Cortical processing of musical pitch as reflected by behavioural and electrophysiological evidence

Elvira Brattico

Academic dissertation to be publicly discussed,
with the permission of the Faculty of Behavioural Sciences,
University of Helsinki, Auditorium XIV
on the 11th of November, 2006, at 10 o'clock.

Cognitive Brain Research Unit
Department of Psychology
University of Helsinki
Finland

Helsinki 2006

Supervised by

Doc. Mari Tervaniemi
Cognitive Brain Research Unit
Department of Psychology
University of Helsinki, Finland

Academy Professor Risto Näätänen
Cognitive Brain Research Unit
Department of Psychology
University of Helsinki, Finland

Reviewed by

Dr. Thomas Jacobsen
Institute of Psychology
University of Leipzig, Germany

Professor Tim D. Griffiths
Medical School
University of Newcastle upon Tyne, United Kingdoms

ISBN 952-92-1124-4 (paperback)
ISBN 952-10-3450-5 (PDF)
(http://ethesis.helsinki.fi)
Yliopistopaino
Helsinki 2006

Ai miei due tesori Emil e Sofi

CONTENTS

ACKNOWLEDGMENTS

The experimental work of this thesis was carried out at the Cognitive Brain Research Unit (CBRU) of the Department of Psychology, University of Helsinki, at the Rotman Research Institute, Baycrest Centre for Geriatric Care, Toronto, Canada, and at the Department of Psychology, University of Montreal, Canada.

I am greatly indebted to my supervisors, Dr. Mari Tervaniemi and Academy Professor Risto Näätänen, for their guidance, advice and support during all stages of my work. Without their continuous support and scientific excellence, this work would have not been possible. In particular, Mari determined my entering the field of cognitive neuroscience when she invited me at CBRU in the spring term of the year 1998 and again in the fall term of the year 1999 after a teaching experience in Italy.

Along with my supervisors, Professors Terence W. Picton and Isabelle Peretz have contributed to my scientific development. The months that I worked in Canada in 2002 and the collaboration with the exceptional colleagues there are still fresh in my mind and have been a source of inspiration ever since.

The teachers and colleagues of the Pythagoras Graduate School represented the ideal scientific environment for developing my ideas. Thanks to them all!

Many colleagues at CBRU, at the Department of Psychology and at the Helsinki Brain Research Centre (HBRC) have contributed to my research. In particular, I wish to express my gratitude to my co-authors, Dr. Petri Paavilainen, Dr. Rika Takegata and Professor Istvan Winkler for their enlightening and precious collaboration, and to the colleagues who permitted the access to lab facilities and gave generous advice on practical and scientific matters, like Professors Kimmo Alho, Synnöve Carlson and Christina Krause, Dr. Minna Huotilainen, Dr. Risto J. Ilmoniemi and Dr. Teija Kujala. I extend my appreciation to Sampo Antila, Dr. Valentina Gumenyuk, Dr. Oleg Korzyukov, Anu Kuijala, Dr. Teemu Rinne, Marieke Saher, Päivi Sivonen, Dr. Maria Uther and Dr. Titia van Zuijen, who helped me to orientate myself in the CBRU labs when I started my experimental enterprise. My thanks are owned to the following persons who made the working atmosphere so pleasant and inspiring at CBRU and at the other labs with which I have collaborated: Dr. Irina Anourova, Dr. Anne Caclin, Dr. Tom Campbell, Alexander Dagerman, Dr. Takako Fujioka, Dr. Titta Ilvonen, Eva Istok, Virpi Kalakoski, Dr. Alexandra Khromatidi, Tuulia Lepistö, Erin McMullen, Dr. Vincenzo Monitillo, Nikolai Noviski, Darja Osipova, Satu Pakarinen, K. Johanne Pallesen, Essi Rossi, Dr. Marc Schönwiesner, Miia Seppänen, Dr. Anna Shestakova, Veerle Simoens (my cheerful and understanding officemate), Dr. Olga Varyagina, Dr. Peter Vuust, Dr. Sari Ylinen. Warmest thanks also to Kaisu Krohn for being such a good friend and colleague through the years and for all the stimulating discussions on life, music and science. I am indebted to Marja Junnonaho, Miika Järvenpää, Markus Kalske, Piiu Lehmus, Teemu Peltonen, Kalevi Reinikainen in Helsinki and to the research assistants and programmers in Toronto and Montreal for valuable support work.

I am grateful to the reviewers of my thesis, Dr. Thomas Jacobsen and Professor Tim D. Griffiths, and to the language editor, Carol Pelli, for their careful and most useful comments which substantially improved the text of the thesis.

I wish to thank Professor Laurel J. Trainor for agreeing to act as my opponent at the public defense of this dissertation.

I am indebted to my Italian relatives, mamma, papá, Luisella, Giuseppe, Lucilla, Enrico, Angelo, zio Michele, zia Lucrezia, Isa, Dino, Annarita, and to all my friends, in particular, Angela, Fiorella, Grazia, Luciana, Maria, Maria Francesca, Mariangela, Roxy, Suvi, the Italians in Helsinki (Barbara, Flavia, Roxie and Zeina) and the Triggiano 'villosi' for their patience and for cheering me up in so many occasions. I extend my thanks to all my Finnish relatives, and particularly to Outi, for taking care of Emil a number of times and giving me a chance to still have some spare time from work and family duties.

A special thanks goes to my husband Pauli, for his help, encouragement and for the challenging discussions on philosophy, language, music, art and science that kept me enthusiastic about research, even during the times of children care.

This thesis was made possible by financial support from the following agencies: University of Palermo (Italy), Centre for International Mobility (CIMO, Ministry of Education, Finland), Academy of Finland (projects 77322 and 73038), Pythagoras Graduate School (Ministry of Education, Finland), Consiglio Nazionale delle Ricerche (CNR, Italy), Government of Canada Award (Canada), Jenny and Antti Wihuri Foundation (Finland), Sixth Framework Programme (NEST, European Union).

ABSTRACT

In a musical context, the pitch of sounds is encoded according to domain-general principles not confined to music or even to audition overall but common to other perceptual and cognitive processes (such as multiple pattern encoding and feature integration), and to domain-specific and culture-specific properties related to a particular musical system only (such as the pitch steps of the Western tonal system). The studies included in this thesis shed light on the processing stages during which pitch encoding occurs on the basis of both domain-general and music-specific properties, and elucidate the putative brain mechanisms underlying pitch-related music perception.

Study I showed, in subjects without formal musical education, that the pitch and timbre of multiple sounds are integrated as unified object representations in sensory memory before attentional intervention. Similarly, multiple pattern pitches are simultaneously maintained in non-musicians' sensory memory (Study II). These findings demonstrate the degree of sophistication of pitch processing at the sensory memory stage, requiring neither attention nor any special expertise of the subjects.

Furthermore, music- and culture-specific properties, such as the pitch steps of the equal-tempered musical scale, are automatically discriminated in sensory memory even by subjects without formal musical education (Studies III and IV). The cognitive processing of pitch according to culture-specific musical-scale schemata hence occurs as early as at the sensory-memory stage of pitch analysis. Exposure and cortical plasticity seem to be involved in musical pitch encoding. For instance, after only one hour of laboratory training, the neural representations of pitch in the auditory cortex are altered (Study V). However, faulty brain mechanisms for attentive processing of fine-grained pitch steps lead to inborn deficits in music perception and recognition such as those encountered in congenital amusia (Study VI). These findings suggest that predispositions for exact pitch-step discrimination together with long-term exposure to music govern the acquisition of the automatized schematic knowledge of the music of a particular culture that even non-musicians possess.

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original articles, referred to in the text by their Roman numerals:

I. Takegata, R., Brattico, E., Tervaniemi, M., Varyagina, O., Näätänen, R., & Winkler, I. (2005). Preattentive representation of feature conjunctions for concurrent spatially distributed auditory objects. *Cognitive Brain Research*, *25*, 169-179.

II. Brattico, E., Winkler, I., Näätänen, R., Paavilainen, P., & Tervaniemi, M. (2002). Simultaneous storage of two complex temporal sound patterns in auditory sensory memory. *NeuroReport*, *13*, 1747-1751.

III. Brattico, E., Näätänen, R., & Tervaniemi, M. (2001). Context effects on pitch perception in musicians and nonmusicians: evidence from event-related potential recordings. *Music Perception*, *19*, 199-222.

IV. Brattico, E., Tervaniemi, M., Näätänen, R., & Peretz, I. (2006). Musical scale properties are automatically processed in the human auditory cortex. *Brain Research*, *1117*, 162-174.

V. Brattico, E., Tervaniemi, M., & Picton, T.W. (2003). Effects of brief discrimination-training on the auditory N1 wave. *NeuroReport*, *14*, 2489-2492.

VI. Peretz, I., Brattico, E., & Tervaniemi, M. (2005). Abnormal electrical brain responses to pitch in congenital amusia. *Annals of Neurology*, *58*, 478-482.

These publications have been reprinted with the kind permission of their copyright holders. In addition, some unpublished material is presented.

ABBREVIATIONS

ABR	auditory brainstem response
BA	Brodmann's area
CF	characteristic frequency
ΔF	frequency difference
EEG	electroencephalogram
ERAN	early right anterior negativity
ERP	event-related potential
F_0	fundamental frequency
fMRI	functional magnetic resonance imaging
Hz	Hertz
HG	Heschl's gyrus
HR	hit rate
IRN	iterated rippled noise
LPC	late positive component
LSD	least-significant-difference
MEG	magnetoencephalogram
MLR	middle latency response
MMN	mismatch negativity
MMNm	magnetic mismatch negativity
N1m	magnetic N1
PAC	primary auditory cortex
PET	positron emission tomography
PP	planum polare
PT	planum temporale
RT	reaction time
SE	standard error of the mean
STG	superior temporal gyrus

1 INTRODUCTION: PITCH PROCESSING IN THE AUDITORY SYSTEM

Pitch is a fundamental aspect of music perception. It may be defined as ‘the attribute of auditory sensation in terms of which sounds may be ordered on a musical scale’ (American Standards Association, 1960). By varying the pitch of sounds, it is possible to generate a melody. While listening to a melody, expectations for incoming pitches are formed continuously and unconsciously by the listener according to either universal domain-general perceptual laws not confined to music or even to audition overall, or to domain-specific and culture-specific knowledge based on previous experience of the musical system of a particular culture. Therefore, while the perception of isolated pitch can be accounted for by sensory laws (studied by psychoacoustics and physiological acoustics), the perception of pitch in a musical context (here referred to as musical pitch) is the outcome of a dynamic process affected retroactively by sound events preceding it and proactively by events following it (Krumhansl, 1990). The study of musical pitch, thus, deals also with learning, long-term memory and structured knowledge of the musical system of a particular culture. Musical pitch is assumed to be processed with the help of culturally acquired schemata, organized long-term memory structures that affect pitch information processing flow (Krumhansl, 1990; Leman, 1995; 1999).

The complementary roles of perceptual constraints and acquired music knowledge on ongoing listening to music are, however, only partially accounted for in the commonly held frameworks of pitch analysis. For instance, recent models (Griffiths, 2003a; 2003b; Griffiths et al., 2004; Koelsch & Siebel, 2005), based on brain imaging evidence, propose that only the initial stage of analysis is quickly performed by the auditory cortex, leaving to other brain structures and to subsequent processing stages the modulation of pitch according to previous experience of sounds and the attribution of meaning to incoming sound signals. A revised framework for pitch analysis is hence warranted to determine at which processing stage and how the brain attributes musical schemata to the pitch of incoming sounds. In other words, it has to be clarified with brain imaging methods, how pitches in a musical context are mentally organized according to tonal structures (cf. Leman, 1995; 1999).

This thesis aims to determine whether sophisticated aspects of pitch, such as its integration into multiple auditory objects and its cognitive processing according to culturally acquired music schemata, can occur as early as at the sensory memory stage, before attentional processes come into play. By addressing this issue, the interrelation between the congenital predispositions for the neural skills of pitch analysis and short- and long-term neuroplastic modifications following sound exposure is also covered.

First, a synopsis of the known stages of pitch processing, from the early subcortical analysis to the cognitive processes taking place in associative cerebral cortices, is presented. Section 2 outlines the open research questions related to musical pitch processing in the human brain, and Section 3 presents general and specific study objectives. Section 4 summarizes and discusses relevant findings from Studies I-VI, providing the empirical foundation for this thesis. Finally, in Sections 5 and 6, the contribution of these findings to the revised view of the sensory-memory stage in musical pitch processing as well as related general theoretical considerations are discussed.

1.1 Early analysis of pitch-related features in the auditory pathway

When a sound reaches the eardrum, it triggers a complex cascade of mechanical, chemical and neural events that eventually result in an auditory object, the integrated representation of sound corresponding to its percept by an individual (Griffiths & Warren, 2004; Zatorre, 2005). The formation of an auditory object entails several stages of processing, from transduction of the elementary component features of a sound from the periphery to the centre of the auditory system, to feature integration and the abstraction of the sound-invariant properties, and further to cross-modal generalization and the attribution of meaning (cf. Griffiths & Warren, 2004).

A sound, i.e., pressure waves repeating over time, is characterized by several acoustic features such as duration, intensity, location and frequency. These acoustic features may or may not linearly correspond to perceptual dimensions; the perceptual response to a stimulus can or cannot be predicted from the responses to isolated stimulus components (Moore, 2004). A crucial aspect of sound perception, necessary for the identification and segregation of auditory objects in speech and music, is pitch (e.g., Bendor & Wang, 2005). The pitch of a periodic sound wave is related to its frequency, i.e., to the number of cycles of the wave per second (measured in Hertz, Hz)[1]. However, perceptually, it linearly corresponds to the sound frequency only in the case of sinusoidal tones, formed by a single sine wave. For this reason, the terms pitch and frequency will be used synonymously in the following specifically when referring to a sine tone.

On the other hand, in nature, many sounds have a complex periodic waveform. This means that the acoustic sound spectrum can be decomposed via a Fourier transform into a sinusoid at the fundamental frequency of periodicity (f_0),

---

[1] A pitch percept can be also produced in the iterated rippled noise (IRN) by delaying a copy of a random noise and adding it back to the original. In this case, the frequency of the resulting pitch corresponds to the inverse of the delay time. The pitch can be strengthened by repeating the delay-and-add process (Griffiths et al., 1998).

which is the lowest sine wave that can be extracted, and into other sinusoids at frequencies that are integer multiples of f_0. Those sinusoids are called partials or harmonics (Moore, 2004). In the case of complex periodic sounds, perceptually, pitch corresponds only quasi-linearly to f_0. Sounds having different spectra can still be perceived as having the same pitch if they share the same f_0. The same pitch is perceived even when the acoustic energy at f_0 is removed, giving rise to the phenomenon of the missing fundamental (Bendor & Wang, 2005). This pitch percept results from the estimation of the frequencies of the missing fundamental components (especially when they consist of the lower-order harmonics; Goldstein, 1973).

In the ascending auditory pathway, the lowest partials of a periodic complex sound are individually analysed (up to the first eight; see Moore, 2004) by the neuronal populations of the basilar membrane of the cochlea in the inner ear. The basilar membrane is then connected to the auditory VIII nerve through inner hair cells (Pickles, 1988). However, the resulting percept is that of a single sound with a pitch roughly corresponding to the frequency of a sine tone at f_0 (Moore, 2004). The extraction of pitch from a sinusoidal or a complex sound seems to depend on the neural representation of the sound according to its spatial (tonotopic) and/or temporal (periodicity) encoding, taking place as early as at the first stages of sound analysis. Spatial encoding is likely to be responsible for the pitch of sine tones at low frequencies and temporal encoding for the pitch of sine tones at high frequencies.

For the pitch of complex tones, the pattern recognition models suggest that pitch is centrally derived according to the neural signals tonotopically distributed across the different auditory neurons and corresponding to the individual partials of the sound. The temporal theories, in contrast, claim that the pitch of a complex tone is determined by computing the time intervals between successive nerve firings. Those mechanisms work best with complex sounds having a low f_0 due to limitations in the timing of neuronal firings, and are supposed to be crucial for music perception since they are most effective in the conventional range of a musical instrument (e.g., the piano; Trainor, 2005). The temporal mechanisms also explain why the iterated rippled noise (IRN), a noise composed of repetitions of the same noise portion, may produce a pitch percept in the absence of spectral cues (Griffiths et al., 1998). A model (Moore, 2004) combining both spatial and temporal mechanisms proposes the initial encoding of a complex sound by a filter bank resolving its individual partials, followed by the transduction of the filter outputs into activity in the neurons with the corresponding characteristic frequencies (CFs), and thereafter, by the analysis of time intervals between successive firings and the computation of the common time intervals.

The second stage of sound encoding takes place in the nuclei of the brainstem (reached by the afferent fibres of the auditory VIII nerve), where further analysis of the elementary features of the acoustic signal takes place. In particular, pitch and the other features in which the spectral aspects are important, such as roughness and, to a certain degree, also timbre, evoke characteristic neural response patterns in the superior olivary complex, in the inferior colliculus and in the medial geniculate body. On the basis of the analysis occurring in the above-mentioned nuclei of the brainstem and the thalamus, sounds signalling danger can be very quickly identified, i.e., before reaching the cortex (when this happens, the amygdala and the other structures of the limbic system are activated; LeDoux, 2000). The thalamus sends fibres up to the granular layers of the primary auditory cortex (PAC).

Based on architectonic mapping measurements on awake and anaesthetized macaque monkeys, the PAC has been identified with the 'core' field (receiving inputs from the medial geniculate body of the thalamus), located on the lower bank of the Sylvian sulcus, surrounded by the non-primary belt fields and laterally by the parabelt fields (Kaas & Hackett, 1998, 2000; Hall et al., 2003). The core field is, in turn, subdivided into three primary-like regions: a primary area (or A1), a rostral area and a rostrotemporal area, characterized by neurons having a similar CF and arranged in rows according to frequency tonotopic gradients (Hall et al., 2003). While neurons of the core area respond well to pure tones with a narrow frequency tuning at their CF, the neurons of the belt areas respond less well to pure tones and with a broader frequency tuning, thus also allowing the processing of stimuli with higher spectral complexity. In contrast to the other two regions, the parabelt area is characterized by projections outside the auditory cortex, to other sites of the temporal lobe, the frontal and prefrontal cortex, and to the temporoparietal junction. Even if these PAC areas exhibit a relatively narrow frequency tuning and are tonotopically organized, their responses are less specific and more labile than those of the lower nuclei of the ascending auditory pathway. According to Griffiths et al. (2004), this might underlie the sound-encoding properties of animal and human PAC neurons; rather than a simple physical representation, as in the visual cortex, these neurons may offer an initial stimulus representation of complex features, which needs integration of the temporal and spectral context in the brief timescale (for similar ideas based on research in cats, cf. Nelken et al., 2003).

In humans, the PAC corresponding to the monkey A1 region is located over approximately two-thirds of the anterior transverse temporal Heschl's gyrus (HG) lying deep in the lateral Sylvian fissure (see Fig. 1). The PAC corresponds to BA41/BA42 (according to the citoarchitectonic parcellation of the human cortex by Brodmann, 1909) and is the terminal station for the nerve fibres originating from the medial geniculate body of the thalamus (Kandel et al., 2000; Kahle & Frotscher, 2003). The initial processing of simple spectrotemporal features performed in the structures up to the PAC is essential to permit the processing of the higher-order abstract features occurring in non-primary association cortices. The non-primary auditory area has been subdivided into planum polare (or PP, anterior to the HG) and planum temporale (or PT, posterior to the HG; Hall et al., 2003). These areas are located in the superior temporal gyrus (STG), corresponding to the Brodmann's region BA22. The electrical stimulation of BA22 induces some complex acoustic sensations such as humming, buzzing and ringing (Penfield & Perot, 1963; Kahle & Frotscher, 2003).

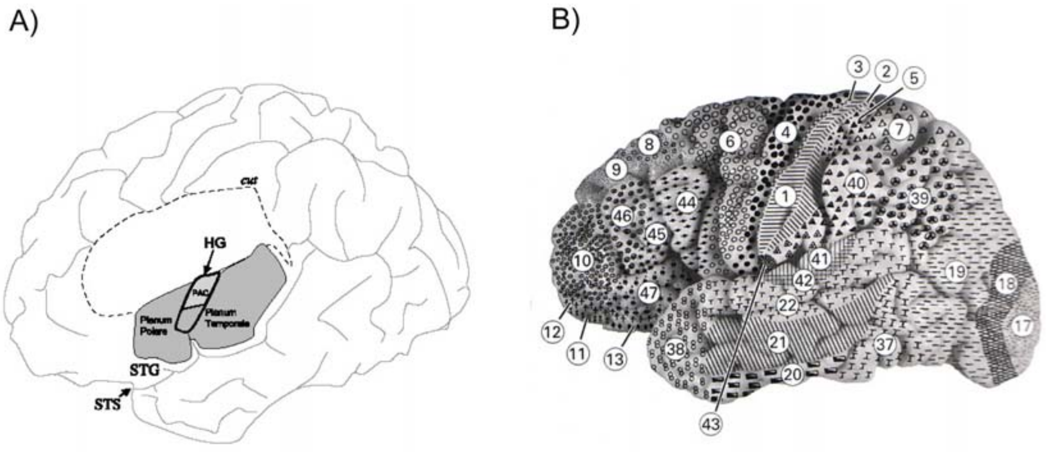

Figure 1. **A)** Dorsolateral view of the human brain (after removal of the overlying parietal cortex; the dashed line depicts the portion of the cortex that has been removed) showing the main brain structures responsible for sound processing: the primary auditory cortex (PAC) lies in the medial part of Heschl's gyrus (HG) and is surrounded by the non-primary auditory structures, i.e., anteriorly by the planum polare (PP) and posteriorly by the planum temporale (PT); these structures are located in the superior temporal gyrus (STG) and bounded by the superior temporal sulcus (STC). Reproduced with permission of the publisher (Karger) from Hall et al. (2003). **B)** A lateral view of the human cortex reporting Brodmann's citoarchitectonical classification (Brodmann, 1909). The primary auditory areas are indicated with numbers 41 and 42, and the non-primary area with number 22.

Pitch processing is not a simple phenomenon. As mentioned previously, the empirical proof that pitch does not linearly correspond to the f0 is provided by the missing fundamental sound. Its harmonics are not resolved by the auditory system; hence, only the temporal information of the sound wave provides cues to determine the pitch. In greater detail, in the inner ear, the representation of the pitch of a complex sound with a missing fundamental consists of a spatial code recording the distribution of the neural discharges over the basilar membrane as well as a temporal code involving the discharge rates and the firing patterns of the auditory nerve fibres (Cariani & Delgutte, 1996; Cedolin & Delgutte, 2005). In the cortex, the spectral and temporal coding strategies are combined, providing a neural correlate for pitch perception. In particular, the neural representation of a missing fundamental pitch is supposed to be computed not in the PAC (as its neurons do not respond to missing fundamentals with a pitch matching their CFs; Fishman et al., 1998) but in a region anterolateral to the low-frequency border of the PAC, as shown by both single-unit intracellular recordings in marmoset primates (Bendor & Wang, 2005) and functional magnetic imaging (fMRI), positron emission tomography (PET) and neuropsychological data in humans (Zatorre, 1988; Patterson et al., 2002; Penagos et al., 2004). In other words, the neurons of the anterolateral region bordering the PAC seem to be responsible for the perception of the same pitch in spite of acoustic variation, and thus, for perceptual constancy in the auditory domain (Zatorre, 2005). The lateral part of the HG (also termed secondary auditory cortex) has hence been qualified as a ‘pitch centre’ since stimulus representation there correlates with pitch percept rather than with simple mapping of frequency features (Griffiths, 2003a).

1.1.1 Early processing stages as reflected by ERPs

This thesis is based on event-related potential (ERP) data. As compared with other brain imaging techniques, such as fMRI and PET, which measure slow metabolic changes in oxygen and other molecule consumption as a consequence of neural activation, the ERPs allow one to study the fast temporal relation between the presentation of a stimulus and its corresponding cortical response (Luck, 2005). The ERPs thus reflect with millisecond accuracy the different stages of neural sound processing, from the encoding of physical stimulus features to discrimination, memorization, retrieval and recognition of cognitive aspects of sounds (Picton, 1988; Näätänen, 1992; Rugg & Coles, 1995). The data derived from the ERP method, lacking accurate spatial resolution but refined in latency information, can be combined with available anatomical and functional knowledge to infer the localization and processing stage of the recorded neural activity.

This electrical activity time-locked to a stimulus was originally termed evoked potential or EP, and later named event-related potential or ERP by Vaughan (1969) to indicate an electric potential having a stable relationship with a definable event, either of sensory-motor or psychological nature (e.g., Luck, 2005). The ERP is derived from electroencephalogram (EEG), which measures the electric fields generated by synchronized neural activity. By averaging the EEG signals time-locked to repeated stimuli, the ERP waveform is obtained, consisting of a sequence of negative and positive voltage deflections called peaks, waves or components (Luck, 2005).

ERP components to a sequence of sine tones are typically measured as the voltage potential between a reference electrode and the electrode placed at the scalp vertex as early as within the first 10 ms from the onset of sound stimulus. The earliest ERPs originate at the initial stages of the auditory pathway, reflecting its neural conductivity from the cochlear nerve to the brainstem, and thus are termed auditory brainstem responses (ABRs). The ABRs are labelled after their temporal order (I–VII) (Pratt et al., 1999). The afferent activation pattern separately transducing each feature of the incoming sounds (Näätänen & Winkler, 1999), mirrored in the ABRs, reaches the PAC at around 10 ms from stimulus onset, as indicated by the small negative wave N0. This wave is followed by other middle-latency responses (MLRs), also labelled after their polarity and temporal order: a positive component at 15 ms (P0 or P15), a negative component at 19 ms (Na or N19), a second positive component at 30 ms (Pa), all originating from the PAC, followed by later component, such as the Pb or P50, reflecting neural activity in areas lateral to the PAC (Liegeois-Chauvel et al., 1994). Both the ABRs and MLRs index the subcortical and cortical stages of feature analysis, with no feature trace being formed, since the neuronal circuits activated by a specific stimulus very quickly revert to their pre-stimulus excitability and hence do not permit persistence in time of the same feature (Näätänen & Winkler, 1999). In other words, the neurons responsible for ABR and MLR elicitation have very short refractory periods. This period depends on the recurrent initiation of neuronal firing, which results in diminished excitability as a function of the repetition rate of a stimulus (Näätänen & Picton, 1987; for description of neural refractoriness following an action potential, see Kandel et al., 2000).

The late ERP components reflect perceptual and cognitive processing of the stimulus, involving, for instance, memory and attention (Näätänen, 1992). The nomenclature of the components is either descriptive (such as the N100 or P300), with the letters ‘P’ and ‘N’ indicating positive or negative polarity of the component, and the number the latency of the peak, or functional (such as error-related negativity, contingent negative variation), related to the mental event associated with the component (Johnson & Baron, 1995).

The N1 (or N100) component of the ERP is the largest obligatory cortical response to incoming sounds[2]. The N1 does not represent a unitary process but is rather the outcome of several neuronal generators acting quasi-simultaneously (Näätänen & Picton, 1987). The main subcomponent of the N1 originating from the auditory cortical source over the supratemporal lobe is recorded at the vertex of the scalp at a latency of around 100 ms. Other N1 sources have been located in the lateral temporal and frontal lobes (Näätänen & Picton, 1987). The supratemporal N1 has a relatively long refractory period, lasting about 10 s, and is characterized by stimulus-feature specificity (even if blurred compared with the

---

[2] The N1 is often followed by the P2 wave, also reflecting obligatory processing of sounds (see, for instance, Shahin et al., 2003).

specificity observed at previous stages of the ascending auditory pathway; Picton et al., 1978; Butler, 1968; Näätänen & Picton, 1987; Näätänen et al., 1988). The N1 is considered to reflect the temporary buffering of the neural representations of sound features, i.e., feature traces, in the central auditory system (Näätänen & Winkler, 1999). It has recently been argued that an incoming stimulus deviating from the repeated one in its frequency feature may be discriminated simply on the basis of the refractoriness state of its afferent neurons (May et al., 1999; Ulanovsky et al., 2003; Jääskeläinen et al., 2004). For instance, single-cell cortical recordings with cats (Ulanovsky et al., 2003) showed a stimulus-specific adaptation of the neurons of the A1; they responded more strongly to sounds when they were rarely presented than when they were frequently repeated. This could provide a bottom-up refractoriness-based mechanism of the initial encoding of pitch invariance in the auditory environment. However, as reviewed in the following, a large body of evidence indicates the existence of brain mechanisms, presumably located in the non-primary auditory cortex associated with the memory comparison between encoded feature traces of the repeated sounds and the incoming sounds or sound successions (Näätänen et al., 2005).

In sum, the supratemporal N1 is considered to be a feature-specific response, reflecting, among other things, the tonotopic representation of the sound frequency in the human auditory cortex (Pantev et al., 1989b, 1995; Tiitinen et al., 1993; for N1 specific to the amplitude feature, see also Pantev et al., 1989a; for N1 to the spatial locus of sound origin, see Elberling et al., 1982; and for feature-specific N1 reflecting the orderly representation of phonemes, see Diesch & Luce, 1997; Shestakova et al., 2004). In magnetoencephalogram (MEG) recordings, the frequency specificity of the N1 is shown by the deeper and more posterior N1m dipole locus to high sound frequencies as compared with low ones, indicating that at least partially separate neuronal populations respond to tones of different frequencies (Romani et al., 1982; Pantev et al., 1988; Tiitinen et al., 1993; for contradictory findings, see Lütkenhöner et al., 2003). Furthermore, the N1m component in missing fundamental tones shows a similar tonotopic organization to that of the N1m to the sine tones corresponding to their f_0 (Pantev et al., 1989b), indicating that the neural representation of the pitch of isolated tones takes place in the regions responsible for N1 generation.

In the EEG recordings, the frequency specificity of the supratemporal N1 is proven by the properties of its refractoriness; the N1 amplitude is increased with larger frequency differences between test tones and decreased with repetition of the same tone frequency, thus demonstrating the tonotopic organization of the generating neuronal populations (Butler, 1968; Picton et al., 1978; Näätänen et al., 1988; Näätänen & Winkler, 1999). Contrasting results have also been obtained, however, indicating that the N1 is a non-specific on-off response to sounds, with large receptive fields, i.e., with a similar morphology and origin of responses to many different sounds (Hari et al., 1987; Lütkenhöner et al., 2006). Consequently, the supratemporal N1 can be considered a complex electrophysiological response indexing brain mechanisms for both transient detection of sounds and feature-trace encoding (Näätänen & Picton, 1987).

1.2 Auditory sensory-memory stage

Auditory objects have a finite extension in time, thus requiring memory processes to operate in order to generate a coherent and unified neural representation of them (Griffiths & Warren, 2004). The first stage in which spectrotemporal features are integrated in the early auditory cortical areas to form a unitary auditory object is sensory memory (Näätänen & Winkler, 1999; Griffiths & Warren, 2004). The 'auditory stimulus representation' in sensory memory is the full outcome of stimulus analysis, it is the neural counterpart of perception, and it can be accessed by top-down operations due to the cortico-cortical connections between associative and sensory areas (Näätänen & Winkler, 1999). In psychological terms, at the sensory-memory stage, internalized schematic models of the auditory world are applied to the incoming sound signals (Griffiths & Warren, 2004). Such schema-based processing (Bregman, 1990) operates through template models of the invariances of the environment, maintained as sensory-memory traces for brief periods of time (Näätänen et al., 2001; Schröger et al., 2004). At a neurophysiological level of analysis, these sensory-memory traces are neural assemblies supposedly distributed in a modality-specific fashion throughout the cortex (Ruchkin et al., 2003; Pülvermüller, 2003). According to recent psychological theories, sensory-memory traces correspond to temporarily activated long-term traces (Cowan, 1995). The number of long-term traces which can be simultaneously activated for short periods of time depends on the finite capacity of voluntary attention (Cowan, 1984, 1995; Crowder, 1993; Ruchkin et al., 2003; Kaernbach, 2004; Schröger et al., 2004).

The main functions of sensory memory are supposed to be the integration of information within brief periods of time and its perseveration to allow further processing (Schröger et al., 2004). According to these functions, two processes have been identified in sensory memory (Cowan, 1984): an initial short sensory-memory phase of about 200-300 ms, and a long sensory-memory phase lasting for about 10 s, in charge of the persistence of information. Auditory sensory memory is considered (at least partially) automatic since the relative encoding, transformation and storage of information occurs independently of the conscious intentions of the subject to perceive that information (Schröger et al., 2004; for a review on automaticity, see Paavilainen, 1996).

At the sensory-memory stage, relations between pitches (or relational pitch) are analysed and stored. Anatomically, the processing of pitch changes in tone patterns, which relies on the short-term temporal persistence of tone information, is supposed to depend on the distributed networks of the superior temporal lobe (in particular, the STG) rather than on the neuronal populations of the HG representing the pitch of individual sounds (Patterson et al., 2002). According to Griffiths & Warren (2002), the auditory non-primary cortex in the superior temporal lobe, and in particular the PT, is responsible for analysing the higher-order features of pitch patterns and for their subsequent segregation into auditory objects by matching the incoming patterns with learned spectrotemporal templates. In general, the non-primary auditory cortex (particularly, area BA 22) is activated by hearing and understanding words, listening to musically structured tone patterns, imagining sounds and accessing melodic representations (for review of fMRI and PET literature, see Besson & Schön, 2003). In the following, related electrophysiological evidence will be reviewed.

1.2.1 Mismatch negativity

The neural basis of sensory memory can be probed by the mismatch negativity (MMN; see Fig. 2) component of the ERP. The MMN is obtained from recording electrical, magnetic or metabolic brain activity by presenting to subjects an infrequent auditory stimulus (“deviant“), differing from the frequently occurring stimulus (“standard“) in one or several physical and abstract parameters (Näätänen et al., 1978; Näätänen & Winkler, 1999). In electric recordings, the MMN corresponds to an enlarged frontocentral negativity (accompanied by a reversed potential at the temporomastoidal regions) to the deviant stimulus as compared with the standard one peaking between 100 and 250 ms from change onset, according to the type of stimulus. In contrast to the N1, the MMN to sound feature changes does not reflect the refractoriness-dependent activation of new afferent neurons responding to the deviant stimulus, but with decreased excitability to the standard stimulus (Jacobsen & Schröger, 2001; Näätänen et al., 2005). Instead, the MMN depends on the presence of a neural memory trace of a brief duration accurately representing each simple and complex feature of the standard stimulus, including its temporal aspects, and on a neurally distinct mechanism comparing the incoming stimulus with the stored trace (Schröger, 1997).

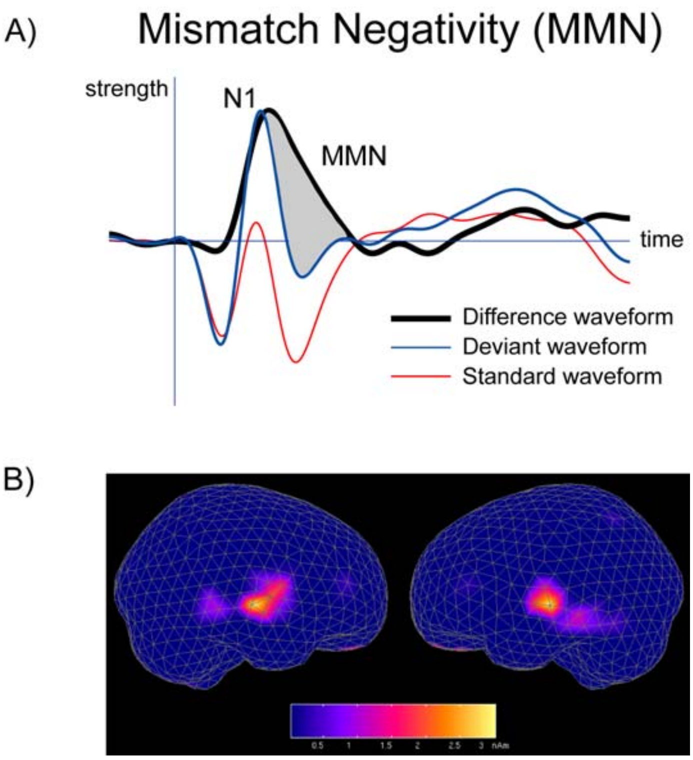

Figure 2. Schematic illustration of the mismatch negativity (MMN). **A)** ERP waveforms evoked at a frontal scalp location by the standard and deviant sounds superimposed on the difference waveform in which the ERP to the standard has been subtracted from that to the deviant. The MMN appears as an enlarged negativity to the deviant sound as compared with the standard sound, following the N1 peak. **B)** Illustration of the supratemporal bilateral sources in the left and right hemispheres of the MMN response to the deviant sound. Modified from Tervaniemi & Brattico (2004).

In psychological terms, the MMN is thus elicited when an invariance encoded into a sensory-memory trace is established, after which a mismatch process can be triggered by the deviant sound. The sensory-memory trace acts as a neuronal model or template to continuously track the auditory environment and promptly respond to an occasional change in it (Winkler et al., 1996). The auditory invariance encoded in the sensory-memory trace can be simple (e.g., a single sound feature), complex (e.g., a repeated pattern) or hypercomplex (e.g., a pattern transposed over several pitch levels; for review, see Picton et al., 2000a). An MMN can also be elicited when the feature processing of the deviant sound does not occur, such as when the deviant event consists of an occasional omission of one of the repeated tone pips separated by a very short regular sound-onset asynchrony (SOA; Yabe et al., 1997, 1998) or an the omission of the second tone of a fastly paced tone pair (Tervaniemi et al., 1994; for review, see Näätänen et al., 2005). The dependence of the MMN on a sensory-memory trace for its elicitation is confirmed by the finding that no MMN can be elicited when sounds are separated by more than 10 s (Sams et al., 1993), which corresponds to the estimated duration of the long phase of auditory sensory memory (Cowan, 1984). However, MMN studies on backward masking (e.g., Winkler & Näätänen, 1992) show that it also probes the short sensory-memory phase. Moreover, the MMN is elicited only by deviant stimuli when they are presented with the intervening standards (Cowan et al., 1993; Korzyukov et al., 1999), and the MMN amplitude is progressively increased with the number of standard stimulus repetitions (Sams et al., 1983; Javitt et al., 1998), thus indirectly probing the strength of the underlying sensory-memory trace.

An MMN can be obtained in response to exact pitch changes (for reviews, see Näätänen et al., 2003; Tervaniemi & Brattico, 2004), changes of exact pitch relations (absolute musical intervals or scale-steps; Brattico et al., 2000), deviations from the pitch contour of short tone patterns (changes in pitch direction regardless of the pitch distance and level between successive tones; Tervaniemi et al., 2001; Trainor et al., 2002) and changes in pitch relations between tones irrespective of the exact pitch value (also termed relative musical intervals or relational pitch; Paavilainen et al., 1999; Trainor et al., 2002). The neural mechanism underlying MMN generation to pitch is supposedly an increase in tonic inhibition of the supragranular neurons of the auditory cortex reacting to the frequencies of the repeated standard stimulus, associated with a transient selective release from inhibition of those neurons with respect to all other frequencies except the repeated one. Consequently, a stimulus with a different pitch would activate this less inhibited neuronal population and lead to the enhanced cortical response visible in the MMN waveform (Näätänen, 1984; 1990; Javitt et al., 1996; Haenschel et al., 2005; Näätänen et al., 2005). According to animal studies (Pincze et al., 2001; 2002), this process occurs in the rostroventral part of the non-primary auditory cortex, distinctly from the N1 neural generators. Human PET studies (Tervaniemi et al., 2000; Müller et al., 2002) and intracranial recordings from the human auditory cortex (Kropotov et al., 1995; Halgren et al., 1995) confirmed that the main MMN generator to pitch changes is located in the bilateral supratemporal plane, slightly more anterior (by about 1 cm) to the N1 generator (Korzyukov et al., 1999; Rosburg et al., 2005). While area BA41 tonotopically registers the frequency features of the incoming sounds and area BA42 responds according to the stimulus repetition rate, thus possibly reflecting the formation of sensory-memory traces (cf. also Haenschel et al., 2005), area BA22 seems to be responsible for the change-detection mechanisms comparing the incoming stimulus with the sensory-memory traces of the preceding sounds (Kropotov et al., 2000). Area BA22 comprises the PT (in the left hemisphere partly coinciding with Wernicke's language area), where the analysis and integration of spectrally and temporally complex sound features are supposed to take place (Griffiths & Warren, 2002). An additional MMN generator seems to be located in the frontal cortex (the inferior frontal gyrus and/or cingulum), possibly bilateral but with a right-hemispheric predominance (Giard et al., 1990; Opitz et al., 2002; Schall et al., 2003; Rosburg et al., 2005). The activation of the frontal generator has been associated with an involuntary attention switch in response to a sound change (Näätänen & Michie, 1979; Giard et al., 1990; Näätänen, 1992;

Rinne et al., 2000). Alternative views propose that the frontal MMN generator may serve to enhance the contrast between stimuli that are difficult to discern (Opitz et al., 2002; Restuccia et al., 2005) or that are more demanding in terms of memory access than isolated tones (Korzyukov et al., 2003).

The MMN and the sensory-memory traces and change detection associated with its elicitation are supposed to index an automatic, modular process, independent and irrespective of the top-down control of the subject (Tervaniemi & Brattico, 2004). The MMN is best recorded when subjects concentrate on a primary task, such as reading material of their own choice, watching a movie, playing a computer game or performing a behavioural test, including visual or auditory stimulation. Voluntary attention to sounds does not seem to modulate the MMN amplitude or latency. This is particularly true for the MMN to large frequency changes (Näätänen et al., 1993a; Paavilainen et al., 1993). In contrast, discrepant evidence has been found for intensity changes, suggesting that the related feature encoding, memory-trace formation and mismatch-registration processes can be suppressed or gated if attention is directed elsewhere (Woldorff & Hillyard, 1991; Woldorff et al., 1991, 1993, 1998). However, the MMN to frequency deviants but not to intensity deviants delivered at the unattended ear was attenuated in a situation in which subjects attentively detected the frequency deviants simultaneously delivered at the other ear, suggesting that voluntary attention in and of itself does not modulate the MMN (Sussman et al., 2003). The independence of the MMN process for pitch features from voluntary attention is further testified by its elicitation in several experimental conditions in which other ERP components associated with voluntary attention are not elicited. Most importantly, the MMN is elicited in REM sleep (Atienza et al., 2001), in sleeping infants (for a review, see Cheour et al., 2000) and coma patients (but only in those who later return to consciousness; Kane et al., 1996). Moreover, the MMN to infrequent frequency-location combinations is not affected by the increasing load of the primary auditory task (Winkler et al., 2005). Altogether, these findings favour the view of the MMN to frequency change as an index of a strongly automatic change-detection system (Näätänen, 1990) rather than of a weakly automatic one, i.e., obligatory but subject to attentional modulation (Schröger, 1997; on automaticity, see Paavilainen, 1996). However, it is important to note that explicit knowledge of the stimulus internal structure, acquired during the experimental session by attentive listening to the sounds according to the instructions of the experimenter, may lead to re-organization of the stimulus-driven representations of the standard stimuli, and thus, to modulation of the MMN response (Sussman et al., 2002).

The sensory-memory traces underlying MMN elicitation are briefly maintained by the central auditory system to enable further processing and decision-making, as testified by the close correspondence between MMN parameters and psycho-acoustical performance. For instance, faster reaction times (RTs) and more accurate hit rates (HRs) in discriminating changes in frequency (Tiitinen et al., 1994), duration (Amenedo & Escera, 2000; Jaramillo et al., 2000) or timbre (Toiviainen et al., 1998) as well as the superior performance of musically talented subjects relative to non-musical ones (Tervaniemi et al., 1997; Seppänen et al., in press) are reflected as an enhanced MMN amplitude and/or a shorter MMN latency. Interestingly, the MMN occurs as late as at 250-280 ms when the stimulus is of higher complexity, as for temporally complex tone patterns (Näätänen et al., 1993b; Tervaniemi et al., 2001).

The MMN also proves the association between sensory-memory encoding of sound features and the matching of these features with long-term memory recognition patterns (Cowan, 1995; Näätänen, 2001; Shestakova, 2004). In the original MMN theory, the full analysis of a stimulus was made only after this stimulus was detected as being of potential behavioural importance and when it triggered an attention switch (Näätänen & Michie, 1979; Giard et al., 1990; Näätänen, 1990; Tiitinen et al., 1994; Rinne et al., 2000). However, the finding by Näätänen et al. (1997) of an enhanced MMN to changes of native language phonemes as compared with those of non-native language phonemes showed that long-term recognition patterns for phonemes are activated at the sensory-memory level outside the attentional focus. In further support, prototype phonemes are encoded faster, with fewer repetitions, than non-prototype phonemes, indicating a facilitation of short-term memory formation by long-term memory traces (Huotilainen et al., 2001; cf. also Shestakova et al., 2002). The feature-analysis and deviance-detection systems used by sensory memory to encode and discriminate the acoustic sound features even when they fall outside the focus of attention, hence, seems to respond faster and more efficiently when those features have entered the system many times because of repeated exposure, leading to strengthened synaptic connections (Shestakova, 2004).

In the music domain, the interplay between long-term memory knowledge and the formed sensory-memory traces is indexed by MMN parameters when contrasting subjects widely familiar with sounds due to their musical expertise with inexpert subjects (for reviews, see Schröger et al., 2004; Brattico & Tervaniemi, 2006). For instance, an MMN was pre-attentively elicited by infrequent slightly mistuned chords (i.e., simultaneously presented tones, having 0.75% change in pitch in the middle tone of a major triad chord, used as a standard) only in professional violinists, not in non-musicians (Koelsch et al., 1999; however, for evidence showing superior pitch discrimination of musicians only under attentional listening, see Tervaniemi et al., 2005). On the other hand, experiments measuring group differences in the processing of temporally or spectrally simple features of sounds in sensory memory tell more about the listeners' auditory abilities than their music-specific capacities (Bigand, 2003; Jackendoff & Lerdahl, 2006). In this respect, untrained listeners may even possess neural skills, e.g. pitch processing according to musical properties, as sophisticated as those of musicians (Bigand, 2003; Bigand & Poulin-Charronnat, 2006). The following section illustrates the subsequent stages of pitch processing, involving high-order cognitive and decision-making operations on the sounds, aided by attentional resources and, further, how these stages are characterized in subjects of varying musical expertise.

1.3 Later stages requiring attention

According to psychological models (Neisser, 1967; Treisman & Gelade, 1980), higher-level stages of information processing permitting cognitive operations on sounds follow early and sensory-memory stages, characterized by an initial non-intelligent analysis. During these later stages, voluntary attention (to be distinguished from involuntary attention, initiated by stimuli in the environment) plays a crucial role. Voluntary attention may be considered a process controlled by the subject, subsequent to the initial feature analysis of the sensory input. In neurophysiological terms, additional neural resources are allocated to selective stimuli of the environment, allowing only relevant information to be further processed by the cerebral cortex (Broadbent, 1958; Näätänen, 1992; Cowan, 1995; Coull, 1998). A proposed neural circuit responsible for voluntary attention includes the parietal lobe, which disengages attention from its present focus, the midbrain, which moves attention to the new target area, and the pulvinar, which restricts the input to the indexed area (Posner, 1995). Subsequently, the neurophysiological signal is amplified in the selected sensory channel as compared with the unselected channel (Posner, 1995; for fMRI evidence concerning the auditory system, see Jäncke et al., 1999; Petkov et al., 2004).

The intervention of voluntary attention allows the sophisticated analysis of pitch according to a particular musical system (e.g., Western tonal music; see the following sections). Part of this analysis can be regarded as being of syntactic nature, as it requires the computation of the relative importance of sound elements and their distribution in time (Krumhansl, 2000; Snyder, 2000). The attribution of meaning to musical sounds, comparable to the semantics of language material, is also supposed to take place during the attentional stage of pitch processing (Besson & Schön, 2003).

1.3.1 Attentional ERP indices of musical pitch processing

The early right anterior negativity (ERAN) component of the ERP (Koelsch et al., 2000; Koelsch & Mulder, 2002; for review, see Koelsch & Siebel, 2005), peaking at around 200-250 ms from sound onset, has been recorded to violations of the music-specific rules of chord sequences (or, in musical terms, cadences) under a paradigm where subjects were required to attentively listen to the sounds and to detect a randomly occurring timbre violation. The ERAN was also elicited under the passive condition of subjects reading a book (Koelsch et al., 2002a). However, in this study, only the largest expectancy violations produced the ERAN response, i.e., when the incongruent chords were located at the end rather than in the middle of the cadence (for the prominent role of the last sound of a musical phrase for evoking feeling of closure, see Palmer & Krumhansl, 1987; Boltz, 1989). Moreover, the ERAN is strongly enhanced in amplitude under conditions of high attentional load to the sounds, and thus, is likely unrelated to automatic sensory-memory operations (Loui et al., 2005). The ERAN generators are located in the bilateral inferior pars opercularis (BA 44/45, corresponding to Broca's area in the left hemisphere; Maess et al., 2001), suggesting the correspondence of the music-related neural mechanisms responsible for the formation of strong expectations with regard to the incoming sounds with language-related syntactic processing. Interestingly, even though the ERAN is larger in amplitude in musicians (Koelsch et al., 2002b), it is also present in subjects without an explicit musical background, thus showing their sensitivity to the musical schemata of harmony, i.e., to the rules of chord combinations (Koelsch et al., 2000). While the ERAN occurs at a relatively early latency (following the MMN with a narrow time margin), it has been elicited by incongruous chords only, which provide multiple acoustic cues for violation detection as compared with those provided by melodic tones. Whether the ERAN truly reflects a cognitive process or whether it is also affected by the acoustic structure of the sounds used for its elicitation remains unclear (cf. Leino et al., submitted).

In electrophysiological recordings, the P3 or P300 (occasionally preceded by the negative wave N200 or N2 or N2b; see, e.g., Schiavetto et al., 1999) has been associated with attentional decision-making processes. The P3 is a parietally maximal positive wave typically peaking at about 300 ms after the presentation of a target tone during a discrimination task (Squires et al., 1975; Donchin, 1981). Intracranial recordings in epileptic patients demonstrated the involvement of the parietal cortex, hippocampus, basal ganglia and thalamus in attentive processing of simple pitch-feature changes as indexed by the P3 (Kropotov et al., 2000). In a musical context, the P3 is elicited by incongruous tones placed at the end of well-known (Verleger, 1990; Paller et al., 1992) or even unfamiliar but repetitive (Tervaniemi et al., 2003) melodies. A frontocentrally maximal P3a (a subcomponent of the P3, associated with the involuntary capture of attention by unexpected stimuli, contrasted to the P3b, reflecting decision-making processes; for reviews, see Escera et al., 2000; Friedman et al., 2001) was also observed when incongruous tones were presented in the opening phrase of a well-known melody (Verleger, 1990). Moreover, the P3a and P3b were larger in amplitude for implausible endings than for harmonically plausible final chords but ones less likely to occur in the primed context (minor chords in a major context; Janata, 1995). The P3b latency also correlated with the behavioural RTs and HRs in a task where subjects were required to rate whether the ending was the best-fitting one, with the RTs being shorter and the HRs higher for the most expected and the most dissonant endings than for minor-chord endings (Janata, 1995). Interestingly, the magnetic counterpart of the P3, the P3m, correlates in musicians with their subjective identification of the musical rules (Beisteiner et al., 1999).

The P600 (also termed the late positive component, LPC), peaking between 600 ms and 900 ms, belongs to the P3 family of waveforms, possibly being a type of P3 with delayed latency due to the characteristics of stimulation. Like the P3, it is typically elicited in an experimental paradigm where subjects actively listen to sounds and evaluate their appropriateness in a given context (Besson & Macar, 1987; Besson et al., 1994; Besson & Faita, 1995; for reviews, see Besson & Schön, 2003; Koelsch & Siebel, 2005). The P600 is supposed to follow general repair processes occurring when the re-analysis of a musical or even a linguistic sequence is needed as a consequence of encountering an incongruity (e.g., Osterhout et al., 1994; Patel et al., 1998a). The P600 is modulated by musical expertise; tones violating the musical expectations in melody or harmony elicit larger amplitudes and earlier responses in musicians than in non-musicians, indicating the facilitation of tonal processing by musical training (Besson et al., 1994; Besson & Faita, 1995).

To sum up, the neural processes underlying the later attentional stages of pitch processing are more widely distributed over the human brain and involve larger neuronal populations than the previous stages. The related main ERP correlates, the P3 and the P600, index decisional processes on sounds, but are elicited under experimental conditions that can be regarded as artificial compared with the everyday experience of music listening (Schmithorst, 2005). Conversely, in order to extract the neural processing specific to the encoding of musical pitch relations, it is preferable to adopt a paradigm in which subjects direct their attention away from the sounds, such as the MMN paradigm. While this experimental procedure does not bear on all of the listening situations in which we find ourselves, it nonetheless includes many of them, at the same time permitting one to avoid contaminations from modulatory processes related to attention. Moreover, the MMN enables one to address questions left open by the current understanding of pitch processing in musical contexts; in particular, whether sensory-memory processing of musical pitch in musicians as well as in laymen is more advanced than thus far proposed.

2 OPEN RESEARCH QUESTIONS

2.1 Pitch processing of multiple auditory objects: attentive or pre-attentive?

As previously mentioned, in audition, as in vision, features of stimuli are transduced and represented in the brain in an orderly fashion in feature maps (Näätänen & Picton, 1987; Pantev et al., 1989a, 1989b; deCharms et al., 1998; May et al., 1999). The integration of these features into a unitary auditory object is thought to occur in the subsequent stage. Behaviourally, this corresponds to the perceptual capacity of properly integrating the features associated with each object even in a situation when multiple objects are presented in an auditory or visual scene. A commonly held view, the feature integration theory (Treisman & Gelade, 1980; Treisman & Sato, 1990; Treisman & Schmidt, 1992), states that feature information is re-synthesized into a holistic representation of a unified object, or into several distinct representations of multiple objects, only with the help of voluntary attention to the spatial locations of the objects.

However, contrasting evidence led to the formulation of a parallel model according to which different stimulus features interact already during the early stages of analysis (Woods et al., 1998). This premise asserts the automaticity of auditory object processing by emphasizing the importance of ‘primitive’ or Gestalt properties in the integration of auditory information before attentional intervention during the sensory-memory phase (Bregman, 1990). Several MMN studies support this view by showing that the central auditory system not only conjoins stimulus features pre-attentively (Gomes et al., 1997; Sussman et al., 1998; Takegata et al., 1999) but also organizes and maintains simple auditory objects automatically for brief periods of time (Ritter et al., 2000).

At the level of sensory memory, single features of sounds are encoded. According to the model of Schröger (1997; see also Schröger et al., 2004), the obligatorily operating deviance-detection system indexed by the MMN encodes the features of the actual stimulus input in an auditory neural memory representation, termed R. This representation compares these features with the representation of the features that were invariant in the recent stimulation registered across the sensory memory span, termed R′. The neural representation of the simple features of repeated sounds seems to be stored in different parts of the auditory cortex, thus reflecting the orderly mapping of these features by cortical neurons, as indexed by MMN source analysis to changes in frequency, intensity and location (Giard et al., 1995) or even in timbre dimensions (Caclin et al., in press). However, the representation of the invariance in the previous stimuli R′ can also be based on the conjunction of simple features between multiple auditory objects. As a consequence, the deviance-detection system may be capable of responding both to deviations of single features and to properties of multiple sound stimuli whose features were infrequently combined (cf. also Takegata et al., 1999, 2001). Feature integration as a domain-general (neither audition- nor music-specific) perceptual skill is relevant for music processing. Within the complex and structured musical soundscape, temporal and spectral features are perceptually integrated (Deutsch, 1982; Krumhansl & Iverson, 1992). Another feature besides pitch where the spectral dimension is important is timbre, crucial for the identification of auditory objects; it distinguishes sounds having the same pitch, loudness, location and duration, hence permitting one to perceive that the same pitch played on a piano and a violin comes from two different sound sources (ANSI, 1973). If the integration of the features belonging to the numerous sound objects in music needed attentional resources, then it would be hard to explain the immediate perception of the multidimensional aspects of music during a master performance, or the common capacity to subconsciously appreciate slight nuances in the interpretation of a skilled performer.

Moreover, to render possible the attribution of features to distinct multiple objects, the central auditory system needs to maintain these objects simultaneously, even when these features are complex (e.g., when several sequential pitches are involved). Similarly to feature integration of a single object, multiple object persistence in auditory sensory memory can be considered a domain-general perceptual skill, i.e., not restricted to audition. Indeed, in the visual modality, complex features are segregated and integrated into multiple objects due to the brief maintenance in iconic memory of afterimages of the visual scene (Henderson & Hollingworth, 1999). The capacity to immediately integrate features into several objects and maintain them in short-term storage is essential in music since it allows one to identify and memorize several brief motifs (e.g., pitch patterns of 4-6 tones), which constitute the building blocks of musical phrases (Schoenberg, 1967). In this thesis, it is hypothesized that pitch and other features, such as timbre, need to be correctly integrated into and simultaneously maintained as multiple auditory objects at the pre-attentive stages of information processing to allow the listener to grasp the composite structure of a musical soundscape.

2.2 At which stage is pitch processed according to musical properties?

The skills illustrated in the previous section are general perceptual principles applied to sensory phenomena and are not specific to audition (Jackendoff & Lerdahl, 2006). Moreover, they are possibly universal and culture-general in the sense that they are available to all individuals, irrespective of age and cultural familiarity with the stimuli (Drake & Bertrand, 2003). However, behavioural and electrophysiological data show that any individual exposed to the musical system of a particular culture possesses capacities for encoding pitch within melodies or cadences according to music- and culture-specific knowledge (Krumhansl, 2000;

Koelsch et al., 2000; Bigand, 2003; Trainor, 2005). This section introduces the open issue of whether music- and culture-specific pitch encoding occurs at the sensory-memory phase or during subsequent stages of processing, and, further, whether it is modulated by musical expertise.

A mistuned or 'sour' note in a melody is detected and perceived as unmusical by almost every adult when inserted in already known musical excerpts (Drayna et al., 2001), or even when it disrupts novel excerpts composed according to the conventions of the listeners' music culture (Cuddy et al., 1981; Trainor & Trehub, 1992). This links music to language, where unconscious principles guide grammaticality judgments and understanding of a sentence (Hauser & McDermott, 2003). Hence, it is possible to speak of 'knowledge of music' in the same way as Chomsky (1957) described 'knowledge of language' (Hauser & McDermott, 2003). Knowledge of a particular music culture is commonly acquired through everyday exposure by, for instance, daily listening to any mass media, hence with no requirement for formal training or education. Long-term exposure to a specific musical system determines whether a succession of sounds is considered more or less musical according to the pitch relations between the tones. As a consequence, many properties of music are implicitly known by individuals exposed to the musical system of a particular culture even when they do not have any music education (Tillman et al., 2000; Bigand, 2003). For instance, when dealing with musical structures rather than with isolated tones, musically untrained listeners use the same principles as music experts in organizing their music representation (Bigand, 2003). This phenomenon has been defined in the literature as implicit learning (Seger, 1994), occurring when a subject learns the structure of a complex stimulus environment without necessarily intending to do so and in such a way that the resulting schematic knowledge is hard to express (alternatively, implicit learning has been considered as an unconscious process giving rise to abstract knowledge; Berry, 1997).

Music schematic knowledge implies a permanently stored repertoire of sound relations and structures of the musical system of a particular culture. Nearly all musical systems attribute a crucial role to the octave interval; notes that are an octave apart are considered functionally equivalent and are usually labelled with the same name. The octave is divided into a finite discrete number of unequally spaced pitches that form the notes of the musical scales adopted in the various musical systems around the world. In the Western tonal system, the chromatic scale includes all the twelve notes from which music is composed. These notes are structurally organized according to specific rules, defining subsets of seven notes and chords, i.e., the musical keys (also termed diatonic scales). The seven notes or chords of the scale are in-key, and the remaining five notes or chords are out-of-key. This set of pitches, each at a specific musical interval to each other, can also be referred to as the ‘pitch space’ (Jackendoff & Lerdahl, 2006). Each note and chord of the tonal scale has a hierarchical function, which determines its frequency and best place of occurrence. For example, the tonic notes or chords having the most central musical function are used more often than are notes or chords in the lower positions of the tonal and harmonic hierarchies (Krumhansl, 2000). The tendency to hear every note in relation to the tonic or tonal centre is a fundamental aspect of tonality (Jackendoff & Lerdahl, 2006).

A musical context and its underlying music schemata thus exert an influence on the way we consciously perceive the sounds embedded in it. For instance, while listening to tonal music (even when with residual attentional resources), we immediately detect any ‘wrong’ notes that do not match our expectations based on the prior context. When we identify, with or without conscious awareness, a tonal centre, we have expectations that the subsequent events of the musical piece will be tonally related, and these expectations affect the way we perceive the incoming sounds (Bharucha & Stoeckig, 1986; Bigand & Parncutt, 1999). It remains to be determined whether the detailed information of pitch space in a melody context is encoded at the sensory-memory stage, without the intervention of the additional resources associated with voluntary attention triggering to sounds.

2.3 Putative neural mechanisms responsible for musical pitch processing

2.3.1 Cortical plasticity and sound exposure

Electrophysiological evidence has demonstrated that even the adult primary auditory cortex displays neurophysiological plasticity (Weinberger & Bakin, 1998). Plasticity of the adult human cortex consists of neural changes following intensive and prolonged exposure to a specific stimulus environment, provided that such exposure is behaviourally significant to the subject and is accompanied by the involvement of attentional resources (Münte et al., 2002). According to Hebb’s rule, new connections between neurons are formed on the basis of synchronous activation: ‘Cells that fire together, wire together’ (Hebb, 1949). Experience involving attentive, continuous, repetitive engagement with complex tasks and stimuli results in several measurable changes in the brain, including increases in dendritic length, increases or decreases in spine density, synapse formation, increased glial activity and modified metabolic activity. These structural changes are correlated with behavioural differences between subjects affected and those unaffected by the changes (Kolb & Whishaw, 1998).

Invasive electrophysiological recordings in animals permit one to measure single-unit discharges or local field potentials, as well as their modifications after peripheral excitation. By using animal models, listening experience was demonstrated to shape the size of cortical networks in the PAC by either expansion or reduction. For instance, a stronger response to a sound may indicate the activation of a greater volume of neural tissue, and thus, sensitization to that particular sound, whereas a smaller response may underlie the recruitment of fewer neurons or differently synchronized firing patterns, or different connectivity with other regions (Weinberger & Bakin, 1998). In a frequency-training study, owl monkeys were trained to make fine frequency discrimination during 60-80 daily sessions. Subsequent invasive multiple-unit recordings in the trained monkeys showed that the neuronal populations of the A1 tuned to the behaviourally relevant frequencies were enlarged two- to threefold those of monkeys untrained or only passively exposed to the same sounds. In addition, the increased amplitudes of the neural responses to the trained frequencies were correlated with improved behavioural performance. These results demonstrate that attended stimulation can modify the tonotopic organization of the PAC in animals, and further, that this plastic change is correlated with perceptual acuity to those stimuli (Recanzone et al., 1993). A second type of frequency-training studies employed a classical conditioning protocol (Weinberger et al., 1993; for review, see Weinberger & Bakin, 1998); after the coupling of a tone with a given frequency (the conditioned stimulus) with an aversive electric shock, the CF of the tuning curves of the PAC neurons shifted towards the frequency of the conditioned stimulus after training. This shift lasted for a few weeks after training and could be reversed by extinction training.

In humans, the macroscopic cortical activity of tens of thousands of neurons can be recorded from the surface of the scalp with EEG or MEG. In the auditory modality, by measuring the N1 amplitude and latency, it is possible to determine how the size of the neuronal populations responding to sound features and their temporal organization are shaped by previous experience of those sounds, which can be easily induced with laboratory training (Pantev et al., 1999). Menning and colleagues' (2000) subjects were trained for 15 days (1.5 hours per session) to detect progressively smaller differences between a 1 kHz standard sine tone and deviant tones of slightly higher frequencies. The behavioural frequency discrimination quickly improved during the first week of training and slowly but constantly thereafter reaching a limit of 2 Hz by the third week. MEG measurements with 1000 Hz standard tones and 1050, 1010 and 1005 Hz deviant tones showed that both the N1m strength for the standard and deviant stimuli and the MMNm strength for the deviant stimuli were enhanced during training. The N1m enhancement during training may reflect changes in the frequency representation in the auditory cortex, i.e., in the involvement of additional neurons responding to those sounds, the more synchronous activation of these neurons or attentional modulation. Improvement of perceptual acuity was obtained after several training sessions, and this was most likely based on slow neural changes resulting from cortical reorganization. Those changes might represent the neural mechanisms responsible for the consolidation of information in long-term memory (Atienza et al., 2002; Galván & Weinberger, 2002). However, the learning effect observed by Menning et al. (2000) diminished as early as three weeks after training.

Furthermore, Näätänen et al. (1993b) found that frequency-discrimination performance can evolve even within a single training session[3]; an MMN gradually emerged to deviants within temporally complex tone patterns (differing from the 8-tone standard patterns only in the decreased frequency of the sixth tone) only in those subjects whose discrimination performance in the training tests also improved over the course of the experimental session. In addition, in one group the MMN did not emerge gradually, but rather was elicited at the beginning of the experiment. The improvement in perceptual sensitivity to complex sounds (as indexed by the MMN) was measured even when subjects were engaged in reading a book. The authors of the study concluded that attention, while crucial during the learning phase, is not needed during the retrieval phase of newly learned information and that this phase relies on previously developed memory traces for sounds that can be automatically activated (cf. Atienza et al., 2002).

The non-invasive studies on humans reviewed above do not, however, fully tackle the temporal properties of the plastic modifications of frequency-specific responses after laboratory training. For instance, whether learning a specific frequency would affect the tonotopic frequency representation in the human auditory cortex in the same brief time as demonstrated in invasive studies with animals is unknown. In other words, the speed of the learning of a simple frequency feature in humans has never been assessed.

2.3.2 Role of exact pitch discrimination in music perception

The developmental sequence for processing of musical pitch from childhood to adulthood shows that exact pitch-step (interval) discrimination is necessary for the acquisition of scale structure knowledge. Infants aged 5 months can reproduce single pitch levels and notice changes in melodic contour while treating the transposition of melodies over different keys as equivalent. However, at this age, infants still possess a poor discriminatory ability for frequency, which improves first for high, then for low frequencies (for review, see Trainor, 2005; for recent electrophysiological data on frequency discrimination in newborns, see Novitski et al., in press). Children 5 years of age are, in turn, able to detect one note change within a Western tonal melody that goes outside the scale of the melody; however, only later (at 7 years of age) reaching the adult's ability to detect within-scale changes that go outside the implied harmony of the melody (Trainor & Trehub,

---

[3] The experimental session consisted of a sequence of 26 stimulus blocks in which subjects were alternatively intent in either reading while ignoring the sounds or performing a sound discrimination task.

1994). Furthermore, 5-year-old children can recognize the tonal centre of melodies, but cannot use consistently the musical scale structure to transpose melodies in an accurate way to new keys (Dowling, 1982). In sum, the acquisition of the sensitivity to musical scale structure is subsequent to the development of contour processing and fine-grained pitch discrimination skills, reaching adults' levels within the first elementary school years (Dowling, 1982; Trainor, 2005).

Contour information is easier to encode and to maintain in long-term memory than interval (scale-step) information; however, exact interval encoding is necessary for melody recognition (Dowling, 1982; cf. also Peretz, 1990; Schiavetto et al., 1999; Trainor et al., 2002; Foxton et al., 2004a; Fujioka et al., 2004). The neural network devoted to the recognition of pitch contours seems separated from that devoted to fixed pitches and intervals. If the former is lesioned, deficits in intonation result in both music and language, but if the latter is damaged, then the impairment is music-specific and spares language, showing the relevance of interval processing neural skills for music perception (Peretz, 2001).

The hypothesis of the perceptual predisposition to process exact pitch relations is supported by the existence of a hardwired 'pitch centre' in the human brain (Peretz, 2003; Peretz & Coltheart, 2003). The first evidence emerged from the neuropsychological studies of patients with focal lobectomy (Milner, 1962; Shankweiler, 1966); right temporal damage resulted in a greater music processing deficit than left temporal damage, and further, this impairment affected the processing of melodies rather than the discrimination of isolated pitches. Moreover, as mentioned previously, the lateral part of the HG in the right superior temporal lobe is supposed to be responsible for the pitch percept even when this cannot be linearly extracted from the frequency of the sound but is derived from the temporal properties of the acoustic stimulation (as in the case of the IRN; Griffiths et al., 1998). Additional support for the hypothesis that abilities of fine-grained pitch discrimination are hardwired in the human brain comes from genetic research (Drayna et al., 2001).

Congenital amusia sheds further light on the role of abilities of fine-grained pitch discrimination and exact interval encoding for music perception. This term refers to the condition of individuals with lifelong musical disabilities (despite efforts to learn it) that cannot be explained by sensory or intellectual problems. Amusic individuals are impaired in appreciating, perceiving, memorizing and recognizing musical material (Peretz, 2003; Peretz et al., 2003). In the UK, congenital amusia is estimated to be present in about 4-5% of the population, according to a pitch-discrimination test in a melodic context (Kalmus & Fry, 1980). The lack of fine-grained pitch discrimination, leading to problems in acquiring perceptual knowledge of musical scales, has been proposed to be at the root of amusics' disabilities (Hyde & Peretz, 2003). This would then lead to their difficulty to grasp the pattern of tension and attraction derived from the use of pitches within a pitch space and, in turn, to the amusics' lack of musical affect, which is partially based on these tension and attraction contours (Jackendoff & Lerdahl, 2006).

The presence of a psychoacoustic defect for fine pitch resolution was first documented in a case study where an amusic individual could barely discriminate a pitch change of a whole tone (1/8 of an octave) and performed worse than at chance level for a change of one semitone (1/12 of an octave), with both changes being inserted in a monotonic (i.e., with constant pitch) and isochronous (i.e., with constant time intervals) 5-tone sequence (Peretz et al., 2002). A similar test with 11 amusics showed that they had difficulty in detecting pitch changes smaller than a tone. Normally, adults and children can easily detect a quartertone pitch change (Olsho et al., 1982). Congenital amusics, however, performed like controls in a comparable time asynchrony-detection task (Hyde & Peretz, 2003), although they performed worse than controls in a tapping task with musical material (Dalla Bella & Peretz, 2003), showing the generalization of their pitch difficulty to other aspects of music.

Amusics showed inferior performance to controls in discriminating pitch variations inserted in either melodies or intonation patterns (speech-intonation contours where all linguistic cues were filtered out; Ayotte et al., 2002), suggesting that congenital amusia is not a music-specific disorder, but rather it is music-relevant. According to the authors, the pitch defect at the root of the disorder is more relevant to music requiring fine-grained pitch discrimination to track pitch changes of a semitone. In contrast, speech prosody relies on coarse (larger than half an octave) pitch shifts (Patel et al., 1998b). Alternatively, top-down processes using more efficiently speech rather than music cues may lead to the different performance in music vs. speech contexts of the intonation task (Ayotte et al., 2002; Patel et al., 2005). Furthermore, amusics showed deficits in determining the pitch direction of pattern tones, but not in segregating sound streams, thus hinting at the presence of a problem in pitch-pattern perception, rather than in primitive sound scene organization, along with difficulty in fine-grained pitch-step discrimination (Foxton et al., 2004b). Generally, congenital amusia highlights discrimination of exact scale-step changes as an essential component around which music cognition develops in a normal brain, since a faulty processing system for it leads to cascade effects on many aspects of music cognition, including beat tracking (Peretz et al., 2002). However, the neural determinants of this pitch problem need still to be elucidated.

3 AIMS OF THE THESIS

The present understanding of the cortical processing of musical pitch lacks specifications on the role of the pre-attentive sensory-memory stage for pitch analysis of multiple auditory objects according to context and past music knowledge. To this end, Studies I and II assessed the abilities of the human brain to simultaneously integrate and maintain in sensory memory simple and complex features of multiple musical objects without attentional focus. Studies III and IV aimed at determining whether pitch can be encoded according to the Western musical scale at the sensory-memory level. Study III additionally ascertained whether musicians show a superior ability in this music-specific pitch encoding relative to subjects with no formal musical education.

Moreover, two putative building blocks of music knowledge induction (Cohen, 2000) were examined: brain plasticity and fine-grained pitch resolution. Study V was carried out to determine how the tonotopical representation of pitch in the auditory cortex is plastically modified over a very short time span. The goal of Study VI, in turn, was to uncover the neural and behavioural correlates of the defect for fine-grained pitch-step discrimination in congenital amusics, further highlighting the relevance of this pitch skill for normal music perception.

4 SUMMARY OF THE ORIGINAL PUBLICATIONS

4.1 Multiple pitch objects in sensory memory

4.1.1 Pre-attentive integration of pitch and timbre

On the basis of evidence of ‘conjunction costs’ (longer RTs for the target search of feature conjunctions than for the search of single features) and ‘illusory feature conjunctions’ (wrong combination of features when stimuli are briefly presented), the feature-integration theory maintains that feature integration is not automatic but rather occurs when voluntary attention is drawn to the spatial locations of the objects (Treisman & Gelade, 1980; Treisman & Sato, 1990; Treisman & Schmidt, 1992). In the auditory modality, an MMN can be pre-attentively elicited by deviants introducing infrequent combinations of two acoustic features already present in the standard sounds (Gomes et al., 1997; Sussman et al., 1998; Takegata et al., 1999; 2001), thus partially challenging Treisman’s theory. However, the lack of illusory conjunctions reflected in the MMN elicitation could be explained by the sounds being sequentially delivered, thus allowing the integration of single sound features from the feature-analysis maps in serial time order.

Study I addressed pre-attentive auditory feature integration by using simultaneous spatially distributed auditory stimulation, therefore providing a more appropriate comparison with visual studies than auditory studies with sequential spatially-fixed stimulation (Gomes et al., 1997; Winkler et al., 2005). In the ‘parallel’ condition (see Fig. 3A), frequent combinations of pitch and timbre (e.g., piano-C5 and violin-E5) were randomly replaced by infrequent combinations (e.g., piano-E5 and violin-C5). The pitch of each component of the sound combinations was varied over eight different levels between experimental blocks to reduce the effects of learning during the experimental session and to minimize the contribution of the N1 enhancement on the ERP to the deviants. The MMN was expected to be elicited by the infrequent combinations if the hypothesis suggesting the pre-attentive integration of features is true. During sound presentation, subjects performed primary tasks differing in the amount of attentional load (1-back and 3-back working-memory tasks). If attention were required to integrate the sound features, then the MMN would be absent or diminished in the low-load primary task. This MMN absence or attenuation would result from the increased deviant-stimulus probability, known to affect the MMN, and the averaging of all deviants together, even when they do not elicit an MMN. In the ‘serial’ condition, the same sounds were sequentially presented (e.g., tone pairs including piano-C#5 and violin-F5 as standards and tone pairs with piano-F5 and violin-C#5 as deviants). In a separate behavioural session (auditory target-search task), subjects also judged whether a single sound probe appeared among two simultaneously presented sounds.

As expected, the visual 3-back task was more demanding for subjects than the visual 1-back task, hence requiring more of the subjects' attention than the 1-back task, whereas no behavioural differences were present between the serial and parallel modes of auditory stimulation. MMNs were elicited in both the parallel and serial conditions and with both the 1-back and 3-back working-memory tasks (see Fig. 3B). These MMNs did not differ in amplitude, latency or scalp distribution between the 1-back and 3-back conditions, thus showing that the attentional load of the primary task had no effect on MMN parameters. However, the MMN was larger, marginally significantly later and was more anterior over the scalp in the parallel than in the serial condition (see Fig. 3C). In contrast, scalp-distribution differences were not obtained when analysing ERPs in response to standard stimuli, suggesting that the topographic differences between the parallel and serial conditions were specific to the MMN process. Behavioural performance in the target-search task did not differentiate between invalid and valid trials, indicating instead frequent illusory conjunctions of timbre and pitch.

The MMN elicitation to rare combinations of pitch and timbre, unmodulated by attentional load, demonstrates that the integration of musical features does indeed occur automatically in the sensory-memory phase of auditory processing. This MMN result contradicts the behavioural data, showing illusory conjunctions in the target-search task, always subsequent to the MMN recording session. It is hence likely that musical features, such as pitch and timbre, are bound together and pre-attentively encoded as an object representation in the sensory-memory stage, but that additional factors coming into play during voluntary access to those representations or in a subsequent attentive feature-binding stage may affect the correct feature integration and lead to illusory conjunctions. The difference in the MMN amplitude and scalp distribution between the parallel and serial conditions may be attributed to an additivity effect of the MMN if separate MMN responses were elicited by the two concurrent sounds. Alternatively, the observed MMN differences may result from partly distinct brain mechanisms responsible for change detection in simultaneously vs. serially presented sound objects (Friedman-Hill et al., 1995; Robertson et al., 1997; Shafritz et al., 2002).

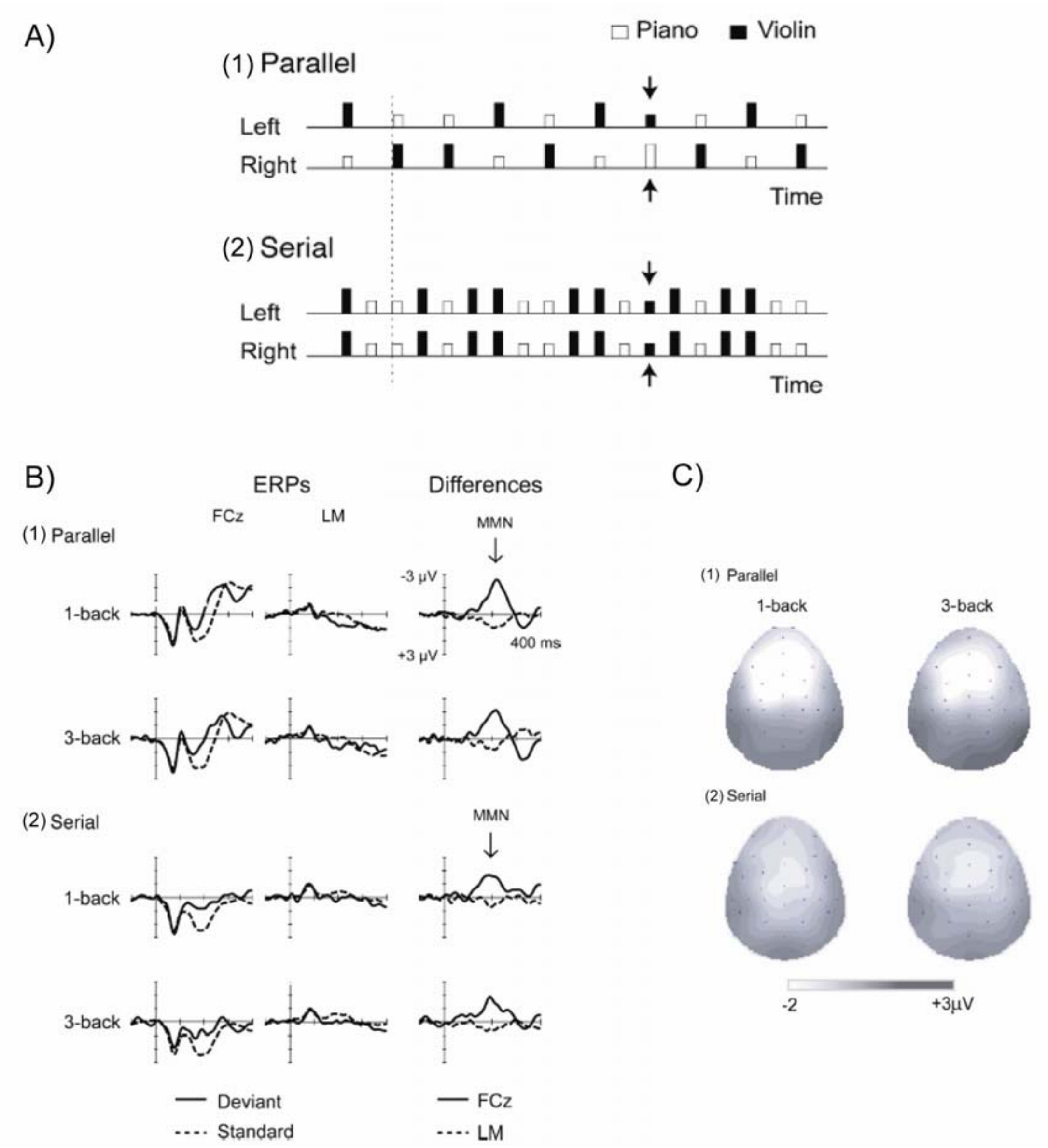

Figure 3. **A)** Schematic representation of the stimulation used in Study I. In each stimulus block and for both the parallel and serial conditions, piano and violin sounds with two pitch levels (one selected from the C5-G5 pitches and the other from the E5-B5 pitches in 1-semitone steps, and so that they were always separated by a major third interval) were presented as coming from the left or right loudspeakers and alternated so that the combination of timbre and loudspeaker was kept for half of the experimental blocks. In each block, 87.5% of the sound pairs had the same pitch-timbre combination, whereas in 12.5% of the pairs, the pitch-timbre combination was randomly reversed, as indicated by the arrows. In the serial condition, the piano and violin sounds were separately presented so that the same timbre was delivered by both loudspeakers, whereas in the parallel condition, the two timbres were simultaneously delivered by the two loudspeakers. **B)** Grand-averaged ERP responses (*N*=12) elicited by the standard and deviant stimuli with the corresponding difference waveforms, where the standard ERP was subtracted from the deviant ERP, recorded at FCz and the left mastoid. In both the parallel and serial conditions and for both attentional loads (low for the 1-back visual task and high for the 3-back visual task; performed by subjects during ERP recordings), an MMN was elicited at around 200 ms from stimulus onset, as indicated by the frontocentral negativity and the corresponding mastoidal positivity. **C)** Voltage isopotential maps of the MMN responses in the two experimental conditions (parallel and serial)

and for the two task loads (1-back and 3-back). The maps were drawn at the peak latencies of the corresponding grand-averaged MMN responses. While the task load did not affect the scalp distribution of the MMN responses, the condition did; the MMN in the parallel conditions was more anterior to that obtained in the serial conditions. Modified from Study I.

4.1.2 Pitches of multiple patterns are maintained in sensory memory

Study II addressed the question of whether two complex temporal auditory objects could be simultaneously represented in sensory memory as indexed by MMN elicitation. For this purpose, three 4-tone patterns were randomly mixed together in a sound sequence. Two of the patterns appeared frequently, whereas the third, the deviant pattern, which started out as one of the standard patterns and ended as the other standard pattern, was infrequently presented (see Fig. 4A).

As visible from Fig. 4B, in the MMN latency window (around 250 ms after deviance onset), the ERP at Fz to the deviant pattern was more negative than that to the first or second standard pattern. At the mastoid electrodes, the ERP to the deviant pattern was also more positive than that to the first or second standard pattern. Its morphology, scalp topography and latency correspond to the notion of the MMN. An MMN was also elicited by both deviant patterns following a homogeneous sequence of only the first standard or mixed sequences of both standards. However, the MMN measured at the mastoids in response to deviants following the mixed sequences of the standard patterns was prolonged compared with the MMN to deviants following the homogeneous sequences, possibly indicating difficulty in the comparison process at the temporal regions (corresponding to the temporal subcomponent of the MMN; Näätänen, 1992) when multiple sensory-memory traces are simultaneously maintained.

Thus, the two patterns were simultaneously encoded in auditory sensory memory. The elicitation of an MMN to a deviation consisting of the conjunction of the two standard patterns without the introduction of any new tone indicated that the two patterns were encoded in auditory sensory memory as stimulus representations (or objects) rather than as individual tones. The findings further suggest that the exact pitch relations in the two repeated patterns were automatically encoded and concurrently maintained. Such a process may shed light on the neural mechanisms underlying the quick and simultaneous memorization of several motifs during listening to music.

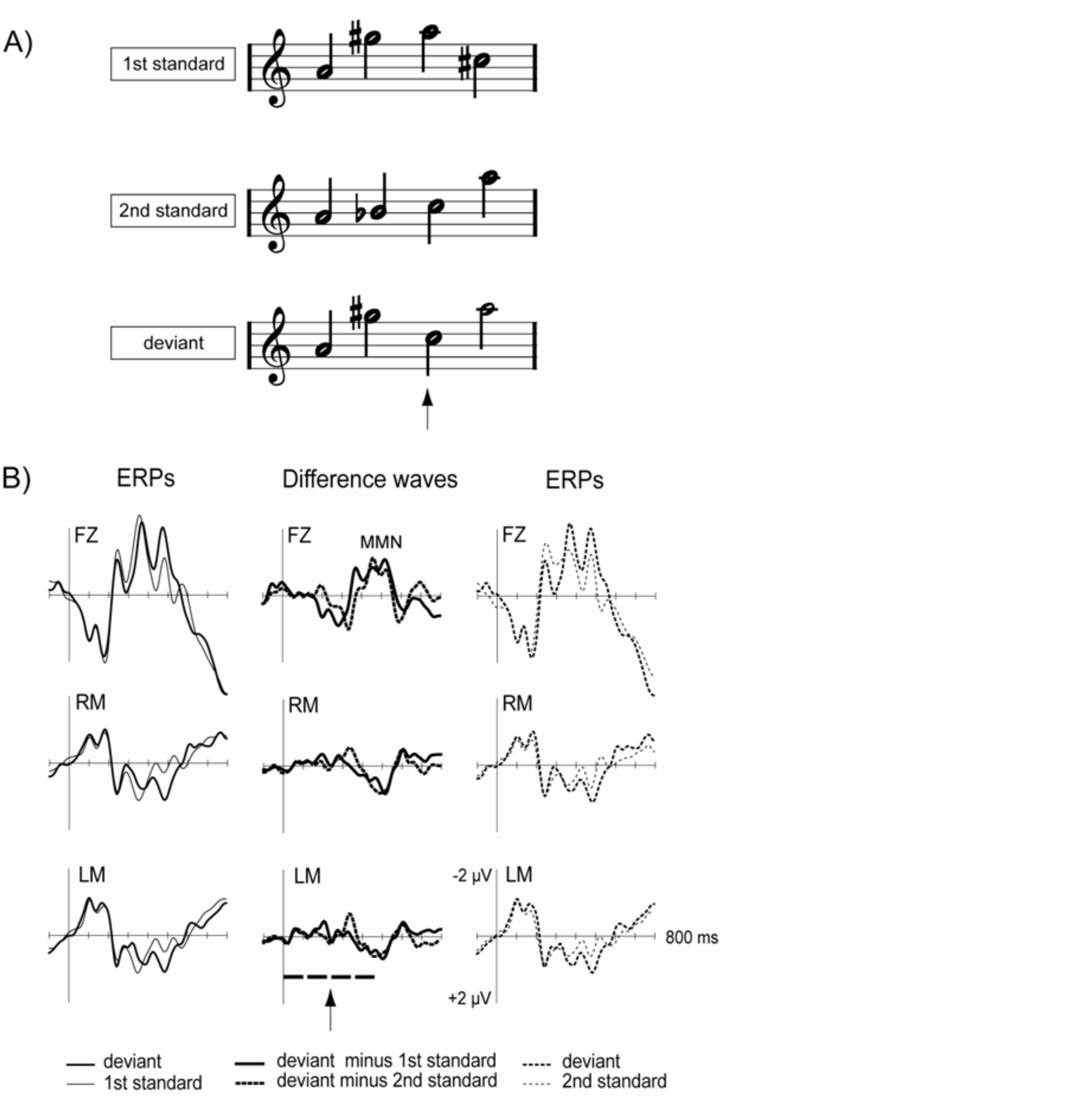

Figure 4. **A)** Stimuli employed in Study II written in musical notation. The deviant pattern (occurring with a frequency of 6%) consisted of the first two pitches of the first standard pattern combined with the last two pitches of the second standard pattern (each standard pattern occurred with a frequency of 47%). Thus, the deviant pattern introduced no new pitch. The deviation from the first standard pattern is indicated by the arrow. **B)** Grand-averaged ERP responses (*N*=10) to the deviant and the first standard (left column) and to the deviant and the second standard (right column) recorded from the Fz and mastoidal electrodes. The middle column shows the two corresponding difference waveforms, where the ERP to each standard was subtracted from that to the deviant. The deviant pattern violated the pre-attentively encoded sensory memories of both standard patterns, as demonstrated by the MMN elicitation. The time course of the stimuli is indicated at the bottom of the middle column with the arrow showing the onset of the deviation (240 ms). Modified from Study II.

4.2 Pitch processing is affected by music knowledge

4.2.1 Familiarity with sounds facilitates pitch-change processing

Behavioural and psychoacoustic literature has demonstrated superior performance of subjects for the detection of mistuning in melodies when familiar musical intervals are employed relative to when uncommon intervals (e.g., derived from non-Western musical systems) are used (Lynch & Eilers, 1991; Lynch et al., 1991; Trainor & Trehub, 1993). Trainor and Trehub (1993), for example, showed that both adults and 9- to 11-month-old infants more rapidly discriminated a semitone change within prototypical Western melodies based on the major triad than a semitone change within non-prototypical melodies based on the augmented triad. Lynch & Eilers (1991) also found that mistunings in a melody are better discriminated when melodies are based on Western rather than non-Western scales. Study III aimed at determining the effects of musical-scale familiarity and context on neural pitch discrimination at the sensory-memory level.

Musicians and non-musicians were presented with three conditions where infrequent pitch changes were either embedded in five-tone patterns or presented in isolation (see Fig. 5A). In the Familiar condition, the third-position deviant tone changed the mode of the pattern, while in the Unfamiliar condition the patterns consisted of five arithmetically determined tones, with the deviant causing no musically meaningful change. The No-context condition included third-position tones only.

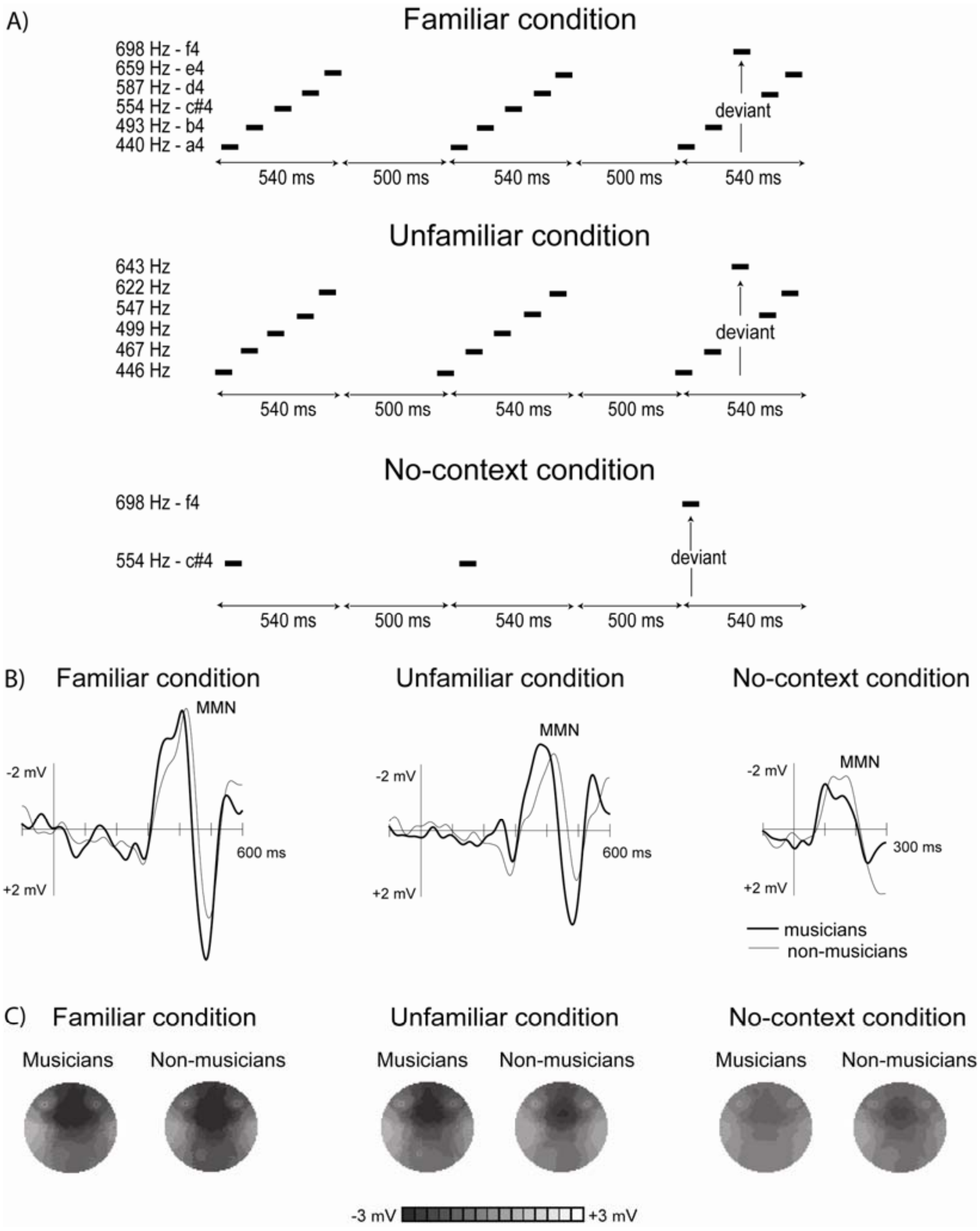

Figure 5. **A)** Schematic illustration of the experimental stimuli of Study III. The patterns of the Familiar and Unfamiliar conditions consisted of 5 sinusoidal tones of 100 ms in duration separated by 10 ms of silence, whereas the 100-ms tones of the No-context conditions were separated by a 1040 ms sound onset asynchrony (SOA), as indicated by the horizontal axis. The vertical axis depicts the tone frequencies and their corresponding musical pitches (for the Familiar and No-context conditions only). The deviant pattern of the Familiar and Unfamiliar conditions had the third tone shifted by 144 Hz, whereas the deviant tone of the No-context condition was higher than the standard tone by 144 Hz. The arrows indicate the loci of the deviant stimuli. **B)** Grand-averaged difference waveforms, where the ERP (recorded at Fz) to the standard was subtracted from that to the deviant for each experimental condition and each group. The difference waveforms for musicians (*N*=19) are drawn with a solid line, and the waveforms for non-musicians (*N*=10) with a dashed line. The MMN peak is labelled in each plot. The MMN to the pitch shift was larger in amplitude for the Familiar than for the Unfamiliar condition, and, further, it was the smallest in the

No-context condition in both musicians and non-musicians. **C)** Voltage isopotential maps of the MMN responses in the three experimental conditions and for the musicians and non-musicians. The maps were drawn at the peak latencies of the corresponding grand-averaged MMN responses. In both musicians and non-musicians, the MMN was slightly right-lateralized for the two pattern conditions, whereas it was bilaterally distributed for the No-context condition. Sections A and B modified from Study III; section C from unpublished data of Study III.

The electric brain responses to the pitch change (subtracted from the responses to the corresponding standard stimulus) in the three experimental conditions are presented in Fig. 5B. In both musicians and non-musicians, the pitch change embedded in the familiar scale context elicited a larger MMN amplitude than that elicited by the comparable pitch change embedded in the unfamiliar scale context, the latter, in turn, eliciting a larger MMN amplitude than the pitch change of the No-context condition. The MMN amplitude did not differ between the groups, however. In addition, the scalp distribution of the MMN responses was modulated by the experimental conditions such that there was a larger MMN amplitude over the right than over the left hemisphere for the Familiar and Unfamiliar conditions, but not for the No-context condition (see Fig. 5C). Furthermore, the MMN peak latency differed between the experimental groups. The MMN of the musicians was earlier than that of the non-musicians over the left hemisphere when the pitch change was embedded in the temporally complex contexts, but not when the pitch change intervened in the No-context condition.

Moreover, the musicians' auditory system reacted faster than that of non-musicians to sequential pitch patterns resembling a temporally complex musical material but not to isolated pitch. A similar result differentiating the musicians' and non-musicians' brain responses to simple vs. complex sounds was recently also obtained by Fujioka et al. (2004) with magnetic MMN recordings. Moreover, the MMN enhancement to pitch change within the Familiar condition in both musicians and non-musicians suggests that automatic pitch-change discrimination is influenced by long-term memory traces for the musical-scale steps of the Western musical system. In addition, the lack of an amplitude difference between the groups in the pitch-pattern processing indicates that explicit learning as a result of formal music training is not needed for superior processing of the pitch relations of Western tonal music.

4.2.2 Automatic processing of the musical scale

Study IV further tested the hypothesis that pitch is processed according to musical-scale properties at the sensory-memory phase by analyzing electric brain responses to pitch incongruities of the chromatic scale (out-of-tune pitch) or to those of the diatonic scale (out-of-key pitch) inserted in random locations within unfamiliar melodies (for examples of the experimental stimuli written in musical notation, see Fig. 6A). The stimuli were presented to subjects in two experiments: in the first passive experiment, subjects were asked to concentrate on watching a movie and to ignore the sounds, whereas in the subsequent active experiment, the subjects rated the melodies according to congruousness or incongruousness on a 7-point scale. Note that the invariance in the musical stimuli, needed for the development of a sensory-memory representation in the auditory cortex, was not physical since the melodies presented were repeated only four times through the passive ~2-h experiment, thus not resulting in learning of the exact pitches or intervals present in the melodies (six repetitions of unknown melodies may lead to laboratory learning of their in-key pitches but not of the out-of-key incongruities placed at the end of the melodies; Kuriki et al., 2005). Rather, the invariance consisted of the schema of the Western musical scale used to compose the melodies. The hypothesis was that if the encoding of the pitch relations of the musical scale occurs as early as at the sensory-memory phase without the intervention of attentional resources, then an MMN to pitch violations would be elicited, with an amplitude not modulated by the attentional load of the primary task (low in the passive, high in the active experiment).

The ERP results (see Fig. 6B and 6C) showed that the incongruous out-of-tune and out-of-key pitches elicited an enhanced long-lasting negativity as compared with the congruous pitch, which was maximal at the frontocentral electrodes between 180 and 280 ms, thus resembling the MMN response. This MMN-like response was unaffected by the attentional load, not differing in amplitude between the passive and active experiments. Furthermore, the source modelling of the out-of-tune and out-of-key negative deflections showed that they originated from the non-primary auditory cortex, with the predominant contribution of the right hemisphere. In the passive experiment only, the enhanced negative ERP to the two incongruous pitches as compared with that to the congruous one was observed until 580 ms. In the active experiment, by contrast, positive deflections were obtained at long latencies (see Fig. 6B and 6C). In particular, at 480-780 ms, parieto-occipital positivities (corresponding to the P600 component) were larger in amplitude to both incongruities compared with the ERP response to the congruous pitch. Interestingly, the parieto-occipital P600 to the out-of-tune pitch did not differ from that to the out-of-key pitch at 480-580 ms, but was enhanced at longer latencies (from 580 to 780 ms) only. These results reflected the subjects' behavioural ratings, attributing the highest degree of congruousness to the melodies containing the congruous pitch and the lowest to the melodies containing the out-of-tune pitch (while the melodies containing the out-of-key pitch were rated in between; see Fig. 6D).

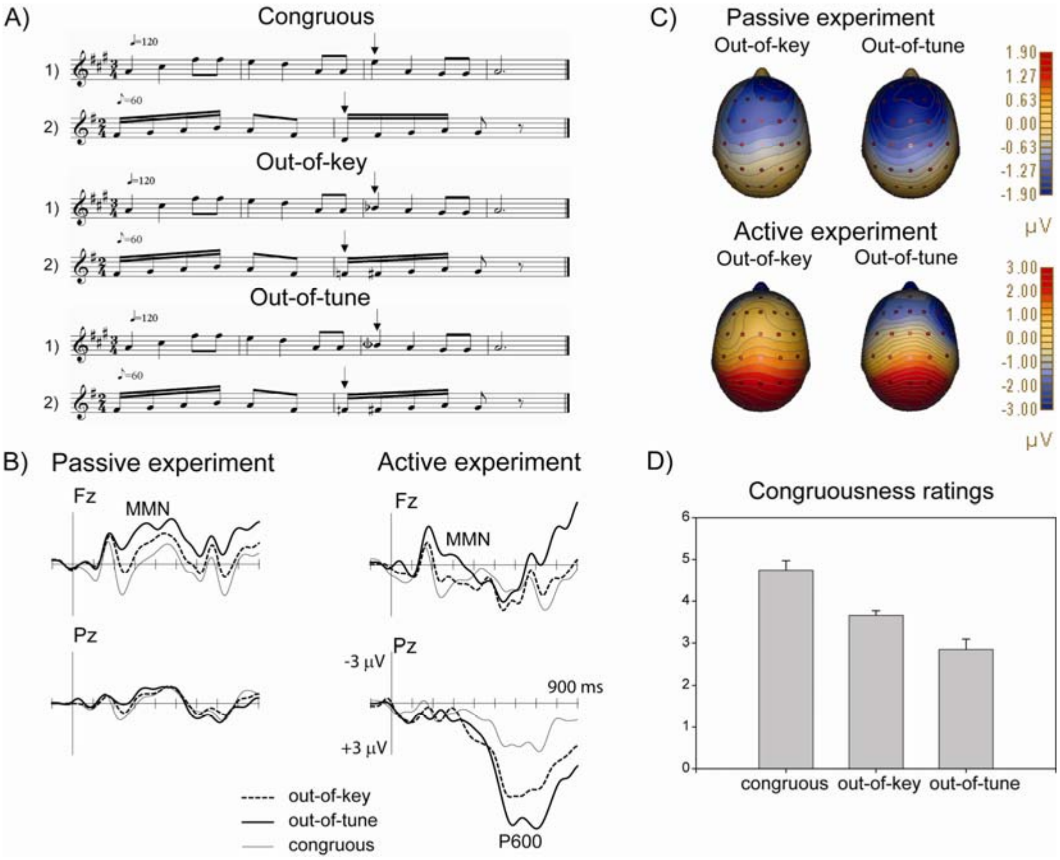

Figure 6. **A)** Two examples in musical notation of the melodies employed as stimuli in Study IV. Embedded in the melodies were the congruous, out-of-key and out-of-tune pitches, as indicated by the arrows. **B)** Grand-averaged ERP waveforms (referenced to the mastoids; *N*=9) to the congruous standard pitch and to the two incongruous (out-of-key and out-of-tune) deviant pitches for the passive (left) and active (right) experiments, recorded at the Fz and Pz electrodes. The incongruous pitches elicited an MMN response in both experiments and a P600 response in the active experiment only. **C)** Voltage isopotential 3-D maps of the MMN responses for the passive experiment and of the P600 responses for the active experiment to the out-of-key and out-of-tune pitch deviants. The maps were drawn from the nose-referenced difference waveforms and calculated at the peak latencies of the corresponding MMN and P600 responses. The maps depict the frontocentral topography of the negative MMN peak and the parieto-occipital distribution of the positive P600 peak. **D)** A histogram illustrating the results of the behavioural ratings on a 7-point scale of melody congruousness/incongruousness carried out by subjects during the active experiment (with the bars showing the standard errors of the mean, SE): the melodies containing the out-of-tune pitch were rated on average as the most incongruous and the melodies with the congruous pitch as the most congruous. Modified from Study IV.

The results indicate that relational musical-scale information is represented at the level of the non-primary auditory cortex. Hence, sensory-memory traces for pitch relations belonging to the schema of the Western musical scale are quickly and automatically formed during an early stage of processing that does not require the intervention of voluntary attention. The subsequent attentional stage, in turn, may reflect the reanalysis and repair of the melody scale schema disrupted by an incongruous pitch. In the next, the putative mechanisms responsible for pitch-related music processing will be addressed.

4.3 The brain mechanisms of musical pitch processing

4.3.1 Short-term cortical plasticity of pitch representations

Study V aimed at determining whether the neural refractoriness, as reflected by the N1 to several sound frequencies, could be affected by a brief discrimination training at a specific frequency. During the training, subjects were asked to judge whether a 1062-Hz tone was same or different from the other tone of the pair for about 1 h. Pre- and post-training, subjects were presented with one of seven different tones (among them a tone of 1062 Hz) occurring every 5 repeated 1000-Hz tones for about 25 min.

Subjects reached Δf steps of 8.7 ± 0.8 (standard errors of the mean, SE) Hz after the first block and 5.0 ± 0.7 Hz after the fourth block (see Fig. 7A). The Δf steps differed from each other during the training session, with larger Δf steps at trial 20 and in the first block compared with the others. In addition, the Δfs discriminated at trials 20 and 40 in the second block were smaller than those in the first block, the Δfs in the second block were smaller than those in the third block (possibly as a consequence of the inversion of the task introduced with the third block) and those in the fourth block were also smaller than those in the third block. Moreover, at trial 100, the Δfs in the third and fourth blocks were smaller than those in the first block. These data prove that subjects learned to discriminate small frequency steps at the end of the training session, hence developing an accurate perception of tone frequencies around 1062 Hz.

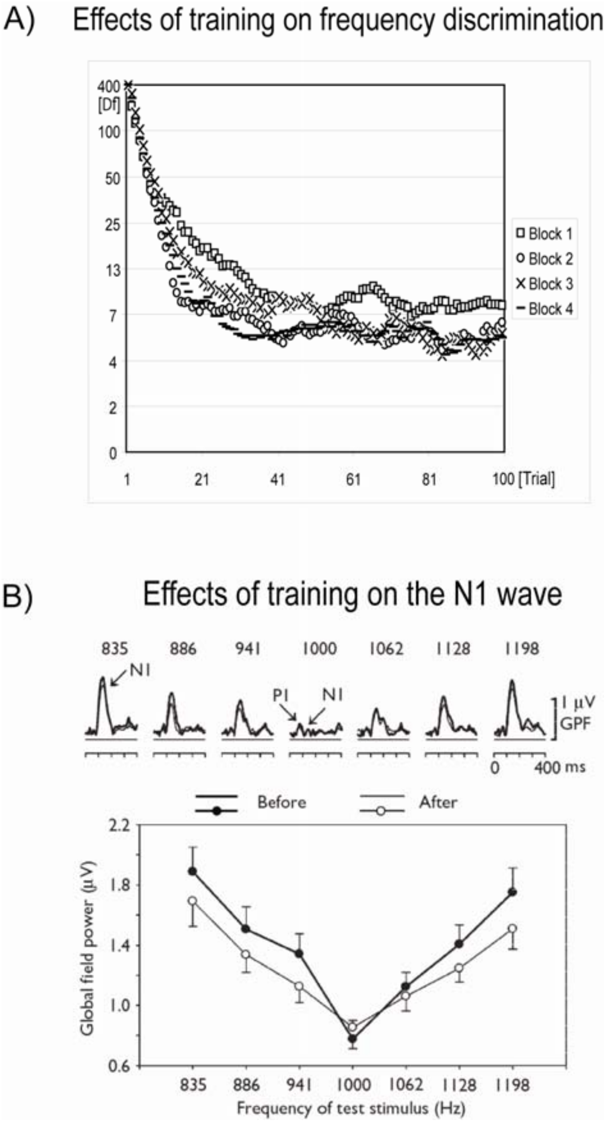

Figure 7. **A)** Plot of the discrimination training task showing the average discriminable frequency across the trials. The frequency difference (Δf) between the tones in the pair to be discriminated is expressed in Hz. Subjects reached Δf steps of about 8 Hz after the first block and about 5 Hz after the fourth block, demonstrating that they learned to discriminate small frequency steps at the end of the 1-h training session. **B)** The upper part illustrates the grand-averaged ERPs (*N*=18) elicited by each of the stimulus frequencies before and after training. The bottom graph represents the mean values of the N1 amplitudes for each tone frequency (the bars show the SE). All frequencies, except for 1000 Hz and 1062 Hz, elicited an N1 with a reduced amplitude after training, possibly indicating perceptual learning of the trained and repeated frequencies. Modified from Study V (the erratum in Figure 2 of the original publication has been corrected).

The N1 amplitudes recorded in the EEG session subsequent to the discrimination task showed training effects which significantly differed between the different frequencies (see Fig. 7B). Whereas all the N1s to the intervening tones decreased in amplitude after training, the N1 to the repeated 1000-Hz test tone and that to the 1062-Hz trained tone did not differ in amplitude when comparing the values obtained before and after training. In addition, changes in the N1 frequency-specificity pattern were found between the first and second blocks of each experimental session, indicating an effect of long-term habituation[4] (or unspecific diminished neuronal reactivity) on the N1 amplitude for some but not all of the tone frequencies. The N1 peak latencies only differed according to the tone frequency, with an earlier N1 to the 1000-Hz tone than to all other tones,

---

[4] Habituation means that a dishabituation could, in principle, be shown.

which is, possibly, a mere consequence of the signal being near to the background noise level.

These results indicate a long-term habituation in the second experimental session of the N1 to the frequencies that were not salient or trained, and a resistance to habituation of those behaviourally important frequencies (such as the trained and repeated ones). In relation to this, Kraut & Smothergill's (1978) model posits two distinct processes engendered by familiarity with a stimulus: enhanced encoding associated with a refined neural model, and decreased orienting to the stimulus or habituation of its neural response. Due to its brief time course, the present finding could be considered a transient phenomenon taking place after continuous exposure to a particular pitch. Additional studies need to be conducted to resolve this issue, by, for instance, testing the reversibility of the neural learning a few hours after laboratory training and further neural changes after a prolonged sound exposure. The perceptual learning of the tones was implicit since all subjects (except for one) did not realize that the test tone was the same than that discriminated in the training session. In sum, the cortical responses to pitch were modified as a result of laboratory training in a time span comparable with that previously observed with invasive electrophysiological measurements in animal models (Weinberger & Bakin, 1998). This implies that even the adult human brain is capable of sensitization for absolute frequencies after brief training, inducing (possibly transient) changes in the cortical representations of sounds.

4.3.2 Fine-grained pitch discrimination in congenital amusics

Study VI aimed at determining, for the first time, the electrophysiological correlates of the pitch discrimination defect in congenital amusics, proposed as the possible cause of their difficulties with music (Peretz et al., 2002; Ayotte et al., 2002). The EEG of congenital amusics and control subjects was recorded while they tried to detect when the fourth tone within isochronous (i.e., with same SOA between tones) and monotonic (i.e., with constant pitch) 5-tone patterns was displaced in pitch. The behavioral results (see Fig. 8D) showed that congenital amusics could detect large pitch changes as well as controls, but could hardly notice changes of 50 and 25 cents (where 1 cent corresponds to an interval size of 1/100 of the equal-temperament semitone).

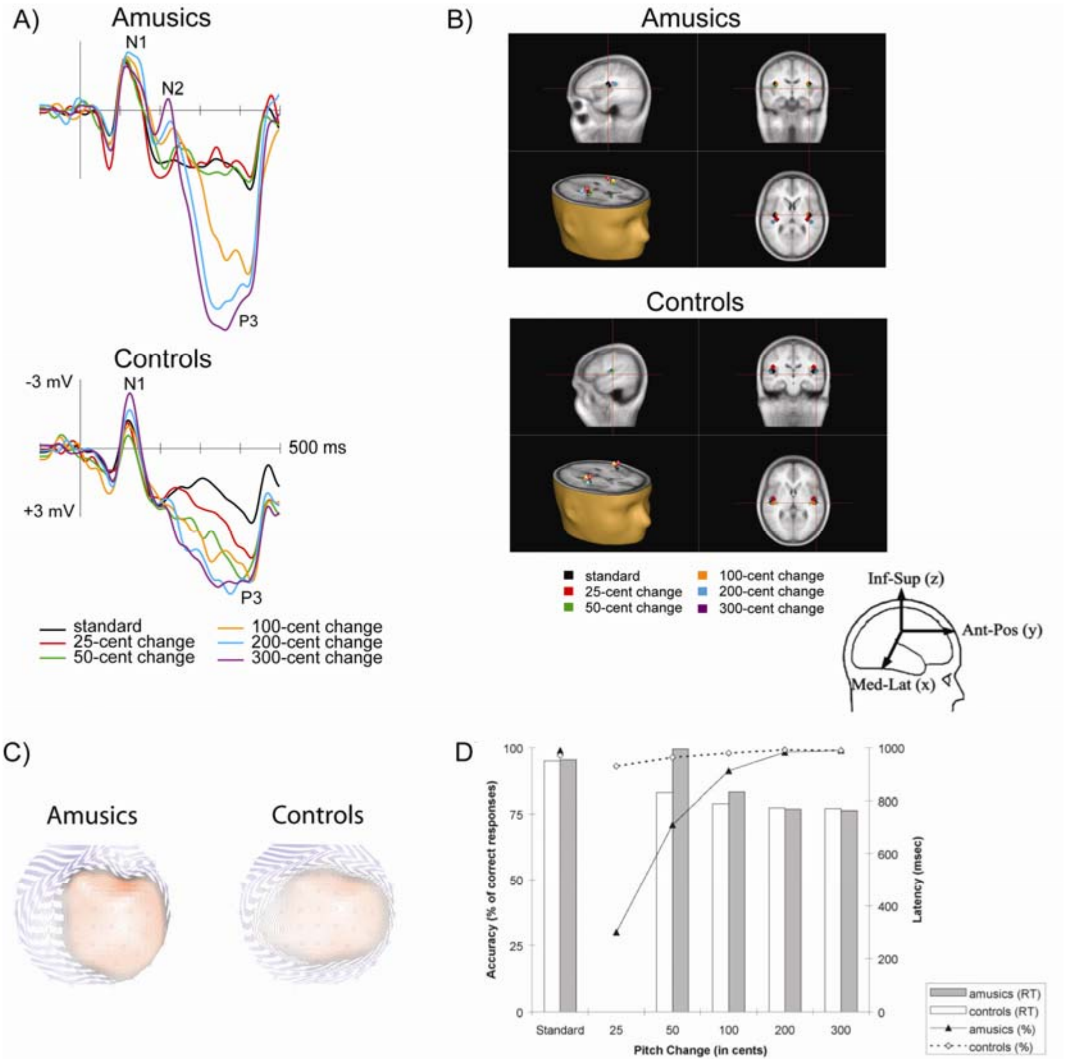

Figure 8. **A)** Grand-averaged ERPs to the standard and to the five levels of pitch change in amusic (*N*=8) and control (*N*=10) subjects, recorded at the Cz electrode. While controls showed a progressively increasing P3 amplitude as a function of the magnitude of pitch change, the amusics showed an N2 response to the largest pitch changes only (which is absent in controls), an enhanced P3 to the largest pitch changes and no P3 to the small changes. **B)** Regional source models of the N1 to the 300-cent pitch change in amusic and control subjects. The brain template was generated by averaging the Talairach-transformed magnetic resonance images of 24 adults (BESA software). The N1 sources in amusics were slightly more posterior to those in controls. **C)** Voltage isopotential maps of the P3 responses in both amusic and control subjects to the 300-cent pitch change. The maps were calculated at the peak latencies of the corresponding P3 responses. They illustrate the right-hemispheric lateralization of the P3 in amusics as compared with controls. **D)** Mean percentages (shown as solid and dashed lines for the amusics and controls, respectively) and mean reaction times (RTs; shown as white and grey bars for the amusics and controls, respectively) of the correct responses in the amusic and control subjects as a function of the magnitude of the pitch change. The mean RTs of the 25-cent change are not reported since there were too few measures for this condition in amusic subjects. Congenital amusics exhibited difficulty in the discrimination of pitch changes smaller than a semitone and could barely detect 25-cent and 50-cent changes. Sections A, B and D modified from Study VI; section C from unpublished data of Study VI.

N1 amplitude and latency did not differentiate the groups, but the regional source models of the N1 were more posterior in the amusics to those in the controls (see Fig. 8A and 8B). As visible from Fig. 8A, the next major electrophysiological event, the N200, was only present in the amusic waveforms. The subsequent P3 varied in amplitude at Pz in the two groups according to the magnitude of the pitch change, with an enhanced P3 to large pitch changes and a diminished one to small pitch changes in amusics relative to those in controls. Moreover, the P3 wave was more right-lateralized, more posterior and superior in control subjects than in amusics (see Fig. 8C).

These findings suggest that the amusic brain is impaired in detecting pitch deviances smaller than a semitone as compared with the normal brain and is sensitized to large pitch changes, as indicated by the N2 present only in amusics and by the P3 enlarged in amplitude almost by a factor of 2 as compared with that of controls. In contrast, the feature-trace stage reflected by the N1 brain response seems to be intact in amusics. Further studies are needed to pinpoint in more structural detail the neural source of the amusics’ pitch processing anomaly, which, according to the current findings, seems to lie outside the auditory cortex and beyond processing of isolated pitch features.

5. GENERAL DISCUSSION

5.1 A revised view of the sensory-memory phase in pitch processing

This thesis demonstrated the sophistication of musical pitch processing at the sensory-memory stage, implying automaticity and rapidity of neural analysis of musical features in the auditory cortex. First, multiple pitch objects are simultaneously encoded as sensory-memory representations, as indicated by the MMN elicitation to the infrequent combination of the simple or temporally complex features of multiple musical objects (Studies I and II). In other words, sensory memory may simultaneously hold auditory stimulus representations of simple as well as temporally complex pitch features. This capacity is possibly based on perceptual skills that are primitive, domain- and culture-general, i.e., available to any individual irrespective of their musical culture, and not restricted to a particular musical system or even to the auditory modality (Bregman, 1990).

Second, pitch is also processed according to properties that are characteristic of a particular musical system (i.e., they are music- and culture-specific or musical-system-specific; see Trainor, 2005), such as the Western equal-tempered musical-scale properties (Studies III and IV). In psychological terms, this implies that a musical context creates automatic expectations for incoming pitch objects without the need for the intervention of attention. A revised framework of pitch processing should therefore incorporate a processing stage of musical-scale pitch that occurs temporally in parallel with, rather than subsequently to, auditory object formation. A very recent comprehensive framework of music perception, which takes into account both the time course and the location of the related neural activity (Koelsch & Siebel, 2005; see Fig. 9), presupposes an initial stage of pitch processing, including feature extraction (taking place in the brainstem, thalamus and PAC from 10 to 100 ms), followed at 100-200 ms by the formation of coherent auditory objects in the supratemporal and the dorsolateral prefrontal cortices, and by the analysis of musical intervals or scale pitch relations. According to the authors, the neural underpinnings of this latter stage still need to be specified. The present thesis provides empirical evidence for the involvement of sensory-memory mechanisms in relational scale-pitch analysis (as illustrated by the shaded box dedicated to auditory sensory memory in Fig. 9). The following stages of musical pitch processing also addressed in this thesis include musical structure building from 200 to 400 ms and structural reanalysis and repair from 600 to 900 ms.

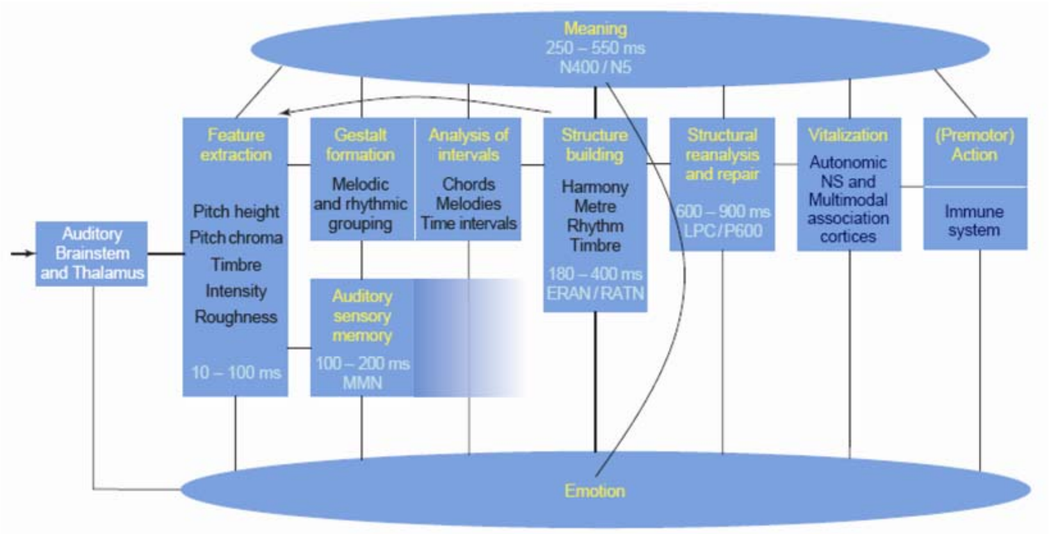

Figure 9. A neurocognitive model of music perception modified from Koelsch and Siebel (2005). Each box represents a separate stage of processing related to the different aspects of music. The shaded box dedicated to auditory sensory memory exemplifies the generalization of sensory-memory operations to other aspects of music than previously supposed, involving the analysis of the exact pitch relations and the musical-scale properties, as demonstrated by this thesis. Modified from Koelsch and Siebel (2005).

The electrophysiological results were valid for all subjects participating in the experiments irrespective of formal explicit education in music academies or conservatories, thus supporting the view about the importance of implicit musicality of the human brain (Koelsch et al., 2000; Bigand, 2003). Passive exposure seems sufficient to produce similar music-specific capacities of automatic processing of pitch according to the Western musical scale in non-musicians as well as in musicians (cf. also Krohn et al., in press). Alternatively, the present findings can be accounted for by the acoustic content of the sounds employed in the experiments. For instance, the unfamiliar pattern in Study III also contained more complex frequency ratios between the tone elements than the familiar pattern containing the quasi-integer ratios of the equal-tempered chromatic scale. Likewise, the neural response to mistuned tones within melodies in Study IV may be explained by the dissonance formed with the previous tones when they are maintained in sensory memory (on dissonance of melodic intervals, see Schellenberg & Trehub, 1994). However, the MMN-like response to the out-of-key pitches in Study IV is hard to account for solely on the basis of general perceptual principles; it is based more on the extraction of the invariant musical-scale properties of the previous melody context. This process could possibly be facilitated by music acculturation (for related MMN findings, see Koelsch et al., 1999; Tervaniemi et al., 2001). This interpretation is further supported by recent findings (Jacobsen et al., 2005) obtained with environmental sounds showing an MMN enhancement when deviants were presented among familiar standards relative to when the deviants were inserted in a context of unfamiliar standards.

The source modelling results of Study IV and those of the topographic analysis of Study III suggest that mechanisms extracting relational pitch information according to the musical scale might be located in the non-primary auditory cortex. Frontal neural circuits might also be involved, as could be inferred from previous source localization findings of the MMNm to simple abstract rule violations (Korzyukov et al., 2003), but additional studies with more spatially accurate brain-imaging techniques are required to solve this issue. Nevertheless, on the basis of the present data, it is reasonable to propose that relational pitch is extracted in the auditory cortex according to the long-term schemata of the Western musical system, thus strengthening the processing of incoming pitches when they match those schemata.

5.2 A neuronal model for musical-scale MMN

According to the model by Schröger (1997) described in the Introduction, the R´ representation of the invariance in the preceding stimulation is compared with the R representation of the encoded features of the incoming sound. Further, when these do not match, an MMN is generated. The frequency or pitch features of the repeated sounds are supposed to be encoded as sensory-memory traces in primary auditory regions, such as the HG, as well as in non-primary ones, like the cortex anterior and lateral to the HG, and the PT (Kropotov et al., 2000; Ulanovsky et al., 2003). This has been demonstrated in humans by a slow positive deflection termed the repetition positivity (RP; Haenschel et al., 2005) of the ERPs starting at 50 ms and lasting until 250 ms post-stimulus. According to the model by Näätänen (1984), the frequency feature of the repeated sounds inhibits the neuronal activity to that feature (as reflected by the RP) and releases from inhibition all other frequencies (as reflected by the MMN to the deviating stimuli). The comparison process seems to occur in separate neural circuits of auditory non-primary BA22 areas (Kropotov et al., 2000; Näätänen et al., 2005).

In Study IV, no specific sounds were repeated, but instead musical-system-specific pitch relations between the sounds. A modification of the model is thus called for. In this model, all of the relational pitch representations belonging to the diatonic musical scale would be inhibited, whereas all other pitch relations would be released from tonic inhibition. This process would not take place in the PAC, but rather in higher-level areas, such as the BA22, and would rely on backward connections between these higher-level areas and lower sensory regions responsible for simple feature encoding. These connections would allow a first comparison process between the incoming pitch features and the feature traces of the preceding sounds (called R` in the model by Schröger, 1997) and would be responsible for the extraction of a second-order feature trace (that can be termed

R``) of the musical scale relations. The R`` representations are reinforced by past experience of music. If the R` representations of the relations between the incoming pitches and the preceding sounds do not conform to the scale relations of Western tonal music extracted from the previous context and stored in long-term memory (the R`` representations), then an MMN would be generated within non-primary auditory areas such as BA22.

At the microscopic level, the complex comparison process described here could rely on modulatory backward neurons connected by N-methyl-D-aspartate (NMDA) receptors, which allow for contextual, divergent effects (Friston & Price, 2001). Supportive evidence comes from previous intracortical recordings and pharmacological manipulations in awake monkeys, demonstrating that competitive and noncompetitive NMDA antagonists selectively damp the MMN elicitation, but not the obligatory ERP components indexing initial sensory analysis of sounds in the PAC (Javitt et al., 1996; for similar results in humans obtained with intravenous injections of ketamine, an NMDA antagonist, see also Umbricht et al., 2000, 2002; Kreitschmann-Andermahr et al., 2001). Glutamate, a neurotransmitter opening the NMDA receptor channels, is involved in long-term potentiation (LTP), which, in turn, represents a model for brain plasticity (Buonomano & Merzenich, 1998). Thus, with the blocking of the NMDA channels, the formation of short-term memory traces and the corresponding activation of long-term representations consisting of reinforced synaptic connections are not possible. Electrophysiologically, the MMN is dampened or not elicited, and behaviourally, the assessment of pitch familiarity is impaired.

To sum up, the MMN can be elicited not only according to a feed-back mechanism, i.e., a comparison of the incoming stimulus with the short-term sensory memory traces formed during the experimental session, but also according to a feed-forward mechanism, involving a prediction of the incoming stimulus on the basis of musical-scale properties (cf. also Lopez et al., 2003).

5.3 Pitch space as a music-specific module

Music, as organized structured sounds leading to an aesthetic experience, is ubiquitous, i.e., it is found in all modern societies (Wallin et al., 2000). It is very common in the industrialized world for individuals to listen to music while driving or to be passively exposed to it in department stores. We also know that music has existed since the beginning of the human race. A bone flute was calculated to be from 43000 to 82000 years old (Huron, 2003). Estimations place music between 250000 and 50000 years of age (Huron, 2003). The centrality of music for human life is also proven by ethological evidence. For instance, Mekranoti Indians sing 2 h every day. Furthermore, mothers worldwide sing to their infants while providing care (Trehub & Trainor, 1998). Pragmatically, the wide use of music in modern society can be quantified by the financial resources devoted to it; e.g., in USA, the music industry is as powerful as the pharmaceutical industry, the former not having the self-evident effects of the latter on human health (Huron, 2003).

Due to this everyday and universal presence of music in all historical epochs and human societies, it is reasonable to hypothesize that the human brain has phylogenetically developed mechanisms to automatically extract other aspects of music besides isolated pitch already at the level of the auditory cortex. In other words, the modulation of pitch encoding by musical context shown in the present thesis may derive from innate predispositions of the human brain for relational pitch processing or may be an outcome of neural skills acquired from music exposure. Neural predispositions for musical pitch have been postulated in the modularity theory for music processing (originally by Fodor, 1983, and later modified for the music domain by Peretz & Coltheart, 2003). According to this theory, separate modules exist for each sensory, motor or cognitive function, with their own procedures and knowledge base. In particular, a mental function may be defined as modular if it has one of the following properties: rapidity of operation, automaticity, domain-specificity, informational encapsulation, neural specificity or innateness (Peretz & Coltheart, 2003). Each module can be subdivided into smaller units, or submodules, serving for the analysis of particular aspects of a cognitive domain. If this hypothesis is accepted, then modularity would imply that the auditory cortex automatically processes sounds not only according to their simple acoustic features but also integrating and cognitively processing them by, e.g., comparison with previously occurring ones. Moreover, this processing could be compared to a reflex, being automatic and hence allowing fast recognition and identification of sound events. Music has been proposed (Peretz, 2003; Peretz & Coltheart, 2003) to comprise a set of neurally isolable processing components, such as melody processing, which can be selectively impaired after brain lesions, thus determining a specific music deficit. In this sense, music can be considered a modular cognitive architecture, comparable to language (Peretz & Coltheart, 2003).

We may hence speculate, on the basis of the current data showing the automaticity of pitch processing according to the Western musical scale and its independence from top-down attentional modulation (Studies III and IV), about a possible status of musical pitch as a mental submodule. In line with this view, Study III showed no MMN amplitude difference between subjects with a formal education in music and musically uneducated subjects in pitch discrimination within a conventional musical scale or an artificial scale context (however, a shorter-latency MMN to pattern changes in musicians than in non-musicians was observed, proving the musicians' enhanced speed of processing for temporally complex sounds). These findings may derive from a predisposition for pitch processing according to musical rules, triggered by long-term exposure to music. Behavioural studies with children show predispositions for perceptual skills related to musical scale pitch structure. For instance, 5-7 months old infants performed better when a mistuning was inserted in the Western major scale or in the Western augmented scale, being both based on the semitone interval, than when the mistuning was inserted in a non-Western (Indonesian pelog) scale, not built on the single semitone interval (Lynch & Eilers, 1992)[5]. These results suggest an innate preference for the semitone relation in the musical scale. As put by Jackendoff & Lerdahl (2006), "although general Gestalt principles of proximity and good continuation lie behind a preference for small melodic intervals, they do not explain why the particular intervals of the whole step and half step are so prevalent in melodic organization across the musical idioms of the world. We conclude that the mind/brain must contain something more specialized than psychoacoustic principles that accounts for the existence and organization of tonality" (p. 20). The congenital amusics studied in this thesis (Study VI) showed attenuated or abolished brain responses to pitch changes equal to or smaller than a semitone (half step), and abnormally enhanced responses to larger musical intervals. These findings coupled with the amusics' cognitive difficulties with music may be interpreted as indirectly probing the relevance of inborn neural pitch-discrimination mechanisms for the normal acquisition of music competence as well as the centrality of the semitone interval in Western tonal music.

Alternatively, the automatization of musical-scale pitch processing could be a consequence of continuous passive (or active) exposure to music since childhood (for a discussion about the consequence of continuous practice on the development of fast, accurate and automatic skills, see Prasher, 1998). Such a view is based on the assumption that music processing does not depend on specific cognitive or neural properties, but involves properties also shared by other domains (Brust, 2003; Altenmüller, 2003). Consequently, according to this view, the brain may respond in a fully adaptive and changeable manner according to the listening strategies involved. In brain measurements, it would imply large differences between individuals in the brain mechanisms and structures for pitch processing according to the quantity of music exposure that they had received in their life. However, this view is only marginally supported by empirical evidence (Altenmüller et al., 1997; Schuppert et al., 2000).

In contrast, most brain-imaging findings suggest innate predispositions for pitch processing in the brain, mirrored by the existence of specific brain structures devoted to this processing (Peretz, 2003). These structures are proposed to be

---

[5] One-year old children, instead, showed a more accurate recognition of mistuning when embedded in the Western major scale than when embedded in the less familiar augmented and pelog scales.

sufficiently plastic to be at least partially shaped by the experience of sounds within the anatomo-physiological constraints of the human brain (Rauschecker, 1999; Schlaug, 2001; Pantev et al., 2003; Brattico & Tervaniemi, 2006; Brattico, 2006). This view is supported by electrophysiological evidence. For instance, sound feature analysis in the primary and non-primary auditory cortex is reinforced by past music experience, as demonstrated by musicians' enhanced obligatory N1 and P2 responses to the sounds of their own instrument (Pantev et al., 1998; 2001; Shahin et al., 2003; for relevant studies on musical training effects on children, see Shahin et al., 2004; Fujioka et al., 2006).

In line with this, Study V demonstrated that a 1-h laboratory training of a specific pitch differentiates the tonotopic auditory-cortex response to this pitch from responses to adjacent pitches. These findings reveal the neural adaptability of pitch representations within the anatomical constraints of the auditory cortex even after a very short sound exposure (Rauschecker, 1999; Schlaug, 2001; Pantev et al., 2003). The complementary role of cortical plasticity and innate predispositions for musical pitch processing needs to be further investigated with experiments on infants (subjects with minimal music exposure) and musicians (subjects with maximal music exposure). The prediction is that the ability to perceive sounds according to a musical scale is the result of the interplay between the inherited perceptual and cognitive predispositions and the acquired knowledge of a particular musical system (Trainor, 2005; Brattico & Tervaniemi, 2006; for informative views on the topic, cf. Peretz & Zatorre, 2003, reviewed in Brattico, 2006).

6 CONCLUSIONS

During the sensory-memory stage indexed by the MMN and mostly occurring at the level of the non-primary auditory cortex, the acoustic input, encoded into a neural representation according to its pitch features, is compared with the stored short-term and long-term traces of complex and abstract pitch invariances of preceding sounds. This process occurs before the intervention of attention to, and conscious awareness of, the sounds. Moreover, the process is automatic and not mediated by attentional control; inevitably, the music-specific schemata acquired through the long-term exposure to music affect the pre-attentive processing of pitch.

In the sensory-memory stage, the automatic memory trace formation for the pitch relations between musical sounds is performed at a considerable level of sophistication. This is proven by the MMN elicitation to infrequent combinations of multiple temporally complex tonal patterns (Study II) or of several spectrally complex features of concurrent musical sounds (Study I). Finally, the automatic sensory-memory trace formation for pitch relations is affected by previous knowledge of the musical sounds of a particular culture. This was demonstrated by the MMN enhancement to pitch changes within familiar musical-scale patterns (Study III) or by the elicitation of an MMN to a pitch incongruity in a melody (Study IV).

Specific neural mechanisms are devoted to process pitch relations. When those mechanisms are impaired, then music cognition and appreciation are not possible, as in the congenital amusic subjects investigated in Study VI. The fast plastic cortical processes leading to the modification of the feature traces to pitch were addressed in Study V.

Musical experience is thus based on early, largely automatic, functions of the central auditory system that dynamically store, in a way that is affected by past experience, the sequential pitches of sounds as integrated regularities of the auditory environment.

APPENDIX: GENERAL METHODS

Subjects

Eighty-two subjects (all right-handed except for one left-handed subject and one ambidextrous subject in Study V) participated in the studies included in this thesis. The subjects participating in Study I (right-handed; *N*=12; mean age 27 years; 4 males; 4 amateur musicians), Study II (right-handed; *N*=10; mean age 20.6 years; 2 males), Study IV (right-handed; *N*=9; mean age 23 years; 5 males) and Study V (1 left-handed, 1 ambidextrous; *N*=18; mean age 31 years; 6 males) were healthy volunteers with normal hearing and with no record of neurological disease. They gave formal or informal consent and received monetary compensation for their participation (except for the subjects of Study V who were all laboratory personnel).

Studies III and VI comprised special groups of subjects defined according to their musical skills or education. In Study III, subjects were separated into two groups according to their level of musical expertise. The 'musicians' (*N*=10; mean age, 21.3 years; 2 males) were selected on the basis of their daily musical training and their attending a formal music school. All of the musicians had started their musical education at the age of 5-7 years. The controls were termed 'non-musicians', since they had no formal musical education (*N*=10; mean age 23.3 years; 4 males). The MMN recorded in musicians was compared with that recorded in non-musicians.

The subjects participating in Study VI were individuals classified as congenital amusics (*N*=8; mean age 58 years; 2 males; mean education 17 years) according to a series of tests designed to assess the presence and specificity of musical disorders, termed the Montreal Battery of Evaluation of Amusias (or MBEA; Ayotte et al., 2002; Peretz et al., 2003). The control group included subjects (N=10; mean age 59 years; 2 males; mean education 17 years) with neither musical education nor musical impairment according to their scores on the MBEA.

Studies I, II and III were approved by the Ethics Committee of the Department of Psychology, University of Helsinki, Study V by the Ethics Committee of Baycrest Centre for Geriatric Care and Studies IV and VI by the Ethics Committee of the University of Montreal.

Stimulation paradigms

Studies II, III and V used tone sequences consisting of sinusoidal tones. The other studies employed spectrally complex sounds, mainly with a piano-like timbre. Study I also included sounds with a violin-like timbre. Moreover, the belongingness of the sound stimuli to the equal-tempered musical scale and their familiarity were manipulated according to the experimental purposes. For instance, to facilitate discrimination between the three tonal patterns used in Study II, each of the patterns was built with pitches belonging to three different diatonic scales (see Fig. 4A). In Study V, by contrast, the frequencies of the stimuli were chosen to not correspond to the fundamental frequencies of the pitches of the diatonic scale to avoid interference of previous musical knowledge with the short-term laboratory perceptual training. In Studies III and IV, unfamiliar tonal patterns or melodies (see Fig. 5A and Fig. 6A, respectively) specifically composed for the experiments were preferred over familiar ones (e.g., 'Frere Jacques') to test expectations for pitch based on the learned schemata of Western tonal music rather than the veridical expectations based on the episodic memories of the pitches of a specific melody (cf. Bharucha & Todd, 1989; Bigand & Dusinberre, 1993).

Experimental procedures

The EEG recordings lasted approximately 2-3 h for each experimental session. During all EEG recordings, except for in Study VI, the subjects were presented with the sounds through earphones while performing a primary task not directly involving sound processing such as performing a demanding visual task (Study I), watching a silent subtitled film (Studies II, IV and V) or reading a book (Study III). The primary task was purposively designed to direct subjects' attention away from the sounds. Study VI instead comprised a primary task in which the subjects listened to tone patterns and tried to discriminate the deviating pitch within them.

Some studies also included conditions in which subjects were required to perform a task related to the sounds, such as the rating of the congruousness or incongruousness of the melodic stimulation (Study IV), or one unrelated to the sounds, such as a visual task involving the memorization of spatial loci on the computer screen (Study I), performed by subjects while their EEG was recorded. In Studies I and V, additional behavioural sessions were also conducted (besides EEG recordings), aimed at training subjects in frequency discrimination or at testing their skill in identifying a target sound from a sound pair, respectively.

EEG recordings and data analysis

The brain activity of healthy human volunteers was recorded with the EEG (10 scalp electrodes for Study II) and with high-resolution EEG (HR-EEG; electrode caps with 64 channels – 60-64 of which were used for scalp recordings – for Studies IV, V and VI, and with 32 channels – 30 of which were scalp electrodes – for Studies I and III). The EEG allows one to follow the exact time course of neuronal activation during music perception with a millisecond resolution. In Studies I, II, III and V, Synamps amplifiers (Neuroscan, El Paso, Texas, USA), and in Studies IV and VI, InstEP amplifiers (Ottawa, Canada) were employed. The averaging and quantification of ERP responses were always performed with Edit software (Neuroscan, El Paso, Texas, USA).

In Study V, eye-movement artefacts were corrected by extracting the ocular source components with the Brain Electric Source Analysis (BESA) software (Picton et al., 2000c). In Studies IV and VI, eye-movement artefacts were corrected by analyzing the dynamic regression between the EOG and EEG in the frequency domain (Woestenburg et al., 1983). In the rest of the studies, eye and other movement-related artefacts were rejected from the averaging procedure by applying voltage limits.

In most studies, the amplitudes and latencies of the ERP components were quantified for each subject and experimental condition from the values recorded at each electrode. In Study V, the global field power (GFP) of the N1 response was first calculated. The GFP approximately corresponds to the total strength of cortical activation, that is, the number of synchronously active neurons; when more neurons fire, the GFP value increases (Pantev et al., 1999; Skrandies, 2003). This reference-independent measure of the potential field strength, which represents the combined activity over a group of selected EEG channels, is computed as the mean potential deviation of all electrodes in the recording array. It is based on the assumption that synchronous neuronal activation yields potential fields with pronounced peaks and troughs, and the resulting steep gradients are associated with the GFP maxima. In all studies, the amplitudes of the ERP components of interest were then calculated as the mean value over a time interval around the ERP peak, individually determined or determined from the grand-average waveforms.

For Studies I, III and IV, the topographic maps of the ERP voltage amplitudes on the scalp surface were also calculated to visualize the scalp distribution of the ERP components at a given instant in time. Differences between the conditions or experimental groups in the shapes of the topographic maps could indicate that different combinations of current sources were active.

Electric source modelling

The locus and timing of the current sources that contribute to the ERP scalp recordings can be approximately inferred by solving an inverse problem with no unique solution. The commercial software BESA, used in Studies IV and VI, offers algorithms to reach this solution. The electric field that is measurable from the scalp originates from the pyramidal neurons in the cortex; current flow renders the surface more positive relative to the depth. At a distance from an activated region, the electric field can be illustrated as an equivalent dipole (Picton et al., 1995). Spatiotemporal source analysis involves the modelling of a source, i.e., the compound electric activity of a particular brain area. In Study VI, regional source models were employed, consisting of three orthogonal dipoles which can image the current flow of a particular brain region in any direction (Picton et al., 1995).

In Study IV, a distributed rather than a local approach to source analysis was adopted since the measured electric field to be modelled could have been generated from a large patch rather than a small patch of the cortex. Particularly, L2 minimum-norm current estimates (MCE) were calculated, providing a distributed current image at each time sample on the basis of the recorded voltage distribution, with the constraint that it should be equal to the smallest amplitude of the overall activity (Hämäläinen & Ilmoniemi, 1984; 1994). As an additional constraint, the source analysis was limited to the sulci and gyri of the cortex, accounting for most of the recorded signal due to their open-field structure and proximity to the scalp (Dale & Sereno, 1993).

Statistical analyses

In most studies, the presence of an ERP component was first tested by comparing its amplitude at a specific electrode with 0 μV (corresponding to baseline) in two-tailed *t*-tests. This was done as a preliminary step in order to avoid investigating noise-related peaks of the ERPs before the comparisons between the conditions and groups.

In most studies, the ERPs elicited by the deviant stimuli overlapped the descending portion of the N1 associated with the standard stimuli, being separated in the negative direction from the standard ERPs at around 100 ms. This additional negativity corresponds to the notion of the MMN. In some studies, the presence of a difference between the ERP elicited by the deviant sounds and that elicited by the standard sounds, corresponding to the MMN or the P3 components, was assessed by using two-tailed *t*-tests (when the comparison only included the ERPs to deviant and standard sound categories at single electrodes) or repeated measures analysis of variance (ANOVA, when the comparison involved more complex factors). ANOVA was also employed to compare differences between experimental conditions with each other, and between ERP scalp distributions for each group and/or condition with each other. In the latter case, the ERP amplitudes recorded at several electrodes were considered, and in Studies I and III also normalized (by dividing the amplitude at each electrode by the square root of the sum of the squared across-subjects averaged amplitudes; McCarthy & Wood, 1985; Picton et al., 2000b). Such a normalization permitted one to isolate possible differences in scalp distributions from large ERP amplitude differences between conditions.

To examine the directions of the effects obtained in the ANOVAs, further post-hoc tests were conducted such as Fisher's least-significant-difference (LSD) test or the Newman-Keuls test. In addition, the Greenhouse-Geisser correction was applied to control for type I errors associated with inhomogeneity of variance when studying factors with more than two levels. For all statistical tests, the alpha level was chosen to correspond to $p < 0.05$.

REFERENCES

Altenmüller, E.O. (2003). How many music centers are in the brain? In I. Peretz & R. Zatorre (Eds.), *The Cognitive Neuroscience of Music* (pp. 346-357). Oxford - New York: Oxford University Press.

Altenmüller, E.O., Gruhn, W., Parlitz, D., & Kahrs, J. (1997). Music learning produces changes in brain activation patterns: A longitudinal DC-EEG-study. *International Journal of Arts Medicine*, *5*, 28--34.

ANSI (1973). *American National Standard Psychoacoustical Terminology*: American National Standards Institute, Inc.

American Standards Association (1960). *Acoustical Terminology SI* (Vol. 1-1960). New York: American Standards Association.

Amenedo, E., & Escera, C. (2000). The accuracy of sound duration representation in the human brain determines the accuracy of behavioural perception. *European Journal of Neuroscience*, *12*, 2570-2574.

Atienza, M., Cantero, J.L., & Dominguez-Marin, E. (2002). The time course of neural changes underlying auditory perceptual learning. *Learning and Memory*, *9*, 138-150.

Atienza, M., Cantero, J.L., & Escera, C. (2001). Auditory information processing during human sleep as revealed by event-related brain potentials. *Clinical Neurophysiology*, *112*, 2031-2045.

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia - A group study of adults afflicted with a music-specific disorder. *Brain*, *125*, 238-251.

Beisteiner, R., Erdler, M., Mayer, D., Gartus, A., Edward, V., Kaindl, T., Golaszewski, S., Lindinger, G., & Deecke, L. (1999). A marker for differentiation of capabilities for processing of musical harmonies as detected by magnetoencephalography in musicians. *Neuroscience Letters*, *277*, 37-40.

Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature*, *436*, 1093-1094.

Berry, D.C. (Ed.). (1997). *How Implicit is Implicit Learning?* Oxford: Oxford University Press.

Besson, M., Faita, F., & Requin, J. (1994). Brain waves associated with musical incongruities differ for musicians and non-musicians. *Neuroscience Letters*, *168*, 101-105.

Besson, M., & Faita, F. (1995). An event-related potential (ERP) study of musical expectancy: Comparison of musicians with nonmusicians. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1278-1296.

Besson, M., & Macar, F. (1987). An event-related potential analysis of incongruity in music and other non-linguistic contexts. *Psychophysiology*, *24*, 14-25.

Besson, M., & Schön, D. (2003). Comparison between language and music. In I. Peretz & R. Zatorre (Eds.), *The Cognitive Neuroscience of Music* (pp. 269-293). New York, NY: Oxford University Press.

Bharucha, J.J., & Stoeckig, K. (1986). Reaction time and musical expectancy: priming of chords. *Journal of Experimental Psychology: Human Perception & Performance*, *12*, 403-410.

Bharucha, J.J., & Todd, P. (1989). Modeling the perception of tonal structures with neural nets. *Computer Music Journal*, *13*, 44-53.

Bigand, E. (2003). More about the musical expertise of musically untrained listeners. *Annals of the New York Academy of Sciences*, *999*(304-312).

Bigand, E., & Dusinberre, D. (1993). Contributions of music to research on human auditory cognition. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 231-277). New York: Clarendon Press - Oxford University Press.

Bigand, E., & Parncutt, R. (1999). Perceiving musical tension in long chord sequences. *Psychological Research*, *62*, 237-254.

Bigand, E., & Poulin-Charronnat, B. (2006). Are we “experienced listeners”? A review of the musical capacities that do not depend on formal musical training. *Cognition*, *100*, 100-130.

Boltz, M. (1989). Rhythm and ‘good endings’: Effects of temporal structure on tonality judgments. *Perception & Psychophysics*, *46*, 9-17.

Brattico, E. (2006). Review of the book *The Cognitive Neuroscience of Music* edited by I. Peretz and R. Zatorre. *Musicae Scientiae*, *10*, 130-137.

Brattico, E., Näätänen, R., Verma, T., Välimäki, V., & Tervaniemi, M. (2000). Processing of musical intervals in the central auditory system: An event-related potential (ERP) study on sensory consonance. Paper presented at the Sixth International Conference on Music Perception and Cognition, Keele (UK).

Brattico, E., & Tervaniemi, M. (2006). Musical creativity and the human brain. In I. Deliege & G.A. Wiggins (Eds.), *Musical Creativity: Current Research in Theory and Practice* (pp. 290-321). London: Psychology Press.

Bregman, A.S. (1990). *Auditory Scene Analysis: The perceptual organization of sound*. Cambridge (MA): MIT Press.

Broadbent, D.E. (1958). *Perception and Communication*: Elsevier Science Ltd.

Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Grosshirnrinde*. Leipzig: Barth.

Brust, J.C.M. (2003). Music and the neurologist: A historical perspective. In I. Peretz & R. Zatorre (Eds.), *The Cognitive Neuroscience of Music* (pp. 181-191). Oxford: Oxford University Press.

Buonomano, D.V., & Merzenich, M.M. (1998). Cortical plasticity: From synapses to maps. *Annual Review of Neuroscience*, *21*, 149-186.

Butler, R.A. (1968). Effects of changes in stimulus frequency and intensity on habituation of the human vertex potential. *Journal of the Acoustical Society of America*, *44*, 945-950.

Caclin, A., Brattico, E., Tervaniemi, M., Näätänen, R., Morlet, D., Giard, M.-H., & McAdams, S. (in press). Separate neural processing of timbre dimensions in auditory sensory memory. *Journal of Cognitive Neuroscience*.

Cariani, P.A., & Delgutte, B. (1996). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *Journal of Neurophysiology*, *76*, 1698-1716.

Cedolin, L., & Delgutte, B. (2005). Pitch of complex tones: Rate-place and interspike interval representation in the auditory nerve. *Journal of Neurophysiology*, *94*, 347-362.

Cheour, M., Leppänen, P.H.T., & Kraus, N. (2000). Mismatch negativity as a tool for investigating auditory discrimination and sensory memory in infants and children. *Clinical Neurophysiology*, *111*, 4-16.

Chomsky, N. (1957). *Syntactic Structures*. The Hague, the Netherlands: Mouton and Co.

Cohen, A.J. (2000). Development of tonality induction: Plasticity, exposure, and training. *Music Perception*, *17*, 437-459.

Coull, J.T. (1998). Neural correlates of attention and arousal: insights from electrophysiology, functional neuroimaging and psychopharmacology. *Progress in Neurobiology*, *55*, 343-361.

Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin*, *96*, 341-370.

Cowan, N. (1995). *Attention and Memory: An integrated framework* (Vol. N. 26). New York: Oxford Psychology Series.

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, & Cognition*, *19*, 909-921.

Crowder, R.G. (1993). Short-term memory: Where do we stand? *Memory and Cognition*, *21*(142-145).

Cuddy, L.L., Cohen, A., & Mewhort, D.J. (1981). Perception of structure in short melodic sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 869-883.

Dale, A.M., & Sereno, M.I. (1993). Improved localization of cortical activity by combining EEG and MEG with fMRI with cortical surface reconstruction: A linear approach. *Journal of Cognitive Neuroscience*, *5*, 162-176.

Dalla Bella, S., & Peretz, I. (2003). Congenital amusia interferes with the ability to synchronize with music. *Annals of the New York Academy of Sciences*, *999*, 166-169.

deCharms, R.C., Blake, D.T., & Merzenich, M.M. (1998). Optimizing sound features for cortical neurons. *Science*, *280*, 1439-1443.

Deutsch, D. (1982). The processing of pitch combinations. In D. Deutsch (Ed.), *The Psychology of Music* (pp. 271-316). New York: Academic Press.

Diesch, E., & Luce, T. (1997). Magnetic fields elicited by tones and vowel formants reveal tonotopy and nonlinear summation of cortical activation. *Psychophysiology*, *34*, 501-510.

Donchin, E. (1981). Surprise! ... Surprise? *Psychophysiology*, *18*, 493-513.

Dowling, W.J. (1982). Melodic information processing and its development. In D. Deutsch (Ed.), *The Psychology of Music* (pp. 413-429). New York - London: Academic Press.

Drake, C., & Bertrand, D. (2003). The quest for universals in temporal processing in music. In I. Peretz & R. Zatorre (Eds.), *The Cognitive Neuroscience of Music* (pp. 21-31). Oxford - New York: Oxford University Press.

Drayna, D., Manichaikul, A., de Lange, M., Snieder, H., & Spector, T. (2001). Genetic correlates of musical pitch recognition in humans. *Science*, *291*, 1969-1972.

Elberling, C., Bak, C., Kofoed, B., Lebech, J., & Saermark, K. (1982). Auditory magnetic fields: Source localization and ‘tonotopical organization’ in the right hemisphere of the human brain. *Scandinavian Audiology*, *11*, 61-65.

Escera, C., Alho, K., Schröger, E., & Winkler, I. (2000). Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiology & Neuro-Otology*, *5*, 151-166.

Fishman, Y.I., Reser, D.H., Arezzo, J.C., & Steinschneider, M. (1998). Pitch vs. spectral encoding of harmonic complex tones in primary auditory cortex of the awake monkey. *Brain Research*, *786*, 18-30.

Fodor, J. (1983). *The Modularity of Mind*. Cambridge (MA): MIT Press.

Foxton, J.M., Brown, A.C.B., Chambers, S., & Griffiths, T.D. (2004a). Training improves acoustic pattern perception. *Current Biology*, *14*, 322-325.

Foxton, J.M., Dean, J.L., Gee, R., Peretz, I., & Griffiths, T.D. (2004b). Characterization of deficits in pitch perception underlying ‘tone deafness’. *Brain*, *127*, 801-810.

Friedman, D., Cycowicz, Y.M., & Gaeta, H. (2001). The novelty P3: An event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience and Behavioral Reviews*, *25*, 355-373.

Friedman-Hill, S.R., Robertson, L.C., & Treisman, A. (1995). Parietal contributions to visual feature binding: Evidence from a patient with bilateral lesions. *Science*, *269*, 853-855.

Friston, K.J., & Price, C.J. (2001). Dynamic representations and generative models of brain function. *Brain Research Bulletin*, *54*, 275-285.

Fujioka, T., Ross, B., Kakigi, R., Pantev, C., & Trainor, L.J. (2006). One year of musical training affects development of auditory cortical-evoked fields in young children. *Brain*, *129*, 2593-2608.

Fujioka, T., Trainor, L.J., Ross, B., Kakigi, R., & Pantev, C. (2004). Musical training enhances automatic encoding of melodic contour and interval structure. *Journal of Cognitive Neuroscience*, *16*, 1010-1021.

Galván, V.V., & Weinberger, N.M. (2002). Long-term consolidation and retention of learning-induced tuning plasticity in the auditory cortex of the guinea pig. *Neurobiology of Learning and Memory*, *77*, 78-108.

Giard, M.H., Lavikainen, J., Reinikainen, K., Perrin, F., Bertrand, O., Pernier, J., & Näätänen, R. (1995). Separate representation of stimulus frequency, intensity, and duration in auditory sensory memory: An event-related potential and dipole-model analysis. *Journal of Cognitive Neuroscience*, *7*, 133-143.

Giard, M.H., Perrin, F., Bertrand, O., Pernier, J., & Bouchet, P. (1990). Brain generators implicated in the processing of auditory stimulus deviance: A topographic event-related potential study. *Psychophysiology*, *27*, 627-640.

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

Gomes, H., Bernstein, R., Ritter, W., Vaughan, Jr., H.G., & Miller, J. (1997). Storage of feature conjunctions in transient auditory memory. *Psychophysiology*, *34*, 712-716.

Griffiths, T.D. (2003a). Functional imaging of pitch analysis. *Annals of the New York Academy of Sciences*, *999*, 40-49.

Griffiths, T.D. (2003b). The neural processing of complex sounds. In I. Peretz & R. Zatorre (Eds.), *The Cognitive Neuroscience of Music* (pp. 168-177). Oxford - New York: Oxford University Press.

Griffiths, T.D., Buchel, C., Frackowiak, R.S., & Patterson, R.D. (1998). Analysis of temporal structure in sound by the human brain. *Nature Neuroscience*, *1*, 422-427.

Griffiths, T.D., & Warren, J.D. (2002). The planum temporal as a computational hub. *Trends in Neurosciences*, *25*, 348-353.

Griffiths, T.D., & Warren, J.D. (2004). What is an auditory object? *Nature Reviews Neuroscience*, *5*, 887-892.

Griffiths, T.D., Warren, J.D., Scott, S.K., Nelken, I., & King, A.J. (2004). Cortical processing of complex sound: A way forward? *Trends in Neurosciences*, *27*, 181-185.

Haenschel, C., Vernon, D.J., Dwivedi, P., Gruzelier, J.H., & Baldeweg, T. (2005). Event-related brain potential correlates of human auditory sensory memory-trace formation. *Journal of Neuroscience*, *25*, 10494-10501.

Halgren, E., Baudena, P., Clarke, J.M., Heit, G., Liégeois, C., Chauvel, P., & Musolino, A. (1995). Intracerebral potentials to rare target and distractor auditory and visual stimuli: I. Superior temporal plane and parietal lobe. *Electroencephalography and Clinical Neurophysiology*, *94*, 191-220.

Hall, D.A., Hart, H.C., & Johnsrude, I.S. (2003). Relationships between human auditory cortical structure and function. *Audiology and Neuro-Otology*, *8*, 1-18.

Hari, R., Pelizzone, M., Mäkelä, J.P., Hällström, J., Leinonen, L., & Lounasmaa, O.V. (1987). Neuromagnetic responses of the human auditory cortex to on- and offsets of noise bursts. *Audiology*, *26*, 31-43.

Hauser, M.D., & McDermott, J. (2003). The evolution of the music faculty: A comparative perspective. *Nature Neuroscience*, *6*, 663-668.

Hebb, D.O. (1949). *The organization of behavior*. New York: Wiley.

Henderson, J.M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, *50*, 243-271.

Huotilainen, M., Kujala, A., Alku, P., & Näätänen, R. (2001). Long-term memory traces facilitate short-term memory formation in audition. *Neuroscience Letters*, *310*, 133-136.

Huron, D. (2003). Is music an evolutionary adaptation? In I. Peretz & R. Zatorre (Eds.), *The Cognitive Neuroscience of Music* (pp. 57-75). Oxford - New York: Oxford University Press.

Hyde, K., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science*, *15*, 356-360.

Hämäläinen, M.S., & Ilmoniemi, R.J. (1984). *Interpreting Measured Magnetic Fields of the Brain: Minimum Norm Estimates of Current Distributions* (Vol. Technical Report TKK-F-A559). Helsinki: Helsinki University of Technology.

Hämäläinen, M.S., & Ilmoniemi, R.J. (1994). Interpreting magnetic fields of the brain: Minimum norm estimates. *Medical & Biological Engineering & Computing*, *32*, 35-42.

Jackendoff, R., & Lerdahl, F. (2006). The capacity for music: What is it, and what's special about it? *Cognition*, *100*, 33-72.

Jacobsen, T., & Schröger, E. (2001). Is there pre-attentive memory-based comparison of pitch? *Psychophysiology*, *38*, 723-727.

Jacobsen, T., Schröger, E., Winkler, I., & Horvath, J. (2005). Familiarity affects the processing of task-irrelevant auditory deviance. *Journal of Cognitive Neuroscience*, *17*, 1704-1713.

Janata, P. (1995). ERP measures assay the degree of expectancy violation of harmonic contexts in music. *Journal of Cognitive Neuroscience*, *7*, 153-164.

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. *Neuroscience Letters*, *290*, 101-104.

Javitt, D.C., Grochowski, S., Shelley, A.M., & Ritter, W. (1998). Impaired mismatch negativity (MMN) generation in schizophrenia as a function of stimulus deviance, probability, and interstimulus/interdeviant interval. *Electroencephalography and Clinical Neurophysiology*, *108*, 143-153.

Javitt, D.C., Steinschneider, M., Schroeder, C.E., & Arezzo, J.C. (1996). Role of cortical N-methyl-D-aspartate receptors in auditory sensory memory and mismatch negativity generation: Implications for schizophrenia. *Proceedings of the National Academy of the United States of America*, *93*, 11962-11967.

Johnson, Jr., R., & Baron, J.C. (Eds.). (1995). *Handbook of Neuropsychology* (Vol. 10). Amsterdam: Elsevier.

Jäncke, L., Mirzazade, S., & Shah, N.J. (1999). Attention modulates activity in the primary and the secondary auditory cortex: a functional magnetic resonance imaging study in human subjects. *Neuroscience Letters*, *266*, 125-128.

Jääskeläinen, I.P., Ahveninen, J., Bonmassar, G., Dale, A.M., Ilmoniemi, R.J., Levänen, S., Lin, F.-H., May, P., Melcher, J., Stufflebeam, S., Tiitinen, H., & Belliveau, J.W. (2004). Human posterior auditory cortex gates novel sounds to consciousness. *Proceedings of the National Academy of Sciences*, *101*, 6809-6814.

Kaas, J.H., & Hackett, T.A. (1998). Subdivisions of auditory cortex and levels of processing in primates. *Audiology and Neuro-Otology*, *3*, 73-85.

Kaas, J.H., & Hackett, T.A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences*, *97*, 11793-11799.

Kaernbach, C. (2004). Auditory sensory memory and short-term memory. In C. Kaernbach, E. Schröger, & H. Mueller (Eds.), *Psychophysics beyond Sensation - Laws and Invariants of Human Cognition* (pp. 331-348). Mahwah, NJ: Lawrence Erlbaum Associates.

Kahle, W., & Frotscher, M. (2003). *Nervous System and Sensory Organs*. Stuttgart - New York: Thieme.

Kalmus, H., & Fry, D.B. (1980). On tune deafness (dysmelodia): Frequency, development, genetics and musical background. *Annals of Human Genetics*, *43*, 369-382.

Kandel, E.R., Schwartz, J.H., & Jessel, T.M. (Eds.). (2000). *Principles of Neural Science* (Fourth edition). New York: McGraw-Hill.

Kane, N.M., Curry, S.H., Rowlands, C.A., Manara, A.R., Lewis, T., Moss, T., Cummings, B.H., & Butler, S.R. (1996). Event-related potentials-neurophysiological tools for predicting emergence and early outcome from traumatic coma. *Intensive Care Medicine*, *22*, 39-46.

Koelsch, S., Gunter, T.C., Friederici, A.D., & Schröger, E. (2000). Brain indices of music processing: “Non-musicians” are musical. *Journal of Cognitive Neuroscience*, *12*, 520-541.

Koelsch, S., & Mulder, J. (2002). Electric brain responses to inappropriate harmonies during listening to expressive music. *Clinical Neurophysiology*, *113*, 862-869.

Koelsch, S., Schröger, E., & Gunter, T.C. (2002a). Music matters: Preattentive musicality of the human brain. *Psychophysiology*, *39*, 38-48.

Koelsch, S., Schmidt, B.-H., & Kansok, J. (2002b). Effects of musical expertise on the early right anterior negativity: An event-related brain potential study. *Psychophysiology*, *39*, 657-663.

Koelsch, S., Schröger, E., & Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *NeuroReport*, *10*, 1309-1313.

Koelsch, S., & Siebel, W.A. (2005). Towards a neural basis of music perception. *Trends in Cognitive Sciences*, *9*, 578-584.

Kolb, B., & Whishaw, I.Q. (1998). Brain plasticity and behavior. *Annual Review of Psychology*, *49*, 43-64.

Korzyukov, O., Alho, K., Kujala, A., Gumenyuk, V., Ilmoniemi, R.J., Virtanen, J., Kropotov, J., & Näätänen, R. (1999). Electromagnetic responses of the human auditory cortex generated by sensory-memory based processing of tone-frequency changes. *Neuroscience Letters*, *276*, 169-172.

Korzyukov, O.A., Winkler, I., Gumenyuk, V.I., & Alho, K. (2003). Processing abstract auditory features in the human auditory cortex. *NeuroImage*, *20*, 2245-2258.

Kraut, A.G., & Smothergill, D.W. (1978). A two-factor theory of stimulus-repetition effects. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 191-197.

Kreitschmann-Andermahr, I., Rosburg, T., Demme, U., Gaser, E., Nowak, H., & Sauer, H. (2001). Effect of ketamine on the neuromagnetic mismatch field in healthy humans. *Brain Research Cognitive Brain Research*, *12*, 109-116.

Krohn, K., Brattico, E., Valimäki, V., & Tervaniemi, M. (in press). Neural representations of hierarchical scale pitch structure. *Music Perception*.

Kropotov, J.D., Alho, K., Näätänen, R., Ponomarev, V.A., Kropotova, O.V., Anichkov, A.D., & Nechaev, V.B. (2000). Human auditory-cortex mechanisms of preattentive sound discrimination. *Neuroscience Letters*, *280*, 87-90.

Kropotov, J.D., Näätänen, R., Sevostianov, A.V., Alho, K., Reinikainen, K., & Kropotova, O.V. (1995). Mismatch negativity to auditory stimulus change recorded directly from the human temporal cortex. *Psychophysiology*, *32*, 418-422.

Krumhansl, C.L. (1990). *Cognitive Foundations of Musical Pitch*. New York – Oxford: Oxford University Press.

Krumhansl, C.L. (2000). Rhythm and pitch in music cognition. *Psychological Bulletin*, *126*, 159-179.

Krumhansl, C.L., & Iverson, p. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception & Performance*, *18*, 739-751.

Kuriki, S., Isahai, N., & Ohtsuka, A. (2005). Spatiotemporal characteristics of the neural activities processing consonant/dissonant tones in melody. *Experimental Brain Research*, *162*, 46-55.

LeDoux, J.E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience*, *23*, 155-184.

Leino, S., Brattico, E., Tervaniemi, M., & Vuust, P. (submitted). Representation of harmony rules in the human brain: Further evidence from event-related potentials.

Leman, M. (1995). *Music and Schema Theory*. Berlin: Springer.

Leman, M. (1999). Naturalistic approaches to musical semiotics and the study of causal musical signification. In I. Zannos (Ed.), *Music and Signs – Semiotics and Cognitive Studies in Music* (pp. 11-38), Bratislava: ASCO Art and Science.

Liegeois-Chauvel, C., Musolino, A., Badier, J.M., Marquis, P., & Chauvel, P. (1994). Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. *Electroencephalography & Clinical Neurophysiology*, *92*, 204-214.

Lopez, L., Jürgens, R., Diekmann, V., Becker, W., Ried, S., Grözinger, B., & Erné, S.N. (2003). Musicians versus nonmusicians: A neurophysiological approach. *Annals of the New York Academy of Sciences*, *999*, 124-130.

Loui, P., Grent-'t-Jong, T., Torpey, D., & Woldorff, M. (2005). Effects of attention on the neural processing of harmonic syntax in Western music. *Cognitive Brain Research*, *25*, 678-687.

Luck, S.J. (2005). *Event-Related Potentials: A Methods Handbook*. Cambridge, Mass.: MIT Press.

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.

Lütkenhöner, B., Seither-Preisler, A., & Seither, S. (2006). Piano tones evoke stronger magnetic fields than pure tones or noise, both in musicians and non-musicians. *NeuroImage*, *30*, 927-937.

Lynch, M.P., & Eilers, R.E. (1991). Children's perception of native and non-native musical scales. *Music Perception*, *9*, 121-132.

Lynch, M.P., & Eilers, R.E. (1992). A study of perceptual development for musical tuning. *Perception & Psychophysics*, *52*, 599-608.

Lynch, M.P., Eilers, R.E., Oller, K.D., Urbano, R.C., & Wilson, P. (1991). Influences of acculturation and musical sophistication on perception of musical interval patterns. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 967-975.

Maess, B., Koelsch, S., Gunter, T.C., & Friederici, A.D. (2001). Musical syntax is processed in Broca's area: An MEG study. *Nature Neuroscience*, *4*, 540-545.

May, P., Tiitinen, H., Ilmoniemi, R.J., Nyman, G., Taylor, J.G., & Naatanen, R. (1999). Frequency change detection in human auditory cortex. *Journal of Computational Neuroscience*, *6*, 99-120.

McCarthy, G., & Wood, C.C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology*, *62*, 203-208.

Menning, H., Roberts, L.E., & Pantev, C. (2000). Plastic changes in the auditory cortex induced by intensive frequency discrimination training. *NeuroReport*, *11*, 817-822.

Milner, B. (1962). Laterality effects in audition. In V.B. Mountcastle (Ed.), *Interhemispheric Relations and Cerebral Dominance* (pp. 177-195). Baltimore: Johns Hopkins.

Moore, B.C.J. (2004). *An Introduction to the Psychology of Hearing*. London - San Diego - New York: Academic Press.

Müller, B.W., Jüptner, M., Jentzen, W., & Müller, P.S. (2002). Cortical activation to auditory mismatch elicited by frequent deviant and complex novel sounds: A PET study. *NeuroImage*, *17*, 231-239.

Münte, T.F., Altenmüller, E., & Jäncke, L. (2002). The musicians' brain as a model of neuroplasticity. *Nature Reviews Neuroscience*, *3*, 473-478.

Neisser, U. (1967). *Cognitive Psychology*. New York: Appleton-Century-Crofts.

Nelken, I., Fishbach, A., Las, L., Ulanovsky, N., & Farkas, D. (2003). Primary auditory cortex of cats: Feature detection of something else? *Biological Cybernetics*, *89*, 397-406.

Novitski, N., Huotilainen, M., Tervaniemi, M., Näätänen, R., & Fellman, V. (in press). Neonatal frequency discrimination in 250-4000 Hz range: Electrophysiological evidence. *Clinical Neurophysiology*.

Näätänen, R. (1984). In search of a short-duration memory trace of a stimulus in the human brain. In L. Pulkkinen & P. Lyytinen (Eds.), *Perspectives to Human Action and Personality: Essays in Honour of Martti Takala* (pp. 22-36). Jyväskylä: University of Jyväskylä.

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences*, *13*, 201-288.

Näätänen, R. (1992). *Attention and Brain Function*. Hillsdale (N.J.): Erlbaum.

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, *38*, 1-21.

Näätänen, R., Brattico, E., & Tervaniemi, M. (2003). Mismatch negativity (MMN): A probe to auditory cognition and perception in basic and clinical research. In A. Zani & A. Mado Proverbio (Eds.), *The Cognitive Electrophysiology of Mind and Brain* (pp. 343-355). San Diego (CA): Academic Press.

Näätänen, R., Gaillard, A.W., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *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., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R.J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, *385*, 432-434.

Näätänen, R., & Michie, P.T. (1979). Early selective attention effects on the evoked potential. A critical review and reinterpretation. *Biological Psychology*, *8*, 81-136.

Näätänen, R., Paavilainen, P., Tiitinen, H., Jiang, D., & Alho, K. (1993a). Attention and mismatch negativity. *Psychophysiology*, *30*, 436-450.

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., Sams, M., Alho, K., Paavilainen, P., Reinikainen, K., & Sokolov, E.N. (1988). Frequency and location specificity of the human vertex N1 wave. *Electroencephalography and Clinical Neurophysiology*, *69*, 523-531.

Näätänen, R., Schröger, E., Karakas, S., Tervaniemi, M., & Paavilainen, P. (1993b). Development of a memory trace for a complex sound in the human brain. *NeuroReport*, *4*, 503-506.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). Pre-attentive cognitive processing ("primitive intelligence") in the auditory cortex as revealed by the mismatch negativity (MMN). *Trends in Neurosciences*, *24*, 283-288.

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, *125*, 826-859.

Olsho, L., Schoon, C., Sakai, R., Turpin, R., & Sperduto, V. (1982). Auditory frequency discrimination in infancy. *Developmental Psychology*, *18*, 721-726.

Opitz, B., Rinne, T., A., M., Cramon, D.Y.v., & Schröger, E. (2002). Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. *NeuroImage*, *15*, 167-174.

Osterhout, L., Holcomb, P.J., & Swinney, D.A. (1994). Brain potentials elicited by garden-path sentences: evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 786-803.

Paavilainen, P. (1996). *Mismatch Negativity: An Automatic Brain Response Reflecting the Neural Mechanisms of Auditory Sensory Memory*. Doctoral Dissertation. Helsinki: University of Helsinki.

Paavilainen, P., Jaramillo, M., Näätänen, R., & Winkler, I. (1999). Neuronal populations in the human brain extracting invariant relationships from acoustic variance. *Neuroscience Letters*, *265*, 179-182.

Paavilainen, P., Tiitinen, H., Alho, K., & Näätänen, R. (1993). Mismatch negativity to slight pitch changes outside strong attentional focus. *Biological Psychology*, *37*, 23-41.

Paller, K.A., McCarthy, G., & Wood, C.C. (1992). Event-related potentials elicited by deviant endings to melodies. *Psychophysiology*, *29*, 202-206.

Palmer, C., & Krumhansl, C.L. (1987). Independent temporal and pitch structures in determination of musical phrases. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 116-126.

Pantev, C., Bertrand, O., Eulitz, C., Verkindt, C., Hampson, S., Schuierer, G., & Elbert, T. (1995). Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. *Electroencephalography and Clinical Neurophysiology*, *94*, 26-40.

Pantev, C., Engelien, A., Candia, V., & Elbert, T. (2003). Representational cortex in musicians. In I. Peretz & R. Zatorre (Eds.), *The Cognitive Neuroscience of Music* (pp. 382-395). Oxford: Oxford University Press.

Pantev, C., Hoke, M., Lehnertz, K., & Lütkenhöner, B. (1989a). Neuromagnetic evidence of an amplitopic organization of the human auditory cortex. *Electroencephalography and Clinical Neurophysiology*, *72*, 225-231.

Pantev, C., Hoke, M., Lehnertz, K., Lütkenhöner, B., Anogianakis, G., & Wittkowski, W. (1988). Tonotopic organization of the human auditory cortex revealed by transient auditory evoked magnetic fields. *Electroencephalography and Clinical Neurophysiology*, *69*, 160-170.

Pantev, C., Hoke, M., Lütkenhöner, B., & Lehnertz, K. (1989b). Tonotopic organization of the auditory cortex: Pitch versus frequency representation. *Science*, *246*, 486-488.

Pantev, C., Oolstenveld, R., Engelien, A., Ross, B., Roberts, L.E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature*, *392*, 811-814.

Pantev, C., Roberts, L.E., Schultz, M., Engelien, A., & Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *NeuroReport*, *12*, 169-174.

Pantev, C., Wollbrink, A., Roberts, L.E., Engelien, A., & Lütkenhöner, B. (1999). Short-term plasticity of the human auditory cortex. *Brain Research*, *842*, 192-199.

Patel, A.D., Foxton, J.M., & Griffiths, T.D. (2005). Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech. *Brain and Cognition*, *59*, 310-313.

Patel, A.D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P.J. (1998a). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, *10*, 717-733.

Patel, A.D., Peretz, I., Tramo, M., & Labrecque, R. (1998b). Processing prosodic and musical patterns: A neuropsychological investigation. *Brain and Language*, *61*, 123-144.

Patterson, R.D., Uppenkamp, S., Johnsrude, I.S., & Griffiths, T.D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, *36*, 767-776.

Penagos, H., Melcher, J.R., & Oxenham, A.J. (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *Journal of Neuroscience*, *24*, 6810-6815.

Penfield, W., & Perot, P. (1963). The brain's record of auditory and visual experience: A final summary and discussion. *Brain*, *86*, 595-696.

Peretz, I. (1990). Processing of local and global musical information by unilateral brain-damaged patients. *Brain*, *113*, 1185-1205.

Peretz, I. (2001). Music cognition in the brain of the majority: Autonomy and fractionation of the music recognition system. In B. Rapp (Ed.), *The Handbook of Cognitive Neuropsychology* (pp. 519-540). Hove: Psychology Press.

Peretz, I. (2003). Brain specialization for music: New evidence from congenital amusia. In I. Peretz & R. Zatorre (Eds.), *The Cognitive Neuroscience of Music* (pp. 192-203). Oxford: Oxford University Press.

Peretz, I., Ayotte, J., Zatorre, R.J., Mehler, J., Ahad, P., Penhune, V.B., & Jutras, B. (2002). Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron*, *33*, 185-191.

Peretz, I., Champod, A.S., & Hyde, K. (2003). Varieties of musical disorders. The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences*, *999*, 58-75.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6*, 688-691.

Peretz, I., & Zatorre, R. (Eds.). (2003). *The Cognitive Neuroscience of Music*. Oxford: Oxford University Press.

Petkov, C.I., Kang, X., Alho, K., Bernard, O., Yund, E.W., & Woods, D.L. (2004). Attentional modulation of human auditory cortex. *Nature Neuroscience*, *7*, 658-663.

Pickles, J.O. (1988). *An Introduction to the Physiology of Hearing*. (2nd edition). London - San Diego: Academic Press.

Picton, T. (Ed.). (1988). *Handbook of Electroencephalography and Clinical Neurophysiology*. Amsterdam: Elsevier.

Picton, T.W., Alain, C., Otten, L., Ritter, W., & Achim, A. (2000a). Mismatch negativity: Different water in the same river. *Audiology & Neuro-Otology*, *5*, 111-139.

Picton, T.W., Bentin, N., Berg, P., Donchin, E., Hillyard, S.A., Johnson, R.J., Miller, G.A., Ritter, W., Ruchkin, D.S., Rugg, M.D., & Taylor, M.J. (2000b). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, *37*, 127-152.

Picton, T., Lins, O.G., & Scherg, M. (1995). The recording and analysis of event-related potentials. In F. Boller & J. Grafman (Eds.), *Handbook of Neuropsychology* (Vol. 10, pp. 3-73). Amsterdam: Elsevier.

Picton, T.W., van Roon, P., Armilio, M.L., Berg, P., llle, N., & Scherg, M. (2000c). The correction of ocular artifacts: A topographic perspective. *Clinical Neurophysiology*, *111*, 53-65.

Picton, T., Woods, D.L., & Proulx, G.B. (1978). Human auditory sustained potentials. II. Stimulus relationships. *Electroencephalography and Clinical Neurophysiology*, *45*, 198-210.

Pincze, Z., Lakatos, P., Rajkai, C., Ulbert, I., & Karmos, G. (2001). Separation of mismatch negativity and the N1 wave in the auditory cortex of the cat: A topographic study. *Clinical Neurophysiology*, *112*, 778-784.

Pincze, Z., Lakatos, P., Rajkai, C., Ulbert, I., & Karmos, G. (2002). Effect of deviant probability and interstimulus/interdeviant interval on the auditory N1 and mismatch negativity in the cat auditory cortex. *Cognitive Brain Research*, *13*, 249-253.

Posner, M.I. (1995). Attention in cognitive neuroscience: An overview. In M.S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (pp. 615-624). Cambridge, MA: The MIT Press.

Prasher, H.E. (1998). *The Psychology of Attention*. Cambridge, MA: MIT Press.

Pratt, H., Aminoff, M., Nuwer, M.R., & Starr, A. (1999). Short-latency auditory evoked potentials. In G. Deuschl & A. Eisen (Eds.), *Recommendations for the Practice of Clinical Neurophysiology* (Vol. 52). New York - Amsterdam: Elsevier Science.

Pülvermüller, F. (2003). *The Neuroscience of Language. On Brain Circuits of Words and Serial Order*. Cambridge, UK: Cambridge University Press.

Rauschecker, J.P. (1999). Auditory cortical plasticity: A comparison with other sensory systems. *Trends in Neurosciences*, *22*, 74-80.

Recanzone, G.H., Schreiner, C.E., & Merzenich, M.M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *Journal of Neuroscience*, *13*, 87-103.

Restuccia, D., Della Marca, G., Marra, C., Rubino, M., & Valeriani, M. (2005). Attentional load of the primary task influences the frontal but not the temporal generators of mismatch negativity. *Cognitive Brain Research*, *25*, 891-899.

Rinne, T., Alho, K., Ilmoniemi, R.J., Virtanen, J., & Näätänen, R. (2000). Separate time behaviors of the temporal and frontal MMN sources. *NeuroImage*, *12*, 14-19.

Ritter, W., Sussman, E., & Molholm, S. (2000). Evidence that the mismatch negativity system works on the basis of objects. *NeuroReport*, *11*, 61-63.

Robertson, L., Treisman, A., Friedman-Hill, S., & Grabowecky, M. (1997). The interaction of spatial and object pathways: Evidence from Balint’s syndrome. *Journal of Cognitive Neuroscience*, *9*, 295-317.

Romani, G.L., Willianson, S.J., & Kaufman, L. (1982). Tonotopic organization of the human auditory cortex. *Science*, *216*, 1339-1340.

Rosburg, T., Trautner, P., Dietl, T., Korzyukov, O.A., Boutros, N.N., Schaller, C., Elger, C.E., & Kurthen, M. (2005). Subdural recordings of the mismatch negativity (MMN) in patients with focal epilepsy. *Brain*, *128*, 819-828.

Ruchkin, D.S., Grafman, J., Cameron, K., & Berndt, R.S. (2003). Working memory retention systems: A state of activated long-term memory. *Behavioral and Brain Sciences*, *26*, 709-777.

Rugg, M.D., & Coles, M.G.H. (Eds.). (1995). *Electrophysiology of Mind: Event-Related Brain Potentials and Cognition* . Oxford: Oxford University Press.

Sams, M., Alho, K., & Näätänen, R. (1983). Sequential effects on the ERP in discriminating two stimuli. *Biological Psychology*, *17*, 41-58.

Sams, M., Hari, R., Rif, J., & Knuutila, J. (1993). The human auditory sensory memory trace persists about 10 s: Neuromagnetic evidence. *Journal of Cognitive Neuroscience*, *5*, 363--370.

Schall, U., Johnston, P., Todd, J., Ward, P.B., & Michie, P.T. (2003). Functional neuroanatomy of auditory mismatch processing: An event-related fMRI study of duration-deviant oddballs. *NeuroImage*, *20*, 729-736.

Schellenberg, E.G., & Trehub, S.E. (1994). Frequency ratios and the discrimination of pure tone sequences. *Perception & Psychophysics*, *56*, 472-478.

Schiavetto, A., Cortese, F., & Alain, C. (1999). Global and local processing of musical sequences: an event-related brain potential study. *NeuroReport*, *10*, 2467-2472.

Schlaug, G. (2001). The brain of musicians: A model for functional and structural adaptation. In R.J. Zatorre & I. Peretz (Eds.), *The Biological Foundations* *of Music* (Vol. 930, pp. 281-299). New York: Annals of the New York Academy of Sciences.

Schmithorst, V.J. (2005). Separate cortical networks involved in music perception: Preliminary functional MRI evidence for modularity of music processing. *NeuroImage*, *25*, 444-451.

Schoenberg, A. (1967). *Fundamentals of Musical Composition*. London: Faber and Faber Limited.

Schröger, E. (1997). On the detection of auditory deviations: A pre-attentive activation model. *Psychophysiology*, *34*, 245-257.

Schröger, E., Tervaniemi, M., & Huotilainen, M. (2004). Bottom-up and top-down flows of information within auditory memory: electrophysiological evidence. In C. Kaernbach, E. Schröger, & H. Müller (Eds.), *Psychophysics beyond Sensation: Laws and Invariants of Human Cognition* (pp. 389–407). Hillsdale, NJ: Erlbaum.

Schuppert, M., Münte, T.F., Wieringa, B.M., & Altenmüller, E. (2000). Receptive amusia: Evidence for cross-hemispheric neural networks underlying music processing strategies. *Brain*, *123*, 546-559.

Seger, C.A. (1994). Implicit learning. *Psychological Bulletin*, *115*, 163-169.

Seppänen, M.S.A, Brattico, E., & Tervaniemi, M. (in press). Practice strategies in musicians modulate the neural processing and learning of sound-patterns. *Neurobiology of Learning and Memory*.

Shafritz, K.M., Gore, J.C., & Marois, R. (2002). The role of the parietal cortex in visual feature binding. *Proceedings of the National Academy of Sciences*, *9*, 10917-10922.

Shahin, A., Bosnyak, D.J., Trainor, L.J., & Roberts, L.E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *Journal of Neuroscience*, *23*, 5545-5552.

Shahin, A., Roberts, L.E., & Trainor, L.J. (2004). Enhancement of auditory cortical development by musical experience in children. *NeuroReport*, *15*, 1917-1921.

Shankweiler, D. (1966). Effects of temporal-lobe damage on the perception of dichotically presented melodies. *Journal of Comparative Physiological Psychology*, *62*, 115-118.

Shestakova, A. (2004). *'Sensory Intelligence' in the Auditory Cortex: Brain Responses to Native and Non-native phonetic Stimuli*. Doctoral Dissertation. Helsinki: Yliopistopaino.

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

Shestakova, A., Brattico, E., Huotilainen, M., Klucharev, V., & Solovjev, A. (2004). Orderly cortical representation of vowel categories presented by multiple exemplars. *Cognitive Brain Research*, *21*, 342-350.

Skrandies, W. (2003). Topographical analysis of electrical brain activity: Methodological aspects. In A. Zani & A. Mado Proverbio (Eds.), *The Cognitive Electrophysiology of Mind and Brain* (pp. 401-416), San Diego: Academic Press.

Snyder, B. (2000). *Music and memory*. Cambridge, MA: MIT Press.

Squires, N.K., Squires, K.C., & Hillyard, S. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, *38*, 387-401.

Sussman, E., Bregman, A.S., Wang, W.J., & Khan, F.J. (2005). Attentional modulation of electrophysiological activity in auditory cortex for unattended sounds in multi-stream auditory environments. *Cognitive, Affective and Behavioral Neuroscience*, *5*, 93-110.

Sussman, E., Gomes, H., Nousak, J.M., Ritter, W., & Vaughan, Jr., H.G. (1998). Feature conjunction and auditory sensory memory. *Brain Research*, *18*, 95-102.

Sussman, E., Winkler, I., Huotilainen, M., Ritter, W., & Näätänen, R. (2002). Top-down effects can modify the initially stimulus-driven auditory organization. *Cognitive Brain Research*, *13*, 393-405.

Sussman, E., Winkler, I., & Wang, W. (2003). MMN and attention: Competition for deviance detection. *Psychophysiology*, *40*, 430-435.

Takegata, R., Huotilainen, M., Rinne, T., Näätänen, R., & Winkler, I. (2001). Changes in acoustic features and their conjunctions are processed by separate neuronal populations. *NeuroReport*, *12*, 525-529.

Takegata, R., Paavilainen, P., Näätänen, R., & Winkler, I. (1999). Independent processing of changes in auditory single features and feature conjunctions in humans as indexed by the mismatch negativity. *Neuroscience Letters*, *266*, 109-112.

Tervaniemi, M., & Brattico, E. (2004). From sounds to music: Towards understanding the neural basis of musical sound perception. *Journal of Consciousness Studies*, *11*, 9-27.

Tervaniemi, M., Huotilainen, M., Brattico, E., Reinikainen, K., Ilmoniemi, R.J., & Alho, K. (2003). Event-related potentials to expectancy violation in musical context. *Musicae Scientiae*, *7*, 241-261.

Tervaniemi, M., Ilvonen, T., Karma, K., Alho, K., & Näätänen, R. (1997). The musical brain: Brain waves reveal the neurophysiological basis of musicality in human subjects. *Neuroscience Letters*, *226*, 1-4.

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schröger, E. (2005). Pitch discrimination accuracy in musicians vs. nonmusicians: an event-related potential and behavioral study. *Experimental Brain Research*, *161*, 1-10.

Tervaniemi, M., Medvedev, S.V., Alho, K., Pakhomov, S.V., Roudas, M.S., van Zuijen, T., & Näätänen, R. (2000). Lateralized automatic auditory processing of phonetic versus musical information: A PET study. *Human Brain Mapping*, *10*, 74-79.

Tervaniemi, M., Rytkönen, M., Schröger, E., Ilmoniemi, R.J., & Näätänen, R. (2001). Superior formation of cortical memory traces of melodic patterns in musicians. *Learning and Memory*, *8*, 295-300.

Tervaniemi, M., Saarinen, J., Paavilainen, P., Danilova, N., & Näätänen, R. (1994). Temporal integration of auditory information in sensory memory as reflected by the mismatch negativity. *Biological Psychology*, *38*, 157-167.

Tiitinen, H., Alho, K., Huotilainen, M., Ilmoniemi, R.J., Simola, J., & Naatanen, 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.

Tillmann, B., Bharucha, J.J., & Bigand, E. (2000). Implicit learning of tonality: A self-organizing approach. *Psychological Review*, *107*, 885-913.

Toiviainen, P., Tervaniemi, M., Louhivuori, J., Huotilainen, M., Saher, M., & Näätänen, R. (1998). Musical timbre: Convergence of neural, behavioral, and computational approaches. *Music Perception*, *16*, 223-241.

Trainor, L.J. (2005). Are there critical periods for musical development? *Developmental Psychobiology*, *46*, 262-278.

Trainor, L.J., McDonald, K.L., & Alain, C. (2002). Automatic and controlled processing of melodic contour and interval information measured by electric brain activity. *Journal of Cognitive Neuroscience*, *14*, 430-442.

Trainor, L.J., & Trehub, S.E. (1993). What mediates infants' and adults' processing of the major over the augmented triad? *Music Perception*, *11*, 185-196.

Trainor, L.J., & Trehub, S.E. (1994). Key membership and implied harmony in Western tonal music: Developmental perspectives. *Perception & Psychophysics*, *56*, 125-132.

Trehub, S.E., & Trainor, L.J. (1998). Singing to infants: Lullabies and play songs. *Advances in Infancy Research*, *12*, 43-77.

Treisman, A.M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97-136.

Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 459-478.

Treisman, A., & Schmidt, N. (1992). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, *14*, 107-141.

Ulanovsky, N., Las, L., & Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature Neuroscience*, *6*, 391-398.

Umbricht, D., Koller, R., Vollenweider, F.X., & Schmid, L. (2002). Mismatch negativity predicts psychotic experiences induced by NMDA receptor antagonist in healthy volunteers. *Biological Psychiatry*, *51*, 400-406.

Umbricht, D., Schmid, L., Koller, R., Vollenweider, F.X., Hell, D., & Javitt, D.C. (2000). Ketamine-induced deficits in auditory and visual context-dependent processing in healthy volunteers: implications for models of cognitive deficits in schizophrenia. *Archive of General Psychiatry*, *57*, 1139-1147.

Vaughan, Jr., H.G. (1969). The relationship of brain activity to scalp recordings of event-related potentials. In E. Donchin & D.B. Lindsley (Eds.), *Average Evoked Potentials: Methods, Results and Evaluations* (pp. 45-75). Washington, DC: U.S. Government Printing Office.

Verleger, R. (1990). P3-evoking wrong notes: Unexpected, awaited, or arousing? *International Journal of Neuroscience*, *55*, 171-179.

Wallin, N.L., Meyer, B., & Brown, S. (2000). *The Origins of Music*. Cambridge, MA: The MIT Press.

Weinberger, N.M., & Bakin, J.S. (1998). Learning-induced physiological memory in adult primary auditory cortex: Receptive field plasticity, model, and mechanisms. *Audiology & Neuro-Otology*, *3*, 145-167.

Weinberger, N.M., Javid, R., & Lepan, B. (1993). Long-term retention of learning-induced receptive field plasticity in the auditory cortex. *Proceedings of the National Academy of Sciences*, *90*, 2394-2398.

Winkler, I., Czigler, E., Sussman, E., Horvath, J., & Balazs, L. (2005). Pre-attentive binding of auditory and visual stimulus features. *Journal of Cognitive Neuroscience*, *17*, 320-339.

Winkler, I., Karmos, G., & Näätänen, R. (1996). Adaptive modeling of the unattended acoustic environment in the mismatch-negativity event-related potential. *Brain Research*, *742*, 239-252.

Winkler, I., & Näätänen, R. (1992). Event-related potentials in auditory backward recognition masking: a new way to study the neurophysiological basis of sensory memory in humans. *Neuroscience Letters*, *140*, 239-242.

Woestenburg, J.C., Verbaten, M.N., & Slangen, J.L. (1983). The removal of the eye-movement artifact from the EEG by regression analysis in the frequency domain. *Biological Psychology*, *16*, 127-147.

Woldorff, M.G., Gallen, C.C., Hampson, S.A., Hillyard, S.A., Pantev, C., Sobel, D., & Bloom, F.E. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Sciences*, *90*, 8722-8726.

Woldorff, M.G., Hackley, S.A., & Hillyard, S.A. (1991). The effects of channel-selective attention on the mismatch negativity wave elicited by deviant tones. *Psychophysiology*, *28*, 30-42.

Woldorff, M., & Hillyard, S.A. (1991). Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalography and Clinical Neurophysiology*, *79*, 170-191.

Woldorff, M.G., Hillyard, S.A., Gallen, C.C., Hampson, S.R., & Bloom, F.E. (1998). Magnetoencephalographic recordings demonstrate attentional modulation of mismatch-related neural activity in human auditory cortex. *Psychophysiology*, *35*, 283-292.

Woods, D.L., Alain, C., & Ogawa, K.H. (1998). Conjoining auditory and visual features during high-rate serial presentation: Processing and conjoining two features can be faster than processing one. *Perception & Psychophysics*, *60*, 239-249.

Yabe, H., Tervaniemi, M., Reinikainen, K., & Näätänen, R. (1997). Temporal window of integration revealed by MMN to sound omission. *NeuroReport*, *8*, 1971-1974.

Yabe, H., Tervaniemi, M., Sinkkonen, J., Huotilainen, M., Ilmoniemi, R.J., & Näätänen, R. (1998). Temporal window of integration of auditory information in the human brain. *Psychophysiology*, *35*, 615-619.

Zatorre, R.J. (1988). Pitch perception of complex tones and human temporal-lobe function. *Journal of the Acoustical Society of America*, *84*, 566-572.

Zatorre, R.J. (2005). Finding the missing fundamental. *Nature*, *436*, 1093-1094.

Study I reprinted from *Cognitive Brain Research*, *25*, Takegata, R., Brattico, E., Tervaniemi, M., Varyagina, O., Näätänen, R., & Winkler, I., Preattentive representation of feature conjunctions for concurrent spatially distributed auditory objects, 169-179, Copyright (2005), with permission from Elsevier.

Study II reprinted from *NeuroReport*, *13*, Brattico, E., Winkler, I., Näätänen, R., Paavilainen, P., & Tervaniemi, M., Simultaneous storage of two complex temporal sound patterns in auditory sensory memory, 1747-1751, Copyright (2002), with permission from Lippincott Williams & Wilkins.

Study III reprinted from *Music Perception*, *19*, Brattico, E., Näätänen, R., & Tervaniemi, M., Context effects on pitch perception in musicians and nonmusicians: evidence from event-related potential recordings, 199-222, Copyright (2001), with permission from University of California Press.

Study IV reprinted from *Brain Research*, *1117*, Brattico, E., Tervaniemi, M., Näätänen, R., & Peretz, I., Musical scale properties are automatically processed in the human auditory cortex, 162-174, Copyright (2006), with permission from Elsevier.

Study V reprinted from *NeuroReport*, *14*, Brattico, E., Tervaniemi, M., & Picton, T.W., Effects of brief discrimination-training on the auditory N1 wave, 2489-2492, Copyright (2003), with permission from Lippincott Williams & Wilkins.

Study VI reprinted from *Annals of Neurology*, *58*, Peretz, I., Brattico, E., & Tervaniemi, M., Abnormal electrical brain responses to pitch in congenital amusia, 478-482, Copyright (2005), with permission from Wiley.