Pihko E, (2006) Newborns discriminate novel from harmonic sounds: a study using magnetoencephalography. *Clin Neurophysiol*; **117**: 496-503.

Sams M, Paavilainen P, Alho K, Näätänen R, (1985) Auditory frequency discrimination and event-related potentials. *Electorencephalogr Clin Neurophysiol;* **62**: 437-448.

Scarff CJ, Reynolds A, Goodyear BG, Ponton CW, Dort JC, Eggermont JJ, (2004) Simultaneous 3-T fMRI and high-density recording of human auditory evoked potentials. *Neuroimage*; **23**: 1129-1142.

Scharf B, Buus S, (1986). Audition I: Stimulus, Physiology, Thresholds, In: K. R. Boff, L. Kaufman, and J. P. Thomas, eds. Handbook of Perception and Human Performance. Wiley, New York, 14.11-14.71.

Schneider P, Sluming V, Roberts N, Scherg M, Goebel R, Specht HJ, Dosch HG, Bleeck S, Stippich C, Rupp A, (2005) Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. *Nat Neurosci;* **8**: 1241-1247.

Schönwiesner M, von Cramon DY, Rubsamen R, (2002) Is it tonotopy after all? *Neuroimage*; **17**: 1144-1161.Sek A, Moore BC, (1995) Frequency discrimination as a function of frequency, measured in several ways. *J Acoust Soc Am*; **97**: 2479-2486.

Shestakova A, Brattico E, Huotilainen M, Galunov V, Soloviev A, Sams M, Ilmoniemi RJ, Näätänen R, (2002) Abstract phoneme representations in the left temporal cortex: magnetic mismatch negativity study. *Neuroreport*; **13**: 1813-1816.

Shmuel A, Augath M, Oeltermann A, Logothetis NK, (2006) Negative functional MRI response correlates with decreases in neuronal activity in monkey visual area V1. *Nat Neurosci;* **9**: 569-577.

Shtyrov Y, Kujala T, Ilmoniemi RJ, Näätänen R, (1999) Noise affects speech-signal processing differently in the cerebral hemispheres. *Neuroreport*; **10**: 2189-2192.

Stansfeld SA, Matheson MP, (2003) Noise pollution: non-auditory effects on health. *Br Med Bull*; **68**: 243-257.

Stapells DR, Kurtzberg D, (1991) Evoked potential assessment of auditory system integrity in infants. *Clin Perinatol*; **18**: 497-518.

Talavage TM, Edmister WB, (2004) Nonlinearity of FMRI responses in human auditory cortex. *Hum Brain Mapp*; **22**: 216-228.

Talavage TM, Ledden PJ, Benson RR, Rosen BR, Melcher JR, (2000) Frequency-dependent responses exhibited by multiple regions in human auditory cortex. *Hear Res*; **150**: 225-244.

Tervaniemi M, (2001) Musical sound processing in the human brain. Evidence from electric and magnetic recordings. *Ann NY Acad Sci*; **930**: 259-272.

Tervaniemi M, Alho K, P. P, Sams M, Näätänen R, (1993) Absolute pitch and event-related brain potentials. *Music Perception*; **10**: 305-316.

Tervaniemi M, Huotilainen M, (2003) The promises of change-related brain potentials in cognitive neuroscience of music. *Ann N Y Acad Sci;* **999**: 29-39.

Tervaniemi M, Ilvonen T, Sinkkonen J, Kujala A, Alho K, Huotilainen M, Näätänen R, (2000a) Harmonic partials facilitate pitch discrimination in humans: electrophysiological and behavioral evidence. *Neurosci Lett*; **279**: 29-32.

Tervaniemi M, Medvedev SV, Alho K, Pakhomov SV, Roudas MS, van Zuijen TL, Näätänen R, (2000b) Lateralized automatic auditory processing of phonetic versus musical information: a PET study. *Hum Brain Mapp*; **10**: 74-79.

Tervaniemi M, Schröger E, Saher M, Näätänen R, (2000c) Effects of spectral complexity and sound duration on automatic complex-sound pitch processing in humans - a mismatch negativity study. *Neurosci Lett*; **290**: 66-70.

Tiitinen H, Alho K, Huotilainen M, Ilmoniemi RJ, Simola J, Näätänen R, (1993) Tonotopic auditory cortex and the magnetoencephalographic (MEG) equivalent of the mismatch negativity. *Psychophysiology*; **30**: 537-540.

Tiitinen H, May P, Reinikainen K, Näätänen R, (1994) Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature*; **372**: 90-92.

Tomasi D, Caparelli EC, Chang L, Ernst T, (2005) fMRI-acoustic noise alters brain activation during working memory tasks. *Neuroimage*; **11**: 11.

Tomasi DG, Ernst T, (2003) Echo planar imaging at 4 Tesla with minimum acoustic noise. *J Magn Reson Imaging*; **18**: 128-130.

Trehub SE, (1973) Infants' sensitivity to vowel and tonal contrast. *Dev Psychol*; **9**: 91-96.

Tremblay S, Jones DM, (1999) Change of intensity fails to produce an irrelevant sound effect: implications for the representation of unattended sound. *J Exp Psychol Hum Percept Perform*; **25**: 1005-1015.

Ulmer JL, Biswal BB, Mark LP, Mathews VP, Prost RW, Millen SJ, Garman JN, Horzewski D, (1998) Acoustic echoplanar scanner noise and pure tone hearing thresholds: the effects of sequence repetition times and acoustic noise rates. *J Comput Assist Tomogr*; **22**: 480-486.

Werner LA, Gray A, 1998. Behavioral studies of hearing development, In: E. W. Rubel, A. N. Popper, and R. R. Fay, eds. Development of the Auditory System. Springer-Verlag, New York, 12-79.

Wessinger CM, Buonocore M, Kussmaul CL, Mangun GR, (1997) Tonotopy in human auditory cortex examined with functional magnetic resonance imaging. *Human Brain Mapping*; **5**: 18-25.

Wessinger CM, VanMeter J, Tian B, Van Lare J, Pekar J, Rauschecker JP, (2001) Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *J Cogn Neurosci*; **13**: 1-7.

Wier CC, Jesteadt W, Green DM, (1977) Frequency discrimination as a function of frequency and sensation level. *J Acoust Soc Am*; **61**: 178-184.

Winkler I, Cowan N, (2005) From sensory to long-term memory: evidence from auditory memory reactivation studies. *Exp Psychol*; **52**: 3-20.

Winkler I, Kushnerenko E, Horvath J, Ceponiene R, Fellman V, Huotilainen M, Näätänen R, Sussman E, (2003) Newborn infants can organize the auditory world. *Proc Natl Acad Sci U S A*; **100**: 11812-11815.

Woods DL, (1995) The component structure of the N1 wave of the human auditory evoked potential. *Electroencephalogr Clin Neurophysiol Suppl*; **44**: 102-109.

Zatorre RJ, Belin P, (2001) Spectral and temporal processing in human auditory cortex. *Cereb Cortex*; **11**: 946-953.

Zwicker E, Fastl H, (1999) Psychoacoustics: facts and models, Springer-Verlag, Berlin-Heidelberg.