Budapest University of Technology and Economics
PhD School in Psychology – Cognitive Science

Gábor Péter Háden

PROBING PERCEPTUAL CAPABILITIES UNDERLYING MUSIC PERCEPTION

PhD thesis

Supervisor:
Dr. István Winkler

Budapest, 2011

Contents

Acknowledgements

The single most important person responsible for me ever finishing this thesis, toward whom I feel the deepest gratitude, is my supervisor Dr. István Winkler. He invited me to the Department of General Psychology (now Department of Experimental Psychology) at the Institute for Psychology of the Hungarian Academy of Sciences and has been ever since an inexhaustible source of inspiration, help, and advice.

I am greatly indebted to all coauthors of the papers presented in the thesis (listed alphabetically): Drs. László Balázs, Anna Beke, Vineta Fellman, Henkjan Honing, Minna Huotilainen, Olivia Ladinig, Gábor Stefanics, and István Sziller.

Dr. János Horváth and Lívia Gabriella Pató taught me most of what I know about EEG recording and analysis. Thank you for the great tools you placed at my disposal.

Special thanks to Erika Józsa Váradiné and Zsuzsa D’Albini, who collected the neonatal and adult ERP data, respectively.

Judit Roschéné Farkas provided great help in compiling the list of references.

I thank all my colleagues at the Institute for Psychology: You are largely responsible for the joy of working at the institute.

The European Commission supported the experimental work reported in the thesis through the funding of the project “EmCAP: Emergent Cognition through Active Perception” (Contract 013123) as part of the 6th Framework Programme ‘‘Information Society Technologies’’. Special thanks to Dr. Susan L. Denham, who conceived and coordinated the project.

I thank my family and friends for their love and support. Getting this far would have been impossible without you.

Finally, I would like to dedicate this thesis to my grandfather Dr. Ferenc Háden. He gave me light.

Glossary of abbreviations

ANOVA	Analysis of variance
EEG	Electroencephalography/Electroencephalogram
ERP	Event related brain potential
MEG	Magnetoencephalography/Magnetoencephalogram
MMN	Mismatch negativity
MMF	Mismatch field
RT	Reaction time
SOA	Stimulus-onset asynchrony (onset-to-onset interval)

Abstract

Music is present in all human cultures, suggesting that it is deeply rooted in the perceptual and cognitive processes available to all humans. The general processes of audition form the foundation upon which music perception and, viewed more broadly, all communication by sounds is based. During ontogenesis, the ability to perceive music unfolds through interactions between innate predispositions, environmental constraints, and learning. Assessing the abilities underlying music perception at the time of birth and their development later in life provide important information for shedding light on these interactions. Electrophysiology offers the possibility to study auditory processes in newborn infants. The mismatch negativity (MMN) event-related potential (ERP) observable in adults and the functionally similar discriminative responses in neonates allow the investigation of various auditory perceptual abilities. The thesis is based on five ERP experiments, four of which investigated neonatal abilities underlying the perception of musical pitch, timbre, and rhythm. One combined ERP and behavioral experiment investigated the perception of musical meter in adults. Based on the results obtained in these experiments, it appears that babies are born well equipped for music perception, with sound processing abilities comparable to those found in adults.

Kivonat

A zene alapvető jelentősége minden ismert kultúrában, feltételezi, hogy az általános észlelési és kognitív feldolgozásban gyökeredzik. A hallás általános folyamatai olyan alapot képeznek, amelyre a zeneészlelés, illetve tágabban értelmezve, minden akusztikus kommunikáció épül. Az egyedfejlődés során a zeneészlelés képessége a veleszületett hajlamok, környezeti korlátok és a tanulás folyamatának interakciójából bomlik ki. Az interakciók feltárásához fontos információkat biztosít a zeneészlelést megalapozó képességek vizsgálata újszülött korban és e képességek fejlődésének vizsgálata a későbbi élet során. Az elektrofiziológia lehetővé teszi a hallási feldolgozás vizsgálatát újszülött gyermekeknél. A felnőtteken mérhető mismatch negativity (MMN – eltérési negativitás) kiváltott komponens és a funkcionális hasonló diszkriminatív válaszok újszülöttekben alkalmasak változatos hallási észlelési képességek vizsgálatára. A disszertáció öt, kiváltott válaszokat használó, kísérleten alapszik, amelyek közül négy az újszülöttek zenei hangmagasság-, hangszín- és ritmusészlelését megalapozó képességeit vizsgálja. Egy kombinált viselkedéses és elektrofiziológiai kísérlet pedig a felnőttek zenei ütemészlelését vizsgálja. A kísérletek eredményei alapján úgy tűnik, hogy a csecsemők már születésükkor jól felkészültek a zeneészleléshez szükséges információ feldolgozására és e képességeik hasonlóak a felnőttek képességeihez.

1. Introduction

What is the similarity between soothing lullabies sung by a mother, work-songs, chamber music and a rock concert? Since the days of Pythagoras numerous theorists tried to understand and give account of what is music and how it "works".[1]

Music cognition, as any other human ability, can be studied on many hierarchically embedded levels starting from basic neural structures, through computational processes implemented in cortical networks, simple and more complex behavior, ontogenetic and philogenetic history and finally, social and cultural constructs. Virtually all human cultures have some kind of musical tradition and in all cultures, individuals are surrounded by music from an early age (Merriam, 1964). The universal prevalence of music strongly suggests that music is deeply rooted in the perceptual, cognitive and emotional processes of the human species. The cognitive study of music should focus on an intermediate level between cells and culture, but should be able to integrate data from the full spectrum of available methodologies.

How do innate cognitive and emotional predispositions and environmental information interact to create music? How is music realized in the human brain? What are the relations between musical ability and other human abilities, especially language processing? Is music a mental faculty? How did music evolve in humans and is it uniquely human? All of these issues can be subdivided to further questions and answering all of them will certainly prove to be a long scientific endeavor. The aim of this thesis is to give some insights into a small subset of the questions related to the perception of music.

With the advent of psychology and later neurosciences, several techniques were developed that allows unprecedented access to the inner workings of the human brain. Based on the vast

---

[1] Throughout the thesis *music* refers to Western tonal music almost exclusively. It is beyond the scope of this thesis to tackle the difficulties of defining *music* therefore the reader is kindly referred to the *Music* entry in The New Grove Dictionary of Music and Musicians (Sadie; 2001) where the problems of definition are elaborated at length.

amount of findings accumulated in the field of sound perception, behavioral and electro-physiological methods can be utilized in assessing important prerequisites of music perception from birth to adulthood. Identifying the perceptual and cognitive abilities underlying music perception can impact a multitude of subjects. By studying the human abilities present at birth, the complex interactions of nature and nurture can be disentangled. This in turn may lead to explanations of cultural similarities and differences in music. Comparative studies also benefit from identifying those human traits which are shared with animals. Computational models trying to mimic the developing and mature musical abilities of humans also utilize experimental findings. Finally, the application of experimental paradigms inspired by cognitive science to the field of music theory may yield a psychologically and physiologically grounded understanding of music cognition and performance.

The structure of thesis is the following: The *Introduction* begins with an examination of the broader context surrounding the cognition of music and the application of cognitive investigations within the field of music (*Chapter 1.1*). It is followed by a short review of the concepts and results emerging from the field of music perception (*Chapter 1.2*), arranged according to the main research topics of the field. The rest of the introduction is dedicated to an important tool for auditory research, the mismatch negativity (MMN) event related potential (ERP) (*Chapter 1.3*), detailed in connection with discriminative responses in infants and music perception. Under the title *Theses* the background, hypotheses and results of each study is presented in a short form and their relation to the overall goals is spelled out. In *Studies* the research articles are included in their published form, followed by a *General Discussion* integrating the present findings with earlier ones. In the final part *Conclusions and further directions* are elaborated.

1.1. Questions about music and its origins

The issue of abilities underlying music in general and the perception of music in specific can be construed as deeply interrelated questions about evolution and brain organization. Does music constitute a separate domain of human cognition by itself? Even if domain-specific processes for music are assumed, domain-general mechanisms are needed to complement a working system of music processing and music can also be seen as an emergent property of multiple domain-general processes working together. To separate multiple domain-general and possibly domain-specific processes it may be helpful to look at the evolution of these processes via comparative studies, examine the responses proposed in answer to similar questions in the language domain and integrate the findings of neuropsychology, neurophysiology, neuroimaging and other fields. The aim of this chapter is to give a short introduction to these topics and to provide a wider context for present thesis.

The idea that the brain should be subdivided into faculties stems from the phrenology of Gall (Fodor, 1983) that saw the mind as the collection of functions linked to anatomically separate areas. This idea had a tremendous influence on cognitive science through Fodor’s theory of modularity (1983) and Chomsky’s faculty of language and generative grammar (1957).

Fodor (1983) gives criteria for mental modules that include domain-specificity, information encapsulation, fixed neural architecture and well-defined ontogenesis. Whereas Fodor (1983, 2001) argues that modules work on the lower levels of information processing, close to perception, and cannot by themselves explain the working of the whole mind, several authors (e.g. Pinker, 1997; Carruthers, 2006) expanded his theory to a ‘massively modular’ account of the mind. One of the key motivations for this expansion is that if the mind is built up from separate (but highly interconnected) and innately determined modules (at least to some degree), then evolutionary psychology (e.g. Barkow, Tooby & Cosmides, 1995) has a much better chance for carrying out its program of identifying adaptations. In a variant of the modular view, Karmiloff-Smith (1992) introduced ‘modularization’, a process by which separate modules emerge during development based on interactions between innate and environmental factors. The modular approach to the brain resulted in research identifying specialized brain areas for specialized functions, such as face perception (Kanwisher, 2000) and reading (Cohen et al., 2000) and generated much debate (see Haxby et al., 2001 and Price & Devlin 2003, respectively).

The above mentioned criteria for cognitive modules, that is, innateness, domain specificity and brain localization are often conflated and Fodor (1983) himself notes that not all criteria must be met to define a module. One motivation for this conflation might be that, in order to make a strong argument for an adaptation, a trait must be both innate and domain specific (Justus & Hutserl, 2005). Brain localization can be related to innateness by arguing on the basis of genetic determination of the brain structure, or to domain specificity by assuming that specific task-related computational needs may be accommodated by separate areas (Johnson, 2001).

Chomsky’s (1953) statements about the innateness and human-specificity of language were refined by Hauser, Chomsky and Fitch (2002). Based on a review of comparative data, the authors distinguish between a broad sense of the language faculty, containing abilities that are necessary for language and are shared with nonhuman species, and a narrow sense of the language faculty containing abilities (identified as the ability of using recursive structures according to their analysis) unique to humans. This treatment of the language faculty is compatible with both the modular and the massively modular account of the mind and provides a good framework for integrating comparative results and evolutionary hypotheses. An adaptationist approach to language (or other mental abilities) can concentrate on identifying homologous (same ancestry) or homoplasic (same function) adaptations that are present in both animals and humans and seek human-specific adaptations. The other benefit of the approach used by Hauser, Chomsky and Fitch (2002) is that it emphasizes the fact that complex abilities like language rely on a myriad of simpler abilities, some of which are shared between different species, breaking up the monolithic approach originally advocated by Chomsky.

Music is similar to language in many ways. Both appear in all human cultures, both are acquired without formal training and both can be identified as such by members of a culture (Hauser & McDermott, 2003). Music and language are generative. A practically infinite number of valid expressions can be created from a finite set of rules. Questions asked about language can also be asked about music (Hauser & McDermott, 2003; Jackendoff & Lerdahl, 2006). Fitch (2006) advocates comparisons of musical and linguistic abilities and use of human-animal comparative data for insights into the biological basis and evolution of music. There are, however, also some marked differences between language and music. An apparent one is that music does not convey meaning in a propositional way, however this does not preclude that it carries some kind of meaning (Koelsch & Siebel, 2005; Koelsch & Sammler, 2008). Music is more restricted than speech both spectrally and temporally, because, unlike speech it usually requires the use of a predefined set of tones and more or less tight conformity to isochronous pulses. There is archeological evidence of music production in the form of bone-flutes. In contrast, using artifacts, language can only be traced back to the time writing appeared (Kunej & Turk, 2001).

Most of the abilities underlying music cognition have some phylogenetic precursors (for an extensive discussion, see the articles in Wallin, Merker & Brown, 2001.) A question already pondered by Darwin (1879/2004) is whether there are special adaptations for music and if so whether they are found only in humans? The views regarding music-related adaptation range from total skepticism to endorsement of adaptationist explanations of varying credibility.

Pinker (1997) for example, in a passage much quoted by critics, calls music ‘auditory cheesecake’ (p. 534), the input that garners responses from sound-sensitive systems evolved to perform other functions. Based on his opinion, music is a good candidate for exaptation (Gould & Vrba, 1982), a trait not in itself adapted but taking over functions adapted for other traits. Darwin (1871) himself noted that there is no straightforward adaptive value in music, and he explained its presence in terms of sexual selection. Other theorists explain the possible adaptive function of music in terms of mother-infant interactions (Dissanayake, 2001), intergroup coordination of behavior (Hagen & Bryant, 2003), emotion (Roederer, 1984) or emotional regulation (Bispham 2006), etc. (for a review, see Huron, 2003). Most theories that describe the adaptive value of music do so by proposing one or a few central adaptations around which multi-purpose systems crystallize forming music perception and production. Trehub and Hannon (2006), for example, propose a music-oriented motivational system that is itself an adaptation and acts as a catalyst for other domain-general, non-human-specific systems establishing music perception. There is no guarantee however, that music relies on one or only a few key adaptations. Fitch (2006) proposes that several adaptive forces must have been at work during the evolution of music as no single force provides explanation for all musical phenomena. This, in his view, renders arguments about past functions of music futile (but not invalid).

It has been shown that both songbirds and non-songbirds can perceive relative pitch (Cynx, 1995), one of the prerequisites of perceiving melodies. To achieve this ability, however, intensive training was necessary and the training did not generalize to frequency ranges not part of the training set, in which birds tended to fall back to the use of absolute pitch cues. Similar results were obtained from Japanese monkeys (Izumi, 2001). With respect to relative pitch perception, Wright et al. (2000) showed that rhesus monkeys can be trained to perform octave generalizations, that is, recognize the identity of transposed sound sequences.

Unfortunately, this ability is limited to transpositions using integer octaves; transpositions of half octaves were not recognized by the monkeys. Taken together, these results show that some animals are able to perceive relative pitch, but this ability might be realized through different mechanisms than in humans. There are also examples of human and animal abilities using similar mechanisms. Both Fitch (2006) and Patel (2006) emphasize the importance of vocal learning abilities, which enable a flexible vocal behavioral repertoire in contrast to the rigid vocal call systems found in species without vocal learning ability. Vocal learning evolved independently in humans, some birds, cetaceans, and pinnipeds, but it is not present e.g., in nonhuman primates. Vocal learning abilities require a high degree of perceptual and motor coordination and may be related to the ability of synchronizing to a musical beat (Patel et al., 2009a). Some evidence supporting this relation is already available (Patel et al., 2009a) and further studies are proposed (Patel et al. 2009b; Schachner et al., 2009). The auditory environment in which animals are reared may also have influence on their perceptual abilities as for example an infant chimpanzee reared as human showed preference of consonant over dissonant music similar to human infants (Sugimoto et al., 2010). The above examples show that comparative studies can meaningfully contribute to the understanding of music perception by pinpointing crucial similarities with animals and possibly identifying some structural and representational constraints on music.

Modular accounts of music processing propose an information processing system that is exclusive to the processing of music (Peretz, 2006; Peretz & Coltheart, 2003; Peretz & Morais, 1989). The general music processing system is not monolithic, rather it consists of separate sub-modules that get their input from general acoustic processing and belong to the larger systems of pitch processing (including spectral pitch, melody and harmony) and temporal processing (including rhythm and meter), which in turn drive emotional analysis of music and a musical lexicon similar to the phonological lexicon (Peretz & Coltheart, 2003).

Evidence in support of the modular view comes mostly from neuropsychological studies on the selective impairment of musical abilities either as result of brain injury (acquired amusia) or a hereditary disorder (congenital amusia or tone deafness in older terminology).

Dissociation of music and language has been found in patients with unilateral as well as bilateral temporal lesions (Peretz et al. 1994; Peretz, 1996; Dalla Bella & Peretz, 1999; Piccirilli, Sciarma & Luzzi, 2000). A reverse pattern of dissociation was found in patients with Alzheimer's disease, who showed preserved musical memory despite severe speech and memory deficits (Vanstone & Cuddy, 2010), and in a case of cortical atrophy where music production was spared despite speech production difficulties (Polk & Kertesz, 2003). Dissociations between processing of pitch (melody, harmony) and processing of temporal (rhythm, tempo, meter) information has been related, respectively, with right and left temporal lesions (Kester et al, 1991; Midorikawa et al., 2003; Murayama et al., 2004; Di Pietro et al., 2004). The lateralization is not so clear-cut, as elements of both temporal processing systems, which have been associated with the left hemisphere, and pitch processing systems, which have been associated with the right hemisphere can dissociate after lesions occurring in either hemisphere (Liégeois-Chauvel et al., 1998). This in turn may hint at a bi-hemispheric network necessary for both systems to work properly (Samson & Zatorre, 1988; Schuppert et al., 2000).

Congenital amusia is a hereditary (Peretz, Cummings & Dubé, 2007) musical disorder characterized mainly by a deficit in processing melodic pitch variation (Peretz, 2001) also extending to musical memory, singing and tapping along music (Ayotte, Peretz & Hyde, 2002; Gosselin, Jolicoeur & Peretz, 2009) with otherwise intact functioning in other domains. Subjects are able to automatically discriminate even small pitch changes (Moreau, Jolicoeur & Peretz, 2009). However, they show abnormal electric activation patterns during the subsequent attentive processing of these changes (Peretz, Brattico & Tervaniemi, 2005; Peretz et al., 2009). The abnormal activation seems to originate from anatomical and functional anomalies found in the right inferior frontal gyrus (Hyde et al., 2006; Hyde et al., 2007; Hyde, Zatorre & Peretz, 2010).

Further evidence for the modular view comes from studies investigating the brain location of music and language systems. Comparing the processing of linguistic and musical meaning Steinbeis and Koelsch (2008) found that although both elicited an N400 (an indicator of semantic processing) distinct areas of the right temporal lobe were activated which was interpreted as a sign of similar processing on distinct networks. A similar activation difference was observed in the left temporal lobe during performing verbal vs. musical semantic memory tasks (Groussard et al., 2010). These results fit well the idea of a domain-specific musical lexicon proposed by Peretz and Coltheart (2003). However, localization studies show that music-processing areas can overlap language processing areas, indicating domain-general processing, for example for musical syntax (Maess et al., 2001) and working memory (Koelsch et al., 2009).

Despite compelling evidence pointing toward a domain-specific and modular organization of at least some aspects of music processing, it must be noted that the vast majority of the results come from studies dealing with adults, showing the endpoint of development. These results are compatible with the view that music undergoes modularization during development and initially domain-general processing mechanisms form the basis of domain-specific modules in adulthood (McMullen & Saffran, 2004). Indeed, music can induce structural and functional changes in the brain, most apparent in the case of musical training (Schlaug et al., 1995; Ohnishi et al., 2001; Gaser & Schlaug, 2003). One possible argument weakening the modularization hypothesis could come from finding music processing at birth (e.g. Perani et al., 2010).

1.2. Music perception

The following chapter introduces those concepts and findings of music perception, which are relevant for discussing the studies reported in the thesis. Detailed treatment of music perception can be found in the books edited by Riess Jones, Fay and Popper (2010) and Deliège and Sloboda (1997); Krumhansl (2000) provides a shorter introduction to music perception. In accord with the aims and structure of the thesis the developmental aspects of music perception are highlighted; some of the related electrophysiological results are discussed later in *Sections 1.3.1 and 1.3.2.*

1.2.1. Pitch and melody

Musical sounds are generally complex tones produced by the vocal tract of a singer or by an instrument. The spectral energy distribution peaks at the fundamental frequency and at integer multiples of it, termed harmonics. The perceived pitch extracted by the auditory system corresponds to the fundamental frequency. The extraction of the fundamental frequency however is not straightforward. This is indicated by the ability to infer the fundamental frequency even when it is removed from a sound, termed the missing fundamental phenomenon (Fastl & Zwicker, 2007, p. 123.). The ontogenetically earliest signs of missing fundamental extraction have been observed at the age of 4 months (He & Trainor, 2009). However, newborns show clearer responses to spectrally rich stimuli than to pure tones (Kushnerenko et al., 2007), which may indicate the use of additional spectral information. Fetuses can discriminate different pitches (Hepper & Shahidullah, 1994) when they are at least half an octave apart (Draganova et al., 2007). Perceptual pitch resolution reaches the smallest difference used in Western music, the semitone (6% difference), by the age of 3 months (Olsho, Koch & Halpin, 1987). These results suggest that adult-like musical pitch perception is possible by the age of 4 months and even newborns have the abilities to access pitch information in typical Western music.

A sequence of tones forms a melody which is the basis of establishing the identity of a song. For example the song *Happy birthday to you* can be recognized across a broad range of starting notes and tempos. It would require enormous effort to memorize the absolute pitches for just this one song with all its possible transpositions, so that the song's identity can be established in a reliable fashion if sung by a bass or soprano voice. Some birds utilize absolute pitch encoding (Cynx, 1995) for storing a finite set of songs, however most adults (Ross, Olson & Marks, 2004; Levitin & Rogers, 2005) and infants (Plantinga & Trainor, 2005; Plantinga & Trainor, 2008) are not too good at encoding absolute pitch. Humans encode melodic information both as a sequence of rising and falling pitch, called the melodic contour, as well as the size of the intervals between subsequent sounds, called relative pitch (Dowling, 1978). Contour and interval representations are separate and both play a role in recognizing melodies (Edworthy, 1985). Infants as young as 2 months old can recognize a melody after a short familiarization (Plantinga & Trainor, 2009). Interval information enables the recognition of transpositions of melodies, both in adults (Cuddy & Cohen, 1976) and in infants (Plantinga & Trainor, 2005, Trainor & Trehub, 1992, Chang & Trehub, 1977a), but infants can also discriminate sound sequences based on contour alone (Trehub, Thorpe & Morrongiello, 1987). A general preference for consonant intervals (corresponding to small integer ratios, e.g. 4:3-perfect fourth, 2:1-octave) over dissonant intervals (corresponding to large integer ratios, e.g. 15:8-major seventh, 16:17-minor second) and better detection of dissonant intervals among consonant intervals, compared to the reverse case in infants and adults (Schellenberg & Trainor, 1996; Trainor, 1997; Trainor, Tsang & Cheung, 2002) hint at the possibility of innate preference for consonant intervals. This possibility is strengthened by results showing that enculturation is indeed needed for the acquisition of another special set of intervals corresponding to the Western major and minor scales (Trainor & Trehub, 1992).

Pitch was found to be represented independently from timbre in memory (Semal & Demany, 1991, 1993; Krumhansl & Iverson, 1992, Experiments 2 and 3), although some studies showed interactions. These interactions however appeared in cases where isolated tones were presented (Melara & Marks, 1990; Singh & Hirsh, 1992; Krumhansl & Iverson, 1992, Experiment 1; Pitt, 1994) or the pitch difference was below one semitone (Singh & Hirsh, 1992; Warrier & Zatorre, 2002) which renders these findings less useful for music perception. Trained musicians were less affected by timbre in their judgments of pitch (Pitt, 1994), further supporting the separability of pitch and timbre, at least in the context of music.

1.2.2. Timbre

The accepted standard definition of timbre is to a large part negative, specifying timbre as the property differentiating between sounds which are equal in pitch, loudness and duration (American Standards Association, 1960). This definition allows one to use timbre to differentiate for example between instruments playing the same note, however tells us more about what is not timbre than what it is. Further specification added to this definition rather adds to the confusion by stating that timbre is dependent mostly on spectral parameters of the sound, but also on the waveform, sound pressure and temporal characteristics (American Standards Association, 1960), because these are the defining properties of pitch, loudness and duration, respectively. This problem has led to timbre being referred to as a "multidimensional waste-basket category" (McAdams & Bregman, 1979, p. 34.). Indeed timbre has been shown to be a multidimensional property by psychophysical identification, verbal and dissimilarity ratings. The inherent difficulty to experimentally control multidimensional attributes as well as the lack of good operational definitions has led to timbre becoming a less studied property of sound (Hajda et al., 1997).

Dissimilarity rating studies (e.g. Grey, 1977; McAdams et al., 1995; and other studies reviewed in Hajda et al., 1997; Handel, 2006) used multidimensional scaling methods to analyze the complex relationship between acoustic parameters and timbre. Based on the ratings, a perceptual timbre space can be defined by representing the perceived dissimilarity as spatial distance with usually 2 to 4 dimensions enabling the best fit. Dimensions of the perceived timbre space can then be correlated with acoustic parameters (Caclin et al., 2006). The two parameters consistently reported to correlate with perceptual dimensions are attack time (more precisely its logarithm) and spectral centroid[2]. Additionally, in a well controlled study, the fine spectral structure was also found to correlate with one of the perceptual dimensions (Caclin et al., 2005).

A different approach to timbre perception (Patterson, Gaudrain & Walters, 2010) defines the sources of information specifying a sound. The source-filter model of sound production posits that sounds hold information about: (1) The *source* of regular acoustic pulses (e.g. a string on a violin or the vocal cords) that determines fundamental frequency and harmonic content of a sound. (2) The *filter*, that is the collective name of the resonances produced by the source, for example in the body of an instrument or in the vocal tract, determining the shape of the spectral envelope[3]. It has been shown that the scale of the source and the scale filter[4] can be manipulated within a sound to modify the perceived size of the instrument or person producing that sound (Smith et al., 2005; van Dinther & Patterson, 2006). This model of the information content of a sound proposes that the pitch of a sound is determined by the scale of the source; the identity of a sound (e.g. as being a member of a family of instruments or a formant defining a speech sound) is determined by the shape of the spectral envelope; and the perceived size of the sound-source (e.g. coming from a large or small instrument or

---

[2] Attack time is usually the time between onset and maximal amplitude of a sound. Spectral centroid is the weighted mean of spectrum energy. (see Handel, 2006, p. 349. and 350. respectively)

[3] This is a simplification. The shape of the spectral envelope depends on how energy is distributed over the harmonics, of which the filter is an important determinant, but not the only one.

[4] The scale of the source and the scale of the filter are directly related to physical variables. For example, for human voices the scale of the source and the filter can be linked to glottal pulse rate and vocal tract length (Patterson, Gaudrain & Walters, 2010).

person) is determined by both the scale of the source and the scale of the filter (Patterson, Gaudrain & Walters, 2010).

Although only a small number of studies investigated timbre processing in infants, it has been shown, that by the last quarter of the firs year they can process features relevant for timbre perception (Tsang & Trainor, 2002), have long-term memory for timbre (Trainor, Wu & Tsang, 2004), and are able to categorize sounds based on their timbre (Trehub, Endman & Thorpe, 1990).

1.2.3. Rhythm, grouping and meter

Rhythm, in the sense used for describing perception of music (as opposed to musical notation), is the temporal organization of perceived events and rests. According to the generative theory of tonal music (Lerdahl & Jackendoff, 1983; Jackendoff & Lerdahl, 2006), rhythm is the product of two independent constituent structures: grouping and meter.

Grouping can be defined as the grouping or concatenation of adjacent notes into a single unit or group. Grouping is based on cues derived from intensity, duration, or pitch variation, as well as from inter–sound-onset intervals. Grouping can be usually traced back to *Gestalt* principles of perceptual organization. The cues for grouping structure are not absolute and can either strengthen or contradict each other resulting in stable or ambiguous groupings. The smaller groups or fragments can form larger groups of phrases, which in turn can be grouped to sections, eventually describing the grouping structure of an entire musical piece. Grouping structure is an important cue for the memory of rhythm (Handel, 1998), and even young infants are able to perceive simple forms of grouping (Demany, McKenzie & Vurpillot, 1977; Chang & Trehub, 1977b). However grouping seems also to be affected by culture (Iversen, Patel & Ohgushi, 2008).

The basic unit of metrical structure is the beat, that usually, but not necessarily, corresponds to the onset of notes. Beats are perceived in periodic, isochronous, time intervals, that allows synchronization e.g. in the form of tapping. Meter is the hierarchical organization of beats that correspond to different time scales, with lower levels of the hierarchy corresponding to faster tempos. Beats aligned on the different levels of the metrical hierarchy are perceived stronger or more accented; other perceptual cues can also accent beats (Hannon et al., 2004). In Western music, the higher hierarchical levels of meter are generally subdivided into two or three equal parts, setting the ratio between the period of the higher level and the lower level as 2:1 or 3:1. Typical examples are marches and waltzes respectively. Higher integer ratios and unequal subdivisions (e.g. 7:4) are possible and used extensively e.g. in jazz and Balkan folk music. However, listeners not familiar with this type of music often give simplified interpretations of the meter (Hannon & Trehub, 2005). Metrical structure is extracted from stimulation based on perceptual cues and top-down information available to listeners. It is not determined by either alone, as illustrated by e.g., syncopation, in which a strong sense of beat arises on a silent interval between notes. The extraction of meter is important for music as it creates the expectation structure that allows coordination between individuals playing music or dancing.

Even musically untrained adults have a sense of meter and are able to tap to music (Repp, 2005). They do this with similar precision as 7 year old children (Drake, 1993). The perception of metrical structure is only possible within constraints on the tempo (SOA) of the sound sequence ranging from ca. 200 ms to 2 s (London, 2002; Bolton, 1894), with the preferred tempo being 600 ms for adults and somewhat lower for infants and young children (Drake, Jones & Baruch, 2000). Rough discrimination between tempos is already available at the age of 2 months (Baruch & Drake, 1997). Infants are able to perceive meter violations even in the complex meter characteristic of Balkan folk music. This ability, however, is only preserved in adults exposed to these complex meters. This finding suggests that enculturation plays a role in the perception of meter (Hannon & Trehub, 2005). Infants are also able to derive simple metric categories (march or waltz) based on the statistical properties of stimulus sound sequences (Hannon & Johnson, 2005). The perception of these metrical categories is biased by vestibular information in both infants and adults (Philips-Silver & Trainor, 2005, 2007, 2008), which is only a small example of auditory-motor interactions in music perception and performance (for a review of these interactions see Zatorre, Chen & Penhune, 2007).

1.3. The mismatch negativity

In this chapter the current theories of interpreting the auditory ERP component mismatch negativity (MMN) are introduced. The studies presented in this thesis utilized MMN paradigms. Because four out of the five studies included in the thesis were carried out on newborn infants, the literature on the development of the MMN is reviewed in *Section 1.3.1*. The focus will be on studies utilizing MMN to probe music perception in *Section 1.3.2*.

For a general introduction to EEG and ERP theory and methods see Zani and Proverbio (2002) Fabiani, Gratton & Federmeier (2007). This section relies mainly on reviews of MMN in general (Näätänen et al., 2007, Kujala, Tervaniemi & Schröger, 2007; Winkler 2007), the development of MMN (Kushnerenko, 2003; Čeponienė, 2001; Cheour, Leppänen & Kraus, 2000; He, Hotson & Trainor, 2007) and MMN in music perception (Tervaniemi, 2006; Tervaniemi & Brattico, 2004; Tervianemi & Huotilainen, 2003).

MMN is an ERP component[5] elicited by sounds that violate some regularity present in the preceding sound sequence. The classic MMN experiments used the auditory oddball paradigm, in which a repeating sound (often called standard) is from time to time exchanged for another sound (often called deviant). The MMN appears as a negative deflection in the ERP waveform peaking 100–250 ms after the onset of the deviation with a fronto–central maximum over the scalp. The primary MMN generator is located in auditory cortex. Because

---

[5] A simultaneous event related magnetic field, usually noted as MMNm, can be measured using magnetoencephalography. (Hari et al., 1984; Alho, 1995)

of this, MMN often appears with a reversed polarity over the other side of the Sylvian fissure, e.g., at the mastoids. Another generator of the MMN is located in frontal cortex. To mitigate the effect of overlapping exogenous components seen on ERPs elicited by the deviant sounds, usually, the ERP elicited by control sounds is subtracted from the deviant response. The control sounds are ideally physically identical to the deviant sounds and are presented with the same sequential probability and temporal distribution (For details see Kujala, Tervaniemi & Schröger, 2007).

MMN is an ERP response elicited by sounds deviating from some acoustic regularity of a sound sequence. Initially it was considered as reflecting deviation from an auditory memory trace established by a repeated sound (Näätänen, Gaillard & Mäntysalo in 1978). The auditory memory trace was thought to correspond to auditory sensory memory as observed in behavioral studies (Cowan, 1984; Näätänen, 1990; Winkler et al., 1990). However since then a large body of evidence has accumulated that ties MMN to representing auditory regularities (Winkler, 2007; Winkler & Czigler, 1998; Winkler, Karmos & Näätänen, 1996). These regularity representations encode not only features of individual sounds, but also complex inter–sound relationships. For example, rule representation was shown by the study of Saarinen et al. (1992), who found that MMN was elicited by infrequent sound pairs in which the direction of frequency or intensity change within the pair was reversed compared to most of the pairs, while the absolute frequencies and intensities randomly varied (for similar results see Paavilainen et al., 2003). In another study, Paavilainen, Arajärvi, and Takegata (2007) constructed a stimulus sequence in which one stimulus feature (duration) predicted another feature (pitch) of the next stimulus. Occasional deviant stimuli, violating the above rules, elicited MMN. The latter experiment also points to the predictive nature of the memory underlying MMN generation. The formation of predictions was directly tested by Bendixen, Schröger & Winkler (2009) who contrasted omissions of (a) sounds that were fully predictable based on the preceding sounds, (b) sounds whose timing was predictable, but some other features were not and (c) sounds whose features could be discovered from the subsequent sound. They found that ERPs to the omission of fully predictable sounds were very similar to ERPs elicited by the actual sounds up to ca. 50 ms from the excepted onset. These results fit well with computational models of processing expected sounds in the human auditory system (Friston & Kiebel, 2009).

The generation of MMN has been initially assumed to be an automatic process that does not require attention. However MMN can be modulated by attention (Haroush et al., 2010; Näätänen et al., 1993; Woldorff et al., 1993). MMN can be elicited by deviants in a sound stream when the subjects are actively performing a task on a concurrent sound stream (Näätänen et al., 2007), but if the deviance is in the same feature in the two streams competition for processing resources occurs (Sussman, Winkler & Wang, 2003). Based on a review of attention effects on MMN, Sussman (2007) proposes a distinction between effects of attention on forming a standard, and detecting a deviant. The existence of attentional effects on the detection of deviants is not entirely clear. However, deviance detection relies on the formation of a regularity representation (a standard is needed to detect something as being deviant). Attentional effects on the formation of regularity representations can (indirectly) influence the deviance detection processes (Sussman et al., 2002a; Háden et al., 2009). Considering the above results, arguably, some functions within the complex system underlying MMN generation can be affected by attention, but MMN generation *per se* probably does not require selective attention. This relative independence from attention allows MMN to be used to study auditory processing in populations otherwise inaccessible for behavioral paradigms: for example, comatose patients, patients under anesthesia, sleeping adults (see Näätänen et al., 2007) and infants regardless of arousal or sleep stage (e.g.

Hirasawa, Kurihara & Konishi, 2002; Kushnerenko et al., 2007; Draganova et al., 2005; Čeponienė et al., 2002; Cheour, Leppänen & Kraus, 2000; see also in Section 1.3.1).

MMN is sensitive to the amount of separation between the standard and deviant stimuli. Larger separations generally elicit an MMN with higher amplitude and shorter peak latency. Effects of the magnitude of difference on MMN parameters have been shown for separation in frequency (Sams et al., 1985; Lang et al. 1990; Novitski et al., 2004) and intensity (Rinne et al., 2006), differences in spectral complexity (Tervaniemi et al. 2000), duration (Amendo & Escera, 2000), as well as in many other auditory features (see Näätänen et al., 2007). When deviants differ from the repeated standard sound in multiple features, then, depending on the combination of the features, the MMN amplitude is fully or partly additive (Wolff & Schröger 2001; Takegata et al., 2001; Paavilainen, Valppu & Näätänen, 2001). The minimal separation between standard and deviant sounds needed to elicit an MMN is about the same as the just noticeable differences (Kraus et al., 1999). Numerous studies have shown a connection between MMN and behavioral measures of discrimination sensitivity (e.g. Novitski et al., 2004; Kujala et al., 2001; Amendo & Escera, 2000; for a review see Näätänen & Alho, 1997). Novitski et al. (2004), for example, found that both the amplitude and peak latency of the MMN response to frequency deviation are strongly (r=0.8 and r=-0.71 for amplitude) and moderately (r=-0.56 and r=0.59 for latency) correlated with hit rate and reaction time data obtained for discriminating the same sounds.

MMN was used to assess discrimination abilities in newborns (e.g Čeponienė et al 2002; see Section 1.3.1 for details) and in various clinical populations such as dyslexic children and aphasic and schizophrenic patients (for a review see Näätänen, 2003). Most of the studies found absent or diminished MMN responses in tasks also showing impaired behavioral performance, but there is also some evidence for normal MMN responses accompanying poor performance (Kujala et al., 2006). It is important to mention that some studies failed to show a relationship between performance and MMN amplitude or latency (Bazana & Stelmack, 2002) MMN responses were found in the absence of conscious discrimination (Paavilainen; Arajärvi, & Takegata, 2007; van Zuijen et al., 2006; Kozou et al., 2005; Allen, Kraus and Bradlow, 2000; Bradlow et al., 1999) and, conversely, discrimination performance without MMN responses (Tervaniemi et al., 2005; Savela et al., 2003; Sussman et al., 2002a). Taken together these results suggest that early auditory processing, generating MMN, and later processing, establishing conscious discrimination, have access to the same perceptual information, but do not rely on each other and can be dissociated by factors such as attention or impairments (Kujala, Tervaniemi & Schröger, 2007)

Discrimination training can enhance MMN amplitudes as shown by Näätänen et al. (1993). Subjects were presented with a sequence composed of a repeating temporal tone pattern. The sequence also included infrequent patterns containing a minor deviation in one element of the repeating pattern. In the passive listening condition delivered at the beginning of the experiment, no MMN was elicited in some of the subjects. In those of these subjects whose discrimination performance improved during the following active discrimination training, MMN was elicited in the passive listening blocks presented at the end of the training. In contrast, no MMN was elicited after the training in those subjects, whose initial poor discrimination performance did not improve during the training blocks. Subsequently the enhancing effects of discrimination learning on MMN amplitude was shown for pure tones (Menning, Roberts & Pantev, 2000) and speech sounds (Kraus et al., 1995). These results spurred further research on how auditory expertise such as language learning (Winkler et al., 1999) or musical training (Tervaniemi et al., 2001; Koelsch, Schröger & Tervaniemi, 1999) affects discrimination sensitivity, as indexed by MMN.

Even this short introduction shows quite clearly that the MMN component can be a versatile tool for studying auditory perception and this is further elaborated as the development of MMN and it's utility for studying music is the topic of the next two sections.

1.3.1 Development of the MMN before birth and during the first year of life

Development of the mismatch negativity plays out against the background of rapid structural changes that characterize prenatal and postnatal brain development in the first years of life. All components of the auditory system can be identified by the end of the first trimester (Moore & Linthicum, 2007) and responses to auditory stimulation can be reliably evoked by the 27th week of gestation, with the frequency range and sensitivity of hearing increasing, as the fetus matures (Hepper & Shahidullah, 1994). By the time of birth the cochlea and subcortical auditory pathways are well developed and resemble their adult form in both structure and function (for reviews see Moore & Linthicum, 2007; Johnson, 2001). The majority of cortical development takes place after birth. A large number of new synapses appear that increases synaptic density well over adult levels (Huttenlocher, 1979; 1984; Levitt, 2003). This synaptic proliferation can be observed in all brain areas. Increase in synaptic density parallels functional development, reaching its maximum by 3 months in auditory cortex, at 8 months in visual cortex, whereas only during the second year of life in frontal cortex (Huttenlocher & Dabholkar 1997). Environmental input allows functional maturation of the cortex as synapses most active in processing information are strengthened while a large number of unused synapses disappear in the process of synaptic pruning (Huttenlocher & Dabholkar 1997). By full term birth, subcortical auditory pathways are fully myelinated (Moore & Linthicum, 2007), but myelin density increases up to the end of the first year (Moore, Perazzo & Braun, 1995). Myelination of cortical pathways proceeds from the primary sensory and motor areas during the first years of life, reaching completion in frontal associative areas well into adulthood (Vaughan & Kurtzberg, 1992; O'Hare & Sowell, 2008).

The rapid changes in brain structure affect the electrophysiological measures of brain activity (Eggermont, 1988). Myelination affects axonal transmission speed and plays an important role in establishing the latency of ERP signals. To a lesser degree, ERP latencies are also affected by synaptic transmission speed. This is well illustrated by the decrease in the latency of auditory brainstem responses as a function of development during the first year (Moore et al., 1996; Ponton et al., 1996). Synaptic density affects both the amplitude and topography of ERPs, but its effects are confounded by numerous other factors, such as the degree of synchronized activation, the alignment of the generator cells, the volume of activated part of the cortex, and changes in the conductance of the tissues between generators and scalp electrodes (Picton & Taylor, 2007). These maturational factors underlie the diverse findings of studies investigating MMN, discussed below.

In addition to the biological factors, the auditory environment also plays an important role in the development of the auditory system. Inside the womb, both internal (i.e. the heartbeat and voice of the mother) and environmental sounds can be heard (for a review see Lecanuet 1996). The womb acts as a low-pass filter attenuating frequencies over 300 Hz and reaching the maximum attenuation of 10–35 dB SPL for frequencies over 8000 Hz (Lecanuet 1996). External stimulation can trigger behavioral responses and heart rate changes from the second trimester, with increased responsiveness as the fetus gets older (Hepper & Shahidullah, 1994). Responses to music also show age-related effects that may indicate more complex processing closer to birth (Kisilevsky et al., 2004). Absolute hearing thresholds decrease during the first 6 months of life to 15 dB above adult thresholds (Trehub, Schneider & Endman, 1980; Olsho et al., 1988; Tharpe & Ashmead 2001) slowly reaching adult levels afterwards (Werner, 2007).

Sound discrimination based on spectral and temporal features is a basic step in auditory information processing and it is imperative to assess discrimination abilities in order to understand the development of both music cognition and linguistic skills. Evidence for fetal discrimination abilities come from behavioral studies investigating the postnatal effects of fetal sound exposure, indicating that newborns show preference to their mothers' voice (DeCasper & Fifer 1980) and language (Moon, Cooper & Fifer 1993) as well as for familiar melodies sung by the mother during pregnancy compared to new melodies (Panneton, 1985 cited in Lecanuet, 1996). The threshold for frequency discrimination decreases rapidly after birth. For tones under 4000 Hz is about 5% at 3 months, decreases below 2% by 12 months and reaches adult levels of under 1% during childhood (Olsho, Koch & Halpin, 1987; Spetner & Olsho 1990). Temporal discrimination threshold decreases from 20 ms at 6 months to 15 ms at 5 and a half years to 10 ms in adulthood (Morrongiello & Trehub, 1987).

Electrical brain responses can be conveniently measured from birth and used to study the state at birth and the development of discriminative and other, more complex processing, abilities of infants. The MMN component is well-suited for studying these abilities in infants, because it does not require the observation of behavioral responses and the sound representation inferred from MMN corresponds well to that underlying conscious sound perception in children and adults (Näätänen and Alho, 1997; Näätänen and Winkler, 1999). The MMN-like discriminative responses (from here onwards, MMN) are rather stable developmentally and they can be recorded both in awake and sleeping infants as well as in adults. There are also important discrepancies between infants and adults: e.g., the absence of a robust N1 component under the age of five (Ponton et al., 2000; Shahin, Roberts & Trainor, 2004), which can confound the MMN (see e.g., Jääskeläinen, 2004 and May & Tiitinen, 2010, but see Näätänen, Jacobsen & Winkler, 2005 and Näätänen, Kujala & Winkler, 2011).

MMN was first recorded in newborns to infrequent changes in frequency of sine-wave tones by Alho and colleagues (1990). The MMN observations during the first year of life are, however, not completely unequivocal (for reviews see Cheour, Leppänen & Kraus, 2000; He, Hotson & Trainor, 2007). The interpretation of the MMN in infants is not as clear-cut as in adults and comparison between recordings obtained from adults and infants should be made with caution, because of morphological and functional differences in the ERP responses (Kushnerenko et al., 2007). *Table 1* summarizes infant studies of frequency discrimination during the firs year of life introduced in the following paragraphs.

Using MEG and advanced artifact-suppression methods it is possible to record brain activity *in utero*. It has been established, that fetuses can not only detect sounds, but also show discriminative responses to changes in sound frequency, functionally similar to the adult MMN (Draganova et al., 2007; Draganova et al. 2005; Huotilainen et al., 2005). All three studies used an oddball paradigm with complex tones (500 Hz vs. 750 Hz), and SOAs ranging from 600-1200 ms. MMN-like deviant minus standard difference waveforms with average peak latencies between 307 and 332 ms were reported for 42-54% (Draganova et al., 2007), 80% (Draganova et al. 2005) and 70% (Huotilainen et al., 2005) of the fetuses. No significant effects of gestation age were observed on the amplitude or peak latency of the responses, although peak latency showed a tendency to decrease with increased gestational age. The percentage of infants showing differential responses significantly increases after birth (Draganova et al., 2007; Draganova et al. 2005).

Delivering sounds with the same stimulus parameters as in Huotilainen et al. (2005), Huotilainen et al. (2003) recorded auditory cortical and temporal lobe MMF (“mismatch field” response, the magnetic counterpart of MMN) responses in newborns with a mean peak latency of 262 ms. This finding was partially replicated by Sambeth et al. (2006), who also localized the MMF activity in the temporal lobe, but found longer peak latencies of 350 ms, probably due to the longer (300 ms long) sounds delivered to neonates. Field distributions showed, that electrical potentials (ERP) would display positive polarity over the frontal and central scalp (Sambeth et al., 2006)

With stimuli and procedures similar to the above reviewed experiments, Čeponienė et al. (2002) recorded EEG in neonates finding a marked fronto-centrally negative deflection in both the deviant minus standard and the deviant minus same-stimulus-control difference waveforms. A full 81% of the infants showed clear MMN responses in this study. The mean peak latency of the MMN component was 171 ms, with maximal peaks appearing at frontal and central electrodes. Using the same stimulus parameters Kushnerenko et al. (2002) found a negative deflection at frontal electrodes with a peak latency of about 190 ms, followed by a positive deflection peaking at about 250 ms in ~75% of newborns tested. Alho et al. (1990) used similar frequency differences (1000 Hz vs. 1200 Hz, sine tones) and an SOA of 610 ms. The deviant–minus-standard difference waveforms showed a fronto-central negativity peaking at about 220 ms with longer component duration (300 ms) than the typical adult MMN. Similar responses were reported by Čeponienė et al. (2000) to smaller acoustic deviance (1000 Hz vs. 1100 Hz). Somewhat longer latencies were reported by Tanaka et al. (2001), who also showed a marked decrease in MMN peak latency (from over 500 ms to 300 ms) as a function of conception age.

In contrast to the above reviewed findings of a negative difference waveform resembling the adult MMN, other studies testing responses to auditory deviance in neonates found discriminative responses of positive polarity in a somewhat later latency range. Leppänen, Eklund and Lyytinen (1997) found a frontocentrally positive response peaking in the 250–300 ms latency range when comparing the response elicited by deviant tones with those of the standard tone (1100 vs. 1300 Hz). Tones were presented with a shorter SOA (425 ms) than in the previously mentioned studies. The authors proposed that the positivity observed in this study could have been a sign of the release from refractoriness of the neuronal circuits processing the standard tone. They also pointed out a decrease in the observed positive waveform shown by most infants, which hinted at the presence of a small negative response, which could have been obscured by the larger positive difference. Using the same paradigm, Leppänen et al. (2004) established that the combined non-ERP measures of neonatal maturity (gestation age, vagal tone and heart period) accurately predicted the polarity of discriminative response. The discriminative response became more positive with increasing maturity of the infant. Fellman et al. (2004), using complex tones of 500 Hz vs. 750 Hz at a 800 ms SOA found a small negative deflection in newborns peaking at about 100 ms, a latency much shorter than any previously reported. This negative difference waveform was followed by a larger positive one peaking at about 230 ms from stimulus onset. Novitski et al. (2006) also reported frontocentral discriminative responses of positive polarity for infrequent frequency deviants of 20% magnitude at 250, 1000, and 4000 Hz with the tones delivered with 800 ms SOA. However, based on the grand-averaged difference waveforms, the reported positive peak had a latency of ~300 ms, which was, at least in the 1000-Hz condition, preceded by a negative waveform peaking at about 150 ms. It is possible that by measuring the amplitudes in intervals of 100 ms starting at stimulus onset, the narrow negative peak observable on the grand average difference waveform was obscured. Novitski et al. (2006) found no significant differences 5% deviation. This result is in accord with those obtained in behavioral studies testing frequency discrimination thresholds in the same age group (Olsho, Koch & Halpin, 1987).

Newborns showed no significant response difference or a small positive waveform when the SOA was either 450 ms or 1500 ms, but a significant negative difference was observed with SOAs of 800 ms (Cheour et al., 2002a) and 1000 ms (Hirasawa, Kurihara & Konishi, 2002). These results may be explained on one hand by immature auditory processing requiring longer (compared with adults) time for encoding the sounds and on the other hand, faster decay of the memory traces (again, compared with adults). It should be noted that the small positivity observed most clearly during quiet sleep is similar to the positivity found by

Leppänen, Eklund and Lyytinen (1997), who also delivered sounds with a short SOA during quiet sleep. Because neither study (Cheour et al., 2002a; Hirasawa, Kurihara & Konishi, 2002) employed control sound sequences, it is possible that the relatively small reduction in the positivity reported and interpreted as a weak negative mismatch response by Leppänen, Eklund and Lyytinen (1997) was also elicited in these studies. A second set of important results obtained in these two studies was that alertness (waking and sleep stages) did not significantly affect the observed MMN-like negative difference. This was replicated in a study using linguistic stimuli (Martynova, Kirjavainen & Cheour, 2003) and it also fits well with results of other studies which controlled for the general arousal state of newborn infants; most of these studies failed to show any effects of alertness on the neonatal MMN response. In 2 month olds one study (Friderici, Friedrich & Weber, 2002) showed larger positive mismatch responses to longer syllables presented among short syllables during sleep compared to awake state. Sleep stages however might have an effect on the detection of the MMN waveform within the EEG signal, because they may modulate the positive waveform obscuring the MMN elicited at short (and possibly also longer) SOAs.

Kushnerenko et al. (2007) did not find a deviance related negative difference when a 500 Hz standard was compared with both intensity (+10 dB) and frequency (750 Hz) deviants. In these conditions, a positive deflection was observed in the 150–350 ms latency range, peaking at about 250 ms from tone onset. In contrast, deviant white-noise bursts and environmental sounds embedded in series of complex tones elicited a significant negative difference in the 100–200 ms latency range, which was followed by a large positive waveform. The authors attributed these findings mainly to the immaturity of frequency tuning in newborns. They suggested that clearer responses can be obtained by presenting spectrally rich sounds, which activate more neuronal circuits compared with narrow–band stimuli.

From the studies above one can conclude that MMN-like discriminative responses can be recorded for deviation in sound frequency from most of newborns and even fetuses. Many of the failures to find an MMN-like response may be due to the low signal-to-noise ratio seen in fetuses and newborns as a result of an immature cortex, and the unfeasibility of longer recordings which would be necessary for improving the signal-to-noise ratio.

Compared to the number of studies investigating MMN to frequency deviance at birth, there are much fewer studies testing the development of the MMN during the first year of life. Among these studies is an almost complete agreement in finding a relatively small MMN at the age of 2-3 months, that becomes more robust and adult-like, showing shorter latencies and higher amplitudes with age (Kushnerenko et al., 2002; Morr et al., 2002; Fellman et al., 2004; Jing & Benasich, 2006; He, Hotson & Trainor, 2007; He, Hotson & Trainor, 2009a). Kushnerenko and colleagues (2002) reported that the elicitation of MMN was not fully consistent within individuals, appearing at birth, but disappearing between 3 and 6 months in some cases. One study failed to find a detectable MMN response to a relatively large frequency deviation (1000 Hz vs. 1200 Hz; Experiment 1 of Morr et al., 2002) in any of the age groups tested. However the same authors reported significant MMN responses (Experiment 2, Morr et al., 2002) to a much larger (1000 Hz vs. 2000 Hz) frequency difference. Most authors describe the presence of a positive wave overlapping the latency range of the MMN at 2-3 months of age (Kushnerenko et al., 2002; He, Hotson & Trainor, 2007), at 6 months of age (Kushnerenko et al., 2002) and during the entire first year of life (Morr et al., 2002). This positive wave, which gradually diminishes with maturation may explain the apparent absence of MMN responses (Kushnerenko et al., 2002; Morr et al., 2002, Experiment 1) and the smaller MMN amplitudes found at younger ages (He, Hotson & Trainor, 2007). Kushnerenko and colleagues (2002) suggests that this positivity is as an immature form of P3a, signaling involuntary orientation towards novel stimuli, whereas others (He, Hotson & Trainor, 2007; Morr et al., 2002) posit that the positive wave represents a separate discriminative process, which has a separate maturation trajectory from MMN (He, Hotson & Trainor, 2007). He, Hotson and Trainor (2009a) directly tested the effects of two different presentation rates (SOAs of 400 and 800 ms) and two magnitudes of frequency change (523.25 Hz standard compared to 554.37 Hz and 689.46 Hz deviants) on the positive discriminative wave in 2 and 4 month olds. Results showed that the positive wave is insensitive to the magnitude of frequency change and it increases with increasing the presentation rate. The latter effect is mostly due to changes in the standard-stimulus response, which suggests different refractoriness in the somewhat different neural populations encoding standards and deviants (He, Hotson & Trainor, 2009a). A recent hypothesis of the function of adult P3a (Horváth, Winkler & Bendixen, 2008), namely processing significant events signaled by change detection in the sensory system, is compatible with the assumptions above.

Table 1 Infant studies of frequency discrimination from conception to the first year of life. (n.a.: not applicable; n.s.: not significant; MMR: mismatch response) Continued on next page.

Article	Age	Method	Stimuli	SOA	Results	Polarity
Alho et al., 1990	newborn	EEG	Pure tones, 40 ms, 1000 Hz standard, 1200 Hz deviant	610 ms	negativity at 296 ms (Fz), 270 ms (Cz)	negativity
Čeponienė et al., 2000	newborn 6 months	EEG	Pure tones, 100 ms, 1000 Hz standard, 1100 Hz deviant	800 ms	negativity peaking ~200-220 ms negativity peaking ~120-150 ms	negativity
Čeponienė et al., 2002	newborn	EEG	100 ms tone, three partials, 500 Hz standard, 750 Hz deviant	800 ms	in 81% of subjects negativity at 168 ms (F3), 174 (F4)	negativity
Cheour et al., 2002a	newborn	EEG	Pure tones, 100 ms, 1000 Hz standard, 1100 Hz deviant	450 ms	no marked peaks, slow positivity in 200-700 ms range	positivity
				800 ms	no marked peaks, slow negativity in 200-700 ms range	negativity
				1500 ms	no marked peaks, slow positivity in 200-700 ms range	positivity
Draganova et al., 2005	in utero, week 33 newborn	MEG	100 ms tone, three partials, 500 Hz standard, 750 Hz deviant	900 ms, random 600-1200 ms	no effect of SOA, MMR at 321 ms in 48% of fetuses, 307 ms in 80% of newborns	n.a.
Draganova et al., 2007	in utero, week 28-36 newborn	MEG	100 ms tone, three partials, 500 Hz standard, 750 Hz deviant	random 800-1000 ms	MMR at 322 ms in 46% of fetuses, 345 ms in 56% of newborns	n.a.
Fellman et al., 2004	newborn	EEG	100 ms tone, three partials, 500 Hz standard, 750 Hz deviant	800 ms	negativity in 50-150 ms range in newborns, in 150-250 ms range at 3, 12 months, not significant in 6, 9 months; positivity in 250-350 ms range in all ages	negativity
	3 months					negativity
	6 months					negativity (n.s.)
	9 months					negativity (n.s.)
	12 months					negativity

Table 1 (continued) Infant studies of frequency discrimination from conception to the first year of life. (n.a.: not applicable; n.s.: not significant; MMR: mismatch response) Continued on next page.

Article	**Age**	**Method**	**Stimuli**	**SOA**	**Results**	**Polarity**
He, Hotson & Trainor, 2007	2 months	EEG	600 ms piano tones, 523.25 Hz standard, 740 Hz deviant	800 ms	with 3-20 Hz filter negativity in all age groups, peak latency decreases with age (214 ms, 212 ms, 199 ms, respectively); with 0.5-3 Hz filter slow positivity at 2 and 3 months in 100-400 ms range	both (negativity only on frontal channels)
	3 months					both (positivity on left frontal channels)
	4 months					negativity
He, Hotson & Trainor, 2009a	2 months	EEG	300 ms piano tone, 523.25 Hz standard, 554.37 Hz (1 ST) small deviant	400 ms and 800 ms	with 0.5-3 Hz filter faster and smaller positivity for 400 ms SOA in both age groups, with 3-20 Hz filter negativity only at 4 months, no effect of presentation rate	positivity (2 months), both (4 months)
	4 months		600 ms piano tone, 523.25 Hz standard, 554.37 Hz (1 ST) small deviant, 698.46 (6 ST) Hz large deviant	800 ms	with 0.5-3 Hz filter broad positivity in both ages, no effect of magnitude of change; with 3-20 Hz filter small frontal negativity at 2 months, negativity at 4 months increasing with amplitude of change	both
Hirasawa, Kurihara & Konishi, 2002	newborn	EEG	Pure tones, 100 ms, 1000 Hz standard, 1200 Hz deviant	450 ms	only with 1000 ms SOA prominent negativity peaking at 230 ms	no negativity
				1000 ms		negativity
				1500 ms		no negativity
Huotilainen et al., 2003	newborn	MEG	100 ms tone, three partials, 500 Hz standard, 750 Hz deviant	800 ms	MMR in 91% of subjects peaking at 247 ms	n.a.

Table 1 (continued) Infant studies of frequency discrimination from conception to the first year of life. (n.a.: not applicable; n.s.: not significant; MMR: mismatch response)

Article	Age	Method	Stimuli	SOA	Results	Polarity
Huotilainen et al., 2005	in utero, week 35-40	MEG	100 ms tone, three partials, 500 Hz standard, 750 Hz deviant	800 ms	MMR in 70% of subjects peaking at 332 ms	n.a.
Kushnerenko et al., 2002	newborn 3 months 6 months 9 months 12 months	EEG	100 ms tone, three partials, 500 Hz standard, 750 Hz deviant	800 ms	negativity in 83% of newborns, peaking at ~150 ms, grand average shows negativity in all ahes, not stable within individuals and age	negativity
Kushnerenko et al., 2007	newborn	EEG	100 ms tone, three partials, 500 Hz standard, 750 Hz deviant, +10 dB intensity deviant, white-noise deviant, novel sounds	800 ms	positivity in 150-250 range for frequency and intensity deviant, negativity in 100-200 range for white-noise deviant and novel sounds, followed by large positivity	positivity for frequency and intensity, negativity to white-noise
Leppänen et al., 2004	newborn	EEG	Pure tones, 74 ms, 1000 Hz standard, 1100 Hz deviant	425 ms	33% of subjects showed negativity, others positivity; correlated with gestation age and vagal tone (smaller = more negative)	both
Leppänen, Eklund & Lyytinen, 1997	newborn	EEG	Pure tones, 74 ms, 1000 Hz standard, 1100 Hz deviant, 1300 Hz deviant	425 ms	no marked peaks, slow positivity in 200-700 ms range	positivity

Continued on next page.

Table 1 (continued) Infant studies of frequency discrimination from conception to the first year of life. (n.a.: not applicable; n.s.: not significant; MMR: mismatch response)

Article	**Age**	**Method**	**Stimuli**	**SOA**	**Results**	**Polarity**
Morr et al., 2002, Experiment 1	2-7 months 8-12 months	EEG	Pure tones, 150 ms, 1000 Hz standard, 1200 Hz deviant	900 ms	positivity peaking ~300 ms in both age groups	positivity
Morr et al., 2002, Experiment 2	2-7 months 8-12 months	EEG	Pure tones, 150 ms, 1000 Hz standard, 2000 Hz deviant	900 ms	negativity peaking at 177 ms (Fz), 165 ms (Cz) for younger, 160 ms (Fz), 150 ms (Cz) for older	negativity
Novitski et al., 2006	newborn	EEG	100 ms pure tone, 250, 1000, 4000 Hz standards (separate blocks), 5%, 20% frequency deviants	800 ms	positivity in 200-300 interval for the 20% deviants, most marked with 1000 Hz	positivity
Sambeth et al., 2006	newborn	MEG	300 ms tone, three partials, 500 Hz standard, 750 Hz deviant, novel sounds	1000 ms	larger response for novel than deviant, no clear peaks, difference in 200-700 ms range	n.a
Tanaka et al., 2001	newborn	EEG	100 ms pure tone, 750 Hz standard, 1000 Hz deviant	1000 ms	negativity peaking at 500 ms at 35 weeks, at 300 ms at 38 weeks	negativity (?)

Discrimination abilities for temporal sound features are important for both music and language perception. *Table 2* summarizes infant studies of temporal discrimination. Two studies (Čeponienė et al., 2002; Cheour et al., 2002b) found MMN in newborns to infrequent changes in tone duration where the deviant tones were either shorter (40 ms vs. 100 ms; 100 ms vs. 200 ms) or longer (200 ms vs. 100 ms) than the standard tones. The MMN was similar to the negativities found in frequency discrimination paradigms, but with longer latencies compared to the frequency-deviation related MMN (Čeponienė et al., 2002) as well as to the duration-deviation related MMN observed in 4 year old, 8 year old, and adult subjects (Cheour et al., 2002b). Whereas MMN was not elicited in all subjects (Čeponienė et al., 2002) and an experiment using MEG (Cheour et al., 2004) failed to find consistent magnetic responses analogue to these ERP responses, these results are consistent with a report of adult-like sustained responses to sounds of different duration in newborns (Kushnerenko et al., 2001). Furthermore, adult-like processing of interval timing was found by Brannon et al. (2008) in 10 month old infants.

The temporal resolution of the auditory system was measured in 2-6 month old babies using the MMN elicited by infrequently inserting short silent gaps into sounds (Trainor et al., 2001; 2003). It has been established, that gap-detection thresholds, as indicated by MMN, are near or at adult levels by 6 months of age (Trainor et al., 2001). In contrast, silent gaps only modulated a slow positivity in 2-month olds and the majority of 3-month olds, with MMN appearing in 33% of the 3-month olds, and the percentage of babies showing MMN rising to 58% by 4 months (Trainor et al., 2003). These results point to a developmental trajectory of gap detection that is compatible with the previously described neurodevelopmental findings.

In summary, few studies tested temporal discrimination abilities during the first year of life. However we can conclude from them that even newborns already represent temporal relations between stimuli, and near-adult performance is reached during the first year of life for some of the temporal processing abilities.

Table 2 Infant studies of temporal discrimination from conception to the first year of life. (n.a.: not applicable)

Article	**Age**	**Method**	**Stimuli**	**SOA**	**Results**	**Polarity**
Brannon et al., 2008	10 months	EEG	50 ms pure tones, 1000 Hz	1550 ms standard, 425, 550, 800, 1050 ms deviant SOA, random +/- 50 ms jitter in all stimuli	negativity in 144-240 range, amplitude increases with magnitude of change	negativity
Čeponienė et al., 2002	newborn	EEG	500 Hz tone, three partials, 200 ms standard, 100 ms deviant	800 ms	in 80% of subjects negativity at 256 ms (F3), 257 (F4)	negativity
Cheour et al., 2002b	newborn	EEG	500 Hz tone, three partials, 100 ms standard, 40 ms deviant, 200 ms deviant in separate blocks	800 ms	negativity in 150-500 range for both deviants, peaks ~300 ms	negativity
Cheour et al., 2004	newborn	MEG	500 Hz tone, three partials, 100 ms standard, 40 ms deviant	800 ms	positive difference in 3 out of 4 subjects in the 200-300 ms range	n.a.
Trainor et al., 2001	6-7 months	EEG	deviant is two 2000 Hz tone pips with 4, 8, 12 ms gap between, standards without gap	800 ms	negative peak at ~220 ms for all deviants	negativity
Trainor et al., 2003	2 months 3 months 4 months	EEG	deviant is two 2000 Hz tone pips 16 ms gap between, standard without gap	800 ms	slow positivity peaking ~200 ms 31% show negativity 58% show negativity	positivity positivity both

The infant MMN responses discussed so far were elicited by deviation in simple acoustic features. MMN, however, is known to be also elicited by violations of higher-order acoustic regularities. Discriminative responses observed for violations of abstract rules in infants provided evidence for adult-like advanced sound processing abilities at birth, and also strengthened the analogy between responses seen in infants and the adult MMN by (in most cases) eliminating the possibility of explaining the differential responses in terms of refractory effects. These studies are summarized in *Table 3*. Ruusuvirta and colleagues (2003, 2004) showed that newborns represent conjunctions of auditory features discriminating between frequent and infrequent conjunctions of two or three features. This is an important prerequisite of representing auditory objects. Moreover, Winkler and colleagues (2003) found that newborns segregate sequences of interleaved high and low tones into separate streams. MMN was elicited in this study in a paradigm in which deviance could only be detected when the sequence was processed as two separate sound streams in the brain. In these studies, the discriminative responses elicited by deviants were positive waveforms.

Evidence that newborns can represent abstract inter-sound relationships was obtained by Carral et al. (2005), who replicated the previously reviewed experiment of Saarinen et al. (1992). Series of tone-pairs ascending in frequency were played to sleeping newborns. Ascending pairs were infrequently replaced by descending ones with both ascending and descending pairs chosen randomly from a set of seven distinct frequency-pairs. Deviant tone-pairs elicited a long-lasting slow positivity signaling their differential processing. He, Hotson and Trainor (2009b) also investigated the effect of direction changes in a sound pair, but used only 2 piano tones, infrequently reversing their order. They found that some 2 month olds show a clear positive discriminative response to deviant tone pairs and responses to deviant tones were generally higher in amplitude compared to standard pairs. Negative differential responses were observed to deviant pairs in both 4 month olds and adults. He and Trainor

(2009) ingeniously used the differential processing of ascending and descending tone-pairs to test whether 3 month old babies are able of extracting the missing fundamental of complex tones. These authors constructed sound pairs that could be heard as either ascending or descending depending on whether or not the missing fundamental was extracted. They found no clear evidence for detecting the missing fundamental in 3 month olds, whereas clear negative responses in 4 and 7 month olds and adults indicated an integrated adult-like pitch perception by the age of 4 months.

Table 3 Infant studies of abstract rule violation from conception to the first year of life. (n.a.: not applicable)

Article	Age	Method	Stimuli	SOA	Results	Polarity
Carral et al., 2005	newborn	EEG	ascending and descending pure tone pairs 40 ms, 50 ms gap, 7 levels between 553-1661 Hz	540 ms	slow positive wave in 50-450 ms range	positivity
He & Trainor, 2009	3 months	EEG	ascending and descending complex tone pairs, 150 ms, 50 ms gap, in 259-456 Hz range, descending deviants emerg only if missing fundamental is processed	750 ms	no significant difference	no response
	4 months				negative peak at 192 ms	negativity
	7 months				negative peak at 166 ms	negativity
He, Hotson & Trainor, 2009b	2 months	EEG	400 ms piano tones (523 Hz, 740 Hz), in either ascending or descending trials	850 ms	no significant difference	no response
	4 months				negative peak at ~250 ms	negativity
Jing & Benasich, 2006	3-12 months	EEG	complex tone pairs, all partials, 70 ms tone, 300 ms intertone interval, 100 Hz - 100 Hz standard, 100 Hz - 300 Hz deviant	1140 ms	negativity first recorded at 4 months, becomes robust at 6 months, amplitude higher with age, frontal shift in distribution with age	negativity form 6 months on
Ruusuvirta et al., 2003	newborn	EEG	50 ms tones, two frequency levels (750, 1000 Hz) and two intensity levels (50, 70 dB) form frequent and infrequent conjunctions	300 ms	slow positive wave in 50-450 ms range for infrequent conjunctions	positivity
Ruusuvirta et al., 2004	newborn	EEG	two duration levels (50, 80 ms), two frequency levels (750, 1000 Hz) and two intensity levels (50, 70 dB) form frequent and infrequent conjunctions, single features equally distributed	300 ms	slow positive wave in 50-450 ms range for infrequent conjunctions	positivity
Winkler et al., 2003	newborn	EEG	50 ms pure tones, varied in frequency (250-1986 Hz) and intensity (61-86 dB), control intensity oddball, one stream, two streams sequences	250 ms, 750 ms	in control and two stream condition, intensity deviants elicited positivity in 150-400 ms range, no difference in one stream condition	positivity

Based on the experiments outlined above, it is evident that newborns show discriminative responses to deviants violating higher-order acoustic regularities. The relationship between these responses and the MMN responses seen in older babies and adults is not clear, however. As mentioned above in the discussion of responses to frequency changes, three alternative explanations exist for the positive discriminative waveform. The first explanation suggests that the positive wave is not related to discrimination, it only overlaps the time range of MMN diminishing the assumedly present negative responses (Morr et al., 2002). This assumption does not explain the responses observed for violations of higher-order regularities. The second explanation suggests that the observed positivity as an immature form of P3a (Kushnerenko et al., 2002). P3a has been linked with orientation towards distinctive stimuli. The traditional interpretation of P3a (see, e.g., Escera et al., 2000; Friedman, Cycowicz, Gaeta, 2001) assumes that some form of discrimination based on inter-sound relationships or temporal distinctiveness precedes the processes reflected in the P3a. However, some more recent interpretations of P3a suggest that at least its first peak (Escera et al., 1998) reflects further processing of stimulus mismatch deemed significant in some regard by the processing system and not just a general shift in attention (Horváth, Winkler & Bendixen, 2008; Rinne et al., 2006). This also ties in well with the third proposed interpretation of the discriminative positive waveform, which suggests that it is separate from MMN and can be distinguished during the first months of postnatal development (He, Hotson & Trainor, 2007; He, Hotson & Trainor, 2009a). It is also important to note the parallel between previous findings for frequency discrimination (Cheour et al., 2002a; Hirasawa et al., 2002), where the low (below 1000 ms) SOAs of stimuli were accompanied by small positive responses and the experiments testing violations of higher-order regularities. All of the experiments using abstract regularities (Carral et al., 2005; He, Hotson & Trainor, 2009b; He & Trainor 2009) had either low SOA (540 ms) or low interstimulus interval (450 and 400 ms; respectively). Therefore effects of the presentation rate on the discriminative responses cannot be ruled out as a confounding factor. These effects in the experiment of Cheour and colleagues (2002a) manifest only on the responses to deviant sounds, which cannot be interpreted in the framework proposed by He, Hotson and Trainor (2009a) that separates the MMN and the positive wave based on refractoriness seen in responses to standard stimuli. Indeed, a fourth explanation proposing differential processing of deviant stimuli as a function of the SOAs may also merit further investigation. Differences in processing as a function of the SOA may stem from differences in the refractoriness of two neural populations, one reacting to deviants by generating a fronto-centrally positive and the other a negative response. To resolve the issues of MMN-like ERP responses in newborns and infants, further experiments, including a systematic exploration of different SOAs, are needed. Utilizing high density EEG and MEG recordings coupled with recent fMRI (e.g. Perani et al., 2010; Saccuman & Scifo, 2009; Seghier, Lazeyras & Huppi, 2006) and anatomical (e.g. Moore & Linthicum, 2007) findings will make source localization of MMN-like responses possible, hopefully establishing a veridical link between infant and adult MMN responses on a structural level. Another possible avenue of linking infant and adult MMN together is finding functional similarities in the processing of abstract acoustic regularities, using paradigms less dependent on physical stimulus properties.

In summary, even before birth, auditory stimuli elicit responses in the fetus/infant brain which clearly demonstrate their ability to discriminate sounds and to process higher-order auditory regularities. Whereas the responses are noisy and in some cases show marked morphological differences compared to the corresponding adult responses, it is evident that infants extract many regularities from their acoustic environment and detect the violations of said regularities. Even if one does not accept the infantile discriminative ERP responses as proper MMNs, these responses can be used to test what is stored in the auditory memory representations of infants and thus they can help to understand the developmental roots of music perception.

1.3.2 MMN as a tool for assessing music perception

Music can be seen as complex spectral information unfolding in time structured in several hierarchical levels spanning from a single sound to a complete musical piece. The greatest advantage of electrophysiological methods in studying the neural mechanisms underlying music perception is the ability to track neural changes caused by musical input with millisecond precision. Several ERP components have been found in the study of various aspects of music including MMN, ERAN/RATN, P300, N400, LPC/P600 (for reviews see Koelsch & Siebel, 2005; Tervianemi, 2001; Besson, 1999). Specifically the MMN has been used to study the automatic processing of musical pitch and timbre discrimination, melodic contours, harmony, musical expectations, meter and rhythm, often in the context of musical training and brain plasticity. Compared to the pure tones used in numerous experiments, spectrally complex sounds elicit stronger MMN responses in listeners, which may be attributed to the fact that unlike complex sounds pure tones are rarely encountered in everyday life or used in music, and, further, complex sounds carry more spectral and temporal information compared with pure tones (Tervianemi et al., 2000). In this respect, musical stimuli are well suited for MMN experiments. The aim of this section is to give a short review of relevant literature pointing out the utility of MMN in understanding music perception and processing.

Even though music, in general, relies on complex relations between sounds and the musical context built from these relations, simple oddball paradigms using isolated sounds can be utilized to study pitch discrimination abilities in musically trained and untrained subjects, providing insight to the different levels of processing involved in perceiving music. Studies have shown no differences between musicians and non-musicians in the amplitudes and latencies of MMNs elicited by deviants in pure tone sequences (Tervaniemi et al., 2006; Brattico, Näätänen & Tervaniemi, 2002; Koelsch, Schröger & Tervaniemi, 1999). Some studies found a small difference between musicians and non-musicians when using a near-threshold (>1% difference) deviance (Koelsch, Schröger & Tervaniemi, 1999). However, even this small difference vanished when complex tones with four harmonics were presented (Tervaniemi et al., 2005), whereas another study using piano tones reported slightly lower latencies in musicians than in non-musicians (Nikjeh, Lister & Frisch, 2008). Further, no MMN difference was observed between control subjects and subjects diagnosed with amusia in response to piano tones (Moreau, Jolicoeur & Peretz 2009). In contrast, musicians performed faster and more accurately on behavioral tests (Nikjeh, Lister & Frisch, 2008; Tervaniemi et al., 2006; Tervaniemi et al., 2005) and showed higher MMN amplitude increase and latency decrease when chords were used and either subjects attended the stimuli (Koelsch, Schröger & Tervaniemi, 1999) or some musical context was provided (Brattico, Näätänen & Tervaniemi, 2002). These results show similar early processing of simple pitch changes in musicians and non-musicians. Processing advantage is only apparent when additional information is available which musicians might better utilize (see, however Tervaniemi et al., 2006; Tervaniemi et al., 2005; Neuloh & Curio, 2004) or the task requires attentive processes. Non-musicians, however, can also use contextual information, although perhaps to a smaller degree compared to musicians, as they exhibit MMNs to out-of-key and out-of-tune deviants in musical contexts (Brattico et al., 2006) requiring the existence of long-term representations for the hierarchical rules of the western chromatic and diatonic scales (Krumhansl, 2000).

Melody is an abstract property of pitch sequences that is represented in contour code, the up and down pattern of changes in pitch, and interval code, intervals between successive sounds (Dowling, 1978), regardless of absolute pitch levels. MMN is sensitive to changes in the basic constituents of melodic information, namely direction and interval changes between sound pairs (see Saarinen et al., 1992 and Paavilainen et al., 1999, respectively). MMN was also reliably elicited, when contour changes and interval changes were presented as part of short melodies transposed over a set of absolute pitch values even when subjects did not attend the sounds (Trainor, McDonald & Alain, 2002; Fujioka et al., 2004; Tervaniemi et al., 2006). Interestingly, in one experiment (Tervaniemi et al., 2001), presentation of an attended stimulus block was needed before MMN was elicited. Investigation of polyphonic melodies showed that two melodies overlapping in frequency played at the same time can be represented (Fujioka et al., 2005) as two separate auditory streams (Bregman, 1990). Musicians' MMNs to melody violations generally have larger amplitude although one experiment (Tervaniemi et al., 2006) failed to show this difference. Differential responses to interval changes embedded in melodies were also shown in infants as young as 6 months old (Tew et al. 2009), but the authors suggest processing differences between adults and infants. The early appearance of the ability to process pitch contours may hint at the importance of this ability, which is not only used in musical context, but also in speech perception (Kemler Nelson et al., 1989; Thiessen, Hill & Saffran, 2005). For example native speakers of mandarin Chinese, a language in which melody contour conveys important lexical information, show higher-amplitude MMNs to pitch-contour deviants in a non-speech context compared with native English speaking musicians and non-musicians (Chandrasekaran, Krishnan & Gandour, 2009). This result highlights the effects of training on automatic melody processing.

Timbre is important for music perception. Studies have shown MMNs to sounds deviating in various properties related to timbre, for example in rise (attack) time (Lyytinen, Bloomberg & Näätänen, 1992), the spectral centroid (Toiviainen et al., 1998), and more generally for timbre categories (Tervaniemi, Winkler & Näätänen, 1997) and emotional valence attributed to timbre variations (Godyke et al., 2004). In an impeccably designed MEG experiment

Caclin and colleagues (2006) used the additive property of MMNs to multiple concurrent violations (Paavilainen, Valppu & Näätänen, 2001; Wolff & Schröger, 2001) to investigate the independence of three timbre dimensions (Caclin et al., 2005). Their reasoning was that if timbre dimensions are separately processed by anatomically different generators then the MMN amplitudes elicited by concurrent violations in different timbre dimensions should be comparable to the sum of individual violations. The authors were able to reliably separate two dimension contrasts based on MMN latency, amplitude and source localization and provide arguments for the separate processing for the third studied dimension. The results of Calcin and colleagues (2006) did not make anatomical localization of the sources possible, but another study suggests involvement of multiple areas in the superior temporal gyrus and sulcus (Menon et al., 2002). Using MMN, Vestergaard, Fyson and Patterson (2009) showed that auditory size information (see *Section 1.2.2.*) is automatically processed in adults and this ability is also present in newborns. Size information is relevant for differentiating between instruments of the same family e.g. a violin and a cello.

The temporal grouping of sounds has been investigated using MMN paradigms in which deviants were presented as part or a pattern (SSSSD) or with the same probability in random order (Sussman, Ritter & Vaughan, 1998; Sussman & Gumenyuk, 2005). Automatic grouping of sounds was found (indicated by the lack of MMN compared to the random condition) when the elements of the pattern were presented in close temporal succession (SOA=200 ms). The SOA at which grouping occurred could be extended to ca. 1 s when subjects were informed about the presence of a repeating pattern of five elements (Sussman et al., 2002b)[6]. The effect of changing the number of elements was not systematically investigated. Note that these experiments used the MMN as an indicator of automatic group formation. When behavioral measures and musical stimuli were used in active paradigms, subjects were able to detect

---

[6] It is important to note that the concept of grouping in Sussman et al. (2002) involves relatively short time periods corresponding to the basic levels of the grouping concept used in Lerdahl and Jackendoff (1983) which may involve time periods much longer than 1 s.

repeating patterns with SOAs up to 2s (for a review see London, 2002). Grouping can also be based on global probability distribution of possible patterns instead of local memory traces. In this case, the grouping process is less sensitive to the temporal proximity of elements (Herholz, Lappe & Pantev, 2009).

Interesting differences between musicians and non-musicians were found when the basis of grouping was either pitch-similarity or the “good continuation” of pitch ascension in a sequence. A deviant sound extending the common length of the sound groups elicited MMN in musicians in both conditions, but only in the pitch-similarity condition in non-musicians. This result revealed the importance of representing pitch relations in music and the advanced ability of musicians for extracting more complex information from the stimulation (van Zuijen et al., 2004). A similar discrepancy between musicians and non-musicians was found when groups were formed on the basis of a common pitch with each group being either made up of the same number of sounds or they had the same overall duration. Both musicians and non-musicians automatically detected deviations in the duration of the groups, but only musicians detected numerical deviations which may be explained by the need for efficient beat tracking in music (van Zuijen et al., 2005). These results again demonstrated training effects on stimulus-driven grouping processes.

Musicians also showed a stronger left hemispheric lateralization of the MMN to pattern violations hinting perhaps at more advanced syntactic rule-extracting abilities. This lateralization of the MMN is similar to the one Vuust et al. (2005) found in expert jazz musician to metric violations as compared to non-musician whose MMNm responses showed a right hemispheric lateralization. The authors suggest that the left hemispheric lateralization is in relation with the communicative value that meter possesses for jazz musicians (to coordinate during performances) and similar to the language-competence based lateralization of MMN during phoneme processing (e.g., Näätänen et al., 1997).

For subtle rhythmic violations no MMNm was found in non-musicians whereas a left lateralized MMNm was found in musicians. Further analysis of the dataset (Vuust et al., 2009) revealed the same effects on dipole amplitudes and localized the dipole sources in the transverse temporal gyrus near the primary auditory cortex in both hemispheres. The authors interpret their results in the predictive coding framework of auditory perception (Friston & Kiebel, 2009) and regard the effects of expertise as favoring a biocultural (Cross, 2003; Cross, 2006) concept of meter, that is "only meaningful in the interaction between music and subject" (Vuust et al., 2009, p. 90.). Not all evidence supports the advantage of musicians in the automatic detection of meter and rhythm violations. Geiser et al. (2009) found MMN responses to both rhythm and meter violations while subjects were attending to musical stimuli but only to rhythm violations in an unattended condition. Furthermore no group differences were obtained between musicians and non-musicians. The superior auditory processing abilities of musicians (revealed by behavioral results) did not appear to originate from pre-attentive processing, possibly because of the less natural stimulus material used by Geiser et al. (2009) compared with Vuust et al. (2005). Altogether, Geiser et al.'s (2009) results leave open the question of how metrical complexity affects automatic meter processing.

The examples above show that MMN is a versatile tool for investigating the early processing stages of the basic building bricks of music and provides results that can be integrated into more general interpretational frameworks for understanding music and the brain.

2. Synopsis and rationale of theses

The general aim of this thesis is to shed light on the question: Which of the abilities underlying the perception of music are functional at the time of birth. Music perception is obviously a higher cognitive function, which depends on several abilities. However, the exact nature of these dependences is far from being trivial. Music perception, similarly to other higher functions of cognition, can be seen as the product of complex interactions between innate predispositions and constraints on information processing, the physical and cultural constraints determining what information is available in the environment, and the process of learning through which quantitative and qualitative differences in information processing emerge as a function of experience over time. Notwithstanding the complex nature of music, it is possible to identify general processes of auditory perception which serve as the building blocks necessary for music perception. Both the functional architecture and the ontogenesis of these abilities provide crucial cues for understanding the role learning plays in music. By applying electrophysiological methods, sources of information inaccessible to traditional behavioral methods can be tapped. In adults, MMN paradigms allow access to processing steps with relatively low interference from attention and explicit knowledge of music. In newborn infants, similar paradigms can be used to study processing abilities functional at birth that would otherwise remain hidden behind the very limited repertoire of behavioral responses available to newborns.

Examining perception near the time of birth enables one to observe an “initial state” of the system (at least with respect to explicit learning), although one should not downplay the importance of intrauterine experiences. Knowledge of this “initial state” allows for better understanding of how music perception emerges by identifying the basic building bricks and comparisons between the different developmental stages. The innateness of specific abilities can be determined by examining infants shortly after birth. In theory, this approach could provide supporting evidence for modular, domain-specific accounts of music processing. However, most of the perceptual abilities assessed in this thesis represent more basic processes, ones possibly shared by both music and other forms of communication by sounds.

Based on the principles above, the theses examine aspects of two equally important areas of music. Theses I and II are related to the perception of melody and theses III, IV and V are related to the perception of rhythm. All theses will be presented together with a synopsis of immediate background information and supporting evidence as well as a short evaluation of the results yielded by the experiments. The published papers constituting the base of the theses are then presented in full length in the *Studies* part of the dissertation.

2.1. Thesis I: Relative pitch extraction in newborns

A sequence of rising and falling pitch steps can be interpreted as a melodic contour irrespective of the absolute pitch level and to some extent the sizes of pitch steps (Dowling, 1978; Edworthy, 1985). The ability to extract the relative sizes of pitch steps however is necessary to recognize melody in the context of both music and speech prosody. Adults are able to extract relative pitch information (Paavilainen et al., 1999; Tervaniemi et al., 2006) and it has been shown that newborns are sensitive to the direction of pitch intervals (Carral et al., 2005).

The ability of newborn infants to extract the size of pitch intervals irrespective of absolute pitch levels was tested. EEG was recorded while sequences of tone pairs with equal pitch steps were presented. The regular sequence was infrequently broken by tone pairs having a larger pitch step. All tone-pairs were presented at several randomly assigned absolute pitch levels. If newborn infants are able to extract the common size of pitch steps as an abstract regularity (in spite of variation in absolute pitch), then the infrequently presented large pitch steps should elicit an MMN-like ERP component in them. The presence of the MMN-like component would suggest that newborns are able to represent pitch intervals. MMN-like components were elicited by infrequent pitch steps, suggesting that newborns are able to process pitch interval information similarly to adults.

2.2. Thesis II: Timbre-independent extraction of pitch in newborns

Important spectral features other than pitch can be extracted from perceived sounds (e.g. McAdams et al., 1995; Handel, 2006). Adults are able to separate pitch from timbre information (Semal & Demany, 1991, 1993; Krumhansl & Iverson, 1992) and thus can identify different instruments playing the same note, or different speakers speaking the same words. The independent extraction of pitch and timbre information is important in both music perception and language acquisition.

The aim of the study was to show that newborn infants can extract pitch information independently of timbre variations. An aspect of timbre, resonator size, was randomly varied over all sounds in a simple pitch oddball design. ERPs were recorded from newborn infants. If infrequent high pitch deviants presented among low pitch standards elicit an MMN-like response despite the random variation in resonator size, this result can be taken as evidence for timbre-independent processing of pitch in newborn infants. Results indicated that newborns represent pitch and resonator size information independently of each other.

2.3. Thesis III: Auditory temporal grouping in newborns

Sounds in an auditory scene can be grouped based on higher order regularities. These regularities can convey information for example on sound sources and can be used in representing the environment by constructing auditory objects (Bregman, 1990). Previous studies showed that newborn infants represent probabilities of local sound features, that is, they produce a discriminative response to infrequent deviant sounds presented among regular standards (for reviews, see e.g. Cheour, Leppänen & Kraus, 2000; He, Hotson & Trainor, 2007). Adults are also able to extract higher order regularities from sound sequences, such as the cyclical repetition of pitch patterns (Sussman, Ritter & Vaughan, 1998; Sussman et al., 2002b; Sussman & Gumenyuk, 2005). This ability supports the construction of hierarchical sound representations, which is a necessary prerequisite for many complex functions, amongst them rhythm perception (Lerdahl & Jackendoff, 1983).

The objective of the present study was to investigate whether newborn infants form auditory groups from cyclically repeating pitch patterns. To this end, EEG was recorded while newborn infants were presenting with oddball sequences. The sequences were either random or they contained a cyclically repeating pitch pattern. The probability of the oddball sound was uniformly fixed at 20% in both types of sequences. If only local probability is taken into account by the newborn auditory system for deviance detection, then oddballs in both types of sequences should elicit an MMN-like response. If grouping occurs, then the oddballs in the grouped sequence become part of the higher order regularity and should not elicit MMN. No MMN-like response was found for oddballs in the grouped sequence, which provided important, however indirect, evidence in support of the neonatal ability to group sounds based on cyclical repetition.

2.4. Thesis IV: Processing of meter in adults

Rhythm perception can be structured by beat and meter (Lerdahl & Jackendoff, 1983). Beat induction, the ability to extract a regular beat or pulse from auditory stimuli and synchronize with that beat or pulse, underlies meter induction the ability to hierarchically order multiple regular beats. Metrical salience is assigned to specific positions in a rhythmic pattern based on the distance in the hierarchy from the downbeat (viz. the firs beat). Metrical salience shows the importance of a specific beat in defining a rhythm (Lerdahl & Jackendoff, 1983; London, 2002).

The goal of the study was to test with both behavioral and ERP measures whether adults automatically process metrical salience by constructing a rhythmic pattern and omitting beats at positions with different metrical salience. The patterns with beats omitted on significant positions were presented as oddballs in sequences consisting predominantly of patterns in which omissions occurred on less salient positions. Different levels of metrical salience were assumed to have an effect on both the sensitivity and reaction time (RT) measures in an omission-detection task. Differences in metrical salience could then also affect the latency of the MMN elicited by the omissions with no sound-related task as well as when the subjects performed a concurrent auditory task on a different sound stream. The results are not clear about the effect of attention on meter processing because in the concurrent task condition an attention modulated N2b component may overlap the MMN. Sensitivity and ERP latency data supported the differentiation based on metrical salience, whereas support from RT measures and ERP amplitude data was weaker.

2.5. Thesis V: Beat detection in newborns

Beat induction, the ability to synchronize with a regular pulse or beat, plays a significant role in meter perception (Lerdahl & Jackendoff, 1983), but may also help coordinating communicative acts (Jaffe et al., 2001). Both infants and adults can form metric categories (Hannon & Trehub, 2005) and able to synchronize with a regular beat (Drake, 1993; Repp, 2005; Philips-Silver & Trainor, 2005). However it is not clear whether this ability is also present in newborns.

The experiments aimed at showing beat induction in newborn babies. A musically plausible rhythmic pattern was created and presented to sleeping newborns, with occasional omissions in the pattern at the positions with the highest (downbeat) and lowest metrical salience. If beat induction occurs in newborns, higher-amplitude MMN-like components should be elicited by omissions from the pattern at the most metrically salient as compared with the least salient positions. Responses were also compared with a physically identical control. Results of the experiments were in full accord with the notion of beat induction occurring in newborns. Results of a control experiment conducted in adults ruled out a possible alternative explanation of the neonate results. The alternative explanation was based on the assumption that the sounds of different instruments making up the test sound sequence were segregated to separate streams and sound omissions were separately detected for the different streams. However, no such omission responses were obtained in the adult participants.

3. Studies

3.1. Study I: Relative pitch extraction in newborns

Stefanics, G., Háden, G.P., Sziller, I., Balázs, L., Beke, A., Winkler, I. (2009) Newborn infants process pitch intervals. *Clinical Neurophysiology* 120. 304-308. DOI: 10.1016/j.clinph.2008.11.020

Clinical Neurophysiology 120 (2009) 304–308

Contents lists available at ScienceDirect

Clinical Neurophysiology

journal homepage: www.elsevier.com/locate/clinph

Newborn infants process pitch intervals

Gábor Stefanics [a,b], Gábor P. Háden [a,c], István Sziller [d], László Balázs [a], Anna Beke [d], István Winkler [a,e,*]

[a] Institute for Psychology, Hungarian Academy of Sciences, P.O. Box 398, Szondi u. 83–85, H-1394 Budapest, Hungary
[b] Department of Experimental Zoology and Neurobiology, University of Pécs, Hungary
[c] Department of Cognitive Science, Budapest University of Technology and Economics, Budapest, Hungary
[d] First Department of Obstetrics and Gynecology, Semmelweis University, Hungary
[e] Institute of Psychology, University of Szeged, Hungary

ARTICLE INFO

Article history:
Accepted 23 November 2008
Available online 7 January 2009

Keywords:
Pitch perception
Relative pitch
Development
Event-related brain potentials (ERP)
Neonates
Mismatch negativity (MMN)

ABSTRACT

Objective: We investigated whether the auditory system of newborn babies extracts the constancy of a pitch interval from exemplars varying in absolute pitch.
Methods: Event-related brain potentials (ERP) were recorded from healthy newborn infants in an oddball paradigm consisting of frequent standard and infrequent deviant tone pairs. Tone pairs varied in absolute frequency. Standard and deviant pairs differed in the amount of pitch difference within the pairs, but not in the direction of pitch change.
Results: Deviant tone pairs elicited a discriminative ERP response.
Conclusions: This result suggests that the neonate auditory system represents pitch intervals similarly to adults.
Significance: Adult-like processing of pitch intervals allows newborn infants to learn music, speech prosody, and to process various important auditory cues based on spectral acoustic features.

1. Introduction

From the early days of life, humans are strongly interested in music (Trehub, 2003). Speech directed to pre-linguistic infants is characteristically melodious (Fernald, 1991) and infants at 6 months of age are more attentive to maternal singing than to maternal speech (Nakata and Trehub, 2004). The spectral aspects of music include the melodic contour, the sequence of rising and falling pitch steps irrespective of their exact size and the actual pitch intervals (Edworthy, 1985; Peretz and Babaï, 1992). Pitch-change direction, the basic building brick of melodic contour, is extracted from sound pairs of varying absolute pitch in adults (Saarinen et al., 1992) and in newborn infants (Carral et al., 2005a) as well as from longer sound sequences of uniform pitch-change direction (Tervaniemi et al., 1994). In adults, melodic contour is extracted from transposed presentations of the same melody (Tervaniemi et al., 2001; Tervaniemi et al., 2001a) and there are indications that also the size of pitch intervals can be extracted from exemplars varying in absolute pitch (Fujioka et al., 2004; Paavilainen et al., 1999; Tervaniemi et al., 2006). Fujioka and colleagues (2004) also found that the detection of change in melodic contour and pitch intervals is enhanced by musical training. Furthermore, 8-month olds were shown to detect 1-semitone changes in short melodies, irrespective of whether the change resulted in violating the key or not, as opposed to adults, who performed better on out-of-key trials (Trainor and Trehub, 1992). However, little is known about whether or not the neonate auditory system already possesses the same ability. Therefore, the question asked in the current study was whether the neonate auditory system processes pitch intervals (relative pitch) similarly to adults, thus providing a basis for adult-like perception of melodies.

The mismatch negativity (MMN) event-related potential provides a way to study the processing of pitch intervals that can be applied both in adults and in newborn babies, allowing direct comparison between these two widely separated developmental stages. This is because MMN is an electrophysiological brain response, which is elicited when the violation of an acoustic regularity has been detected in the brain whether or not participants focus their attention on the sounds (Näätänen et al., 1978; for recent reviews, see Kujala et al., 2007; Näätänen et al., 2007). An MMN-like discriminative response was observed in sleeping neonates (Alho et al., 1990; for a recent review, see Csibra et al., 2008) and has been previously used to study auditory frequency discrimination (Novitski et al., 2007) and regularity extraction (Stefanics et al., 2007). MMN studies of pitch contour and pitch intervals were based on the observation that sounds delivered in close succession and separated from other sounds by relatively longer silent intervals are treated by the

---

[*] Corresponding author. Address: Institute for Psychology, Hungarian Academy of Sciences, P.O. Box 398, Szondi u. 83–85, H-1394 Budapest, Hungary. Tel.: +36 1 3542296; fax: +36 2 3542416.
E-mail address: iwinkler@cogpsyphy.hu (I. Winkler).

1388-2457/$34.00 © 2008 International Federation of Clinical Neurophysiology. Published by Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.clinph.2008.11.020

auditory system as relatively encapsulated units (Müller and Schröger, 2007; Takegata et al., 2005). Using this principle, several studies presented adult participants with series of tone pairs in which the direction of the within-pair frequency change was kept constant (either always ascending or descending) while absolute tone frequency was randomly varied (Carral et al., 2005b; Korzyukov et al., 2003; Paavilainen et al., 1995, 1998, 1999, 2003; Saarinen et al., 1992). Occasional tone-pairs with the opposite within-pair frequency change direction, or no frequency change evoked the MMN response. The MMN component elicited in the frequency-change direction paradigm has been found to be insensitive to the magnitude of the deviation (Carral et al., 2005b). This result suggests that frequency-change direction was extracted and represented irrespective of the size of the change (i.e., the pitch interval). Using the same method, processing of frequency-change direction has been shown in 8- to 14-year-old children (Gumenyuk et al., 2003) as well as in newborn babies (Carral et al., 2005a). Studies using transposed melodies instead of tone pairs (Fujioka et al., 2004; Tervaniemi et al., 2001, 2006) and even varying individual pitch steps within the melodies (Trainor et al., 2002b) extended the frequency-change direction results to melodic contours. The result obtained with ERP methods are fully compatible with those obtained in behavioral studies, which demonstrated the processing of melodic contour in adults (e.g., Cuddy and Cohen, 1976; Tervaniemi et al., 2001; Trainor et al., 2002a,b) and in infants of 3 months (Demany and Armand, 1984), 8–11 months (Trehub et al., 1984) and 9–11 months of age (Trehub et al., 1987).

However, accurate perception of relative pitch intervals is also necessary for music perception (Dowling, 1978; Fujioka et al., 2004; Trainor and Trehub, 1992). Paavilainen and his colleagues (1999) tested adults with a modified tone-pair paradigm in which standard pairs included the same pitch interval (at varying absolute frequencies) and deviants differed only in the size of the interval, but not in the direction of frequency change. MMN elicitation by deviants in this paradigm indicated that the constancy of pitch intervals was extracted from the sequence of pairs. Furthermore, Trainor and colleagues (2002), Fujioka and colleagues (2004), and Tervaniemi and colleagues (2006) found that occasional changes in the final pitch interval (without changing the pitch contour) in repetitive short transposed melodies also elicited the MMN. However, some studies suggested that detecting changes in pitch intervals is more difficult than that for the melodic contour (Cuddy and Cohen, 1976; Dowling, 1978). Furthermore, accurate representations for pitch intervals may develop later in life than that for melodic contour (Dowling, 1982). In contrast, one study in 3-month old infants showed that they display novelty reactions to variation in the size of pitch intervals in simple three-tone melodies, suggesting that infants also represent the size of pitch intervals (Demany and Armand, 1984). Similar sensitivity to changes in pitch intervals was observed in 8-month old infants using somewhat longer melodies (Trainor and Trehub, 1992). The current study tested whether the newborn auditory system extracts pitch-interval invariance from tone pairs varying in absolute frequency. The intervals used in the current study were based on the interval-deviant condition of Tervaniemi and colleagues (2006), whereas the paradigm is generally similar to that of Paavilainen et al. (1999, Experiments 2a and 3a). Elicitation of a discriminative MMN-like response by occasional pitch-interval deviants would suggest the existence of adult-like representation for pitch intervals at birth.

2. Materials and methods

2.1. Participants

ERP responses were recorded from 15 (5 female) healthy full-term newborn infants on day 2 or 3 postpartum. Their gestational age was 38–41 weeks, birth weight 2900–5400 g and Apgar score 9/10. All infants had normal hearing tested by otoacustic emissions as part of the routine hospital procedures. The recordings were carried out in the hospital ward of the First Department of Obstetrics and Gynecology, Semmelweis University, Budapest, Hungary after informed consent was obtained from one or both parents. The mother of the infant was present during the recording. The study was approved by the Ethics Committee of the Semmelweis University as well as by the Institutional Review Board of the Institute for Psychology, Hungarian Academy of Sciences. One infant's data could not be used due to too few responses in some stimulus categories and an additional four infants' data were discarded due to excessive electric artifacts. Of the remaining 10 infants, three were female.

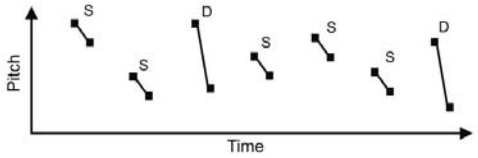

Fig. 1. Schematic illustration of the stimulus paradigm. Tone pairs are represented by connected black squares. The x-axis represents time, the y-axis is scaled in pitch (i.e., in semitones). Standard pairs (2-ST descending pitch intervals) are marked by "S", deviant pairs (7-ST descending pitch intervals) by "D".

2.2. Stimuli and procedure

Sleeping newborns were presented with pairs of sinusoidal tones (50-ms duration with 5-ms rise and 5-ms fall times, 75 dB sound pressure level) descending in frequency. Fig. 1 shows a schematic illustration of the stimulus paradigm. The within-pair interval was 150 ms, the between-pair interval 1250 ms. Standard (small pitch interval, 2 semitones [ST]) and deviant pairs (big pitch interval, 7 ST) differed in the size of the descending pitch step. Deviant pairs equiprobably varied on six absolute frequency levels, separated from each other by 1-ST steps. Standard pairs varied on 11 absolute frequency levels, six of which had the first tone identical to those of the deviant pairs, and six had its second tone identical to those of the deviant pairs (one standard pair had both tones appearing also in deviants; see Table 1 for the exact tone frequencies of all pairs). Thus neither the first nor the second tone of the deviant pairs was unique to deviant pairs; both appeared also in standard pairs. The tone pairs were delivered in a pseudo-randomized order constrained by the requirement of at least 3 standard pairs separating consecutive deviant pairs and consecutive pairs starting at different frequencies. In five stimulus blocks, a total of 1000 tone pairs were delivered of which 150 were deviants ($p = 0.15$) and 850 standards ($p = 0.85$). Each stimulus block began

Table 1
Tone frequencies (in Hz) of the standard (small step) and deviant (big step) pairs.

First tone	Small step (2 ST)	Big step (7 ST)
494	440	
523	466	
554	494	
587	523	
622	554	
659	587	440
698	622	466
740	659	494
784	698	523
830	740	554
880	784	587

with 6 standard pairs. In a control stimulus block presented at random position amongst the five main experimental blocks, 300 tone pairs with the big pitch interval were delivered alone, providing identical-stimulus comparison for the deviant-pair responses. Although the interval separating consecutive pairs with the 7-ST interval differs between the main and the control experimental blocks, the overall inter-pair interval was the same across the two types of sequences. Thus any difference between the ERP responses elicited by deviant and the control tone pairs can be attributed to the specific pitch interval of the deviant, whether by detecting them as deviants within the oddball sequences or by processing them separately from the small pitch intervals and thus becoming sensitive to the temporal density with which these pairs were encountered. In either case, finding a difference between the deviant and the control ERP responses suggests that the neonate auditory system represents different pitch intervals.

Sounds were presented binaurally using the E-Prime stimulus presentation software (Psychology Software Tools, Inc., Pittsburgh, PA, USA) through ER-3A loudspeakers (EtymStic Research, Inc., Elk Grove Village, IL, USA) connected via sound tubes to self-adhesive ear-couplers (Natus Medical, Inc., San Carlos, CA, USA) placed over the babies' ears.

2.3. EEG recording and data analysis

EEG was recorded from electrodes placed at the F3, Fz and F4 scalp locations (set according to the international 10–20 system) with the common reference attached to the tip of the nose. The ground electrode was placed on the forehead. Eye movements were monitored by recording the electrooculogram (EOG) between two electrodes, one placed below the left and another above the right eye. EEG was recorded at 24 bit resolution and a sampling rate of 250 Hz by a direct-coupled amplifier (V-Amp, Brain Products, Munich, Germany). The signals were on-line low-pass filtered at 110 Hz.

EEG was filtered off-line between 1.5 and 16 Hz. For each tone pair, an epoch of 600-ms duration was extracted from the continuous EEG record. The epochs started 100 ms before the onset of the second tone of the pair. Epochs with a voltage change below 0.1 μV or above 75 μV on any EEG or EOG channel were rejected from further analysis. Epochs were baseline-corrected for the 100-ms prestimulus period and averaged separately for standard, deviant, and control pairs. The mean number of artifact-free deviant trials per infant was 133 with a corresponding number of standard responses collected in the control condition.

The ratio of quiet and active sleep was assessed separately for each stimulus block, following the criteria suggested by Anders et al. (1971), using EEG and EOG measures and observations of the infant's behavior. Periods of wakefulness and indeterminate stages were excluded from this analysis. Overall, the infants were more often in active (53% of the time) than in quiet sleep (26% of the time) during the recordings. This result is in line with the similar observations of Novitski and colleagues (2007). The ratio of quiet and active sleep did not significantly differ between the test and the control stimulus blocks ($t(11) = 0.87$, $p = 0.4$).

For amplitude measurements, a 40-ms long time window was selected from the group-averaged deviant-minus-control difference waveform (Fig. 1) at the Fz electrode. The window was centered on the negative peak observed at 212 ms from the onset of the second tone in the group-averaged difference waveform. The effects of stimulus type were analyzed with three-way repeated-measures analyses of variance (ANOVA: Stimulus type [Deviant vs. Control] × Lateral position [Left vs. Central vs. Right]). Greenhouse–Geisser correction of the degrees of freedom was applied where appropriate.

3. Results and discussion

Fig. 2 shows the grand-averaged ERP responses elicited by the second tone of deviant and identical control tone pairs at the three frontal scalp locations (F3, Fz, and F4) together with the corresponding deviant-minus-control difference waveforms. The gray-shaded areas show the time window selected for the amplitude measurements. The ANOVA for the negative difference wave yielded a significant main effect of stimulus-type ($F(1,9) = 5.29$, $p < 0.05$, $\eta^2 = 0.37$). No other main effect or interaction was significant.

The ERP responses elicited by deviant tone-pairs were significantly different from those elicited by control tone-pairs having identical pitch intervals. This result demonstrates that the neonate auditory system shows adult-like capabilities in representing pitch intervals and detects deviations from a constant interval irrespective of variation in absolute frequency (for compatible results in adults, see Fujioka et al., 2004; Paavilainen et al., 1999; Tervaniemi et al., 2006; Trainor et al., 2002a,b). The discriminative ERP response could not stem from infrequent tones appearing only in the deviant pairs, because all tones used in composing the deviants also appear in the same position in one of the standards. An alternative interpretation of the current and similar results in adults is that humans are sensitive to the harmonic dissonance (perceptual roughness) of tone pairs, just as adults are sensitive to this feature in chords (Bigand et al., 1996; Lerdahl and Jackendoff, 1983). In the current study, standards had a dissonant, whereas deviants a consonant interval. The difference in consonance would explain the contrast between results suggesting that forming accurate representation for pitch intervals may be difficult (also for adults: Cuddy and Cohen, 1976; Dowling, 1978) and those showing that certain pairs of intervals are discriminated both by adults and infants even in the absence of musical training or attention focused on the

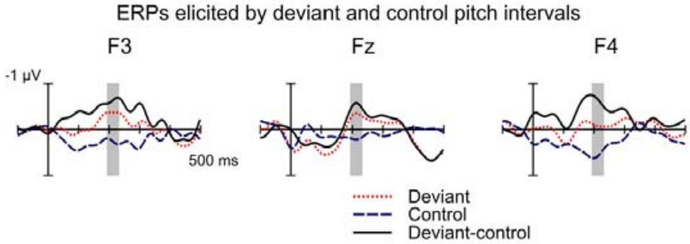

Fig. 2. Group-averaged ERP responses elicited by deviant (dotted line) and control (dashed line) pitch pairs at three frontal electrode locations (F3, Fz, and F4). Deviant-minus-control difference waveforms are plotted in black (solid line). The onset of the second tone of the pairs is at the crossing of the axes. Gray areas mark the time window selected for amplitude measurements (192–232 ms).

sounds (the current results and Demany and Armand, 1984; Fujioka et al., 2004; Paavilainen et al., 1999; Tervaniemi et al., 2006; Trainor et al., 2002a,b). For example, Trainor and her colleagues (2002b) found preference for consonant compared with dissonant two-tone intervals in 2- and 4-month-old infants, suggesting that they are sensitive to interval consonance even before learning musical scales. This alternative interpretation does not argue against the conclusion that newborn babies process pitch intervals similarly to adults. The studies showing good discrimination performance in adults also contrasted consonant and dissonant intervals (e.g., Tervaniemi et al., 2006; Trainor and Trehub, 1992). Thus it is possible that both adults and neonates rely at least partly on consonance/dissonance information in detecting changes in a sequence of transposed pitch intervals.

Carral et al.'s (2005a) results indicated that newborn infants process the direction of frequency change within tone pairs, while our recent results (Háden et al., in press) showed that neonates form separate representations for the pitch and timbre (resonator size, a timbral feature) of complex tones. Taking these and the current results together, they indicate that newborn babies process the spectral cues of melodies similarly to adults, a conclusion strongly supported by the recent finding showing that newborn infants are sensitive to the prosodic cues of speech (Sambeth et al., 2008). However, Novitski et al.'s results (2007) advise caution, as these authors demonstrated that the accuracy of pitch representation in the neonate auditory system is significantly lower than that in adults (also shown by behavioral studies in somewhat older infants, see e.g., Wormith et al., 1975).

Other neonatal studies applying the MMN method demonstrated the operation of higher-level auditory abilities in neonates, such as the segregation of concurrent streams of sound (Winkler et al., 2003), auditory feature binding (Ruusuvirta et al., 2003) and auditory temporal grouping (Stefanics et al., 2007). Whereas neonates appear to be more sensitive to surface sound features (such as spectral width or intensity) than adults and their sound categorization is slower and less elaborate than that in adults (Kushnerenko et al., 2007), it appears that most (perhaps all) basic and higher-level auditory processing abilities are functional at birth (while not necessarily as accurate as in adults). These abilities allow newborn babies to extract invariant relationships from the normally variable acoustic environment and thus to track speakers and extract messages from the composite auditory input. Therefore, these auditory functions underlie attachment to and communication with the caretaker persons as well as learning to speak and to enjoy music.

In summary, our results demonstrated the adult-like functioning of the neural mechanisms underlying the processing of pitch intervals in newborn infants.

Acknowledgments

This work was supported by the European Commission's Sixth Framework Programme for "Information Society Technologies" (Project Title: EmCAP Emergent Cognition through Active Perception, Contract No. 013123). The authors thank Prof. Dr. Zoltán Papp and Dr. János Rigó for supporting the study of neonates at the First Department of Obstetrics and Gynaecology, Semmelweis University, Budapest, Hungary and RN Erika Józsa Váradiné for collecting the data. We are also grateful to Dr. Laurel J. Trainor for finding a design flaw in an earlier version of the study and to Dr. Stefan Koelsch and Dr. Daniel Mietchen for help with interpreting the results.

References

Alho K, Saino K, Sajaniemi N, Reinikainen K, Näätänen R. Event-related brain potential of human newborns to pitch change of an acoustic stimulus. Electroencephal Clin Neurophysiol 1990;77:151–5.

Anders T, Emde R, Parmelee A. A manual of standardized terminology, techniques and criteria for scoring of states of sleep and wakefulness in newborn infants. Los Angeles: UCLA Brain Information Service; 1971.

Bigand E, Parncutt R, Lerdahl F. Perception of musical tension in short chord sequences: the influence of harmonic function, sensory dissonance, horizontal motion, and musical training. Percept Psychophys 1996;58:125–41.

Carral V, Huotilainen M, Ruusuvirta T, Fellman V, Näätänen R, Escera C. A kind of auditory 'primitive intelligence' already present at birth. Eur J Neurosci 2005a;21:3201–4.

Carral V, Corral MJ, Escera C. Auditory event-related potentials as a function of abstract change magnitude. Neuroreport 2005b;16:301–5.

Csibra G, Kushnerenko E, Grossmann T. Electrophysiological methods in studying infant cognitive development. In: Nelson CA, Luciana M, editors. Handbook of developmental cognitive neuroscience. 2nd ed. Cambridge (MA): MIT Press; 2008. p. 247–62.

Cuddy LL, Cohen AJ. Recognition of transposed melodic sequences. Q J Exp Psychol 1976;28:255–70.

Demany L, Armand F. The perceptual reality of tone chroma in early infancy. J Acoust Soc Am 1984;76:57–66.

Dowling WJ. Scale and contour: two components of a theory of memory for melodies. Psychol Rev 1978;85:341–54.

Dowling WJ. Contour in context: comments on Edworthy. Psychomusicology 1982;2:47–8.

Edworthy J. Interval and contour in melody processing. Music Percept 1985;2:375–88.

Fernald A. Prosody in speech to children: prelinguistic and linguistic functions. Ann Child Dev 1991;8:43–80.

Fujioka T, Trainor LJ, Ross B, Kakigi R, Pantev C. Musical training enhances automatic encoding of melodic contour and interval structure. J Cogn Neurosci 2004;16:1010–21.

Gumenyuk V, Korzyukov O, Alho K, Winkler I, Paavilainen P, Näätänen R. Electric brain responses indicate preattentive processing of abstract acoustic regularities in children. Neuroreport 2003;14:1411–5.

Háden G, Stefanics G, Vestergaard MD, Denham SLM, Sziller I, Winkler I. Timbre-independent extraction of pitch in newborn infants. Psychophysiology 2009; in press. doi:10.1111/j.1469-8986.2008.00749.x.

Korzyukov OA, Winkler I, Gumenyuk VI, Alho K. Processing abstract auditory features in the human auditory cortex. Neuroimage 2003;20:2245–58.

Kujala T, Tervaniemi M, Schröger E. The mismatch negativity in cognitive and clinical neuroscience: theoretical and methodological considerations. Biol Psychol 2007;74:1–19.

Kushnerenko E, Winkler I, Horváth J, Näätänen R, Pavlov I, Fellman V, et al. Processing acoustic change and novelty in newborn infants. Eur J Neurosci 2007;26:265–74.

Lerdahl F, Jackendoff R. A generative theory of tonal music. Cambridge (MA): MIT Press; 1983.

Müller D, Schröger E. Temporal grouping affects the automatic processing of deviant sounds. Biol Psychol 2007;74:358–64.

Näätänen R, Gaillard AWK, Mäntysalo S. Early selective attention effect on evoked potential reinterpreted. Acta Psychol (Amst) 1978;42:313–29.

Näätänen R, Paavilainen P, Rinne T, Alho K. The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clin Neurophysiol 2007;118:2544–90.

Nakata T, Trehub SE. Infants' responsiveness to maternal speech and singing. Infant Behav Dev 2004;27:455–64.

Novitski N, Huotilainen M, Tervaniemi M, Näätänen R, Fellman V. Neonatal frequency discrimination in 250–4000 Hz frequency range: electrophysiological evidence. Clin Neurophysiol 2007;118:412–9.

Paavilainen P, Saarinen J, Tervaniemi M, Näätänen R. Mismatch negativity to changes in abstract sound features during dichotic listening. J Psychophysiol 1995;9:243–9.

Paavilainen P, Jaramillo M, Näätänen R. Binaural information can converge in abstract memory traces. Psychophysiol 1998;35:483–7.

Paavilainen P, Jaramillo M, Näätänen R, Winkler I. Neuronal populations in the human brain extracting invariant relationships from acoustic variance. Neurosci Lett 1999;265:179–82.

Paavilainen P, Degerman A, Takegata R, Winkler I. Spectral and temporal stimulus characteristics in the processing of abstract auditory features. Neuroreport 2003;14:715–8.

Peretz I, Babaï M. The role of contour and intervals in the recognition of melody parts: evidence from cerebral asymmetries in musicians. Neuropsychologia 1992;30:277–92.

Ruusuvirta T, Huotilainen M, Fellman V, Näätänen R. The newborn human brain binds sound features together. Neuroreport 2003;14:2117–9.

Saarinen J, Paavilainen P, Schröger E, Tervaniemi M, Näätänen R. Representation of abstract attributes of auditory stimuli in the human brain. Neuroreport 1992;3:1149–51.

Sambeth A, Ruohio K, Alku P, Fellman V, Huotilainen M. Sleeping newborns extract prosody from continuous speech. Clin Neurophysiol 2008;119:332–41.

Stefanics G, Háden G, Huotilainen M, Balázs L, Sziller I, Beke A, et al. Auditory temporal grouping in newborn infants. Psychophysiol 2007;44:697–702.

Takegata R, Roggia SM, Winkler I. Effects of temporal grouping on the memory representation of inter-tone relationships. Biol Psychol 2005;68:41–60.

Tervaniemi M, Maury S, Näätänen R. Neural representations of abstract features in the human brain as reflected by the mismatch negativity. Neuroreport 1994;5:844–6.

Tervaniemi M, Rytkönen M, Schröger E, Ilmoniemi RJ, Näätänen R. Superior formation of cortical memory traces for melodic patterns in musicians. Learn Mem 2001;8:295–300.
Tervaniemi M, Castaneda A, Knoll M, Uther M. Sound processing in amateur musicians and nonmusicians: event-related potential and behavioral indices. Neuroreport 2006;17:1225–8.
Trainor LJ, Trehub SE. A comparison of infants' and adults' sensitivity to western musical structure. J Exp Psychol Hum Percept Perform 1992;18:394–402.
Trainor LJ, McDonald KL, Alain C. Automatic and controlled processing of melodic contour and interval information measured by electrical brain activity. J Cogn Neurosci 2002a;14:1–13.
Trainor LJ, Tsang CD, Cheung VHW. Preference for consonance in 2- and 4-month-old infants. Music Perc 2002b;20:187–94.
Trehub SE. The developmental origins of musicality. Nat Neurosci 2003;6:669–73.
Trehub SE, Bull D, Thorpe LA. Infants' perception of melodies: the role of melodic contour. Child Dev 1984;55:821–30.
Trehub SE, Thorpe LA, Morrongiello BA. Organizational processes in infants' perception of auditory patterns. Child Dev 1987;58:741–9.
Winkler I, Kushnerenko E, Horváth J, Čeponienė R, Fellman V, Huotilainen M, et al. Newborn infants can organize the auditory world. Proc Natl Acad Sci USA 2003;100:1182–5.
Wormith SJ, Pankhurst D, Moffit AR. Frequency discrimination by young infants. Child Dev 1975;46:272–5.

3.2. Study II: Timbre-independent extraction of pitch in newborns

Háden, G. P., Stefanics, G., Vestergaard, M.D., Denham, S.L., Sziller, I., Winkler, I. (2009) Timbre-independent extraction of pitch in newborn infants. *Psychophysiology* 46 (1), 69-74. DOI: 10.1111/j.1469-8986.2008.00749.x

Psychophysiology, 46 (2009), 69–74. Wiley Periodicals, Inc. Printed in the USA.

DOI: 10.1111/j.1469-8986.2008.00749.x

Timbre-independent extraction of pitch in newborn infants

GÁBOR P. HÁDEN,[a,b] GÁBOR STEFANICS,[a] MARTIN D. VESTERGAARD,[c]
SUSAN L. DENHAM,[d] ISTVÁN SZILLER,[e] AND ISTVÁN WINKLER,[a,f]
[a]Institute for Psychology, Hungarian Academy of Sciences, Budapest, Hungary
[b]Department of Cognitive Science, Budapest University of Technology and Economics, Budapest, Hungary
[c]Centre for the Neural Basis of Hearing, Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK
[d]Centre for Theoretical and Computational Neuroscience, University of Plymouth, Plymouth, UK
[e]First Department of Obstetrics and Gynecology, Semmelweis University, Budapest, Hungary
[f]Institute of Psychology, University of Szeged, Szeged, Hungary

Abstract

The ability to separate pitch from other spectral sound features, such as timbre, is an important prerequisite of veridical auditory perception underlying speech acquisition and music cognition. The current study investigated whether or not newborn infants generalize pitch across different timbres. Perceived resonator size is an aspect of timbre that informs the listener about the size of the sound source, a cue that may be important already at birth. Therefore, detection of infrequent pitch changes was tested by recording event-related brain potentials in healthy newborn infants to frequent standard and infrequent pitch-deviant sounds while the perceived resonator size of all sounds was randomly varied. The elicitation of an early negative and a later positive discriminative response by deviant sounds demonstrated that the neonate auditory system represents pitch separately from timbre, thus showing advanced pitch processing capabilities.

Descriptors: Pitch processing, Timbre, Perceived resonator size, Development, Neonates, Event-related brain potentials (ERP), mismatch negativity (MMN)

Pitch is a perceived auditory feature ordering sounds on a scale from low to high. Pure tones resonate at a single frequency, and the classical findings of psychophysics showed that for such tones, perceived pitch is determined by the tone frequency (see, e.g., Zwicker & Fastl, 1990). However, adults perceive pitch even when hearing nonharmonic sounds, and they can compare pitch across spectrally and temporally very different sounds (e.g., telling whether a sound produced by a flute was higher or lower than a sound produced by a violin). This suggests that the auditory system analyzes sound structure in a complex way to arrive at perceiving and representing pitch. Some theorists suggest that learning plays an important role in pitch perception (Terhardt, 1974). Therefore, it is important to test whether or not nontrivial forms of pitch analysis are functional in newborn infants, especially because pitch processing is an important prerequisite of music and speech perception and is also crucial for representing auditory objects (Kubovy & Van Valkenburg, 2001). The current study was designed to test whether neonates can equate pitch across sounds with different timbres.

Sensitivity to spectral sound features appears very early during ontogenesis. Behavioral studies have shown that the fetal auditory system is functional to a degree already at ~ 20 weeks of gestational age (Birnholz & Benacerraf, 1983; Hepper & Shahidullah, 1994). Reliable event-related magnetic responses were obtained to pure tones from fetuses from the 27th week of gestational age onward with the latency of the evoked response decreasing with age (Holst et al., 2005; Draganova et al., 2005). Fetal responses related to spectrum-based sound discrimination have been obtained for pitch-deviant complex tones from ~ 30 weeks of gestational age (e.g., Huotilainen et al., 2005).

In the current study, pitch processing in newborn infants was tested by recording electric brain responses to sounds (auditory event-related brain potentials [ERPs]). In adults, a component of the human auditory ERP has been shown to be sensitive to perceived pitch, as opposed to the raw acoustic parameters underlying pitch perception, such as the frequencies present in a sound. Using the missing fundamental phenomenon (when removing the fundamental frequency from a complex tone does not change its perceived pitch; see de Boer, 1976), Winkler and colleagues (1995) presented healthy young adults with infrequent complex tones of 600-Hz fundamental frequency (pitch deviants) among nine different complex tones having 300 Hz as the missing

---

This work was supported by the European Commission's 6th Framework Programme for "Information Society Technologies" (project title: EmCAP-Emergent Cognition through Active Perception, contract no.: 013123). The authors thank Prof. Dr. Zoltán Papp and Dr. János Rigó for supporting the study of neonates at the First Department of Obstetrics and Gynecology, Semmelweis University, Budapest, Hungary. We also thank RN Erika Józsa Váradinê for collecting the data.

Address reprints requests to: István Winkler, Institute for Psychology, Hungarian Academy of Sciences, 83–85 Szondi u., H-1394 Budapest, Hungary. E-mail: iwinkler@cogpsyphy.hu

fundamental (standards). Thus neither the standard tones nor the deviants contained harmonics below 600 Hz. Pitch-deviant tones elicited an MMNm response, the magnetic counterpart of the mismatch negativity (MMN) ERP component, which is an electrophysiological signal reflecting that the violation of an acoustic regularity was detected in the brain (for a recent review, see Näätänen, Paavilainen, Rinne, & Alho, 2007). Because each frequency present in the deviant was also present in five of the standard tones, these results indicated that the MMN response is based on perceived pitch rather than the spectral composition of the complex tones (Winkler et al, 1995; for a review of other compatible results, see Näätänen & Winkler, 1999). Because the sequence of standard (300-Hz missing fundamental) tones was composed of nine complex tones differing in their harmonic structure, MMN elicited by pitch change also showed that pitch repetition was detected even when other spectral parameters varied in the sound sequence (Näätänen & Winkler, 1999; Winkler et al., 1995). Thus it is feasible to use MMN to test whether pitch-repetition regularity is extracted from a sequence composed of spectrally different sounds.

Similarly to adults, occasional changes in the frequency of pure and complex tones have been found to elicit an ERP component in full- and pre-term newborn infants, and these neonatal electromagnetic responses have been regarded as analogues of the adult MMN and MMNm (Alho, Saino, Sajaniemi, Reinikainen, & Näätänen, 1990; Čeponienė et al., 2002; Draganova et al., 2005). However, the sensitivity of detecting frequency changes in pure sinusoidal tones was found to be substantially lower in neonates than in adults (Novitski, Huotilainen, Tervaniemi, Näätänen, & Fellman, 2007), a result in line with behavioral evidence obtained for 1-month old infants (Wormith, Pankhurst, & Moffit, 1975). More reliable discrimination of spectral changes was shown with complex harmonic tones (Čeponienė et al., 2002), environmental sounds (Sambeth, Huotilainen, Kushnerenko, Fellman, & Pihko, 2006), noise (Kushnerenko et al., 2007), and speech sounds (Kujala et al., 2004), suggesting that changes in sounds activating wider regions of the auditory cortex are more likely to be detected. Using the MMN method, the operation of higher-level auditory abilities was also demonstrated in newborn infants, such as the segregation of concurrent streams of sound (Winkler et al., 2003) and auditory temporal grouping (Stefanics et al., 2007). Very importantly for the current study, Carral and colleagues (2005) demonstrated that the neonate auditory system extracts invariant features from variable sound sequences, providing a basis for testing the processing of higher-order auditory cues.

Timbre is a complex sound property partly based on spectral as well as dynamic cues and having several perceptually distinct components (e.g., McAdams, Winsberg, Donnadieu, De Soete, & Krimphoff, 1995; Saldanha & Corso, 1964). In adults, it has been found that occasional timbre-deviant sounds elicit the MMN response (Goydke, Altenmüller, Möller, & Münte, 2004; Tervaniemi, Winkler, & Näätänen, 1997). Perceived resonator size is an aspect of timbre, which allows adults to estimate the size of the sound source with reasonable accuracy irrespective of the pitch of the emitted sound (Smith, Patterson, Turner, Kawahara, & Irino, 2005, van Dinther & Patterson 2006). Vestergaard, Shtyrov, Pulvermüller, and Patterson (2007) and Vestergaard and colleagues (2008) showed that infrequent perceived resonator size changes elicit MMN both in adults, who did not attend the sounds, and in sleeping newborn infants, suggesting the existence of a neural subsystem for automatic speaker-size processing in the human auditory system.

In the current experiment, timbre variation was implemented by varying the resonator size of an instrument (French horn) sound, using such levels of the resonator-size feature, which are discriminated by the neonate auditory system (Vestergaard et al., 2008). Sounds with two different pitches (distinguishable by neonates; see Novitski et al., 2007) were presented to sleeping newborn infants in an oddball design with resonator size varying independently of pitch. If the neonate auditory system extracts pitch information independently of the resonator size, then occasional pitch deviants will be detected, resulting in the elicitation of the corresponding electrophysiological signals (i.e., pitch invariance will be detected despite variance in timbre; cf. Carral et al., 2005). If, however, pitch and timbre (resonator size) are not processed separately by newborn infants, then pitch deviants will not be detected and no electrophysiological sign of the differential processing of the "standard" and "deviant" sounds will be observed. That is, if pitch repetition is not detected and, because no individual sound appears sufficiently often in the sequence to allow the formation of holistic sound-repetition regularity, then the pitch-deviant sounds do not violate any detected regularity in the sequence and, therefore, will not trigger differential processing in the neonate auditory system.

Methods

Participants

ERP responses were recorded from 12 (5 female) healthy full-term newborn infants on day 2 or 3 postpartum. Their gestational age was 37–40 weeks, birth weight 2730–4330 g, and Apgar score 9–10. Data from 3 of the 12 subjects (2 boys) were discarded because of excessive artifacts. The recordings were carried out in the hospital ward of the First Department of Obstetrics and Gynecology, Semmelweis University, Budapest, Hungary, after informed consent was obtained from one or both parents. The mother of the infant was present during the recording. The study was approved by the Ethics Committee of the Semmelweis University as well as by the Institutional Review Board of the Institute for Psychology, Hungarian Academy of Sciences.

Stimuli and Procedure

Sequences were composed of instrumental sounds with two different pitches, 139 Hz (C#3) and 175 Hz (F3). The sounds, played on a French horn, were taken from the database by Goto, Hashiguchi, Nishimura, and Oka (2003). For both pitches, seven resonator-size variants were created. Different resonator sizes were simulated by setting the resonance scale of the sounds with a vocoder (Kawahara & Irino, 2004). The just noticeable difference (JND) for simulated resonator size for French horn is approximately 6%–9% for adult listeners (van Dinther & Patterson, 2006). For the current study, resonator size was varied over 22% (ca. 6 JNDs), corresponding to the scale factors between 1/1.22 and 1.22. The seven resonator sizes were equally spaced on the logarithmic scale between −22% and +22%, producing a step size of approximately 1 JND for the horn resonance scale (values: 0.82, 0.88, 0.94, 1.00, 1.07, 1.14, 1.22, where 1.00 denotes the base sound). To improve the signal-to-noise ratio of the early ERP components, sound onsets were made more abrupt by removing the 60-ms-long initial period from the original sound and imposing a 10-ms-long raised-cosine

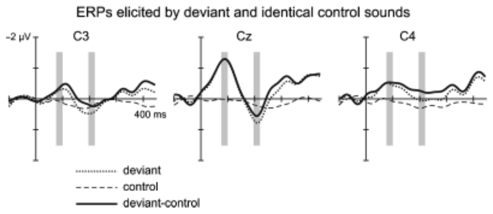

Figure 1. Group-averaged ERPs elicited by deviant (dotted line) and identical control (dashed line) sounds of 175-Hz base frequency (musical note F3) at three central electrode locations (C3, Cz, and C4). Deviant-minus-control difference waveforms are plotted in black (solid line). Stimulus onset is at the crossing of the axes. Time windows for amplitude measurements (78–98 and 198–218 ms) are shaded gray.

ramp on the new sound onset. The total sound duration was 545 ms for both notes. Sounds were presented binaurally using the E-Prime stimulus presentation software (Psychology Software Tools, Inc., Pittsburgh, PA) via ER-3A loudspeakers (EtymStic Research, Inc., Elk Grove Village, IL) connected via sound tubes to self-adhesive ear-couplers (Natus Medical, Inc., San Carlos, CA) placed over the babies' ears. We have previously shown that occasionally changing the resonator size of the above-described sounds by 22% elicits ERP correlates of deviance detection in sleeping neonates (Vestergaard et al., 2008).

Sounds were presented in an oddball sequence (750 ms stimulus onset asynchrony [SOA]) with 87.5% of them (standards) having C#3 and 12.5% (deviants) F3 for pitch. The seven resonator-size variants were equiprobably distributed, separately within the standard and the deviant stimuli (14.3% each). Two stimulus blocks of 600 sounds were delivered (a total of 1,200 sounds). In control stimulus blocks delivered after the main stimulus sequences, the base frequencies for standard and deviant tones were exchanged, allowing comparison between responses elicited by identical (F3) sounds when they served as deviants in the main experimental stimulus blocks and when they were delivered frequently within the control block.

EEG Recording

EEG was recorded from the F3, F4, C3, Cz, and C4 scalp electrodes (according to the International 10 20 System) and from electrodes placed over the left and right mastoids, with the common reference attached to the tip of the nose. The ground electrode was placed on the forehead. Eye movements were monitored by recording the electrooculogram (EOG) between two electrodes, one placed below the left eye and another above the right eye. EEG was recorded at 24-bit resolution and a sampling rate of 250 Hz by a direct-coupled amplifier (V-Amp, Brain Products, Munich, Germany). The signals were on-line low-pass filtered at 110 Hz.

Data Analysis

EEG was filtered off-line between 1.5 and 16 Hz. For each stimulus, an epoch of 600 ms duration including a 100-ms prestimulus period was extracted from the continuous EEG record. Epochs with a voltage change below 0.1 µV or above 70 µV on any EEG or EOG channel were rejected from further analysis. Responses were analyzed for the central line of electrodes showing the best signal-to-noise ratio. Epochs were baseline corrected for the 100-ms prestimulus period and averaged separately for standards and deviants and for the experimental and the control condition. The mean number of artifact-free deviant trials per infant was 136 with over 600 identical-stimulus standard responses collected in the control condition.

For amplitude measurements, two 20-ms-long time windows were selected from the grand-averaged deviant-minus-standard difference waveform (Figure 1) at the Cz electrode. One window was centered on the negative peak in the 78 98-ms latency range and the other on the positive peak in the 198 218-ms latency range. The effects of stimulus type were analyzed with two-way repeated-measures analyses of variance (ANOVA of Stimulus type [deviant vs. control] × Electrode [C3 vs. Cz vs. C4]), separately for the two latency ranges. Greenhouse Geisser correction of the degrees of freedom was applied where appropriate and ε values as well as η^2 effect sizes are given in Results. ANOVA results were further specified by Fisher's LSD post hoc tests.

Although a previous study found no significant effect of sleep stages on the MMN parameters measured in neonates (Martynova, Kirjavainen, & Cheour, 2003), other studies suggested the possibility of a sleep-stage effect (Friederici, Friedrich, & Weber, 2002). Therefore, the ratio of quiet and active sleep was assessed separately for each stimulus block, following the criteria suggested by Anders, Emde, and Parmelee (1971), using EEG and EOG measures and observations of the infant's behavior. Periods of wakefulness and indeterminate stages were excluded from this analysis. Overall, the infants were more often in active (58% of the time) than in quiet sleep (22% of the time) during the recordings. This result is in line with the similar observations of Novitski and colleagues (2007). The ratio of quiet and active sleep did not significantly differ between the test and the control stimulus blocks. Therefore, comparing responses between the test and the control stimulus blocks does not significantly affect the results.

Results

Figure 1 shows grand-averaged ERPs elicited by deviant and identical control sounds at the three central electrode locations (C3, Cz, C4) together with the corresponding deviant-minus-control difference waveforms. The time windows for the amplitude measurements are highlighted. The ANOVA for the negative peak in the 78 98-ms latency range yielded a significant main effect of Stimulus type, $F(1,8) - 6.23$, $p < .05$, $\eta^2 - .44$, and Electrode, $F(2,16) - 6.26$, $p < .02$, $\eta^2 - .44$ $\varepsilon - .89$, and a significant Stimulus type × Electrode interaction, $F(2,16) - 7.54$, $p < .02$, $\eta^2 - .49$ $\varepsilon - .76$. The interaction was explained by significant differences found in the post hoc test between deviant- and control-stimulus responses at Cz ($p < .01$) and a tendency at

C4 (p = .071) but not at C3 and that the deviant response at Cz was significantly more negative (at least $p < .05$) than either deviant or standard responses measured at the two lateral electrodes.

The ANOVA for the positive peak in the 198–218-ms latency range yielded only a significant Stimulus-type × Electrode interaction, $F(2,16) = 3.67$, $p < .05$, $\eta^2 = .31$, $\varepsilon = .86$. Post hoc tests showed a significant difference between the deviant- and control-sound response on Cz ($p < .02$), but not on either of the lateral leads and the deviant response at Cz was significantly more positive (at least $p < .05$) than the standard responses measured at the two lateral electrodes as well as the deviant response measured at C4.

Discussion

Significant differences were found between the ERP responses elicited by deviant and control sounds of identical pitch while resonator size was varied. Our results showed that, despite variation in timbre (resonator size), the neonate auditory system extracted pitch invariance from the sound sequences and detected sounds that deviated from the common pitch.

In all but one of the previous studies of auditory change detection in newborn infants, the standard stimulus was constant throughout the stimulus blocks. Therefore, these studies did not shed light on whether newborn infants process the various spectrum-related perceived auditory features separately or in a global, undistinguished manner. Carral et al.'s (2005) results gave the first indication that some of the spectral features are separately processed. These authors found that variance in one spectral feature (absolute tone frequency) did not interfere with the detection of another spectral feature (pitch-change direction within a tone pair). Their results can be interpreted as showing that absolute frequency and the direction of frequency change within a perceptual unit are extracted and represented separately within the neonate auditory system. In the current experiment, extraction of pitch invariance was tested in the face of variation in a timbral feature (resonator size). The distribution of pitch and timbre levels was distinctly different: One (pitch) allowed detection of a repetition regularity whereas the other (timbre) showed a homogeneous distribution across the feature levels appearing within the sequence. Elicitation of a discriminative response by pitch deviants suggests that pitch repetition was detected under these circumstances. Thus the current results suggest that in newborn infants, a pitchlike auditory feature was analyzed and represented separately from the cues perceived as resonator size by adults. This should not be taken to mean that, in the absence of attention, the neonate auditory system does not conjoin auditory features. Rare combinations of sound features, which separately appear frequently within the stimulus sequence, have been shown to elicit discriminative ERP responses in newborn infants (Ruusuvirta, Huotilainen, Fellman, & Näätänen, 2003, 2004). Although the current results definitely do not rule out the possibility that some processes of pitch processing are learned, they suggest that representation for at least a proto-pitch feature already exists at birth. It can be argued that the fetus is already exposed to sounds within the uterus and thus even the current finding may reflect learned abilities. However, although exposure to sounds within the uterus may be necessary for normal development of the human auditory system, this cannot be regarded as the kind of learning hypothesized by theories of pitch perception (Terhardt, 1974).

The negative response found in the current experiment peaked rather early (at 88 ms in the group average). This is probably due to the use of complex, natural stimulus material, which provides multiple congruent cues, thus allowing early determination of deviance. Similar early discriminative negative responses have been previously shown in neonates (Kushnerenko et al., 2007; peak latency: 75 ms) and preschool-age children (Räikkönen, Birkás, Horváth, Gervai, & Winkler, 2003; peak latency: 66 ms) to rare environmental sounds and in adults (Jacobsen, Schröger, & Alter, 2004; peak latency: ca. 105 ms) to occasional deviant vowels. Previous studies showed that the early negative difference in neonates is mainly elicited by large spectral changes (Kushnerenko, Čeponienė, Balan, Fellman, & Näätänen, 2002; Kushnerenko et al., 2007; Morr, Shafer, Kreuzer, & Kurtzberg, 2002). In adults, similar changes usually elicit high-amplitude N1 responses (e.g., Alho et al., 1998; Escera, Alho, Winkler, & Näätänen, 1998). However, the N1 response appears only later in development (Pang & Taylor, 2000; Ponton, Eggermont, Kwong, & Don, 2000; Sharma, Kraus, N., McGee, & Nicol, 1997). Furthermore, unlike the adult N1, the early neonatal negativity is elicited reliably only by deviant stimuli. This suggests that the early neonatal negativity may be more similar to the adult MMN response. However, in contrast to the adult MMN, the neonatal ERP response to acoustic deviation often appears as a positive difference in the 100–300-ms poststimulus interval (e.g., Winkler et al., 2003). This positive response, which in some studies (such as the current one) follows the early negative ERP difference (whereas in others appears alone), was shown to be sensitive to acoustic energy (Kushnerenko et al., 2007). This feature of the neonate positive difference response contrasts those of the somewhat similar P3a response in adults (Friedman, Cycowicz, & Gaeta, 2001; Polich, 2007). Furthermore, the neonatal positive difference can be elicited even when no P3a response would be expected in adults (e.g., in response to rare feature conjunctions; see Ruusuvirta et al., 2003). This feature of the neonatal positive discriminative response is similar to the adult MMN response. Thus, although the morphology of the current as well as some previously observed neonatal ERP responses to infrequent deviant sounds may resemble the sequence of N1/MMN and P3a in adults (e.g., Kushnerenko et al., 2007), the underlying processes are probably different. Both the early negativity and the somewhat later positivity in neonates show some of the features of the MMN response in adults, but they are far more sensitive to primary surface stimulus features (such as spectral width and acoustic energy) than the adult MMN. In this, they resemble more the adult N1. It is possible that during maturation of the human auditory system, responses to energy change and to deviation from detected regular features of the preceding sound sequence become better separated, forming the basis of passive attention (James, 1890). This development may be reflected by the specialization of ERP responses, the emergence of the well-known N1, MMN, P3a sequence of ERP components, which characterizes the adult response to unexpected salient changes in the acoustic input.

In conclusion, the current results demonstrated the existence of advanced pitch-related processing in newborn infants, and they suggest that separate representation of a pitchlike feature exists at birth. Pitch processing is a necessary prerequisite of normal speech development, including the perception of prosody and emotional contents, speaker identification, and music perception. Therefore, future screening methods for various auditory-related deficits, such as, for example, dyslexia might be based on the testing of pitch discrimination at a very early age (Fellman & Huotilainen, 2006).

REFERENCES

Alho, K., Saino, K., Sajaniemi, N., Reinikainen, K., & Näätänen, R. (1990). Event-related brain potential of human newborns to pitch change of an acoustic stimulus. *Electroencephalography and Clinical Neurophysiology, 77,* 151–155.

Alho, K., Winkler, I., Escera, C., Huotilainen, M., Virtanen, J., Jääskeläinen, I. P., et al. (1998). Processing of novel sounds and frequency changes in the human auditory cortex: Magnetoencephalographic recordings. *Psychophysiology, 35,* 211–224.

Anders, T., Emde, R., & Parmelee, A. (1971). *A manual of standardized terminology, techniques and criteria for scoring of states of sleep and wakefulness in newborn infants.* Los Angeles: UCLA Brain Information Service.

Birnholz, J. C., & Benacerraf, B. R. (1983). The development of human fetal hearing. *Science, 222,* 516–518.

Carral, V., Huotilainen, M., Ruusuvirta, T., Fellman, V., Näätänen, R., & Escera, C. (2005). A kind of auditory 'primitive intelligence' already present at birth. *European Journal of Neuroscience, 21,* 3201–3204.

Čeponienë, R., Kushnerenko, E., Fellman, V., Renlund, M., Suominen, K., & Näätänen, R. (2002). Event-related potential features indexing central auditory discrimination by newborns. *Brain Research Cognitive Brain Research, 13,* 101–113.

de Boer, E. (1976). On the residue and auditory pitch perception. In W. D. Keidel & W. D. Neff (Eds.), *Handbook of sensory physiology* (vol 3, pp. 479–583). New York: Springer.

Draganova, R., Eswaran, H., Lowery, C. L., Murphy, P., Huotilainen, M., & Preissl, H. (2005). Sound frequency change detection in fetuses and newborns—A magnetoencephalographic study. *NeuroImage, 28,* 354–361.

Escera, C., Alho, K., Winkler, I., & Näätänen, R. (1998). Neural mechanisms of involuntary attention switching to novelty and change in the acoustic environment. *Journal of Cognitive Neuroscience, 10,* 590–604.

Fellman, V., & Huotilainen, M. (2006). Cortical auditory event related potentials in newborn infants. *Seminars in Fetal and Neonatal Medicine, 11,* 452–458.

Friederici, A. D., Friedrich, M., & Weber, C. (2002). Neural manifestation of cognitive and precognitive mismatch detection in early infancy. *NeuroReport, 13,* 1251–1254.

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 Biobehavioral Reviews, 25,* 355–373.

Goto, M., Hashiguchi, H., Nishimura, T., & Oka, R. (2003). RWC music database Music genre database and musical instrument database. In Holger H. Hoos & David Bainbridge (eds.), Proceedings of 4th International Conference on Music Informational Retrieval (ISMIR 2003) (pp. 229–230). Baltimore: Johns Hopkins University.

Goydke, K. N., Altenmüller, E., Möller, J., & Münte, T. F. (2004). Changes in emotional tone and instrumental timbre are reflected by the mismatch negativity. *Cognitive Brain Research, 21,* 351–359.

Hepper, P. G., & Shahidullah, B. S. (1994). Development of fetal hearing. *Archives of Disease in Childhood, 71,* F81–87.

Holst, M., Eswaran, H., Lowery, C., Murphy, P., Norton, J., & Preissl, H. (2005). Development of auditory evoked fields in human fetuses and newborns: A longitudinal MEG study. *Clinical Neurophysiology, 116,* 1949–1955.

Huotilainen, M., Kujala, A., Hotakainen, M., Parkkonen, L., Taulu, S., Simola, J., et al. (2005). Short-term memory functions of the human fetus recorded with magnetoencephalography. *NeuroReport, 16,* 81–84.

Jacobsen, T., Schröger, E., & Alter, K. (2004). Pre-attentive perception of vowel phonemes from variable speech stimuli. *Psychophysiology, 41,* 654–659.

James, W. (1890). *The principles of psychology.* New York: Holt.

Kawahara, H., & Irino, T. (2004). Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation. In P. L. Divenyi (Ed.), *Speech separation by humans and machines* (pp. 167–180). Norwell, MA: Kluwer Academic.

Kubovy, M., & Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition, 80,* 97–126.

Kujala, A., Huotilainen, M., Hotakainen, M., Lennes, M., Parkkonen, L., Fellman, V., et al. (2004). Speech-sound discrimination in neonates as measured with MEG. *NeuroReport, 15,* 2089–2092.

Kushnerenko, E., Čeponienë, R., Balan, P., Fellman, V., & Näätänen, R. (2002). Maturation of the auditory change-detection response in infants: A longitudinal ERP study. *NeuroReport, 13,* 1843–1848.

Kushnerenko, E., Winkler, I., Horváth, J., Näätänen, R., Pavlov, I., Fellman, V., et al. (2007). Processing acoustic change and novelty in newborn infants. *European Journal of Neuroscience, 26,* 265–274.

Martynova, O., Kirjavainen, J., & Cheour, M. (2003). Mismatch negativity and late discriminative negativity in sleeping human newborns. *Neuroscience Letters, 340,* 75–78.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities and latent subject classes. *Psychological Research, 58,* 177–192.

Morr, M. L., Shafer, V. L., Kreuzer, J. A., & Kurtzberg, D. (2002). Maturation of mismatch negativity in typically developing infants and preschool children. *Ear & Hearing, 23,* 118–136.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology, 118,* 2544–2590.

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

Novitski, N., Huotilainen, M., Tervaniemi, M., Näätänen, R., & Fellman, V. (2007). Neonatal frequency discrimination in 250–4000 Hz frequency range: Electrophysiological evidence. *Clinical Neurophysiology, 118,* 412–419.

Pang, E. W., & Taylor, M. J. (2000). Tracking the development of the N1 from age 3 to adulthood: An examination of speech and non-speech stimuli. *Clinical Neurophysiology, 111,* 388–397.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology, 118,* 2128–2148.

Ponton, C. W., Eggermont, J. J., Kwong, B., & Don, M. (2000). Maturation of human central auditory system activity: Evidence from multi-channel evoked potentials. *Clinical Neurophysiology, 111,* 220–236.

Räikkönen, K., Birkás, E., Horváth, J., Gervai, J., & Winkler, I. (2003). Test-retest reliability of auditory ERP components in healthy 6-year-old children. *NeuroReport, 14,* 2121–2125.

Ruusuvirta, T., Huotilainen, M., Fellman, V., & Näätänen, R. (2003). The newborn human brain binds sound features together. *NeuroReport, 14,* 2117–2119.

Ruusuvirta, T., Huotilainen, M., Fellman, V., & Näätänen, R. (2004). Newborn human brain identifies repeated auditory feature conjunctions of low sequential probability. *European Journal of Neuroscience, 20,* 2819–2821.

Saldanha, E. L., & Corso, J. F. (1964). Timbre cues and the identification of musical instruments. *Journal of the Acoustical Society of America, 36,* 2021–2026.

Sambeth, A., Huotilainen, M., Kushnerenko, E., Fellman, V., & Pihko, E. (2006). Newborns discriminate novel from harmonic sounds: A study using magnetoencephalography. *Clinical Neurophysiology, 117,* 496–503.

Sharma, A., Kraus, N., McGee, T. J., & Nicol, T. G. (1997). Developmental changes in P1 and N1 central auditory responses elicited by consonant-vowel syllables. *Electroencephalography and Clinical Neurophysiology, 104,* 540–545.

Smith, D. R., Patterson, R. D., Turner, R., Kawahara, H., & Irino, T. (2005). The processing and perception of size information in speech sounds. *Journal of the Acoustical Society of America, 117,* 305–318.

Stefanics, G., Háden, G., Huotilainen, M., Balázs, L., Sziller, I., Beke, A., et al. (2007). Auditory temporal grouping in newborn infants. *Psychophysiology, 44,* 697–702.

Terhardt, E. (1974). Pitch, consonance, and harmony. *Journal of the Acoustical Society of America, 55,* 1061–1069.

Tervaniemi, M., Winkler, I., & Näätänen, R. (1997). Pre-attentive categorization of sounds by timbre as revealed by event-related potentials. *NeuroReport, 8,* 2571–2574.

van Dinther, R., & Patterson, R. D. (2006). The perception of size in musical instruments. *Journal of the Acoustical Society of America, 120,* 2158–2176.

Vestergaard, M. D., Háden, G. P., Shtyrov, Y., Patterson, R. D., Pulvermüller, F., Denham, S. L., et al. (2008). Auditory size-deviant detection in adults and infants. Manuscript submitted for publication.

Vestergaard, M. D., Shtyrov, Y., Pulvermüller, F., & Patterson, R. D. (2007). EEG evidence of normalization of communication sounds in humans. *NeuroImage, 36*(Suppl. 1), S120.

Winkler, I., Kushnerenko, E., Horváth, J., Čeponienë, R., Fellman, V., Huotilainen, M., et al. (2003). Newborn infants can organize the auditory world. *Proceedings of the National Academy of Sciences, USA*, *100*, 1182–1185.

Winkler, I., Tervaniemi, M., Huotilainen, M., Ilmoniemi, R., Ahonen, A., Salonen, O., et al. (1995). From objective to subjective: Pitch representation in the human auditory cortex. *NeuroReport*, *6*, 2317–2320.

Wormith, S. J., Pankhurst, D., & Moffit, A. R. (1975). Frequency discrimnation by young infants. *Child Development*, *46*, 272–275.

Zwicker, E., & Fastl, H. (1990). *Psychoacoustics. Facts and Models*. Berlin: Springer Verlag.

(RECEIVED February 14, 2008; ACCEPTED April 18, 2008)

3.3. Study III: Auditory temporal grouping in newborns

Stefanics, G., Háden, G., Huotilainen, M., Balázs, L., Sziller, I., Beke, A., Fellman, V., Winkler, I. (2007) Auditory temporal grouping in newborn infants. *Psychophysiology* 44 (5), 697-702. DOI: 10.1111/j.1469-8986.2007.00540.x

---

Psychophysiology, 44 (2007), 697–702. Blackwell Publishing Inc. Printed in the USA.

DOI: 10.1111/j.1469-8986.2007.00540.x

Auditory temporal grouping in newborn infants

GÁBOR STEFANICS,[a] GÁBOR HÁDEN,[a,b] MINNA HUOTILAINEN,[c] LÁSZLÓ BALÁZS,[a]
ISTVÁN SZILLER,[d] ANNA BEKE,[d] VINETA FELLMAN,[e,f] AND ISTVÁN WINKLER[a]
[a]Institute for Psychology, Hungarian Academy of Sciences, Budapest, Hungary
[b]Department of Cognitive Science, Budapest University of Technology and Economics, Budapest, Hungary
[c]Cognitive Brain Research Unit, Department of Psychology, University of Helsinki, Helsinki, Finland
[d]First Department of Obstetrics and Gynecology, Semmelweis University, Budapest, Hungary
[e]Hospital for Children and Adolescents, Helsinki University Central Hospital, Helsinki, Finland
[f]Department of Pediatrics, Lund University, Lund, Sweden

Abstract

Adults normally perceive auditory scenes in terms of sound patterns emitted by concurrently active sources. Thus pattern formation is an important process of auditory object perception. The aim of the present study was to determine whether neonates group sounds by repeating pitch patterns. Standard ("S"; $p = 80\%$) and deviant tones ("D", $p = 20\%$) differing only in pitch were delivered either in a randomized order (random condition) or in a repeating SSSSD pattern (grouped condition). Both event-related brain potentials and gamma-band activity differed between the S and D tones in the random condition but not in the grouped condition. These results suggest that in the grouped condition, the S and D tones were processed as part of the same higher order regularity by the neonate auditory system. Also, for the first time, we observed oscillatory gamma-band activity in neonates, which was sensitive to infrequent pitch changes.

Descriptors: Auditory event-related potential, Gamma synchronization, Mismatch negativity (MMN), Neonate, Perceptual development

Several studies have shown that human newborn infants can extract regularities from the acoustic input as was revealed by electric brain responses recorded to infrequent deviants embedded in regular sound sequences (e.g., Alho, Sainio, Sajaniemi, Reinikainen, & Näätänen, 1990; Dehaene-Lambertz & Dehaene, 1994; Friederici, Friedrich, & Weber, 2002). Responses elicited by rare sounds acoustically deviating from a repeating (standard) sound provided evidence that the auditory system of full-term healthy newborn infants is sensitive to local stimulus probabilities, although newborns usually required larger amounts of acoustic deviance than adults for change detection (e.g., Novitski, Huotilainen, Tervaniemi, Näätänen, & Fellman, 2007). However, veridical sound perception in an everyday environment requires detection of higher order regularities encoding not only local stimulus probabilities, but also the regular features of the wider acoustic environment (Bregman, 1990). Recent studies have shown that the neonate auditory system not only keeps track of the probabilities of individual sound features, but also that of their co-occurrence on the same sound (Ruusuvirta, Huotilainen, Fellman, & Näätänen, 2003) and can extract an invariant intersound relationship from a set of acoustically varying sounds (Carral et al., 2005). Moreover, regarding auditory processing abilities directly underlying perception in real-life situations, it was found that newborn infants segregate (Winkler et al., 2003) and select sound sources (Barker & Newman, 2004) at least by simple sound features.

Under ecologically valid circumstances, sounds are encountered within complex patterns. However, in almost all of the previous studies of the neonate auditory system, the stimulus material was composed of individual sounds. The objective of the present study was to test whether pitch patterns are detected by the newborn auditory system. Pitch contour (melody) has been suggested as the representational unit (object) in auditory perception (Kubovy & Van Valkenburg, 2001). Therefore, grouping pitch patterns by repetition is an important method for extracting higher order perceptual units from the acoustic input.

We tested grouping of a repeating pitch pattern using event-related brain potentials (ERPs), which can be recorded from sleeping newborn infants (Vaughan & Kurtzberg, 1991). One of the ontogenetically earliest ERP responses, the mismatch negativity (MMN), is elicited by deviants embedded in a regular

---

The authors thank Prof. Dr. Zoltán Papp for supporting the study of neonates at the First Department of Obstetrics and Gynecology, Semmelweis University, Budapest, Hungary. We also thank RN Erika Józsa Váradiné and RN Tarja Ilkka for collecting the data and István Dániel (Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics) for calibrating the sound stimulation system. This work was supported by the European Commission's 6th Framework Programme for "Information Society Technologies" (project title: EmCAP—Emergent Cognition through Active Perception, contract no.: 013123).

Address reprint requests to: István Winkler, Institute for Psychology, Hungarian Academy of Sciences, 83–85 Szondi u., H-1394 Budapest, Hungary. E-mail: iwinkler@cogpsyphy.hu

sound sequence (Näätänen, Gaillard, & Mäntysalo, 1978; Picton, Alain, Otten, Ritter, & Achim, 2000; in newborns, see Kushnerenko, Čeponienė, Balan, Fellman, & Näätänen, 2002; Morr, Shafer, Kreuzer, & Kurtzberg, 2002). For example, in adults, when a repeating tone (standard, "S") was randomly exchanged for a different tone (deviant, "D") 20% of the time, deviants elicited MMN (Scherg, Vajsar, & Picton, 1989; Sussman, Ritter, & Vaughan, 1998). In contrast, when the same tones with the same probability were presented in a fully regular order (SSSSDSSSSDSSSSD . . .) at short (below 200 ms) stimulus onset asynchronies (SOA), no MMN was elicited by the D tones (Sussman & Gumenyuk, 2005; Sussman et al., 1998). The authors suggested that the five-tone SSSSD cycle had been extracted as the repeating unit of the sequence and, therefore, the D tones became part of the regular sequence. However, with long SOAs, MMN was elicited by D tones in the regular sequence (Scherg et al., 1989) unless subjects were informed about the structure of the sequence (Sussman, Winkler, Huotilainen, Ritter, & Näätänen, 2002). These results confirmed that the ERP response elicited by D tones is an appropriate measure of auditory grouping functions and that automatic grouping occurs only at fast stimulus presentation rates in adults.

The present study tested automatic grouping of the SSSSD pattern in sleeping neonates by comparing the responses to D tones between regular and random-order sequences. In addition to the ERP responses, event-related spectral perturbations (ERSP) were also analyzed, because gamma-band activity related to cognitive processing of visual binding, object representation, and perception of goal-directed action has been reported in 6–8-month-old infants (Csibra, Davis, Spratling, & Johnson, 2000; Kaufman, Csibra, & Johnson, 2003; Reid, Csibra, Belsky, & Johnson, 2007, respectively). However, little is known about deviance-related gamma-band activity in newborns. In adults, oscillatory EEG activity in the gamma-band (>30 Hz) has been observed in response to rare targets (Gurtubay et al., 2001), nontarget deviants (Edwards, Soltani, Deouell, Berger, & Knight, 2005; Kaiser, Lutzenberger, Preissl, Ackermann, & Birbaumer, 2000), and stimulus omissions (Gurtubay, Alegre, Valencia, & Artieda, 2006) in the auditory oddball paradigm. Because rare deviants were reported to induce higher gamma-band activity than standards (Edwards et al., 2005; Kaiser et al., 2000) we expected stronger gamma-band activity in response to rare deviants in the random condition. In the grouped condition, if standards and deviants are processed within a regularly repeating SSSSD pattern, no difference in the gamma-band response to standards and deviants should emerge.

Methods

Participants

We recorded ERP responses from 12 (7 female) healthy full-term newborn infants on day 2 or 3 postpartum. Their gestational age was 38–41 weeks and birth weight was 2870–4200 g. The recordings were carried out in the hospital ward of the First Department of Obstetrics and Gynecology, Semmelweis University, Budapest, Hungary, after informed consent was obtained from the parents. The mother of the infant was present at the recording. The study was approved by the Ethics Committee of the Semmelweis University as well as by the Institutional Review Board of the Institute for Psychology, Hungarian Academy of Sciences. A preliminary pilot study was conducted at the Women's Hospital of the Helsinki University Central Hospital with the approval of the Ethics Committee for Pediatrics, Adolescent Medicine, and Psychiatry, Hospital District of Helsinki and Uusimaa.

The infants' hearing was tested by recording auditory brain stem evoked potentials (Gorga, Reiland, Beauchaine, Worthington, & Jesteadt, 1987). All infants had clearly identifiable wave V and were thus enrolled into the study. During the EEG recording, the neonates were in quiet or active sleep, lying supine in an infant cot, and cared for by their mother and the research nurse. ERPs were averaged across the different sleep stages, because these were approximately evenly distributed across the different conditions. The experimental session took about 2 h, overall. This time included the setup and removal of the electrodes, the experimental stimulus blocks, and the auditory brain stem test.

Stimuli

Stimuli consisted of two complex tones, each with three harmonic frequencies added to the base frequency. The two tones differed only in frequency. The base frequency of the low tone was 500 Hz and that of the high tone 612 Hz. The amplitude of the first, second, and third harmonic components was one-half, one-quarter, and one-eighth of the base harmonic amplitude, respectively. The tones were 50 ms long, including 5-ms rise and 5-ms fall times, and their overall intensity was 75 dB (sound pressure level). The stimuli were presented binaurally using the E-Prime software (Psychology Software Tools, Inc., Pittsburgh, PA) via ER-3A loudspeakers (EtymOtic Research, Inc., Elk Grove Village, IL) connected via sound tubes to self-adhesive ear couplers (Natus Medical, Inc., San Carlos, CA), which were placed over the infants' ears.

Procedure

In two stimulus conditions, standard (S) and deviant (D) tones were delivered with a SOA of 100 ms, the D tone appearing less frequently (20%) than S tones (80%). In the random condition, tones were delivered in a pseudorandomized order with at least two standards separating successive deviants. In the grouped condition, the SSSSD pattern was cyclically repeated. The overall probability of the D tone was identical between the two conditions (20%). To eliminate possible confounds stemming from acoustic differences between the S and D tones, low and high tones serving as standards and deviants were swapped across different stimulus blocks, separately for each condition. In each of the four stimulus blocks, a total of 1000 stimuli were delivered.

EEG Recording

EEG was recorded from F3, F4, C3, Cz, and C4 scalp electrodes (according to the international 10–20 system) and from electrodes placed over the left and right mastoids (Lm and Rm, respectively), with the common reference attached to the tip of the nose. The ground electrode was placed on the forehead. Eye movements were monitored by recording the electrooculogram (EOG) between two electrodes, one placed below the left eye and another above the right eye. EEG was recorded at 24-bit resolution and a sampling rate of 250 Hz by a direct-coupled amplifier (V-Amp, Brain Products, Munich, Germany). The signals were low-pass filtered at 40 Hz.

Data Analysis

Event-related potentials. Raw EEG data were filtered off-line between 2.5 and 16 Hz. For each stimulus, an epoch of 500-ms duration including a 100-ms prestimulus period was extracted from the continuous EEG record. Epochs with a voltage change below 0.1 µV or above 100 µV on any EEG or EOG channel were rejected from further analysis. Responses were analyzed for the central line of electrodes, whose impedance was below 25 kΩ for each infant. Epochs were baseline corrected using the 100-ms prestimulus period and averaged separately for the different stimulus types and conditions. The mean number of artifact-free trials per infant was 376 and 556 for deviants and standards, respectively.

For amplitude measurements, a 40-ms-long time window was selected, which was centered on the peak of the grand-averaged deviant-minus-standard difference waveform in the 172–212 ms latency range of the central (Cz) recording channel of the random condition, where MMN elicitation was expected (see Figure 1A). The effects of stimulus condition and stimulus type on the ERP responses were analyzed with a three-way repeated-measures analysis of variance (ANOVA): Condition (random vs. grouped) × Stimulus Type (Standard vs. Deviant) × Electrode (C3 vs. Cz vs. C4). Greenhouse–Geisser correction of the degrees of freedom was applied where appropriate and ε values are given in the text. The results of the ANOVA were further specified by planned comparisons based on the results of the similar adult studies (see the Introduction). A similar statistical analysis was conducted for the positive wave following the MMN-like negative difference in the 250–350-ms latency range (Figure 1A).

Event-related spectral perturbation (ERSP). The dynamics of oscillatory EEG activity in newborns was studied by computing event-related spectral perturbation transforms (ERSP; Makeig, 1993). ERSP time-frequency matrices show changes from the spectral power baseline, allowing the study of the time course of the EEG signal energy, separately for each frequency. This method generalizes the narrow-band measures of event-related synchronization and desynchronization introduced by Pfurtscheller and Aranibar (1977) and covers phase-locked and non-phase-locked contributions. The principle of calculating the ERSP is to compute the power spectrum of the EEG signal from a sliding time window. For n trials, if $F_k(f,t)$ is the power of trial k at frequency f and time t, the ERSP value is calculated as

$$ERSP(f,t) = \frac{1}{n}\sum_{k-1}^{n} |F_k(f,t)|^2.$$

To obtain the $F_k(f,t)$ function (the signal power at a given frequency and time point), the EEG signal was convolved with Hanning-windowed sinusoidal wavelets. The number of wavelet cycles increased evenly with frequency (starting at three cycles at 6 Hz) for optimal time-frequency resolution. We analyzed epochs extending from 1000 ms before to 2000 ms after stimulus onset in the 6–40-Hz frequency range. The sliding window was 512 ms wide, and it was applied 400 times with an average step size of 6 ms. The ERSP time-frequency matrices were baseline corrected by the average power calculated from the − 1000 to 0 ms prestimulus period. Dynamical changes in oscillatory activity were studied by computing ERSPs for each individual trial, then averaging them separately for the standard and deviant stimuli (Tallon-Baudry & Bertrand, 1999; Herrmann, Munk, & Engel, 2004). EEG signal processing and analysis for ERSP computation were carried out using the EEGLAB interactive Matlab toolbox (Delorme & Makeig, 2004).

For the ERSP calculations, the raw EEG data were filtered off-line with a 1-Hz high-pass filter. Epochs with a voltage change below 0.1 µV or above 100 µV in the 356-ms prestimulus to 656-ms poststimulus interval at any EEG or EOG channel were rejected from further analysis. On average, 172 trials were included from each stimulus block per subject, separately for the standard and deviant stimuli. Two infants' data were discarded from the ERSP analyses due to low numbers of accepted trials ($n < 100$ per block).

For statistical analysis of the gamma-band activity, a window of 140-ms duration was selected in the 30–40-Hz frequency range. The window was centered on the peak of the observable gamma-band oscillations in the grand-averaged central (Cz) ERSP response to deviants delivered in the random condition, where deviance-related gamma-band activity could be expected (Edwards et al., 2005). The effects of stimulus condition and stimulus type on the observed gamma-band activity were analyzed with a three-way repeated-measures analysis of variance (ANOVA): Condition (random vs. grouped) × Stimulus Type (Standard vs. Deviant) × Electrode (C3 vs. Cz vs. C4). Greenhouse–Geisser correction of the degrees of freedom was applied where appropriate and ε values are given in the text. The results of the ANOVA were further specified by planned comparisons.

Results

ERP Results

Grand-average ERPs to deviants in the random condition displayed a fronto-centrally negative wave between 150 and 250 ms from stimulus onset (Figure 1A). No similar wave was observed in the grouped condition. The Condition × Stimulus Type × Electrode ANOVA of the ERP amplitudes measured in the 172–212-ms window revealed a significant Condition × Stimulus Type interaction, $F(1,11) = 6.33$, $p < .05$, $\eta^2 = 0.36$, which was explained by a significant difference between the standard and deviant stimulus responses in the random condition but not in the grouped condition (tested by planned comparisons, $p < .05$). No other effects or interactions reached the level of significance. That is, the D tones elicited a different response when they appeared as part of the repeating pitch pattern than in the randomized sequence.

The central negativity elicited by deviants in the random condition was followed by a positivity with a similar scalp distribution. The ANOVA of the ERP amplitudes measured in the 262–322-ms window yielded no significant results.

ERSP Results

Wavelet-based time-frequency analysis of the single-trial responses revealed peaks of electrical brain activity in the gamma band (Figure 1B), consistent with our hypothesis. In the random condition, at approximately 200 ms from stimulus onset, strong gamma-band oscillations appeared with a center frequency of about 36 Hz in response to deviants. In contrast, standards were followed by a decrease in the gamma-band activity amplitude. However, no such difference can be seen in the grouped condition. The three-way ANOVA (Condition × Stimulus Type × Electrode) of the gamma-band activity revealed a significant main effect of Stimulus Type, $F(1,9) = 10.224$, $p < .05$, $\eta^2 = .53$, which was explained by a significant difference between the standard and deviant stimulus responses in the random condition (tested by planned comparisons, $p < .05$), whereas the same difference was not significant in the grouped condition. No other effects or interactions reached the level of significance.

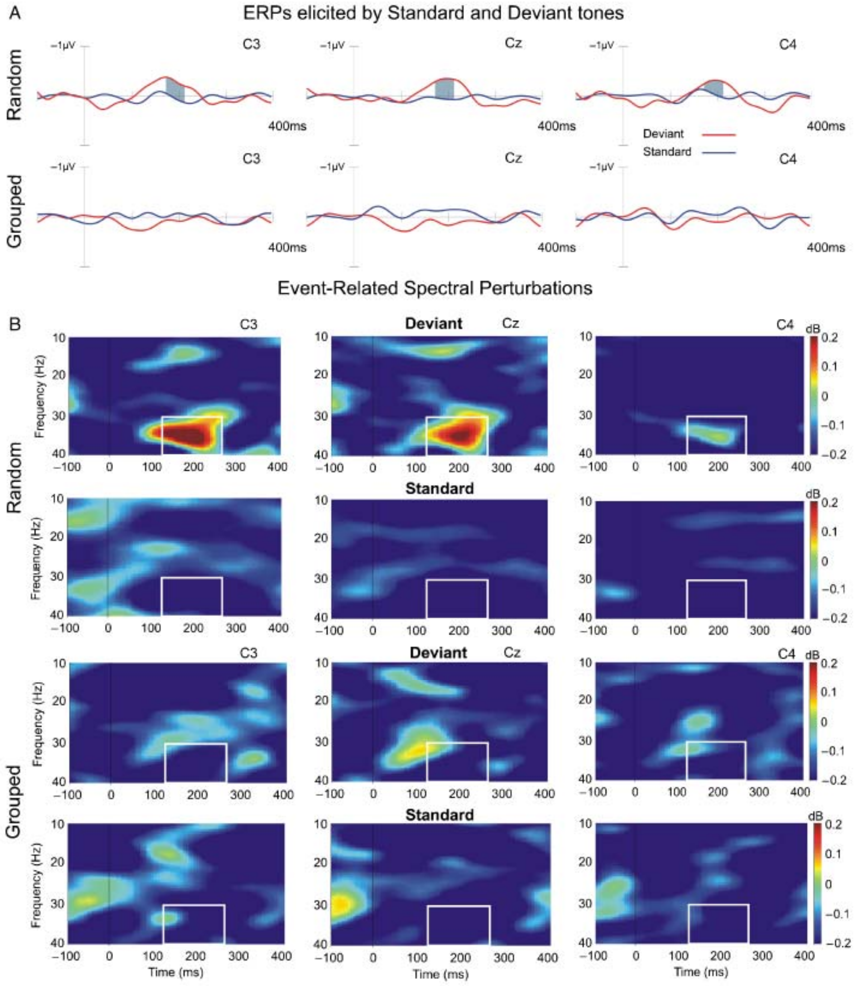

Figure 1. A: Grand-averaged ERPs elicited by frequent ("standard," blue line) and rare ("deviant," red line) sounds in the random condition (top row) and grouped condition (bottom row) at three central electrode locations (C3, Cz, and C4). Stimulus onset is at the crossing of the x (time in milliseconds) and y (amplitude in microvolts) axes. The interval in which deviants elicited a significantly more negative response in the random condition is marked by light blue shading. B: Grand-averaged event-related spectral perturbation (ERSP) time-frequency matrices calculated by wavelet-based analysis of single-trial data from the random condition (top rows) and grouped condition (bottom rows), separately for the frequent sound (upper row) and infrequent sound (lower row) at three central electrode locations (C3, Cz, and C4). The signal power (in decibels) relative to the prestimulus baseline (− 1000 to 0 ms) is shown by color code, with warmer colors representing higher power values (color calibration is at the right side of each row). The time-frequency window used for measurements is marked by white boxes.

DISCUSSION

We found that in the random condition, deviant tones elicited a significantly more negative ERP waveform in the 172 212-ms poststimulus interval than standard tones. Because the newborn infants were asleep during the EEG recordings, this differential negativity reflects the activation of an automatic deviance-detection system, which may be similar to that underlying the MMN response in adults, as was suggested by several studies (e.g., Alho et al., 1990; Kushnerenko et al., 2002). In the grouped condition, no such a difference was observed. The difference between the two conditions suggests that, similarly to adults (Sussman & Gumenyuk, 2005; Sussman et al., 1998, 2002), in the grouped condition, the repetition of the SSSSD pattern was detected and, therefore, S and D tones were processed as part of the same regularity. Thus, it appears that the neonate auditory system detects repeating pitch patterns quite similarly to adults.

Our analysis of gamma-band oscillations triggered by deviants and standards in the random condition showed a clearly discernible burst of gamma activity for deviants, whereas standards were followed by a decrease in amplitude in the 30–40-Hz frequency range 140–270 ms from stimulus onset. Induced oscillatory gamma activity with a similar time course has been observed for occasional frequency deviants (Edwards et al., 2005; Kaiser et al., 2000) and omissions of a repeating tone (Gurtubay et al., 2006). Thus, similarly to adults, newborn infants show gamma-band oscillatory activity that is sensitive to stimulus probability. The gamma-band response corroborates the ERP finding in showing that the D tones were processed differently from the S tones in the random condition, but not in the grouped condition.

The current and some previous results (e.g., Carral et al., 2005; Ruusuvirta et al., 2003; Winkler et al., 2003) provide important glimpses into intermediate levels of auditory processing in human neonates. The fragments we have learned about so far suggest that newborn infants possess basic capabilities for sound organization similar to those of adults. This does not mean that neonates perceive the auditory world similarly to adults. Sensitivity to various types of sound as well as sound discrimination abilities develop until well into childhood (e.g., Allen & Wightman, 1992; Trehub, Schneider, Morrongiello, & Thorpe, 1988). Furthermore, perception is largely dependent on experience (e.g., Bregman, 1990). On the other hand, newborns have the means to extract invariant relationships from varying input, form composite units, and so forth. In short, they are able to organize and represent auditory information. These capabilities are crucial for finding relevant information in realistic environments, and thus they underlie learning, which starts even before birth (DeCasper & Spencer, 1986), because auditory functions emerge very early during prenatal development (Chugani & Phelps, 1986). So, what are the main directions of maturation and learning, beyond improved sensitivity and discrimination thresholds? In a recent study, we compared the ERP responses of adults and neonates to environmental sounds embedded in the repetitive sequence of a simple harmonic tone (Kushnerenko et al., in press). We found that neonates were much more sensitive to the surface features of the environmental sounds than adults (i.e., these sounds covered a wider range of frequencies and were somewhat louder than the repeating tone). In contrast, adults show much faster categorization than what we found in newborns, and their categories are finer than that observed in newborns or in 6-month-old infants (for the latter, see Quinn, Westerlund, & Nelson, 2006). The ability of the neonate auditory system to form groups from a short sequence of sounds, which was revealed by the current results, is an important prerequisite of forming auditory objects (Kubovy & Van Valkenburg, 2001). Therefore, it underlies higher level operations, such as categorization, ultimately enabling infants to learn to speak or understand music.

Conclusion

The ERPs and induced gamma-band responses obtained in this study show that the neonate auditory system is sensitive to the large-scale structure of a sound sequence and that the sequential context can determine the way individual sounds are processed in the newborn brain. Our results demonstrate the operation of a sound grouping function in newborn infants. The ERSP findings suggest that, similarly to adults, newborn infants show gamma-band oscillatory activity, which is sensitive to stimulus probability.

REFERENCES

Alho, K., Sainio, K., Sajaniemi, N., Reinikainen, K., & Näätänen, R. (1990). Event-related brain potential of human newborns to pitch change of an acoustic stimulus. *Electroencephalography and Clinical Neurophysiology*, *77*, 151–155.

Allen, P., & Wightman, F. (1992). Spectral pattern discrimination by children. *Journal of Speeech & Hearing Research*, *35*, 222–233.

Barker, B. A., & Newman, R. S. (2004). Listen to your mother! The role of talker familiarity in infant streaming. *Cognition*, *94*, B45–B53.

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

Carral, V., Huotilainen, M., Ruusuvirta, T., Fellman, V., Näätänen, R., & Escera, C. (2005). A kind of auditory 'primitive intelligence' already present at birth. *European Journal of Neuroscience*, *21*, 3201–3204.

Chugani, H., & Phelps, M. (1986). Maturational changes in cerebral function in infants determined by 18-FDG positron emission tomography. *Science*, *231*, 840–843.

Csibra, G., Davis, G., Spratling, M. W., & Johnson, M. H. (2000). Gamma oscillations and object processing in the infant brain. *Science*, *290*, 1582–1585.

DeCasper, A., & Spencer, M. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development*, *9*, 113–150.

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

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21.

Edwards, E., Soltani, M., Deouell, L. Y., Berger, M. S., & Knight, R. T. (2005). High gamma activity in response to deviant auditory stimuli recorded directly from human cortex. *Journal of Neurophysiology*, *94*, 4269–4280.

Friederici, A. D., Friedrich, M., & Weber, C. (2002). Neural manifestation of cognitive and precognitive mismatch detection in early infancy. *NeuroReport*, *13*, 1251–1254.

Gorga, M. P., Reiland, J. K., Beauchaine, K. A., Worthington, D. W., & Jesteadt, W. (1987). Auditory brainstem responses from graduates of an intensive care nursery: Normal patterns of response. *Journal of Speeech & Hearing Research*, *30*, 311–318.

Gurtubay, I. G., Alegre, M., Labarga, A., Malanda, A., Iriarte, J., & Artieda, J. (2001). Gamma band activity in an auditory oddball paradigm studied with the Wavelet Transform. *Clinical Neurophysiology*, *112*, 1219–1228.

Gurtubay, I. G., Alegre, M., Valencia, M., & Artieda, J. (2006). Cortical gamma activity during auditory tone omission provides evidence for the involvement of oscillatory activity in top-down processing. *Brain Research Experimental Brain Research*, *175*, 463–470.

Herrmann, C. S., Munk, M. H., & Engel, A. K. (2004). Cognitive functions of gamma-band activity: Memory match and utilization. *Trends in Cognitive Science*, *8*, 347–355.

Kaiser, J., Lutzenberger, W., Preissl, H., Ackermann, H., & Birbaumer, N. (2000). Right-hemisphere dominance for the processing of sound-source lateralization. *Journal of Neuroscience*, *20*, 6631–6639.

Kaufman, J., Csibra, G., & Johnson, M. H. (2003). Representing occluded objects in the human infant brain. *Proceedings. Biological Sciences/The Royal Society*, *270*, S140–143.

Kubovy, M., & Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition*, *80*, 97–126.

Kushnerenko, E., Čeponienë, R., Balan, P., Fellman, V., & Näätänen, R. (2002). Maturation of the auditory change detection response in infants: A longitudinal ERP study. *NeuroReport*, *13*, 1843–1848.

Kushnerenko, E., Winkler, I., Horváth, J., Näätänen, R., Pavlov, I., Fellman, V., et al. (in press). Processing acoustic change and novelty in newborn infants. *European Journal of Neuroscience*.

Makeig, S. (1993). Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. *Electroencephalography and Clinical Neurophysiology*, *86*, 283–293.

Morr, M. L., Shafer, V. L., Kreuzer, J. A., & Kurtzberg, D. (2002). Maturation of mismatch negativity in typically developing infants and preschool children. *Ear and Hearing*, *23*, 118–136.

Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*, 313–329.

Novitski, N., Huotilainen, M., Tervaniemi, M., Näätänen, R., & Fellman, V. (2007). Neonatal frequency discrimination in 250–4000-Hz range: Electrophysiological evidence. *Clinical Neurophysiology*, *118*, 412–419.

Pfurtscheller, G., & Aranibar, A. (1977). Event-related cortical desynchronization detected by power measurements of scalp EEG. *Electroencephalography and Clinical Neurophysiology*, *42*, 817–826.

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

Quinn, P. C., Westerlund, A., & Nelson, C. A. (2006). Neural markers of categorization in 6-month-old infants. *Psychological Science*, *17*, 59–66.

Reid, V. M., Csibra, G., Belsky, J., & Johnson, M. H. (2007). Neural correlates of the perception of goal-directed action in infants. *Acta Psychologica (Amsterdam)*, *124*, 129–138.

Ruusuvirta, T., Huotilainen, M., Fellman, V., & Näätänen, R. (2003). The newborn human brain binds sound features together. *NeuroReport*, *14*, 2117–2119.

Scherg, M., Vajsar, J., & Picton, T. W. (1989). A source analysis of the late human auditory evoked potentials. *Journal of Cognitive Neuroscience*, *1*, 336–355.

Sussman, E., & Gumenyuk, V. (2005). Organization of sequential sounds in auditory memory. *NeuroReport*, *16*, 1519–1523.

Sussman, E., Ritter, W., & Vaughan, H. G., Jr. (1998). Predictability of stimulus deviance and the mismatch negativity. *NeuroReport*, *9*, 4167–4170.

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

Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Science*, *3*, 151–162.

Trehub, S., Schneider, B., Morrongiello, B., & Thorpe, L. (1988). Auditory sensitivity in school-age children. *Journal of Experimental Child Psychology*, *29*, 273–285.

Vaughan, H. G. Jr., & Kurtzberg, D. (1991). Electrophysiologic indices of human brain maturation and cognitive development. In M. R. Guunar & C. A. Nelson (Eds.), *Minnesota symposia on child psychology* (vol. 24, pp. 1–36). Hillsdale, NJ: Erlbaum.

Winkler, I., Kushnerenko, E., Horváth, J., Čeponienė, R., Fellman, V., Huotilainen, M., et al. (2003). Newborn infants can organize the auditory world. *Proceedings of the National Academy of Sciences, USA*, *100*, 1182–1185.

(Received February 6, 2007; Accepted April 4, 2007)

3.4. Study IV: Processing of meter in adults

Ladinig, O., Honing, H., Háden, G., Winkler, I. (2009) Probing attentive and pre-attentive emergent meter in adult listeners without extensive music training. *Music Perception* 26(4), 377-386. DOI: 10.1525/MP.2009.26.4.377

Due to errors in data analysis discovered following publication, an errata report was submitted to and accepted by Music Perception, which is reproduced in verbatim below.

Errata for:

Ladinig, O., Honing, H., Háden, G., & Winkler, I. (2009). Probing attentive and pre-attentive emergent meter in adult listeners without extensive music training. *Music Perception, 26*, 377-386.

Due to the authors' error, data analysis was not performed as stated in the published article. Here we present the corrected analysis (as stated in the Method section of the paper). The results are similar to those published with a few exceptions. These are noted and corresponding interpretations are provided. The error affected the MMN amplitudes entered into the Attention (Unattended vs. Passive) × Position (Strong vs. Weak) × Frontality (Frontal vs. Central electrode line) × Laterality (Left vs. Middle vs. Right) ANOVA. MMN amplitudes were, in contrast to the published text and the figure (which are correct), incorrectly measured using the standard as opposed to the control data. MMN amplitudes have been remeasured using 30 ms wide time windows centered on the difference peaks (in contrast to the 60 ms window stated in the published text). This shorter time windows better fit the peaks on the deviant-minus-control difference waveforms. Table 1 with corrected measurements is included below. ANOVAs for these amplitude measures yielded the following results: significant main effects were obtained for Frontality, $F(1, 10) = 19.96$, $p < .01$, $\eta^2 = 0.67$, and Laterality, $F(2, 20) = 4.14$, $p < .05$, $\eta^2 = 0.29$; significant interactions between Attention × Frontality, $F(1, 10) = 15.30$, $p < .01$, $\eta^2 = 0.60$, Frontality × Laterality, $F(2, 20) = 5.00$, $p < .05$, $\eta^2 = 0.33$, and Attention × Position × Laterality, $F(1, 10) = 7.32$, $p < .05$, $\eta^2 = 0.42$. A posthoc Tukey HSD test revealed that the Attention × Frontality interaction was due to significantly higher central MMN amplitudes in the Unattended condition than in any other combination of the two factors ($df = 10$, $p < .01$ for all comparisons).

The lack of a significant Position effect on the deviant-minus-control difference amplitudes somewhat weakens the argument for representing the hierarchical metrical structure in the MMN system. However, the latency effect remains (it was correctly reported in the published work) and MMN peak latency rather than amplitude is the stronger indicator of the perceived magnitude of change (Schröger & Winkler, 1995). Please note that the behavioral results also indicate processing differences for omission of beats of different metrically based salience. The interactions involving the Attention factor probably reflect a partial overlap between the N2b and the MMN components: the more central than frontal scalp distribution of N2b interacting with the more frontal than central scalp distribution of MMN as a function of the attention condition. This result casts some doubt on whether the extraction of the metrical hierarchy is fully automatic. Further experiments are needed to clarify this issue.

References:

Schröger, E., & Winkler, I. (1995). Presentation rate and magnitude of stimulus deviance effects on human pre-attentive change detection. *Neuroscience Letters*, 193, 185-188.

Table 1 (corrected).

Group-Averaged MMN Amplitudes in µV with Standard Errors of the Mean (SEM) in Parentheses.

Attention	Passive		Unattended	
Electrode/Position	Strong	Weak	Strong	Weak
F3	-1.07 (0.28)	-0.47 (0.35)	-0.82 (0.28)	-0.92 (0.28)
Fz	-1.08 (0.31)	-0.73 (0.50)	-1.01 (0.34)	-1.24 (0.29)
F4	-0.82 (0.24)	-0.49 (0.51)	-0.23 (0.13)	-1.14 (0.25)
C3	-0.89 (0.22)	-0.61 (0.33)	-1.44 (0.38)	-1.30 (0.31)
Cz	-1.03 (0.30)	-0.65 (0.42)	-1.57 (0.38)	-1.61 (0.32)
C4	-0.85 (0.24)	-0.63 (0.38)	-1.44 (0.33)	-1.59 (0.32)

PROBING ATTENTIVE AND PREATTENTIVE EMERGENT METER IN ADULT LISTENERS WITHOUT EXTENSIVE MUSIC TRAINING

OLIVIA LADINIG AND HENKJAN HONING
Universiteit van Amsterdam, Amsterdam,
The Netherlands

GÁBOR HÁDEN AND ISTVÁN WINKLER
Hungarian Academy of Sciences, Budapest, Hungary

BEAT AND METER INDUCTION ARE CONSIDERED important structuring mechanisms underlying the perception of rhythm. Meter comprises two or more levels of hierarchically ordered regular beats with different periodicities. When listening to music, adult listeners weight events within a measure in a hierarchical manner. We tested if listeners without advanced music training form such hierarchical representations for a rhythmical sound sequence under different attention conditions (Attend, Unattend, and Passive). Participants detected occasional weakly and strongly syncopated rhythmic patterns within the context of a strictly metrical rhythmical sound sequence. Detection performance was better and faster when syncopation occurred in a metrically strong as compared to a metrically weaker position. Compatible electrophysiological differences (earlier and higher-amplitude MMN responses) were obtained when participants did not attend the rhythmical sound sequences. These data indicate that hierarchical representations for rhythmical sound sequences are formed preattentively in the human auditory system.

Received June 22, 2008, accepted January 12, 2009.

Key words: rhythm, emergent meter, syncopation, event-related potentials, attention

THE CONCEPTS OF *BEAT* AND *METER* ARE well-established terms in music production and perception (Clarke, 1999; London, 2004). Most authors agree that *beat induction*, the cognitive ability that allows one to infer a regular beat (or pulse) from a musical excerpt, is universal in and unique to humans, enabling us to entrain to music, and coordinate our movements with others (Honing, 2002). Meter can be defined as being composed of at least two levels of beat with different periodicities. However, there is little agreement in the literature regarding the perceptual/cognitive reality of meter. Is meter simply a concept facilitating the structuring of written musical scores, introduced by composers and performers, or are there indeed some cognitive faculties reflected in the concept of meter? Beat induction can be considered the simplest case of meter, and refers to the subjective emphasis of certain elements of a rhythm (but also in an isochronous stream of clicks), making some elements more salient than others; the beat or *tactus* (Lerdahl & Jackendoff, 1983) is usually equally spaced in time, and is reflected in spontaneous tapping and dancing, usually with an interbeat interval close to 600 ms (Bolton, 1894; Brochard, Abecasis, Potter, Ragot, & Drake, 2003; London, 2004; Yeston, 1976). Meter, seen here as a more fine-grained differentiation of the elements of a rhythm due to multiple levels of hierarchically ordered regular beats, requires the specification of a fixed entity of duration, in this case one *musical measure*. Theoretical models (Lerdahl & Jackendoff, 1983; Longuet-Higgins & Lee, 1984) specify *metric salience*, a value assigned to each sequential position of a rhythmic sound pattern regarding to its position within that measure, by recursively breaking down a musical pattern (with an initially specified length) into subpatterns of equal length (see the top of Figure 1).

The number of recursive subdivisions needed to arrive at a given point (event) in a rhythmic pattern governs the salience of that point: the more subdivisions needed, the lower the salience of the point. The first position in the measure (referred to as the *downbeat*) receives the highest salience in any pattern. In other words, meter reflects the fact that different events in a musical pattern have different importance for the listener. In general, it holds that the higher the salience of an event compared to other events within the same measure, the more listeners expect it to occur. A high-salience event is more important for processing the measure, as indicated for example by the fact that it gets memorized and recalled easier, and, if it is absent, the measure will be perceived as being more complex (Fitch

Music Perception VOLUME 26, ISSUE 4, PP. 377–386, ISSN 0730-7829, ELECTRONIC ISSN 1533-8312 DOI:10.1525/MP.2009.26.4.377

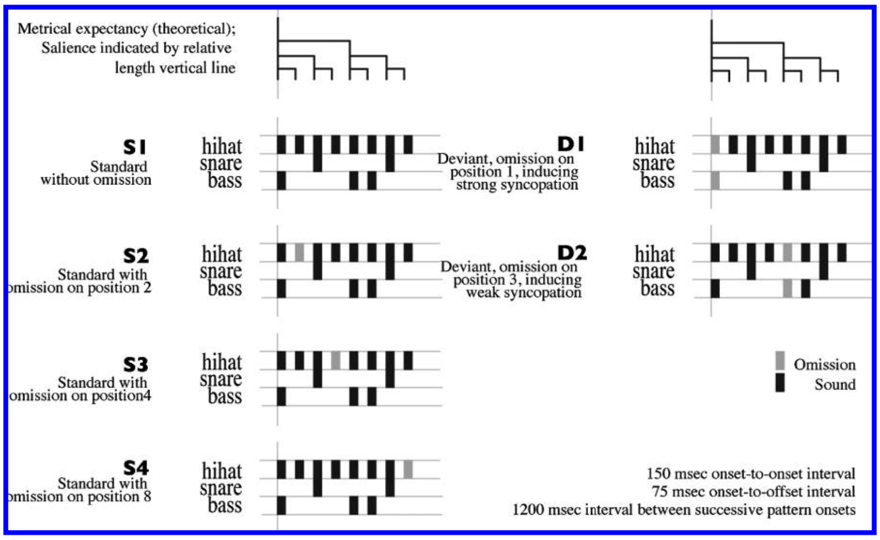

FIGURE 1. Schematic illustration of the stimuli used in the experiment. The top of the figure represents the recursive subdivision of a rhythmic pattern (cf. Martin, 1972) with eight equidistant grid points. The horizontal dimension represents the subdivisions of one musical measure; the vertical dimension represents event salience (i.e., increasing salience with longer lines).

& Rosenfeld, 2007; Pressing, 2002). Supporting this notion, Palmer and Krumhansl (1990) showed, for a corpus of Western classical music, that the average distribution of event occurrences within a measure was highly correlated with the theoretical model proposed by Lerdahl and Jackendoff (1983).

Existing theories disagree whether or not sensitivity to meter is prevalent in all listeners, and where such sensitivity, if any, would come from. Specifically, the question is, whether or not listeners form multilevel hierarchical representations for rhythmic sequences. Expectations in adult listeners with formal music training suggest that they weight events within a measure in a hierarchical manner (Jongsma, Desain, & Honing, 2004; Palmer & Krumhansl, 1990). A study by Ladinig and Honing (2009) shows that this holds irrespective of listener's musical expertise. Furthermore, recent evidence suggests that already at a very early age (e.g., at seven months of age), human infants are sensitive to metric violations (Hannon & Johnson, 2005). Thus it is possible that humans possess some processing predisposition to extract hierarchically structured regularities from complex patterns. Lower–level chunking processes are usually more or less automatic (i.e., they proceed even when one does not attend the given stimuli; e.g., temporal integration, see Cowan, 1984). In contrast, higher-level chunking processes typically require attention to be focused on the stimuli, because they rely on voluntary allocation of limited–capacity resources (e.g., finding sentences in continuous speech). The crucial question is whether or not the hierarchical representation characterizing meter emerges when the rhythmical sound sequence falls outside the focus of attention.

In the current study, we tested whether meter (hierarchical representation for a rhythmical sound sequence) emerges in adults with no extensive music training, and whether meter emergence is modulated by attention. To this end, reactions to meter violations were assessed using behavioral and electrophysiological measures. Reaction time (RT) and discrimination sensitivity (*d'*) measurements served to characterize active detection of meter violations, whereas event-related brain potentials (ERP) were used to assess the detection of meter violations under different task loads while the rhythmic sound sequences were not relevant to the participants' task. The mismatch negativity (MMN) ERP component (Näätänen, Gaillard, & Mäntysalo, 1978; for a recent review, see Näätänen, Paavilainen, Rinne, & Alho, 2007) can be used as a sensitive tool for determining which regular features of a sound sequence the brain has detected, because MMN is elicited by sounds violating detected auditory regularities. Furthermore, MMN is elicited even when participants perform a task that is unrelated to the test sound sequence (for a review of the effects of attention on MMN, see Sussman, 2007).

MMN has been shown to reflect violations of musical regularities and the effects of music training (for a review, see Tervaniemi & Huotilainen, 2003). For example, Trainor, McDonald, and Alain (2002) showed that participants with no formal music training detected occasional pitch interval changes within transposed melodies in the absence of focused attention. Other studies showed sensitivity to musical key (e.g., Brattico, Tervaniemi, Näätänen, & Peretz, 2006), mistuning of chords (Leino, Brattico, Tervaniemi, & Vuust, 2007), etc. Although fewer previous investigations addressed rhythm processing with the MMN method (the exceptions are Pablos Martin, Deltenre, Hoonhorts, Markessis, Rossion, & Colin, 2007; Vuust et al., 2005), the representation of simpler temporal features has been studied in more detail. For example, it was found that occasionally shortening the interstimulus interval in an otherwise isochronous sequence of sounds elicits the MMN (Nordby, Roth, & Pfefferbaum, 1988). Omitting a sound from a sequence delivered at a fast presentation rate also triggers the MMN response (Yabe, Tervaniemi, Reinikainen, & Näätänen, 1997). Regarding more complex temporal patterns, Pablos Martin and colleagues (2007) found faster processing of binary (e.g., 1:2) as opposed to nonbinary (e.g., 1:3) interval ratios. Finally, music training effects have been shown for both melodic (e.g., Brattico & Näätänen, 2002; Fujioka, Trainor, Ross, Kakigi, & Pantev, 2004) and rhythmic patterns (Vuust et al., 2005; van Zuijen, Sussman, Winkler, Näätänen, & Tervaniemi, 2005). The current interpretation of MMN generation suggests that this ERP component is elicited in response to deviations from expected sounds (Baldeweg, 2007; Winkler, 2007). This makes MMN especially appropriate for testing the emergence of musical meter, because it allows one to compare the strength of expectations between violations at different positions of a rhythmical pattern. The strength of expectation is a prime behavioral correlate of the hierarchical metric structure and more salient deviations trigger earlier and possibly larger-amplitude MMN responses (for a review, see Näätänen & Alho, 1997).

Based on these principles, we presented participants with sound sequences consisting of four sound patterns (Figure 1) having strictly metrical rhythms of the same type (Standard patterns; 90% of the patterns overall), and two patterns that were syncopated variants of the same rhythm (Deviant patterns; 10% overall). One deviant violated the standard pattern at the downbeat position (strong syncopation), and the other at the second most salient position (weaker syncopation). If the brain creates a hierarchical representation for the rhythm of the sound sequences, syncopation at the downbeat is expected to elicit stronger responses from participants than syncopation at the metrically less salient position. "Stronger" response means better detection performance when syncopated patterns are designated as targets and earlier and possibly higher-amplitude MMN response when participants ignore the rhythmic sequence. If, however, the sound sequence is represented in terms of a single-level structure, then sounds in all positions are equally expected by the brain and, therefore, the responses to syncopation will not be stronger at the downbeat than in the metrically less salient position.

Effects of attention were tested at three levels: (1) meter violations are task-relevant (Behavioral Experiment); (2) meter violations are task-irrelevant: participants perform an easy concurrent task (watching a muted movie with subtitles; Electrophysiological Experiment, "Passive Condition"); and (3) meter violations are task-irrelevant: participants perform a difficult concurrent task (detecting unpredictable slight intensity changes in a noise stream; Electrophysiological Experiment, "Unattend Condition"). If forming a hierarchical representation of the rhythmical sound sequence required focused attention, then the strength of expectation should only depend on the position of the syncopation within the pattern when participants focus their attention on the sound sequence. If, however, a hierarchical representation of the rhythmical sound sequence is formed even without focused attention, then syncopation is expected to elicit a stronger response at the downbeat than in the metrically less salient position, irrespectively of the attention condition.

Method

Participants

Twelve healthy volunteers (seven male, $M = 22.83$, $SD = 3.93$) participated in the experiment. Participants gave informed consent after the procedures and aims of the experiments were explained to them. The study was approved by the Ethical Committee (institutional review board) of the Institute for Psychology, Hungarian Academy of Sciences. All participants had frequency thresholds not higher than 20 dB SPL in the

250-4000 Hz range and no threshold difference exceeding 10 dB between the two ears (assessed with a Mediroll, SA-5 audiometer). All participants reported to have received less than one year of music training (i.e., playing an instrument, or singing in a choir) after the obligatory music lessons in primary/secondary school in the past, and did not perform music regularly (defined as once a month) for the past two years. Each participant was tested in both experiments (behavioral and electrophysiological), which were carried out in one session on the same day. One participant's (male, age 20) data was excluded from the analyses because of measurement errors. Throughout the experiments, participants sat in a comfortable chair in the sound-attenuated experimental chamber of the Institute for Psychology, Budapest.

Stimuli

Six different sound patterns were constructed (see Figure 1), which were variants of a rhythmic rock pattern (base-pattern, S1) with eight grid points. The rhythmic patterns were presented by a typical rock-drum accompaniment using snare and bass, and with a hihat on every grid point. The base pattern and the three variants (containing omissions on the lowest metrical level) were *strictly metrical*; that is, they contained no syncopation or slurred notes throughout the pattern. Together, these four metric patterns formed the set of standard patterns (S1-S4). In order to avoid the confound of finding responses resulting from simple pattern matching, a set of sound patterns that share the characteristic of being strictly metrical and regular rhythms, instead of a single sound pattern, was employed to constitute the standard ("abstract MMN"). Two deviants were constructed by omitting events on metrically salient positions in the base-pattern, which lead to syncopated patterns: A strongly syncopated pattern was created by omitting the downbeat (D1), and a slightly weaker syncopation by omitting the second most important beat (D2). Sounds were generated using QuickTime's *drum* timbres (Apple Inc.). Sound duration was 50 ms for *hihat*, 150 ms for *snare*, and 100 ms for *bass* sounds. The interval between grid points (onset-to-onset interval) was 150 ms. Thus each pattern lasted 1200 ms, with no extra silence between patterns (i.e., they formed a continuous stream of rhythm).

Procedures for the Behavioral Experiment

In the behavioral experiment, we assessed the effects of different metrical positions on deviance detection by asking participants to listen to two blocks of 300 continuously presented patterns and to indicate when they felt that there was a break in the rhythm by pressing a response button placed in their dominant hand. The instructions given to participants were as follows:

> You will be presented with sequences of a continuous, regular rhythm. From time to time, the rhythm will be disrupted by some irregularity. This irregularity can be described as if the rhythm appeared to break, or stumble, or get syncopated for a moment. Please indicate by pressing the button as soon as you think such an event occurred.

Two stimulus blocks with 90% standard patterns (S1, S2, S3, and S4 with equal probabilities of 22.5% each) were presented. In one block, D1 was the deviant rhythmic pattern (10%) and in the other block, D2 was the deviant rhythmic pattern (10%). Randomization was constrained so that at least three standard patterns intervened between successive deviants and with S4 never preceding a deviant. The latter constraint was necessary to avoid concatenating two gaps, because S4 had an omission at the last grid position, whereas D1 at the first. The stimuli were presented binaurally using MATLAB via headphones (Sennheiser HD-430), 60 dB over the individual hearing threshold. The order of the two stimulus blocks (differing in the deviant pattern) was balanced across participants.

Data Analysis for the Behavioral Experiment

For each participant, *d'* values (a measure of discrimination sensitivity; see Macmillan & Creelman, 1991) and average reaction-times (RT) for correct responses were computed using MATLAB. The *d'* values were calculated separately from the hit rates for the D1 and D2 deviants and the overall false alarm rate. Responses given within 200-2000 ms from the target (omission) onset were regarded as hits; all other responses as false alarms. Paired two-sample *t*-tests were performed to compare *d'* and RT between the two deviants.

Procedures for the Electrophysiological Experiment

The electrophysiological experiment was conducted always before the behavioral experiment. The fixed order was necessary to avoid drawing participants' attention to the rhythmic deviations. Electrodes were removed between the two experiments, thus giving participants approximately 30 minutes of break time between the two experiments.

The rhythmic stimulus sequences were constructed from the same sound patterns as in the behavioral experiment, but they were delivered by two loudspeakers positioned 0.40 m from the side and 0.15 m behind the participants' head. Sound intensity was again 60 dB above the participant's hearing threshold. A continuous white noise with its intensity alternating between 52 and 54 dB above the participant's hearing threshold was presented concurrently with the rhythmic sound sequences. The noise stream was used to direct attention away from the rhythmic sound sequence in the Unattend condition (see below). Intensity changes occurred randomly with 1.5–32.0 s ($M = 16.75$ s) between them. The noise stream was delivered by a third loudspeaker placed directly in front of the participant at a distance of 1.35 m. During the stimulus blocks, participants also watched a self-selected muted movie with subtitles.

Two attention conditions were employed with identical auditory stimulation (rhythmic sequence and continuous noise). In the Unattend Condition, participants were asked to press a response button to the intensity changes in the noise stream. Performance in the intensity change detection task (group-average hit rate $HR = 0.78$, standard deviation $SD = 0.12$, and reaction time $RT = 1035$ ms, $SD = 77$ ms) showed that the task was difficult but possible to perform at a relatively high level. In the Passive Condition, participants were instructed to ignore all sounds (both the rhythmic sequence and the continuous noise) and to follow a muted self-selected movie. Each condition received 10 stimulus blocks of 300 continuously presented rhythmic patterns. Stimulus blocks consisted of 90% standard patterns (S1, S2, S3, and S4 with equal probabilities of 22.5%, each), 5% of the D1, and 5% of the D2 pattern. Presenting both types of deviants within the same stimulus block ensured that they appeared within exactly the same context and thus the deviance-related ERP responses could be compared directly. Randomization was constrained so that at least three standard patterns intervened between successive deviants and, for the same reasons as mentioned above, the S4 pattern never preceded a deviant pattern. Constructing 90% of the sequence from four different frequent patterns was necessary to avoid MMN being elicited by simple pattern deviation and thus to allow us to interpret the ERP responses specific to the D1 and D2 deviants as related to rhythm violations. Occasional changes of a single repeating pattern are known to elicit MMN even when rhythm is not violated (e.g., Winkler & Schröger, 1995). In the current design, the "standard" (the sequences made up of S1, S2, S3, and S4) is the rhythm, not any given sound pattern, and the deviants are the rhythmic violations caused by D1 and D2. In order to be able to directly compare the deviance-related responses elicited by D1 and D2, these responses were derived by separately subtracting the response elicited by the D1 and D2 patterns when they were regular (standard) within a sequence from when they were violating the rhythm of the sequence (deviant). Thus the pattern-specific responses were eliminated from the difference waveforms, which could then be compared with each other. To this end, participants also were presented with two control stimulus blocks of 300 patterns presenting sequences composed of either the D1 or the D2 pattern alone. The responses recorded to the D1 and D2 patterns in the control stimulus blocks (i.e., when they are standard patterns) served to derive the MMN response (see the EEG data analysis section below). The order of the two attention conditions was balanced across participants. Stimulus blocks usually were separated by short 1-2 minutes breaks, with longer breaks allowing the participant to leave the experimental chamber inserted at need.

EEG Recording

The electroencephalogram (EEG) was recorded at the F3, Fz, F4, C3, Cz, and C4 scalp locations (according to the international 10-20 system) and the left and right mastoids (A1 and A2, respectively), with the common reference electrode attached to the tip of the nose. The ground electrode was placed on the forehead. Eye movements were monitored by recording the electrooculogram (EOG) between two electrodes placed above and below the left eye (vertical EOG) and between two electrodes placed lateral to the outer canthi on both sides (horizontal EOG). EEG was recorded with 32 bit resolution at a sampling rate of 250 Hz by a Neuroscan, NuAmps amplifier (Compumedics Neuroscan Inc.). The signals were on-line low-pass filtered at 40 Hz.

EEG Data analysis

EEG was filtered off-line between 0.1 and 20 Hz. For each D1 and D2 pattern (experimental and control stimulus blocks, separately), an epoch of 1200-ms duration was extracted from the continuous EEG record. The epoch started 600 ms before the onset of the deviation. Epochs with a voltage change below 0.1 μV or above 100 μV on any EEG or EOG channel within the—100 to 500 ms time window (relative to the deviation onset) were rejected from further analysis. Epochs were baseline-corrected by the average voltage of the whole analysis period and averaged separately for the two deviants and identical control patterns and in the two attention conditions. Using the whole analysis period as baseline balances possible slow shifts that may appear in the long analysis period. The mean number of artifact-free deviant trials per participant was 130.

MMN peak latencies were established as the central (Cz) negative maximum of the average deviant-minus-control difference waveform in the 100-250 ms *post-deviance* time-range, separately for each participant, deviant, and condition. Peak latencies were established automatically in the target latency range. In cases where two or more negative peaks fell within the 100-250 post-deviance time-window and the amplitude difference between the peaks was small ($<0.5\ \mu V$), selection of the latency was aided by visual inspection of waveforms recorded by the C and F electrodes. The effects of attention and deviance position were analyzed by a repeated-measure analysis of variance (ANOVA) with the structure Attention (Unattend vs. Passive) × Position (Strong vs. Weak).

MMN mean amplitudes were averaged from 60 ms time windows centered on the central (Cz) negative MMN peaks observed from the group-averaged deviant-minus-control difference waveforms, separately for the two deviants and two attention conditions. Thus MMN was derived by subtracting between responses elicited by identical sound patterns presented in different sequences (i.e., when D1 and D2 are deviants among standards and when D1 and D2 form homogenous control sequences). Responses elicited by the standard patterns were not used in the MMN measurements. This derivation of MMN prevents the emergence of confounding differences stemming from pattern-specific ERP responses. The group-averaged central MMN peak latencies were: 160, 140, 196, and 176 ms *from deviation (omission) onset* for the Unattend-Strong, Passive-Strong, Unattend-Weak, and Passive-Weak deviant responses, respectively. The effects of attention, deviance position, and the scalp distribution of the MMN amplitudes were analyzed with a repeated-measure ANOVA of the structure Attention (Unattend vs. Passive) × Position (Strong vs. Weak) × Frontality (Frontal vs. Central electrode line) × Laterality (Left vs. Middle vs. Right). All significant effects and interactions are reported below. Greenhouse-Geisser correction of the degrees of freedom was applied where appropriate and the ε correction factor as well as η^2 effect size are reported.

Behavioral Data

Discrimination sensitivity was significantly higher for Strong than for Weak deviants, $t(10) = 2.80$, $p < .05$; $d'(\text{Strong}) = 2.77$, $d'(\text{Weak}) = 2.13$. There was also a tendency for faster RT's for Strong than for Weak deviants, $t(10) = 1.85$, $p < .10$; RT(Strong) = 536.69 ms, RT(Weak) = 585.68 ms.

Discussion of the Behavioral Data

Higher sensitivity and shorter RT's for Strong as compared to Weak deviants suggest that theoretical metrical salience affected the processing of rhythmic patterns in our participants when they attended the stimulus sequence.

Electrophysiological Data

The D1 and D2 patterns elicited a fronto-centrally more negative response between 100 and 250 ms from the onset of the omissions when the patterns violated the rhythmic context set up by the frequent standard patterns (S1-S4) than when the same patterns were presented alone in the homogeneous control stimulus blocks (Figure 2). The difference between the ERP responses elicited by the deviant and the identical control stimuli can be identified as an MMN response (cf. below). Significantly shorter MMN peak latencies (measured from the onset of deviation; see Figures 2 and 3) were obtained for Strong as compared to Weak deviants, $F(1, 10) = 20.69$, $p < .01$, $\eta^2 = 0.67$ (average peak latencies: Passive[Strong] = 145.45 ms, Passive[Weak] = 165.45 ms, Unattend[Strong] = 149.09 ms, and Unattend[Weak] = 190.18 ms). The ANOVA of MMN amplitudes (see Figures 2 and 3, and Table 1 for mean MMN amplitudes) yielded main effects of Position, $F(1, 10) = 5.62$, $p < .05$, $\eta^2 = 0.36$, Frontality, $F(1, 10) = 10.56$, $p < .01$, $\eta^2 = 0.51$, and Laterality, $F(2, 20) = 13.86$, $p < .001$, $\varepsilon = 0.83$, $\eta^2 = 0.58$. Strong deviants elicited higher-amplitude MMN responses as compared to Weak deviants. MMN was larger over central than frontal electrodes and over midline than lateral electrodes. There was also a significant interaction between Attention and Frontality, $F(1, 10) = 35.24$, $p < .001$, $\eta^2 = 0.78$, stemming from lower frontal MMN amplitudes in the Passive condition than in any other combination of these two factors (Tukey HSD posthoc test with $df = 10$, $p < .001$ for all of the referred comparisons). This result rules out the possibility that the deviant-minus-control difference waveform would contain significant contribution from the N2b ERP component. This is because N2b is elicited only when participants actively detect a stimulus (Novak, Ritter, Vaughan, & Wiznitzer, 1990). Furthermore, the ERP difference cannot reflect difference between two

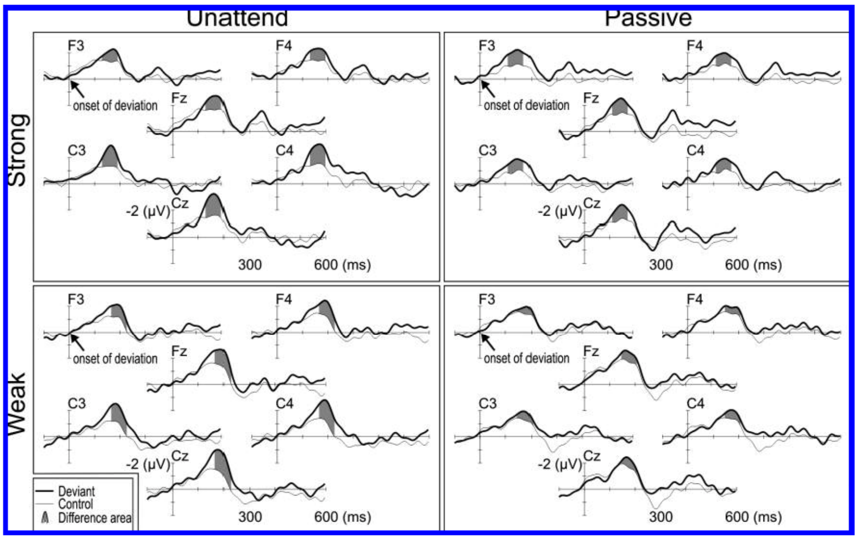

FIGURE 2. Group-averaged ($n = 11$) ERP responses elicited by deviant patterns (experimental stimulus blocks; thick lines) and identical control patterns (control stimulus blocks; thin lines). Left: Unattend condition; right: Passive condition. Upper panels show the responses to Strong, lower panels to Weak metrical position deviants. The area between deviant and control responses within the measurement window is marked by grey shading. Responses are aligned at the onset of deviation (the time point at which the omitted sound appears in the S1 pattern).

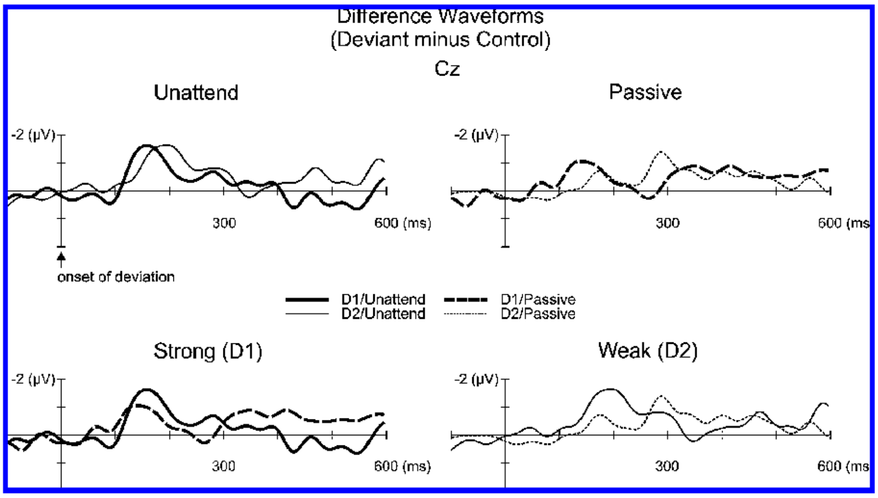

FIGURE 3. Group-averaged ($n = 11$) deviant-minus-control difference waveforms (thick lines for Strong, thin lines for Weak deviants; continuous lines for the Unattend, dashed lines for the Passive condition). Top panels: Comparison between responses elicited by Strong and Weak deviants, separately for the Unattend (left) and Passive (right) conditions. Bottom panels: Comparison between the two attention conditions, separately for Strong (left) and Weak (right) deviants. Responses are aligned at the onset of the deviation.

TABLE 1. Group-Averaged MMN Amplitudes in μV with Standard Errors of the Mean (SEM) in Parentheses.

Attention	Passive		Unattend	
Electrode/Position	Strong	Weak	Strong	Weak
F3	−2.23 (0.40)	−1.20 (0.29)	−2.00 (0.19)	−1.53 (0.40)
Fz	−2.62 (0.47)	−1.70 (0.38)	−2.58 (0.28)	−1.99 (0.47)
F4	−1.93 (0.41)	−1.27 (0.45)	−2.10 (0.31)	−1.68 (0.41)
C3	−2.03 (0.37)	−1.42 (0.35)	−2.72 (0.34)	−2.15 (0.37)
Cz	−2.57 (0.47)	−1.71 (0.41)	−3.29 (0.30)	−2.49 (0.47)
C4	−2.08 (0.41)	−1.48 (0.40)	−2.99 (0.35)	−2.38 (0.41)

N1 components, because it is elicited by sound omissions, which do not elicit the N1 component. Very importantly, the Attention factor did not significantly interact with the Position factor for either peak latencies or MMN amplitudes. This means that Strong deviants elicited earlier and higher-amplitude MMN responses than Weak deviants irrespective of the attention conditions.

Discussion of the Electrophysiological Data

MMN responses were elicited by deviations in both metrical positions and in both attention conditions. This suggests that rhythmic violations are detected even when attention is not focused on the sound sequence. Furthermore, Strong deviants elicited a stronger (earlier and higher-amplitude) response than Weak ones. This result corroborates the behavioral data in suggesting that metric salience affected the detection of rhythm violations. Stronger MMN responses are usually recorded to perceptually more salient deviations (Näätänen & Alho, 1997). Since the amount of raw acoustic deviation did not differ between the two deviant positions, larger perceived deviations suggest sharper (more precise) memory representations for metrically salient elements of rhythmic patterns (a similar effect on the sharpness of the memory representations underlying MMN has been demonstrated by masking studies; see Winkler, Reinikainen, & Näätänen, 1993). Modulation of the memory representations by metric salience strongly argues for the conclusion that the brain formed hierarchical representations for the rhythmic stimulus sequences.

The only effect of attention was lower frontal MMN amplitudes in the Passive compared with the Unattend condition. This effect was not significantly different between MMNs elicited by Strong and Weak deviants. Rather, it probably reflects differences in the general activity of the frontal cortex in the two attention conditions (e.g., difference in the arousal level or between processing simple sound change as opposed to following a movie). Thus it appears that the processing of meter (forming hierarchical representations for rhythmical sound sequences) does not require significant amounts of limited higher-level capacities, a sign that meter may be processed at lower levels of auditory perception. The picture emerging from the electrophysiological data is that meter is extracted more or less automatically from rhythmic sequences, suggesting that it is an "intelligent" low level auditory processing capability, of which more and more are discovered by recent research (Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001).

General Discussion and Conclusion

The behavioral detection of syncopated rhythms data as well as the ERPs recorded under two different attention conditions concerning meter induction were consistent in our group of participants. They were able to detect syncopated rhythms in an active behavioral task (indicated by the accuracy and speed of detection), as well as passively in the ERP experiment when they focused their attention on a task unrelated to the rhythmic sound sequences. Not only did participants distinguish syncopated patterns from strictly metrical ones, but they also showed sensitivity to the position (metric salience) or in other words, to the strength of the syncopation. This result is in full accordance with the Longuet-Higgins and Lee (1984) model, which predicts that the most salient position elicits a significantly stronger response than syncopation on any lower salient position of the rhythm. Furthermore, this result suggests that meter is not only a concept facilitating the structuring of written musical scores, but it corresponds to the structure of memory representations in the human brain.

These results suggest that beat induction, which according to Povel (1981) is an essential first step in the perception of temporal sequences, is functional both in active and passive listening situations. Furthermore, our participants clearly were sensitive to the hierarchical ordering in beat perception (as revealed by the difference in responses between D1 and D2; cf. Figure 3). This provides further evidence for the general perceptual/cognitive capability based interpretation of meter. While earlier research showed only a marginal sensitivity to meter in listeners with little or no formal music training (e.g., Jongsma et al., 2004; Palmer & Krumhansl, 1990), the current study demonstrated that meter is a mental representation that does not require advanced formal music training. This conclusion does not rule out the possibility that, similarly to other music-related processing capabilities, the representation of rhythmic structures can be improved by music training (see for example van Zuijen et al., 2005). It remains a question for future research whether basic sensitivity for meter is a result of learning by exposure to one's musical environment (Huron, 2006), or, as hinted by the current results as well as studies showing sensitivity to meter at a very early age (Hannon & Johnson, 2005), whether it stems from a general cognitive predisposition of the human brain for breaking down complex patterns recursively into equal sized subpatterns (Martin, 1972).

Author Note

Correspondence concerning this article should be addressed to Olivia Ladinig, Music Cognition Group, Institute for Logic, Language and Computation (ILLC), Universiteit van Amsterdam (UvA), The Netherlands. E-MAIL: olivia.ladinig@gmail.com

References

Baldeweg, T. (2007). ERP repetition effects and mismatch negativity generation: A predictive coding perspective. *Journal of Psychophysiology, 211*, 204-213.

Bolton, T. (1894). Rhythm. *The American Journal of Psychology, 6*, 145-238.

Brattico, E., & Näätänen, R. (2002). Context effects on pitch perception in musicians and nonmusicians: Evidence from event-related-potential recordings. *Music Perception, 19*, 199-222.

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.

Brochard, R., Abecasis, D., Potter, D., Ragot, R., & Drake, C. (2003). The 'ticktock' of our internal clock: Direct brain evidence of subjective accents in isochronous sequences. *Psychological Science, 14*, 362-366.

Clarke, E. F. (1999). Rhythm and timing in music. In D. Deutsch (Ed.), *Psychology of music* (2nd ed., pp. 473-500). New York, NY: Academic Press.

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

Fitch, W. T., & Rosenfeld, A. J. (2007). Perception and production of syncopated rhythms. *Music Perception, 25*, 43-58.

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.

Hannon, E. E., & Johnson, S. P. (2005). Infants use meter to categorize rhythms and melodies: Implications for musical structure learning. *Cognitive Psychology, 50*, 354-377.

Honing, H. (2002). Structure and interpretation of rhythm and timing. *Dutch Journal of Music Theory, 7*, 227-232.

Huron, D. (2006) *Sweet anticipation. Music and the psychology of expectation*. Cambridge, MA: The MIT Press.

Jongsma, M. L. A., Desain, P., & Honing, H. (2004). Rhythmic context influences the auditory evoked potentials of musicians and non-musicians. *Biological Psychology, 66*, 129-152.

Ladinig, O., & Honing, H. (2009). *Complexity judgments as a measure for event salience.* Manuscript submitted for publication.

Leino, S., Brattico, E., Tervaniemi, M., & Vuust, P. (2007). Representation of harmony rules in the human brain: Further evidence from event-related potentials. *Brain Research, 1142*, 169-177.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music.* Cambridge, MA: MIT Press.

London, J. (2004). *Hearing in time: Psychological aspects of musical meter.* New York, NY: Oxford University Press.

Longuet-Higgins, H. C., & Lee, C. S. (1984). The rhythmic interpretation of monophonic music. *Music Perception, 1*, 424-441.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.

Martin, J. G. (1972). Rhythmic (hierarchic) versus serial structure in speech and other behaviour. *Psychological Review, 79*, 487-509.

Näätänen, R., & Alho, K. (1997). Mismatch negativity (MMN)—the measure for central sound representation accuracy. *Audiology and Neuro-otology, 2*, 341-353.

Näätänen, R., Gaillard, A. W. K., & Mantysalo, S. (1978). Early selective attention effect on evoked potential reinterpreted. *Acta Psychologica, 42,* 313-329.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology, 118,* 2544-2590.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). 'Primitive intelligence' in the auditory cortex. *Trends in Neurosciences, 24,* 283-288.

Nordby, H., Roth, W. T., & Pfefferbaum, A. (1988). Event-related potentials to time-deviant and pitch-deviant tones. *Psychophysiology, 25,* 249-261.

Novak, G. P., Ritter, W., Vaughan, Jr., H. G., & Wiznitzer, M. L. (1990). Differentiation of negative event related potentials in an auditory discrimination task. *Electroencephalography and Clinical Neurophysiology, 75,* 255-275.

Pablos Martin, X., Deltenre, P., Hoonhorst, I., Markessis, E., Rossion, B., & Colin, C. (2007). Perceptual biases for rhythm: The mismatch negativity latency indexes the privileged status of binary vs non-binary interval ratios. *Clinical Neurophysiology, 118,* 2709-2715.

Palmer, C., & Krumhansl, C. L. (1990). Mental representations for musical meter. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 728-741.

Povel, D. (1981). Internal representation of simple temporal patterns. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 3-18.

Pressing, J. (2002). *Cognitive complexity and the structure of musical patterns.* Retrieved June 2, 2006, from University of Melbourne web site: http://psy.uq.edu.au/CogPsych/Noetica/OpenForumIssue8/Pressing.html

Sussman, E. S. (2007). A new view on the MMN and attention debate: The role of context in processing auditory events. *Journal of Psychophysiology, 21,* 164-175.

Tervaniemi, M., & Huotilainen, M. (2003). The promises of change-related brain potentials in cognitive neuroscience of music. *Neurosciences and Music, 999,* 29-39.

Trainor, L. J., McDonald, K. L., & Alain, C. (2002). Automatic and controlled processing of melodic contour and interval information measured by electrical brain activity. *Journal of Cognitive Neuroscience, 14,* 1-13.

van Zuijen, T. L., Sussman, E., Winkler, I., Näätänen, R., & Tervaniemi, M. (2005). Auditory organization of sound sequences by a temporal or numerical regularity: A mismatch negativity study comparing musicians and non-musicians. *Cognitive Brain Research, 23,* 270-276.

Vuust, P., Pallesen, K. J., Bailey, C., van Zuijen, T. L., Gjedde, A., Roepstoff, A., & Østergaard, L. (2005). To musicians, the message is in the meter. Pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians. *NeuroImage, 24,* 560-564.

Winkler, I. (2007). Interpreting the mismatch negativity (MMN). *Journal of Psychophysiology, 21,* 147-163.

Winkler, I., Reinikainen, K., & Näätänen, R. (1993). Event related brain potentials reflect traces of the echoic memory in humans. *Perception and Psychophysics, 53,* 443-449.

Winkler, I., & Schröger, E. (1995). Neural representation for the temporal structure of sound patterns. *Neuroreport, 6,* 690-694.

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

Yeston, M. (1976). *The stratification of musical rhythm.* New Haven, CT: Yale University Press.

3.5. Study V: Beat detection in newborns

Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., Honing, H. (2009) Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences* 106. 2468-2471. DOI: 10.1073/pnas.0809035106

Newborn infants detect the beat in music

István Winkler[a,b,1], Gábor P. Háden[a,c], Olivia Ladinig[d], István Sziller[e], and Henkjan Honing[d]

[a]Institute for Psychology, Hungarian Academy of Sciences, H-1394 Budapest, P.O. Box 398, Hungary; [b]Institute of Psychology, University of Szeged, H-6722 Szeged, Petőfi S. sgt. 30-34, Hungary; [c]Department of Cognitive Science, Budapest University of Technology and Economics, H-1111 Budapest, Stoczek u. 2, Hungary; [d]Music Cognition Group, Institute for Logic, Language and Computation, Universiteit van Amsterdam, Postbus 19268, 1000 GG Amsterdam; and [e]First Department of Obstetrics and Gynaecology, Semmelweis University, H-1088 Budapest, Baross u. 27, Hungary

Edited by Dale Purves, Duke University Medical Center, Durham, NC, and approved December 23, 2008 (received for review September 10, 2008)

To shed light on how humans can learn to understand music, we need to discover what the perceptual capabilities with which infants are born. Beat induction, the detection of a regular pulse in an auditory signal, is considered a fundamental human trait that, arguably, played a decisive role in the origin of music. Theorists are divided on the issue whether this ability is innate or learned. We show that newborn infants develop expectation for the onset of rhythmic cycles (the downbeat), even when it is not marked by stress or other distinguishing spectral features. Omitting the downbeat elicits brain activity associated with violating sensory expectations. Thus, our results strongly support the view that beat perception is innate.

event-related brain potentials (ERP) | neonates | rhythm

Music is present in some form in all human cultures. Sensitivity to various elements of music appears quite early on in infancy (1, 2–4), with understanding and appreciation of music emerging later through interaction between developing perceptual capabilities and cultural influence. Whereas there is already some information regarding spectral processing abilities of newborn infants (5, 6), little is known about how they process rhythm. The ability to sense beat (a regular pulse in an auditory signal; termed "tactus" in music theory; 7, 8) helps individuals to synchronize their movements with each other, such as necessary for dancing or producing music together. Although beat induction would be very difficult to assess in newborns using behavioral techniques, it is possible to measure electrical brain responses to sounds (auditory event related brain potentials, ERP), even in sleeping babies. In adults, infrequently violating some regular feature of a sound sequence evokes a discriminative brain response termed the mismatch negativity (MMN) (9, 10). Similar responses are elicited in newborns (11) by changes in primary sound features (e.g., the pitch of a repeating tone) and by violations of higher-order properties of the sequence, such as the direction of pitch change within tone pairs (ascending or descending) that are varying in the starting pitch (12). Newborns may even form crude sound categories while listening to a sound sequence (13): an additional discriminative ERP response is elicited when a harmonic tone is occasionally presented among noise segments or vice versa, suggesting a distinction between harmonic and complex sounds.

Neonates are also sensitive to temporal stimulus parameters [e.g., sound duration (14)] and to the higher-order temporal structure of a sound sequence [such as detecting periodical repetition of a sound pattern (15)]. Because the MMN is elicited by deviations from expectations (16), it is especially appropriate for testing beat induction. One of the most salient perceptual effects of beat induction is a strong expectation of an event at the first position of a musical unit, i.e., the "downbeat" (17). Therefore, occasionally omitting the downbeat in a sound sequence composed predominantly of strictly metrical (regular or "nonsyncopated") variants of the same rhythm should elicit discriminative ERP responses if the infants extracted the beat of the sequence.

Results and Discussion of the Neonate Experiment

We presented 14 healthy sleeping neonates with sound sequences based on a typical 2-measure rock drum accompaniment pattern (S1) composed of snare, bass and hi-hat spanning 8 equally spaced (isochronous) positions (Fig. 1 *A* and *B*). Four further variants of the S1 pattern (S2–S4 and D) (Fig. 1 *C–F*) were created by omitting sounds in different positions. The omissions in S2, S3, and S4 do not break the rhythm when presented in random sequences of S1–S4 linked together, because the omitted sounds are at the lowest level of the metrical hierarchy of this rhythm (Fig. 1*A*) and, therefore, perceptually less salient (7). The 4 strictly metrical sound patterns (S1–S4; standard) made up the majority of the patterns in the sequences. Occasionally, the D pattern (Fig. 1*F*, deviant) was delivered in which the downbeat was omitted. Adults perceive the D pattern within the context of a sequence composed of S1–S4 as if the rhythm was broken, stumbled, or became strongly syncopated for a moment (18) (Sound File S1). A control sequence repeating the D pattern 100% of the time was also delivered ("deviant-control").

Fig. 2 shows that the electrical brain responses elicited by the standard (only S2–S4; see *Methods*) and deviant-control patterns are very similar to each other, whereas the deviant stimulus response obtained in the main test sequence differs from them. The deviant minus deviant-control difference waveform has 2 negative waves peaking at 200 and 316 ms followed by a positive wave peaking at 428 ms. The difference between the deviant and the other 2 responses was significant in 40-ms-long latency ranges centered on the early negative and the late positive difference peaks (see Table 1 for the mean amplitudes) as shown by dependent measures ANOVAs with the factors of Stimulus (Standard vs. Deviant control vs. Deviant) × Electrode (C3 vs. Cz vs. C4). The Stimulus factor had a significant effect on both peaks (for the early negative waveform: $F[2,26] = 3.77, P < 0.05$, with the Greenhouse-Geisser correction factor $\varepsilon = 0.85$ and the effect size $\eta^2 = 0.22$; for the positive waveform: $F[2,26] = 8.26, P < 0.01, \varepsilon = 0.97, \eta^2 = 0.39$). No other main effects or interactions reached significance. Posthoc Tukey HSD pairwise comparisons showed significant differences between the deviant and the deviant-control responses in both latency ranges (with $df = 26, P < 0.05$ and 0.01 for the early negative and the late positive waveforms, respectively) and for the positive waveform, between the deviant and the standard response ($df = 26, P < 0.01$). No significant differences were found between the standard and the deviant control responses.

Author contributions: I.W., G.P.H., O.L., and H.H. designed research; G.P.H., O.L., and I.S. performed research; G.P.H. analyzed data; and I.W., I.S., and H.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1] To whom correspondence should be addressed. E-mail: iwinkler@cogpsyphy.hu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0809035106/DCSupplemental.

2468–2471 | PNAS | February 17, 2009 | vol. 106 | no. 7

www.pnas.org/cgi/doi/10.1073/pnas.0809035106

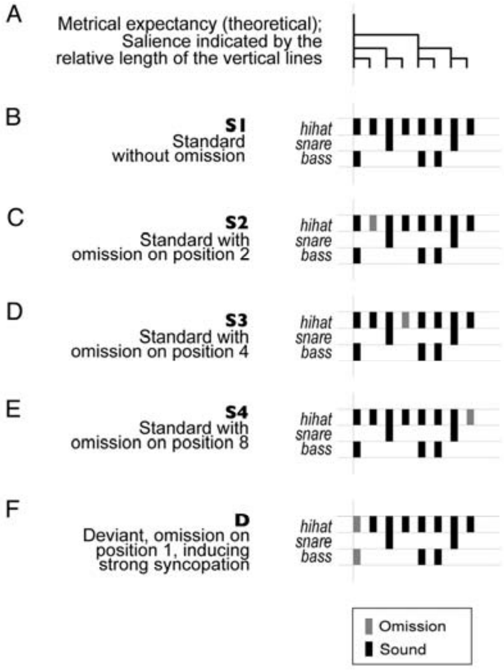

Fig. 1. Schematic diagram of the rhythmic stimulus patterns.

Results showed that newborn infants detected occasional omissions at the 1st (downbeat) position of the rhythmic pattern, but, whereas the S2–S4 patterns omitted only a single sound (the hi-hat), the D pattern omitted 2 sounds (hi-hat and bass). The double omission could have been more salient than the single sound omissions, thus eliciting a response irrespective of beat induction. However, the omission in D could only be identified as a double omission if the neonate auditory system expected both the bass and the hi-hat sound at the given moment of time. Because bass and hi-hat cooccurs at 3 points in the base pattern (see Fig. 1*B*), knowing when they should be encountered together requires the formation of a sufficiently detailed representation of the whole base pattern in the neonate brain. In contrast, beat detection requires only that the length of the full cycle and its onset are represented in the brain. It is possible that neonates form a detailed representation of the base pattern. This would allow them not only to sense the beat, but also to build a hierarchically ordered representation of the rhythm (meter induction), as was found for adults (18). This exciting possibility is an issue for further research.

Another alternative interpretation of the results suggests that newborn infants track the probabilities of the succession of sound events (e.g., the probability that the hi-hat and bass sound event is followed by a hi-hat sound alone). However, in this case, some of the standard patterns (e.g., S2) should also elicit a discriminative response, because the omission has a low conditional probability (e.g., the probability that the hi-hat and bass sound event is followed by an omission, as it occurs in S2, is 0.078 within the whole sequence).

Finally, it is also possible that newborn infants segregated the sounds delivered by the 3 instruments, creating separate expectations for each of them. This explanation receives support from our previous results showing that newborn infants segregate tones of widely differing pitches into separate sound streams (6). If this was the case, omission of the bass sound could have resulted in the observed ERP differences without beat being induced. To test this alternative, we presented the test and the control sequences of the neonate experiment to adults, silencing the hi-hat and snare sounds. All stimulation parameters, including the timing of the bass sounds and the probability of omissions (separately for the test and the control sequences) were identical to the neonate experiment.

Results and Discussion of the Adult Control Experiment

Fig. 3 shows that the ERP responses elicited by the deviant and the control patterns are highly similar to each other. Taking the peaks where the central (C3, Cz, and C4 electrodes) deviant-minus-control difference was largest (132 and 296 ms) within the latency range in which discriminative ERP responses are found in adults, we conducted ANOVAs of similar structure as was done for the neonate measurements [dependent measures factors of Stimulus (Deviant control vs. Deviant) × Electrode (C3 vs. Cz vs. C4); standard patterns could not be used, because they contained no omission in the bass sequence]. We found no significant main effect of Stimulus or interaction between the Stimulus and the Electrode factor ($P > 0.2$ for all tests). The only significant effect found was that of Electrode for the later latency range ($F[2,24] = 7.30$, $P < 0.01$, $\varepsilon = 0.75$, $\eta^2 = 0.38$) However,

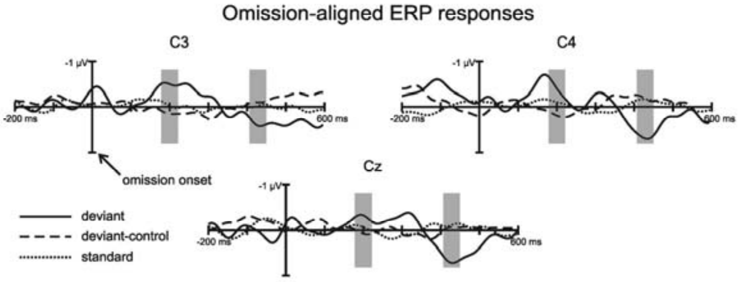

Fig. 2. Group averaged ($n = 14$) electrical brain responses elicited by rhythmic sound patterns in neonates. Responses to standard (average of S2, S3, and S4; dotted line), deviant (D; solid line), and deviant-control patterns (D patterns appearing in the repetitive control stimulus block; dashed line) are aligned at the onset of the omitted sound (compared with the full pattern: S1) and shown from 200 ms before to 600 ms after the omission. Gray-shaded areas mark the time ranges with significant differences between the deviant and the other ERP responses.

Table 1. Group averaged (n = 14) mean ERP amplitudes

	Amplitude, μV					
	180–220 ms interval			408–448 ms interval		
Group	C3	Cz	C4	C3	Cz	C4
Deviant	−0.50 (0.19)	−0.30 (0.22)	−0.41 (0.27)	0.38 (0.17)	0.67 (0.13)	0.67 (0.27)
Deviant-control	0.14 (0.13)	0.06 (0.16)	0.18 (0.25)	−0.10 (0.13)	−0.06 (0.16)	−0.18 (0.15)
Standard	−0.03 (0.09)	−0.06 (0.12)	−0.11 (0.09)	−0.02 (0.09)	−0.12 (0.12)	−0.16 (0.13)

SEM values are shown in parentheses.

because this effect does not include the Stimulus factor, it is not the sign of a response distinguishing the deviant from the control response.

Thus, in adults, omission of the position-1 bass sound does not result in the elicitation of discriminative ERP responses in the absence of the rhythmic context. This result is compatible with those of previous studies showing that stimulus omissions (without a rhythmic structure) only elicit deviance-related responses at very fast presentation rates (<170-ms onset-to-onset intervals; see ref. 19). In our stimulus sequences, the omitted bass sound was separated by longer intervals from its neighbors. It should be noted that adult participants elicited the MMN discriminative ERP response, when they received the full stimulus sequence (all 3 instruments) as presented to newborn babies in the neonate experiment (18).

Discussion

These results demonstrate that violating the beat of a rhythmic sound sequence is detected by the brain of newborn infants. In support of this conclusion we showed that the sound pattern with omission at the downbeat position elicited discriminative electrical brain responses when it was delivered infrequently within the context of a strictly metrical rhythmic sequence. These responses were not elicited by the D pattern per se: When the D pattern was delivered in a repetitive sequence of its own, the brain response to it did not differ from that elicited by the standards. Neither were discriminative responses simply the result of detecting omissions in the rhythmic pattern. Omissions occurring in non-salient positions elicited no discriminative responses (see the response to the standards in Fig. 2). Furthermore, the discriminative ERP response elicited by the D pattern was not caused by separate representations formed for the 3 instruments: only omissions of the downbeat within the rhythmic context elicit this response.

So it appears that the capability of detecting beat in rhythmic sound sequences is already functional at birth. Several authors consider beat perception to be acquired during the first year of life (2–4), suggesting that being rocked to music by their parents is the most important factor. At the age of 7 months, infants have been shown to discriminate different rhythms (2, 3). These results were attributed to sensitivity to rhythmic variability, rather than to perceptual judgments making use of induced beat. Our results show that although learning by movement is probably important, the newborn auditory system is apparently sensitive to periodicities and develops expectations about when a new cycle should start (i.e., when the downbeat should occur). Therefore, although auditory perceptual learning starts already in the womb (20, 21), our results are fully compatible with the notion that the perception of beat is innate. In the current experiment, the beat was extracted from a sequence comprised of 4 different variants of the same rhythmic structure. This shows that newborns detect regular features in the acoustic environment despite variance (12) and they possess both spectral and temporal processing prerequisites of music perception.

Many questions arise as a result of this work. Does neonate sensitivity to important musical features mean that music carries some evolutionary advantage? If so, are the processing algorithms necessary for music perception part of our genetic heritage? One should note that the auditory processing capabilities found in newborn babies are also useful in auditory communication. The ability to extract melodic contours at different levels of absolute pitch is necessary to process prosody. Sensing higher-order periodicities of sound sequences is similarly needed for adapting to different speech rhythms e.g., finding the right time to reply or interject in a conversation (22). Temporal coordination is essential for effective communication. When it breaks down, understanding and cooperation between partners is seriously hampered. Therefore, even if beat induction

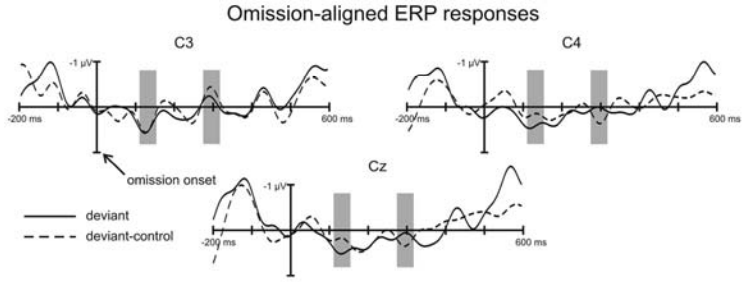

Fig. 3. Group averaged (n = 13) electrical brain responses elicited by the bass sound patterns in adults. Responses to deviant (D; solid line), and deviant-control patterns (D patterns appearing in the repetitive control stimulus block; dashed line) are aligned at the onset of the omitted bass sound (compared with the standard patterns: S1–S4) and shown from 200 ms before to 600 ms after the omission. Gray-shaded areas mark the time ranges in which amplitudes were measured.

is an innate capability, the origin and evolutionary role of music remains an issue for further research.

Methods

Neonate Experiment. Sound sequences were delivered to 14 healthy full term newborn infants of 37–40 weeks gestational age (3 female, birth weight 2650–3960 g, APGAR score 9/10) on day 2 or 3 post partum while the electroencephalogram was recorded from scalp electrodes. The study was approved by the Ethics Committee of the Semmelweis University, Budapest, Hungary. Informed consent was obtained from one or both parents. The mother of the infant was present during the recording.

The experimental session included 5 test sequences, each comprising 276 standard (S1–S4; Fig. 1) and 30 deviant (D) patterns and a control sequence in which the D pattern (termed deviant-control) was repeated 306 times. The control stimulus block was presented at a randomly chosen position among the test stimulus blocks. In the test sequences, S1–S4 appeared with equal probability, 22.5%, each, with the D pattern making up the remaining 10%. This is a prerequisite of the deviance detection method, which requires deviations to be infrequent within the sequence (9). The order of the 5 patterns was pseudorandomized, enforcing at least 3 standard patterns between successive D patterns. The onset-to-onset interval between successive sounds was 150 ms with 75-ms onset-to-offset interval (75-ms sound duration). Patterns in the sequence were delivered without breaks. Loudness of the sounds was normalized so that all stimuli (including the downbeat) had the same loudness.

EEG was recorded with Ag-AgCl electrodes at locations C3, Cz, and C4 of the international 10–20 system with the common reference electrode attached to the tip of the nose. Signals were off-line filtered between 1 and 16 Hz. Epochs starting 600 ms before and ending 600 ms after the time of the omission in the sound patterns (compared with the S1 pattern) were extracted from the continuous EEG record. Epochs with the highest voltage change outside the 0.1–100 μV range on any EEG channel or on the electrooculogram (measured between electrodes placed below the left and above right eye) were discarded from the analysis. Epochs were baseline corrected by the average voltage during the entire epoch and averaged across different sleep stages, whose distribution did not differ between the test and the control stimulus blocks. Responses to the S2–S4 patterns were averaged together, aligned at the point of omission (termed ''Standard''). Responses were averaged for the D pattern separately for the ones recorded in the main test and those in the control sequences (Deviant and Deviant-control responses, respectively). For assessing the elicitation of differential ERP responses, peaks observed on the group-average difference waveforms between the deviant and deviant-control responses were selected. For robust measurements, 40-ms-long windows were centered on the selected peaks. Amplitude measurements were submitted to ANOVA tests (see the structure in *Results and Discussion of the Neonate Experiment*). Greenhouse-Geisser correction was applied. The correction factor and the effect size (partial eta-square) is reported. Tukey HSD pairwise posthoc comparisons were used. All significant results are discussed.

Adult Experiment. Fourteen healthy young adults (7 female, 18–26 years of age, mean: 21.07) participated in the experiment for modest financial compensation. Participants gave informed consent after the procedures and aims of the experiments were explained to them. The study was approved by the Ethical Committee of the Institute for Psychology, Hungarian Academy of Sciences. All participants had frequency thresholds not >20 dB SPL in the 250-4000 Hz range and no threshold difference exceeding 10 dB between the 2 ears (assessed with a Mediroll, SA-5 audiometer). Participants watched a silenced subtitled movie during the EEG recordings. One participant's data were rejected from the analyses due to excessive electrical artifacts.

All parameters of the stimulation, EEG recording, and data analysis were identical to the neonate study except that hi-hat and snare sounds were removed from the stimulus patterns without changing the timing of the remaining sounds.

ACKNOWLEDGMENTS. We thank János Rigó for supporting the study of neonates at the First Department of Obstetrics and Gynecology, Semmelweis University, Budapest and Erika Józsa Váradiné and Zsuzsanna D'Albini for collecting the neonate and adult ERP data, respectively and to Martin Coath and Glenn Schellenberg for their comments on earlier versions of the manuscript. This work was supported by European Commission 6th Framework Program for ''Information Society Technologies'' (project title: EmCAP Emergent Cognition through Active Perception) Contract 013123.

1. Trehub S (2003) The developmental origins of musicality. *Nat Neurosci* 6:669–673.
2. Hannon EE, Trehub S (2005) Metrical categories in infancy and adulthood. *Psychol Sci* 16:48–55.
3. Phillips-Silver J, Trainor LJ (2005) Feeling the beat: Movement influences infant rhythm perception. *Science* 308:1430.
4. Patel A (2008) *Music, Language and the Brain* (Oxford Univ. Press, Oxford).
5. Novitski N, Huotilainen M, Tervaniemi M, Näätänen R, Fellman V (2007) Neonatal frequency discrimination in 250–4000 Hz frequency range: Electrophysiological evidence. *Clin Neurophysiol* 118:412–419.
6. Winkler I, *et al.* (2003) Newborn infants can organize the auditory world. *Proc Natl Acad Sci USA* 100:1182–1185.
7. Lerdahl F, Jackendoff R (1983) *A Generative Theory of Tonal Music* (MIT press, Cambridge, MA).
8. Honing H (2002) Structure and interpretation of rhythm and timing. *Dutch J Music Theory* 7:227–232.
9. Kujala T, Tervaniemi M, Schröger E (2007) The mismatch negativity in cognitive and clinical neuroscience: Theoretical and methodological considerations. *Biol Psychol* 74:1–19.
10. Näätänen R, Tervaniemi M, Sussman E, Paavilainen P, Winkler I (2001) ''Primitive intelligence'' in the auditory cortex. *Trends Neurosci* 24:283–288.
11. Alho K, Saino K, Sajaniemi N, Reinikainen K, Näätänen R (1990) Event-related brain potential of human newborns to pitch change of an acoustic stimulus. *Electroencephal Clin Neurophysiol* 77:151–155.
12. Carral V, *et al.* (2005) A kind of auditory ''primitive intelligence'' already present at birth. *Eur J Neurosci* 21:3201–3204.
13. Kushnerenko E, Winkler I, Horváth J, Näätänen R, Pavlov I, Fellman V, Huotilainen M (2007) Processing acoustic change and novelty in newborn infants. *Eur J Neurosci* 26:265–274.
14. Kushnerenko E, Čeponienė R, Fellman V, Huotilainen M, Winkler I (2001) Event-related potential correlates of sound duration: Similar pattern from birth to adulthood. *NeuroReport* 12:3777–3781.
15. Stefanics G, *et al.* (2007) Auditory temporal grouping in newborn infants. *Psychophysiol* 44:697–702.
16. Winkler I Interpreting the mismatch negativity (MMN) (2007) *J Psychophysiol* 21:147–163.
17. Fitch WT, Rosenfeld AJ (2007) Perception and production of syncopated rhythms. *Music Percept* 25:43–58.
18. Ladinig O, Honing H, Háden G, Winkler I (2009). Probing attentive and pre-attentive emergent meter in adult listeners without extensive music training. *Music Percept*, in press.
19. Yabe H, *et al.* (1998) The temporal window of integration of auditory information in the human brain. *Psychophysiol* 35:615–619.
20. Visser GHA, Mulder EJH, Prechtl HFR (1992) Studies on developmental neurology in the human fetus. *Dev Pharmacol Ther* 18:175–183.
21. Huotilainen M, *et al.* (2005) Short-term memory functions of the human fetus recorded with magnetoencephalography. *NeuroReport* 16:81–84.
22. Jaffe J, Beebe B (2001) *Rhythms of Dialogue in Infancy*, Monographs for the Society for Research and Child Development (Blackwell, Boston), Vol 66, No 2.

4. General discussion

The main goal of the studies presented in this thesis was to identify perceptual processes that provide information for music processing and, with one exception, establish the presence of the functionality of these processes at birth. To this end, the studies described here utilized carefully crafted variations of the auditory oddball paradigm while ERPs were recorded to assess what regularities were represented in the auditory system by testing the discrimination of infrequent stimuli deviating from the target regularities. Discriminative responses to these deviants were obtained in all of the studies. These responses were identified as MMN or MMN-like components. The results, on one hand, offer evidence for an adult-like processing of pitch and sound grouping in newborns. These abilities are important for music perception as well as for verbal communication. On the other hand, the results further strengthen the feasibility of using MMN paradigms as tools for studying the early stages of music processing in both newborns and adults. Additionally, the study examining the processing of musical meter (*Thesis IV*) provides direct evidence in support of the theoretic assumption of metrical hierarchy (Lerdahl & Jackendoff, 1983) and for the extraction of beat in musically untrained subjects.

4.1. Well equipped for music processing already at birth

Human babies are born relatively helpless compared to most other mammalian species, but have an unmatched potential to learn the complex rules and relations governing the physical and social environment. For this learning to occur, however, tools are needed that can efficiently extract relevant information from the environment. From this perspective, it is not surprising that evolution provided infants with abilities necessary for undertaking the enormous task they have at their hand. The best tools adults have for gathering information is communication, which allows the collecting information gathered by individuals from their immediate spatial environment and from a much larger cultural environment. With communication being the most useful source of information, the tools necessary to understand and part-take in it are of the highest importance. Music, similarly to language, is a form of communication ubiquitous to all human cultures (Merriam, 1964) and, as both music and language are based on acoustic signals, the basic tools for tapping these information sources must be found within auditory perception.

Stepping over the unproductive contrast of nature and nurture as a strict dichotomy, there are still interesting questions regarding scale. How elementary are the particles of information from which music is built up? How much of these "information particles" are accessible in a base state of the processing system? Taking birth as an important milestone in development, we are only beginning to answer these questions. The answers gained from the studies presented in this thesis point in the direction suggesting that newborn babies come well equipped with perceptual tools needed to process music.

The ability to discriminate the relative size of pitch steps (*Thesis I*), taken together with the results that show newborns' sensitivity to the direction of a pitch change (Carral et al., 2005; for similar results in 2 month olds, see He, Hotson & Trainor, 2009b), provides the interval information used for encoding melodies (Dowling, 1978). While there is no direct evidence that contour information, also used for melody encoding, is actually recognized by newborn infants, there is clear evidence for contour recognition by the age of 2 month (Plantinga & Trainor, 2009). Neonatal sensitivity for prosodic information (Sambeth et al., 2008) is also relevant here as melody is a prosodic cue. Although it is not clear at this point whether melodies are indeed encoded by newborns, it appears that the information necessary for the encoding melodies are available in the newborn auditory system. In the same manner, whereas it has been shown that newborns are able to encode pitch separately from timbre (*Thesis II*), an ability both useful in the context of music as well as in recognizing speech despite individual voice differences, yet it is uncertain how this ability affects the perception of sounds in infants. Timbre and pitch can interact with each other in perception under certain circumstances (Melara & Marks, 1990; Pitt, 1994; Warrier & Zatorre, 2002), which, however were not investigated in our experiments. On the other hand, the neonatal ability to bind together sound features based only on the probability of co-occurrence (Ruususvirta et al., 2003, 2004) can function on the basis of separate pitch and timbre representations. The results however rule out the hypothesis that a complex post-natal learning process is necessary for adult-like perception of pitch (Terhardt, 1974).

The ability to extract higher-order regularities from the auditory input is essential for processing both speech and music, as well as for performing a deeper analysis of the auditory scene (Bregman, 1990). Sound grouping has been studied in adults (Sussman, Ritter & Vaughan, 1998; Sussman et al., 2002b; Sussman & Gumenyuk, 2005) and it was found to be sensitive to both top-down and bottom-up influences. At low SOA's the formation of groups is automatic, as indicated by the MMN responses (Sussman, Ritter & Vaughan, 1998). Newborns do this kind of automatic grouping (*Thesis III*) similar to adults. As grouping can be based on several spectral and temporal cues and their combinations (McAuley, 2010), which were tested mostly using behavioral methods, it would be premature to conclude that newborns group musical stimuli just like adults, but a basic form of grouping is definitely available to them. In the face of recent results indicating cultural differences in the grouping of musical stimuli (Iversen, Patel & Ohgushi, 2008), further examination of newborns' grouping ability is required.

The ability to detect regular beats or pulses in the auditory signal is a prerequisite of perceiving musical meter (Lerdahl & Jackendoff, 1983; London, 2004). It is also needed for synchronization between partners in dyadic interactions (Jaffe et al., 2001). Newborns are able to extract the main regular beat from a strictly metrical rhythmic sound sequence (*Thesis V*). The ability to detect beats (*Thesis V*) coupled with the ability to group sounds (*Thesis III*) hints at the possibility that newborns can extract meter from music, in a similar fashion as adults do (*Thesis IV*). Therefore, meter detection is a valid candidate for an innate musical ability despite the evidence showing effects of learning on meter perception (Hannon & Trehub, 2005; Philips-Silver & Trainor, 2005).

4.2. Discriminative ERP responses in newborns and their relation to adult MMN

The ERP results obtained in the studies forming this thesis may also provide insights to the electrophysiological correlates of auditory deviance detection in newborns. As mentioned in *Section 1.3.1.* previous studies found discriminative ERP components of both positive and negative polarity in newborns and young infants. Whereas the negative component is generally viewed as an immature form of the adult MMN, only loosely compatible explanations for the positive component has been put forward (Morr et al., 2002; Kushnerenko et al., 2002; He, Hotson & Trainor, 2007; He, Hotson & Trainor, 2009a). The explanations are in agreement in proposing that the positive component overlaps the negative one, but differ in regard of its functional description: 1) a waveform entirely unrelated to deviance detection (Morr et al., 2002), 2) an immature form of P3a, elicited by deviance detection (Kushnerenko et al., 2002; Horváth, Winkler & Bendixen, 2008) and 3) a separate component related to deviance detection, but following a different developmental trajectory, which is independent from the adult MMN (He, Hotson & Trainor, 2007; He, Hotson & Trainor, 2009a). A possible fourth explanation would be to tie the positive wave selectively to SOAs below and above the 800-1000 ms range, especially to SOAs in the 400-540 ms range, which is generated by neural populations separate from those involved in the MMN (see in *Section 1.3.1.*), albeit functionally similar to MMN in discriminating stimuli. Nelken and Ulanovsky (2007) discuss stimulus-rate sensitivity in the context of single cell recordings and local field potentials. It may be a little far-fetched to draw a direct parallel with surface recorded EPRs, however there is a distinct similarity that should be taken into account when studying the generators of MMN-like responses in newborns.

The only significant positive response comparable to those mentioned above come from *Thesis II*, where it is peaking at about 200 ms from stimulus onset and it follows an early negative peak of about 90 ms latency. This dissociation of the positive and negative responses in time might be the result of using an SOA of 750 ms, but with 545 ms long, natural-sounding, spectrally rich tones that are able to elicit the negative response (Kushnerenko et al. 2007), while also evoking the positive response. The relatively large pair-onset asynchrony of 1250 ms produced a negative peak at 200 ms in *Thesis I*. Interestingly, the small SOA of 100 ms, not tested before in the literature, also produced a negative peak in the random condition of *Thesis III* making the true relationship between SOA and discriminative responses even less clear. These findings call for a systematic testing of the deviance-related ERPs for a wide range of SOA. *Thesis V* showed for the first time adult-like responses in newborns to sound omission (Yabe et al., 1997). Whereas the neural substrates might be different, this result shows a striking functional similarity in predictive coding between the adult MMN and the neonatal discriminative ERP response.

4.3. Integrating results into a broader picture

The innate perceptual abilities of newborns discussed above do not provide evidence in direct support of the modularist view of music perception (Peretz & Coltheart 2003), because no purely musical abilities were tested in the experiments. The results, however, do not contradict the possibility of innate musical modules by demonstrating the functionality of such innate perceptual abilities as can provide input to the hypothetical music modules. Neither do the current results either vindicate or rule out the adaptationist view of music (Merker, Wallin & Brown, 2001). The origins of the perceptual abilities found in newborns are unclear in the evolutionary sense. While it can be argued that the general need for representing the auditory environment is the ultimate driving force behind these abilities a more detailed view could perhaps include influences from music and other forms of communication as selective pressures. However in such a detailed view, the number of selective forces that are to be considered may become too large to be concise or even meaningful (Fitch, 2006).

The connection between language and music is evident from the results showing common perceptual basis for these abilities in newborns. The ability to extract a melodic contour is the same ability that is needed for analyzing prosodic cues (*Thesis I*); grouping is necessary whether boundaries of words or boundaries of musical phrases are to be extracted (*Thesis III*); and synchronization in a common time-frame is necessary not only for tapping/moving along music, but also for communication in dyads or larger groups (*Thesis IV*). Interestingly, the similarities between music and language do not end here, as linguistic contrasts (e.g. changes in phoneme category, vowel type or vowel length) presented to newborns in oddball paradigms evoke MMN-like magnetic (Kujala et al., 2004) and ERP responses. The latter can be either of negative (Cheour et al., 1998, 2002c; Cheour-Luhtanen et al., 1995; 1996) or positive polarity (Dehaene-Lambertz & Pena, 2001; Pihko et al., 1999). The polarity of these responses also varies considerably between studies reporting measurements during the first year of life (Dehaene-Lambertz & Balliet, 1998; Dehaene-Lambertz & Dehaene, 1994; Dehaene-Lambertz, 2000; Friderici, Friedrich & Weber, 2002; Thierry, Vihman & Roberts, 2003; Friedrich, Weber & Friderici, 2004; van Leeuwen et al., 2006, for reviews Dehaene-Lambertz & Gliga, 2004 and Bishop, 2007). This functional similarity in the basic processing of musical and speech stimuli raises the question: At which level of processing can language and music really be dissociated? Surely there is a multitude of shared abilities already present in newborns that will be uncovered by further research.

One way of unraveling the phylogeny of music is by comparison with animal models. MMN or MMN-like components have been elicited in several animal species including rats (e.g. Astikainen et al., 2006), guinea pigs (e.g. Okazaki et al., 2010), cats (e.g. Csépe, Karmos & Molnár, 1987, Pincze et al., 2001; 2002), and monkeys (Javitt et al., 1992; 1994). These studies generally found MMN to frequency and amplitude deviants (for a review see Nelken & Ulanovsky, 2007), who pointed out the similarity in rule representations and general discriminative abilities of different species. Animals also exhibit further abilities necessary for music processing: for example, some species represent rare feature conjunctions (Astikainen et al., 2006) or sound duration (Okazaki et al., 2010). The most interesting result in the context of music perception comes from Rusuuvirta and colleagues (2007) who showed pitch-contour discrimination in anesthetized rats. This finding opens the possibility to directly compare between species the perceptual abilities underlying music and test predictions about the music-specificity of some abilities using animals, which presumably lack music cognition.

The broadest way of integrating the results of this thesis with the results of cognitive science can be based on the similarities of basic auditory processing found in adults, newborns and animals using the MMN paradigm, which Näätänen and colleagues (2010) see as a manifestation of a common automatic core of other sound-related cognitive abilities.

5. Conclusions and further directions

Babies are born well-equipped for gathering information from the auditory environment. In many respects, their abilities are comparable to those of adults, if not in precision, certainly in function. The experiments of this thesis are amongst the first steps toward a systematic examination of neonatal auditory processing. Further experiments within the field of music perception should try to test broader theories about the origins and functional organization of music, as well as the underlying neuronal substrates. Abilities, which might be exclusive to music processing, should be one of the focal points of research. A good candidate for such an ability, which was identified in this thesis, is meter perception. The MMN-like response in newborns is a useful and proven tool for this endeavor, albeit one that itself is in need of further research. Finally, integrating music research with the study of language and the broader field of communication should yield a deeper understanding of human cognition.

References

Alho, K., Sainio, K., Sajaniem, N., Reinikainen, K., Näätänen, R. (1990) Event-related brain potential of human newborns to pitch change of an acoustic stimulus. *Electroencephalography, and clinical Neurophysiology*, 77, 151-155.

Alho K. (1995) Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear and Hearing*. 16(1), 38-51.

Allen, J., Kraus, N., Bradlow, A. (2000) Neural representation of consciously imperceptible speech sound differences. *Perception & Psychophysics,* 62(7), 1383-1393.

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

American Standards Association (1960) Acoustical Terminology. American Standards Association, New York.

Astikainen, P., Ruusuvirta, T., Wikgren, J., Penttonen, M. (2006) Memory-based detection of rare sound feature combinations in anesthetized rats. *Neuroreport*, 17(14), 1561-1564.

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

Barkow, J. H., Cosmides, L., Tooby, J. (Editors) (1995) The Adapted mind: evolutionary psychology and the generation of culture. Oxford University Press, Oxford

Baruch, C., Drake, C. (1997) Tempo Discrimination in Infants. *Infant behavior and development*, 20(4), 573-577.

Bazana, P., G., Stelmack, R. M. (2002) Intelligence and Information Processing During an Auditory Discrimination Task With Backward Masking: An Event-Related Potential Analysis. *Journal of Personality and Social Psychology*, 83(4), 998–1008.

Bendixen, A., Schröger, E., Winkler, I. (2009) I Heard That Coming: Event-Related Potential Evidence for Stimulus-Driven Prediction in the Auditory System. *The Journal of Neuroscience,* 29(26), 8447–8451.

Besson, M., (1999) The Musical Brain: Neural Substrates of Music Perception. *Journal of New Music Research*, 28(3), 246-256.

Bishop, D. V. M. (2007) Using Mismatch Negativity to Study Central Auditory Processing in Developmental Language and Literacy Impairments: Where Are We, and Where Should We Be Going? *Psychological Bulletin*, 133(4), 651–672.

Bispham, J. (2006) Rhythm in Music: What is it? Who has it? And why? *Music Perception,* 24, 125-134.

Bolton, T. L., (1894) Rhythm. *The American Journal of Psychology*, 6(2) 145-238.

Bradlow, A. R., Kraus, N., Nicol, T. G., McGee, T. J.,Cunningham, J., Zecker, S. G., Carrell, T. D. (1999) Effects of lengthened formant transition duration on discrimination and neural representation of synthetic CV syllables by normal and learning-disabled children**.** *Journal of the Acoustical Society of America*, 106(4), 2086-2096.

Brannon, E. M., Libertus, M. E., Meck, W. H., Woldorff, M. G. (2008) Electrophysiological Measures of Time Processing in Infant and Adult Brains: Weber's Law Holds. *Journal of Cognitive Neuroscience*, 20(29), 193–203.

Brattico, E., Näätänen, R., Tervaniemi, M. (2002). Context effects on pitch perception in musicians and non-musicians: evidence from ERP recordings. Music Perception, 19(2), 199-222.

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.

Bregman, A. S. (1990) Auditory scene analysis: the perceptual organization of sound. Bradford Books, Cambridge, MA.

Caclin, A., McAdams, S., Smith, B. K, Winsberg, S. (2005) Acoustic correlates of timbre space dimensions: a confirmatory study using synthetic tones. *Journal of the Acoustical Society of America*, 118(1), 471-482.

Caclin, A., Brattico, E., Tervaniemi, M., Näätänen, R., Morlet, D., Giard, M-H., McAdams, S. (2006) Separate Neural Processing of Timbre Dimensions in Auditory Sensory Memory. *Journal of Cognitive Neuroscience*, 18(12) 1959–1972.

Carral, V., Huotilainen, M., Ruusuvirta, T., Fellman, V., Näätänen, R., Escera, C. (2005) A kind of auditory 'primitive intelligence' already present at birth. *.European Journal of Neuroscience*, 21, 3201–3204.

Carruthers, P. (2006) The Achitecture of the Mind, Oxford University Press, Oxford.

Čeponiene, R., Hukki, J., Cheour, M., Haapanen, M-L., Koskinen, M., Alho, K., Näätänen, R. (2000) Dysfunction of the auditory cortex persists in infants with certain cleft types. *Developmental Medicine & Child Neurology*, 42, 258–265.

Čeponiene, R. (2001) Event-related potential (ERP) indices of central auditory development in healthy children and in children with oral clefts. Doctoral dissertation, University of Helsinki, Faculty of Arts, Department of Psychology, Helsinki.

Čeponiene, R., Kushnerenko, E., Fellman, V., Renlund, M., Suominen, K., Näätänen, R. (2002) Event-related potential features indexing central auditory discrimination by newborns. *Cognitive Brain Research*, 13, 101–113.

Chandrasekaran, B., Krishnan, A., Gandour, J. T. (2009) Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain & Language*, 108, 1–9.

Chang, H., Trehub, A. E. (1977a) Auditory Processing of Relational Information by Young Infants. *Journal of Experimental Child Psychology*, 24, 324-331.

Chang, H., Trehub, A. E. (1977b) Infants' Perception of Temporal Grouping in Auditory Patterns. *Child Development*, 48, 1666-1670.

Cheour, M., Alho, K., Čeponiene, R., Reinikainen, K., Sainio, K., Aaltonen, O., Näätänen, R. (1998) Maturation of mismatch negativity in infants, *International Journal of Psychophysiology*, 29, 217-226.

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

Cheour, M., Čeponiene, R., Leppänen, P. H. T., Alho, K., Kujala, T., Renlund, M., Fellman, V., Näätänen, R. (2002a) The auditory sensory memory trace decays rapidly in newborns. *Scandinavian Journal of Psyhology*, 43, 33-39.

Cheour, M., Kushnerenko, E., Čeponiene, R., Fellman, V., Näätänen, R. (2002b) Electric Brain Responses Obtained From Newborn Infants to Changes in Duration in Complex Harmonic Tones. *Developmental Neuropsychology*, 22, 471–479.

Cheour, M., Martynova, O., Näätänen, R., Erkkola, R., Sillanpää, M., Kero, P., Raz, A., Kaipio, M.L., Hiltunen, J., Aaltonen, O., Savela, J., Hämäläinen, H. (2002c). Speech sounds learned by sleeping newborns. *Nature*, 415, 599–600.

Cheour, M., Imada, T., Taulu, S., Ahonen, A., Salonen, J., Kuhl, P. (2004) Magnetoencephalography is feasible for infant assessment of auditory discrimination. *Experimental Neurology*, 190, S44-51.

Cheour-Luhtanen, M., Alho, K., Kujala, T., Sainio, K., Reinikainen, K., Renlund, M., Aaltonen, O., Eerola, O., Näätänen, R. (1995) Mismatch negativity indicates vowel discrimination in newborns. *Hearing Research*, 82, 53-58.

Cheour-Luhtanen, M., Alho, K., Sainio, K., Rinne, T., Reinikainen, K., Pohjavuori, M., Aaltonen, O., Eerola, O., Näätänen, R. (1996) The ontogenetically earliest discriminative response of the human brain. *Psychophysiology*, 33, 478-481.

Chomsky, N. (1957) Syntactic structures. Mouton.

Cohen, L., Dehaene, S., Naccache, L., Lehericy, S., Dehaene-Lambertz, G., Henaff, M., Michel, F., (2000) The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, 123, 291–307.

Cowan, N. (1984) On short and long auditory stores. *Psychological Bulletin*, 96(2), 341-370.

Cross, I. (2003) Music as a Biocultural Phenomenon. *Annals of the New York Academy of Sciences*, 999, 106-111.

Cross, I. (2006) Music, Cognition, Culture, and Evolution. *Annals of the New York Academy of Sciences*, 1060, 28-42.

Csépe, V., Karmos, G., Molnár, M. (1987) Evoked potential correlates of stimulus deviance during wakefulness and sleep in cat–Animal model of mismatch negativity. *Electroencephalography and Clinical Neurophysiology*, 66, 571–578.

Cuddy, J. J., Cohen, A. J. (1976) Recognition of transposed melodic sequences. *Quarterly Journal of experimental psychology*, 28, 255-270.

Cynx, J. (1995) Similarities in Absolute and Relative Pitch Perception in Songbirds (Starling and Zebra Finch) and a Nonsongbird (Pigeon). *Journal of Comparative Psychology*, 109 (3), 261-267.

Dalla Bella, I., Peretz, I., (1999) Music Agnosias: Selective Impairments of Music Recognition After Brain Damage. *Journal of New Music Research*, 28 (3), 209-216.

Darwin, C. (1879/2004) The descent of man, and selection in relation to sex. 2nd edition. John Murray, London./ Penguin Books, London.

DeCasper, A. J., Fifer, W. P. (1980) Of Human Bonding: Newborns Prefer their Mothers' Voices. *Science*, 208(4448), 1174-1176.

Deliège, I., Sloboda, J. (Editors) (1997) Perception and cognition of music. Psychology Press, Hove, UK.

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

Dehaene-Lambertz, G.,Baillet, S. (1998) A phonological representation in the infant brain. *Neuroreport*, 9:1885-1888

Dehaene-Lambertz, G. (2000) Cerebral specialization for speech and non-speech stimuli in infants. *Journal of Cognitive Neuroscience*, 12, 449–460.

Dehaene-Lambertz, G, Pena, M. (2001) Electrophysiological evidence for automatic phonetic processing in neonates. *Neuroreport*, 12, 3155–3158.

Dehaene-Lambertz, G., Gliga, D. (2004) Common neural basis for phoneme processing in infants and adults. *Journal of Cognitive Neuroscience*, 16, 1375–1387.

Demany, L., McKenzie, B., Vurpillot, E. (1977) Rhythm perception in early infancy. *Nature*, 266(5604), 718-719.

Di Pietro, M., Laganaro, M., Leemann, B., Schnider, A. (2004) Receptive amusia: temporal auditory processing deficit in a professional musician following a left temporo-parietal lesion. *Neuropsychologia*, 42, 868–877.

Dissanayake, E. (2001). Antecedents of the temporal arts in early mother-infant interaction. In: Wallin, N. L., Merker, B., Brown, S. (Eds.) *The Origins of Music*. MIT Press, Cambridge, MA.

Dowling, W. J. (1978) Scale and Contour: Two Components of a Theory of Memory for Melodies. *Psychological Review*, 85(4), 341-354.

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

Draganova, R., Eswaran, H., Murphy, P., Lowery, C., Preissl, H. (2007) Serial magnetoencephalographic study of fetal and newborn auditory discriminative evoked responses. *Early Human Development*, 83, 199–207.

Drake, C. (1993) Reproduction of musical rhythms by children, adult musicians, and adult nonmusicians. *Perception & Psychophysics,* 53(1), 25-33.

Drake, C., Jones, M. R., Baruch, C. (2000) The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. *Cognition* ,77 , 251-288.

Edworthy, J. (1985) Interval and Contour in Melody Processing. *Music Perception* 2(3), 375-388.

Eggermont, J. J. (1988) On the rate of maturation of sensory evoked potentials. *Electroencephalography and clinical Neurophysiology*, 70, 293-305.

Escera, C., Alho, K., Winkler, I., & Näätänen, R. (1998). Neural mechanisms of involuntary attention switching to novelty and change in the acoustic environment. *Journal of Cognitive Neuroscience*, 10, 590-604.

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

Fabiani, M., Gratton, G., Federmeier, K. D. (2007) Event-Related Brain Potentials: Methods, Theory, and Applications. In.: Cacioppo, J., Tassinary, L. G., Brentson, G. G. *The Handbook of Electrophysiology*, Cambridge University Press, New York.

Fastl, H., Zwicker, E. (2007) Psychoacoustics – facts and models. 3rd edition, Springer-Verlag, Berlin, Heidelberg, New York.

Fellman, V., Kushnerenko , E., Mikkola, K., Čeponiene, R., Leipälä, J., Näätänen, R. (2004) Atypical Auditory Event-Related Potentials in Preterm Infants during the First Year of Life: A Possible Sign of Cognitive Dysfunction? *Pediatric Research* 56(2) 291-297.

Fitch,W.T. (2006) The biology and evolution of music: A comparative perspective. *Cognition,* 100, 173-215.

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

Fodor, J. A. (2001) The Mind doesn't work that way. MIT Press, Cambridge, MA.

Friederici, A. D., Friedrich, M., Weber, C. (2002) Neural manifestation of cognitive and precognitive mismatch detection in early infancy. *Neuroreport*, 13, 1251–1254.

Friedrich, M., Weber, C., Friederici, A. D. (2004) Electrophysiological evidence for delayed mismatch response in infants at risk for specific language impairment. *Psychophysiology*, 41, 772–782.

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 Biobehavioral Reviews*, 25, 355-373.

Friston, K., Kiebel, S. (2009) Cortical circuits for perceptual inference. *Neural Networks,* 22, 1093-1104.

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(6), 1010–1021.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., Pantev, C. (2005) Automatic Encoding of Polyphonic Melodies in Musicians and Nonmusicians. *Journal of Cognitive Neuroscience*, 17(10) 1578–1592.

Gaser, C., Schlaug, G. (2003) Brain Structures Differ between Musicians and Non-Musicians. *The Journal of Neuroscience,* 23(27), 9240–9245.

Geiser, E., Ziegler, E., Jancke, L., Meyer, M. (2009) Early electrophysiological correlates of meter and rhythm processing in music perception. *Cortex*, 45, 93-102.

Gosselin, N., Jolicoeur, P., Peretz, I. (2009) Impaired Memory for Pitch in Congenital Amusia. *Annals of the New York Academy of Sciences*. 1169, 270–272.

Gould, S. J., Vrba E. S. (1982) Exaptation — a missing term in the science of form. *Paleobiology*, 8 (1), 4–15.

Goydke, K. N., Altenmüller, E., Möller, J., Münte, T. F. (2004) Changes in emotional tone and instrumental timbre are reflected by the mismatch negativity. *Cognitive Brain Research*, 21, 351–359.

Grey, J. M. (1977) Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, 61(5), 1270-1277.

Groussard, M., Viader, F., Hubert, V., Landeau, B., Abbas, A., Desgranges, B., Eustache, F., Platel, H. (2010) Musical and verbal semantic memory: Two distinct neural networks? *NeuroImage*, 49, 2764–2773.

Háden G. P., Sussman E. S., Czigler I., Winkler I. (2009). The MMN attention effect revisited: Forming the standard or detecting the deviant?. *Frontiers in Human Neuroscience. Conference Abstract: MMN 09 Fifth Conference on Mismatch Negativity (MMN) and its Clinical and Scientific Applications*. DOI: 10.3389/conf.neuro.09.2009.05.051

Hagen, E., Bryant, G. (2003). Music and dance as a coalition signaling system. *Human Nature*, 14(1), 21–51.

Hajda, J. M., Kendall, R. A., Carterette, E. C., Harshberger, M. L. (1997) Methodological issues in timbre research. In Deliège, I., Sloboda J. (Eds.), *Perception and Cognition of Music* ,253-306 , Psychology Press, Hove, UK.

Handel, S. (1998) The Interplay Between Metric and Figural Rhythmic Organization. *Journal of Experimental Psychology: Human Perception and Performance*, 24(5), 1546-1561.

Handel, S. (2006) Perceptual coherence: Hearing and seeing, Oxford University Press, New York.

Hannon, E. E., Snyder, J. S., Eerola, T., Krumhansl, C. L. (2004) The Role of Melodic and Temporal Cues in Perceiving Musical Meter. *Journal of Experimental Psychology: Human Perception and Performance*, 30(5), 956–974.

Hannon, E. E., Johnson, S. P. (2005) Infants use meter to categorize rhythms and melodies: Implications for musical structure learning. *Cognitive Psychology* 50, 354–377.

Hannon, E. E., Trehub, S. E. (2005) Metrical Categories in Infancy and Adulthood. *Psychological Science*, 16(1), 48-55.

Hari, R., Hämäläinen, M., Ilmoniemi, R., Kaukoranta, E., Reinikainen, K., Salminen, J., Alho, K., Näätänen, R., Sams, M. (1984) Responses of the primary auditory cortex to pitch changes in a sequence of tone pips: neuromagnetic recordings in man. *Neuroscience Letters*, 50(1-3), 127-32.

Haroush, K., Hochstein, S., Deouell, L. Y. (2010) Momentary fluctuations in allocation of attention: Cross-modal effects of visual task load on the auditory Mismatch Negativity. *The Journal of Neuroscience*, 22(7), 1440-1451.

Hauser, M. D., Chomsky, N., Fitch, W. T. (2002) The Faculty of Language: What Is It, Who Has It and How Did It Evolve? *Science*, 298, 1569-1579.

Hauser, M. D., McDermott, J. (2003) Functional brain development in humans. *Nature Neuroscience,* 6, 663-668.

Haxby, J. V., Gobbini, I. M., Furey, M. L., Ishai, A., Schouten, J. L., Pietrini, P. (2001) Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science*, 293, 2425-2430.

He, C. Hotson, L., Trainor, L. J. (2007) Mismatch Responses to Pitch Changes in Early Infancy. *Journal of Cognitive Neuroscience,* 19(5), 878–892.

He, C. Hotson, L., Trainor, L. J. (2009a) Maturation of cortical mismatch responses to occasional pitch change in early infancy: Effects of presentation rate and magnitude of change *Neuropsychologia,* 47, 218–229.

He, C. Hotson, L., Trainor, L. J. (2009b) Development of infant mismatch responses to auditory pattern changes between 2 and 4 months old. *European Journal of Neuroscience*. 29, 861–867.

He, C., Trainor, L. J. (2009) Finding the Pitch of the Missing Fundamental in Infants. *The Journal of Neuroscience,* 29(24), 7718–8822.

Hepper, P. G., Shahidullah, B. S. (1994) Development of fetal hearing. *Archives of Disease in Childhood*, 71, F81-F87.

Herholz, S. C., Lappe, C., Pantev, C. (2009) Looking for a pattern: An MEG study on the abstract mismatch negativity in musicians and nonmusicians. *BMC Neuroscience*, 10(42), 1-10.

Hirasawa, K., Kurihara, M., Konishi, Y. (2002) The relationship between mismatch negativity and arousal level. Can mismatch negativity be an index for evaluating the arousal level in infants? *Sleep Medicine,* 3, S45–S48.

Horváth, J., Winkler, I., & Bendixen, A. (2008). Do N1/MMN, P3a, and RON form a strongly coupled chain reflecting the three stages of auditory distraction? *Biological Psychology*, 79, 139-147.

Huotilainen, M., Kujala, A., Hotakainen, Shestakova, A., Kushnerenko, E., Parkkonen, L., Fellman V., Näätänen, R. (2003) Auditory magnetic responses of healthy newborns. *NeuroReport*, 14(14) 1871-1875.

Huotilainen, M., Kujala, A., Hotakainen, M., Parkkonen, L., Taulu, S., Simola, J., Nenonen, J., Karjalainen, M., Näätänen, R. (2005) Short-term memory functions of the human fetus recorded with magnetoencephalography. *NeuroReport*, 16(1), 81-84.

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

Huttenlocher, P. R. (1979).Synaptic density in human frontal cortex - Developmental changes and effects of aging. *Brain Research*, 163, 195-205.

Huttenlocher, P. R. (1984) Synapse elimination and plasticity in developing human cerebral cortex. *American Journal of Mental Deficiency.* Vol 88(5), 488-496.

Huttenlocher, P. R., and Dabholkar, A. S. (1997). Regional differences in synaptogenesis in human cerebral cortex. *Journal of Comparative Neurology*, 387(2), 167-178.

Hyde, K. L., Zatorre, R. J., Griffiths, T. D., Lerch, J. P., Peretz, I. (2006) Morphometry of the amusic brain: a two-site study. *Brain* ,129, 2562-2570.

Hyde, K. L., Lerch, J. P., Zatorre, R. J., Griffiths, T. D., Evans, A. C., Peretz, I. (2007) Cortical Thickness in Congenital Amusia: When Less Is Better Than More. *The Journal of Neuroscience*, 27(47), 13028 –13032.

Hyde, K. L., Zatorre, R. J., Peretz, I. (2010, published online) Functional MRI Evidence of an Abnormal Neural Network for Pitch Processing in Congenital Amusia. *Cerebral Cortex*

Iversen, J. R., Patel, A. D., Ohgushi, K. (2008) Perception of rhythmic grouping depends on auditory experience. *Journal of the Acoustical Society of America*, 124(4), 2263–2271.

Izumi, A. (2001) Relative pitch perception in Japanese monkeys (Macaca fuscata). *Journal of Comparative Psychology*, 115 (2), 127-131.

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(17), 6809–6814.

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

Jaffe J., Beebe, B., Feldstein, S., Crown, C. L., Jasnow, M. D. (2001) Rhythms of Dialogue In.: Infancy,Monographs for the Society for Research and Child Development Vol 66, No 2., Blackwell, Boston.

Javitt, D.C., Schroeder, C.E., Steinschneider, M., Arezzo, J.C., Vaughan, H.G., Jr. (1992) Demonstration of mismatch negativity in the monkey. *Electroencephalography and Clinical Neurophysiology*, 83, 87–90.

Javitt, D.C., Steinschneider, M., Schroeder, C.E., Vaughan, H.G., Jr., Arezzo, J.C. (1994) Detection of stimulus deviance within primate primary auditory cortex: Intracortical mechanisms of mismatch negativity (MMN) generation. *Brain Research*, 667, 192–200.

Jing, H., Benasich, A. A. (2006) Brain responses to tonal changes in the first two years of life. *Brain Development*, 28(4), 247–256.

Johnson, M. H. (2001) Functional brain development in humans. *Nature Reviews Neuroscience*, 2, 475-483.

Justus ,T., Hutsler, J. J. (2005) Fundamental Issues in the Evolutionary Psychology of Music : Assessing Innateness and Domain Specificity . *Music Perception,* 23, 1-27.

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, 3(8), 759-763.

Karmiloff-Smith, A. (1992) Beyond modularity. MIT Press, Cambridge, MA.

Kemler Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., Wright Cassidy, K. (1989) How the prosodic cues in motherese might assist language learning. *Journal of Child Language*, 16, 55-68.

Kester, D. B., Saykin, A. J., Sperling, R., O'Connor, M. J., Robinson, L. J., Gur, R. C. (1991) Acute effect of anterior temporal lobectomy on musical processing. *Neuropscyhologia.* 29(1), 703-708.

Kisilevsky, B. S., Hains, S. M. J., Jacquet, A. -Y., Granier-Deferre, C., Lecanuet, J. P. (2004) Maturation of fetal responses to music. *Developmental Science*, 7, (5) 550–559.

Koelsch, S., Schöger, 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, (12) 579-584.

Koelsch, S. ,Sammler, D. (2008) Cognitive Components of Regularity Processing in the Auditory Domain. *PLoS ONE*, 3(7), e2650, 1-7.

Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller K., Gruber, O. (2009) Functional architecture of verbal and tonal working memory: An FMRI study. *Human Brain Mapping*, 30(3), 859–873.

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

Kraus, N., McGee, T., Carrell, T. D., King, C., Tremblay, K., Nicol, N. (1995) Central auditory plasticity associated with speech discrimination training. *Journal of Cognitive Neuroscience*, 7, 25–32.

Kraus, N., Koch, D. B., McGee, T. J., Nicol, T. G., Cunningham, J. (1999) Speech-Sound Discrimination in School-Age Children: Psychophysical and Neuropsychological Measures. *Journal of Speech*, 42,1042-1060.

Krumhansl, C., L. Iverson, P. (1992) Perceptual Interactions Between Musical Pitch and Timbre. *Journal of Experimental Psychology: Human Perception and Performance*. 18(3), 739-751.

Krumhansl, C., L. (2000) Rhythm and Pitch in Music Cognition. *Psychological Bulletin*, 126(1), 159-179.

Kujala, T., Kallio, J., Tervaniemi, M., Näätänen, R. (2001) The mismatch negativity as an index of temporal processing in audition. *Clinical Neurophysiology*, 112, 1712-1719.

Kujala, A., Huotilainen, M., Hotakainen, M., Lennes, M., Parkkonen, L., Fellman, V., Näätänen, R. (2004) Speech-sound discrimination in neonates as measured with MEG. *Neuroreport*, 15, 2089-2092.

Kujala, T., Halmetoja, J., Näätänen, R., Alku, P., Lyytinen, H., Sussman, E. (2006) Speech- and sound-segmentation in dyslexia: evidence for a multiple-level cortical impairment. *European Journal of Neuroscience*. 24, 2420–2427.

Kujala, T., Tervaniemi, M., Schröger, E. (2007) The mismatch negativity in cognitive and clinical neuroscience: Theoretical and methodological considerations. *Biological Psychology*, 74 ,1–19.

Kunej, D., Turk, I. (2001) New Perspectives on the Beginnings of Music: Archeological and Musicological Analysis of a Middle Paleolithic Bone “Flute”. In: Wallin, N. L., Merker, B., Brown, S. (Eds.) *The Origins of Music*. MIT Press, Cambridge, MA.

Kushnerenko, E., Čeponiene, R., Fellman, V., Huotilainen, M., Winkler, I. (2001) Event-related potential correlates of sound duration: similar pattern from birth to adulthood. *NeuroReport,* 12(17), 3777-3781.

Kushnerenko, E., Čeponiene, R., Balan, P., Fellman, V., Näätänen, R. (2002) Maturation of the auditory change detection response in infants: a longitudinal ERP study. *NeuroReport,* 13(15), 1843-1848.

Kushnerenko, E. (2003) Maturation of the cortical auditory event-related brain potentials in infancy. Doctoral dissertation, University of Helsinki, Faculty of Arts, Department of Psychology, Helsinki.

Kushnerenko, E. Winkler, I., Horváth, J., Näätänen, R., Pavlov, I., Fellman, V., Huotilainen, M. (2007) Processing acoustic change and novelty in newborn infants. *European Journal of Neuroscience*, 26, 265–274.

Lang, H. A., Nyrke, T., Ek, M., Aaltonen, O., Raimo, I., Näätänen, R. (1990) Pitch discrimination performance and auditory event-related potentials. In: Brunia, C.H.M., Gaillard, A.W.K., Kok, A., Mulder, G., Verbaten, M.N. (Eds.) *Psychophysiological brain research* 1990, vol. 1. Tilburg University Press, Tilburg.

Lecanuet, J. (1996) Prenatal auditory experience. In.: Deliège, I., Sloboda J. A. (eds.) *Musical beginnings: origins and development of musical competence*. Oxford University Press, New York

Leppänen, P. H. T., Eklund, K. M., Lyytinen, H. (1997) Event-related brain potentials to change in rapidly presented acoustic stimuli in newborns. *Developmental neuropsychology* 13(22), 175-204.

Leppänen, P. H. T., Guttorm T. M., Pihko, E., Takkinen, S., Eklund, K. M., Lyytinen, H. (2004) Maturational effects on newborn ERPs measured in the mismatch negativity paradigm. *Experimental Neurology*, 190, S91–S101.

Lerdahl, F., Jackendoff, R. (1983). A generative theory of tonal music. MIT Press, Cambridge, MA.

Levitin, D. J., Rogers, S. E. (2005) Absolute pitch: perception, coding, and controversies. *Trends in Cognitive Sciences*, 9(1), 26-33.

Levitt, P. (2003) Structural and functional maturation of the developing primate brain. *The Journal of Pediatrics*, 143(4), S35-S45.

Liégeois-Chauvel, C., Peretz, I., Babai, M., Laguitton, V., Chauvel, P. (1998) Contribution of different cortical areas the temporal lobes to music processing. *Brain*, 121, 1853-1867.

London, J. (2002) Cognitive Constraints on Metric Systems: Some Observations and Hypotheses. *Music Perception*, 19(4), 529–550.

London, J. (2002) Hearing in time. Oxford University Press, New York.

Lyytinen, H., Blomberg, A-P., Näätänen, R. (1992) Event-Related Potentials and Autonomic Responses to a Change in Unattended Auditory Stimuli. *Psychophysiology*, 29(5), 523-534.

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(5), 540-545.

Martynova, O., Kirjavainen, J., Cheour, M. (2003) Mismatch negativity and late discriminative negativity in sleeping human newborns. *Neuroscience Letters*, 340, 75–78.

May, P. J. C., Tiitinen, H. (2010) Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology*, 47 (1), 66-122.

McAdams, S., Bregman, A. (1979) Hearing Musical Streams. *Computer Music Journal*, 3(4) 26-43.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., Krimphoff, J. (1995) Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. *Psychological Research*, 58(3),177-192.

McAuley, J. D. (2010) Tempo and Rhythm. In: Riess Jones, M., Fay, R., R., Popper, A., N. (Eds.) Music Perception, Springer Handbook of Auditory Research, Springer, New York.

McMullen, E., Saffran, J. R. (2004) Music and Language: A Developmental Comparison. *Music Perception*, 21(3), 289-311.

Melara, R. D., Marks, L. E. (1990) Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception & Psychophysics*, 48(2), 169-178.

Menning, H., Roberts, L. E., Pantev, C. (2000) Plastic changes in the auditory cortex induced by intensive frequency discrimination training. *NeuroReport,* 11(4), 817-822.

Menon, V., Levitin, D. J., Smith, B. K., Lembke, A., Krasnow, B. D., Glazer, D., Glover, G. H., McAdams S. (2002) Neural correlates of timbre change in harmonic sounds. *Neuroimage*, 17(4), 1742-1754.

Merriam, A. P. (1964) The anthropology of music. Northwestern University Press, Evanston, IL.

Midorikawa, A., Kawamura, M., Kezuka, M., (2003) Musical Alexia for Rhythm Notation: A Discrepancy Between Pitch and Rhythm. *Neurocase* , 9(3), 232–238.

Moon, C., Cooper, R. B., Fifer, W. P. (1993) Two-Day-Olds Prefer Their Native Language. *Infant Behavior and Development*, 16, 495-500.

Moore, J. K., Perazzo, L. M., Braun, A. (1995) Time course of axonal myelination in the human brainstem auditory pathway. *Hearing Research*, 87, 21-31.

Moore, J. K., Ponton, C. W., Eggermont, J. J., Wu, B. J. C., Huang, J. Q. (1996) Perinatal Maturation of the Auditory Brain Stem Response: Changes in Path Length and Conduction Velocity. *Ear & Hearing*, 17(5), 411-418.

Moore, J. K., Linthicum, F. H. (2007) The human auditory system: a timeline of development. *International Journal of Audiology*, 46, 460-478.

Moreau, P., Jolicoeur, P., Peretz, I. (2009) Automatic Brain Responses to Pitch Changes in Congenital Amusia. *Annals of the New York Academy of Sciences*, 1169, 191–194.

Morr, M. L., Shafer, V. L., Kreuzer, J. A., Kurtzberg, D. (2002) Maturation of Mismatch Negativity in Typically Developing Infants and Preschool Children. *Ear & Hearing*, 23,(2) 118-136.

Morrongiello, B. A., Trehub, S. E. (1987) Age-related Changes in Auditory Temporal Perception. *Journal of Experimental Child Psychology*, 44, 413-426.

Murayama, J., Kashiwagi, T., Kashiwagi, A., Mimura, M., (2004) Impaired pitch production and preserved rhythm production in a right brain-damaged patient with amusia. *Brain and Cognition* , 56 ,36–42

Näätänen, R., Gaillard, A. W., Mäntysalo, S. (1978) Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica (Amst)*, 42(4), 313-329.

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 Brain Sciences*, 13, 201–288.

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

Näätänen, R., & Alho, K. (1997). Mismatch negativity (MMN) - the measure for central sound representation accuracy. *Audiology and Neuro-Otology, 2,* 341-353.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Ilvonen, 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. N*ature*, 385, 432-434.

Näätänen, R., & Winkler, I. (1999) The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, 125 (6), 826-859.

Näätänen, R. (2003) Mismatch negativity: clinical research and possible applications. *International Journal of Psychophysiology*, 48, 179–188.

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., Paavilainen , P., Rinne, T., Alho, K. (2007) The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology,* 118, 2544–2590.

Näätänen, R., Astikainen, P., Ruusuvirta, T., Huotilainen, M. (2010) Automatic auditory intelligence: An expression of the sensory-cognitive core of cognitive processes. *Brain Research Reviews*, 64(1), 123-136.

Näätänen, R., Kujala, T., Winkler, I. (2011). Auditory processing that leads to conscious perception: A unique window to central auditory processing opened by the mismatch negativity (MMN) and related responses. *Psychophysiology*, 48(1), 4-22.

Nelken, I., Ulanovsky, N. (2007) Mismatch Negativity and Stimulus-Specific Adaptation in Animal Models. *Journal of Psychophysiology*, 21(3–4), 214–223.

Neuloh, G., Curio, G. (2004) Does familiarity facilitate the cortical processing of music sounds? *NeuroReport*, 15(16), 2471-2475.

Nikjeh, D. A., Lister, J. J., Frisch, S.A. (2008) Hearing of note: An electrophysiologic and psychoacoustic comparison of pitch discrimination between vocal and instrumental musicians. *Psychophysiology*, 45, 994–1007.

Novitski, N., Tervaniemi, M., Huotilainen, M., Näätänen, R. (2004) Frequency discrimination at different frequency levels as indexed by electrophysiological and behavioral measures. *Cognitive Brain Research*, 20, 26–36.

Novitski, N., Huotilainen , M., Tervaniemi , M., Näätänen, R., Fellman, V. (2006) Neonatal frequency discrimination in 250–4000-Hz range: Electrophysiological evidence. *Clinical Neurophysiology*, 118, 412–419.

O'Hare, E. D., Sowell, E. R. (2008) Imaging Developmental Changes in Gray and White Matter in the Human Brain. In.: Nelson, C. A., Luciana, M. (Eds.) Handbook of developmental cognitive neuroscience, 2nd ed., The MIT Press, Cambridge, MA.

Ohnishi, T., Matsuda, H., Asada, T., Aruga, M., Hirakata, M., Nishikawa, M., Katoh, A., Imabayashi, E. (2001) Funtional Anatomy of Musical Perception in Musicans. *Cerebral Cortex*, 11, 754-760.

Okazaki, S., Kanoh, S., Tsukada, M., Oka, K. (2010) Neural substrate of sound duration discrimination during an auditory sequence in the guinea pig primary auditory cortex. *Hearing Research*, 259(1-2), 107-116.

Olsho, L. W., Koch, E. G., Halpin, C. F. (1987) Level and age effects in infant frequency discrimination. *Journal of the Acoustical Society of America* . 82(2), 454-464.

Olsho, L. W., Koch, E. G., Carter, E. A., Halpin, C. F., Spetner, N.B. (1988) Pure-tone sensitivity of human infants. *Journal of the Acoustical Society of America*, 84(4), 1316-1324.

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.

Paavilinen, P., Valppu, S., Näätänen, R. (2001) The additivity of the auditory feature analysis in the human brain as indexed by the mismatch negativity: 1+1≈2 but 1+1+1<3. *Neuroscience Letters,* 301, 179-182.

Paavilainen, P., Degerman, A., Takegata, R., Winkler, I. (2003) Spectral and temporal stimulus characteristics in the processing of abstract auditory features. *NeuroReport*, 14(5), 715-718.

Paavilainen, P., Arajärvi, P., Takegata, R. (2007) Preattentive detection of nonsalient contingencies between auditory features. *Neuroreport*, 18(2), 159-163

Patel, A.D. (2006) Musical Rhythm, Linguistic Rhythm, and Human Evolution. *Music Perception* 24(1), 99-103.

Patel, A. D., Iversen, J. R., Bregman, M. R., Schulz, I.,(2009a) Experimental Evidence for Synchronization to a Musical Beat in a Nonhuman Animal. *Current Biology*, 19, (10) 827-830.

Patel, A. D., Iversen, J. R., Bregman, M. R., Schulz, I. (2009b) Studying synchronization to a musical beat in nonhuman animals. *Annals of the New York Academy of Sciences*, 1169, 459-469.

Patterson, R. D., Gaudrain, E., Walters, T. C. (2010) The Perception of Family and Register in Musical Tones. In.: Riess Jones, M., Fay, R., R., Popper, A., N. (Eds.) Music Perception, Springer Handbook of Auditory Research, Springer, New York.

Perani, D., Saccuman, M. C., Scifo, P., Spada, D., Andreoll, G., Rovelli, R., Baldoli, B., Koelsch, S. (2010) Functional specializations for music processing in the human newborn brain. *Proceedings of the National Academy of Sciences*, 107(10), 4758–4763.

Peretz, I., Morais, J. (1989) Music and modularity. *Contemporary Music Review*, 4 , 279-293.

Peretz, I., Kolinsky, I., Tramo, M., Labrecque, L., Hublet, C., Demeurisse, G., Belleville, S. (1994) Functional dissociations following bilateral lesions of auditory cortex. *Brain*, 117, 1283-1301.

Peretz, I. (1996) Can We Lose Memory for Music? A Case of Music Agnosia in a Nonmusician. *Journal of Cognitive Neuroscience*, 8 (6), 481-496.

Peretz, I. (2001) Brain Specialization for Music New Evidence from Congenital Amusia. *Annals New York Academy of Science*, 930, 153-165.

Peretz, I., Coltheart, M. (2003) Modularity of music processing. *Nature Neuroscience*, 6(7), 688-691.

Peretz, I., Brattico,E., Tervaniemi, M. (2005) Abnormal Electrical Brain Responses to Pitch in Congenital Amusia. *Annals of Neurology*, 58(3), 478-482.

Peretz, I. (2006) The nature of music from a biological perspective. *Cognition*, 100, 1–32.

Peretz, I., Cummings, S., Dubé, M. D. (2007) The Genetics of Congenital Amusia (Tone Deafness): A Family-Aggregation Study. *The American Journal of Human Genetics*, 81, 582-588.

Peretz, I., Brattico, E., Järvenpää, M., Tervaniemi, M. (2009) The amusic brain: in tune,out of key, and unaware. *Brain*, 132, 1277-1286.

Phillips-Silver, J., Trainor, L. J. (2005) Feeling the Beat: Movement Influences Infant Rhythm Perception. *Science*, 308, 1430.

Phillips-Silver, J., Trainor, L. J. (2007) Hearing what the body feels: Auditory encoding of rhythmic movement. *Cognition* 105, 533–546.

Phillips-Silver, J., Trainor, L. J. (2008) Vestibular influence on auditory metrical interpretation. *Brain and Cognition,* 67, 94–102.

Piccirilli M., Sciarma, T., Luzzi, S. (2000) Modularity of music: evidence from a case of pure amusia. *Journal of Neurology, Neurosurgery, and Psychiatry*, 69, 541–545.

Picton, T. W., Taylor, M. J. (2007) Electrophysiological Evaluation of Human Brain Development. *Developmental Neuropsychology*, 31(3), 249–278.

Pihko, E., Leppänen, P.H.T., Eklund, K.M., Cheour, M., Guttorm, T.K., Lyytinen, H. (1999) Cortical responses of infants with a without a genetic risk for dyslexia; I. Age effects. *NeuroReport*, 10, 901-905.

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. *Brain Research: Cognitive Brain Research*, 13, 249–253.

Pinker, S. (1998) How the mind works? The Penguin Press, London.

Pitt, M.A. (1994) Perception of Pitch and Timbre by Musically Trained and Untrained Listeners. *Journal of Experimental Psychology: Human Perception and Performance*, 20 (5) 976-986.

Plantinga, J., Trainor, L. J. (2005) Memory for melody: infants use a relative pitch code. *Cognition*, 98, 1–11.

Plantinga, J., Trainor, L. J. (2008) Infants’ memory for isolated tones and the effects of interference. *Music Perception* 26(2), 121–128.

Plantinga, J., Trainor L. J. (2009) Melody recognition by two-month-old infants. *Journal of the Acoustical Society of America,* 125 (2), 58-62.

Polk, M., Kertesz, A. (2003) Music and Language in Degenerative Diseaseof the Brain. *Brain and Cognition*, 22, 98-117.

Ponton, C. W., Moore, J. K., Eggermont, J. J. (1996) Auditory Brain Stem Response Generation by Parallel Pathways: Differential Maturation of Axonal Conduction Time and Synaptic Transmission *Ear & Hearing,* 17(5), 402-410.

Ponton, C. W., Eggermont, J.,J., Don, M., Kwong, W.,B., Cunningham, J., Trautwein, P. (2000) Maturation of the Mismatch Negativity: Effects of Profound Deafness and Cochlear Implant Use. *Audiology and Neuro-otology*, 5, 167–185.

Price, C. J., Devlin, J.T. (2003) The myth of the visual word form area. *NeuroImage*, 19, 473-481.

Repp, B. H. (2005) Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review*, 12(6), 969-992.

Riess Jones, M., Fay, R. R., Popper, A., N. (Editors)(2010) Music Perception, Springer Handbook of Auditory Research, Springer, New York.

Rinne, T., Särkkä, A., Degerman, A., Schröger, E., & Alho, K. (2006). Two separate mechanisms underlie auditory change detection and involuntary control of attention. *Brain Research*, 1077, 135-143.

Roeder, J. (1984). The search for a survival value of music. *Music Perception*, 1, 350–356.

Ross, D., A., Olson, I., R., Marks, L., E., Gore, J., C. (2004) A nonmusical paradigm for identifying absolute pitch possessors. *Journal of the Acoustical Society of America*, 116(3), 1793-1799.

Ruusuvirta, T., Huotilainen, M., Fellman, V., Näätänen, R. (2003) The newborn human brain binds sound features together. *NeuroReport*, 14(16), 2117-2119.

Ruusuvirta, T., Huotilainen, M., Fellman, V., Näätänen, R. (2004) Newborn human brain identifies repeated auditory feature conjunctions of low sequential probability. *European Journal of Neuroscience*, 20, 2819–2821.

Ruusuvirta, T., Koivisto, K., Wikgren, J., Astikainen , P. (2007) Processing of melodic contours in urethane-anaesthetized rats . *European Journal of Neuroscience*, 26, 701–703.

Saarinen, J., Paavilainen P, Schröger E., Tervaniemi, M., Näätänen, R. (1992) Representation of abstract stimulus attributes in human brain. *NeuroReport*, 3, 1149–51.

Saccuman, M. C., Scifo, P. (2009) Using MRI to Characterize the Anatomy and Function of the Auditory Cortex in Infancy. *Annals of the New York Academy of Sciences*, 1169, 297–307.

Sadie, S. (Editor) (2001) New Grove Dictionary of Music and Musicians. Grove.

Sambeth, A., Huotilainen, M., Kushnerenko, E., Fellman, V., Pihko E. (2006) Newborns discriminate novel from harmonic sounds: A study using magnetoencephalography. *Clinical Neurophysiology*, 117, 496–503.

Sams, M., Paavilainen, P., Alho, K., Näätänen, R. (1985) Auditory frequency discrimination and event-related potentials. *Electroencephalography and Clinical Neurophysiology*, 62(6), 437-48.

Samson, S., Zatorre, R. J. (1988) Melodic and Harmonic Discrimination following Unilateral Cerebral Excision. *Brain and Cognition,* 7, 348-360.

Savela, J., Kujala, T., Tuomainen, J., Ek, M., Aaltonen, O., Näätänen, R. (2003) The mismatch negativity and reaction time as indices of the perceptual distance between the corresponding vowels of two related languages. *Cognitive Brain Research,* 16, 250–256.

Schachner, A., Brady, T. F., Pepperberg, I. M., Hauser, M. D. (2009) Spontaneous Motor Entrainment to Music in Multiple Vocal Mimicking Species. *Current Biology,* 19, 831–836.

Schellenberg, E. G., Trainor, L. J. (1996) Sensory consonance and the perceptual similarity of complex-tone harmonic intervals: tests of adult and infant listeners. *Journal of the Acoustical Society of America*, 100(5), 3321-3328.

Schlaug, G., Jäncke, L., Huang, Y., Steinmetz, H. (1995) In Vivo Evidence of Structural Brain Asymmetry in Musicians, *Science*, 267, 699-701.

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

Seghier, M. L., Lazeyras, F., Huppi, P. S. (2006) Functional MRI of the newborn. *Seminars in Fetal & Neonatal Medicine*, 11, 479-488.

Semal, C., Demany, L. (1991) Dissociation of pitch from timbre in auditory short-term memory. *Journal of the Acoustical Society of America*, 89(5), 2404-10.

Semal, C., Demany, L. (1993) Further evidence for an autonomous processing of pitch in auditory short-term memory. *Journal of the Acoustical Society of America*, 94(3), 1315-1322.

Shahin, A., Roberts, L. E., Trainor, L ,J. (2004) Enhancement of auditory cortical development by musical experience in children . *NeuroReport*, 15(12), 1917-1921.

Singh, P., G. Hirsh, I, J. (1992) Influence of spectral locus and F0 changes on the pitch and timbre of complex tones. *Journal of the Acoustical Society of America*, 92(5), 2651-2661.

Smith, D. R. R., Patterson, R. D., Turner, R. (2005) The processing and perception of size information in speech sounds. *Journal of the Acoustical Society of America*, 117, (1) 305–318.

Spetner, N. B., Olsho, L., W. (1990) Auditory Frequency Resolution in Human Infancy. *Child Development*, 61, 632- 652.

Steinbeis, N., Koelsch, S. (2008) Comparing the Processing of Music and Language Meaning Using EEG and fMRI Provides Evidence for Similar and Distinct Neural Representations. *PLoS ONE*, 3(5), e2226, 1-6.

Sugimoto, T., Kobayashi, H., Nobuyosh, N., Kiriyama, Y., Takeshita, H., Nakamura, T., Hashiya, K. (2010) Preference for consonant music over dissonant music by an infant chimpanzee. *Primates*, 51, 7–12.

Sussman, E., Ritter, W., Vaughan, H.G. (1998) Predictability of stimulus deviance and the mismatch negativity. *NeuroReport*, 9, 4167–4170.

Sussman, E., Winkler, I., Kreuzer, J., Saher, M., Näätänen, R., Ritter ,W. (2002a) Temporal integration: intentional sound discrimination does not modulate stimulus-driven processes in auditory event synthesis. *Clinical Neurophysiology*, 113, 1909–1920.

Sussman, E., Winkler, I., Huoutilainen, M., Ritter, W., Näätänen, R. (2002b) Top-down effects on 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, 30-435.

Sussman, E.S., Gumenyuk, V. (2005) Organization of sequential sounds in auditory memory. *NeuroReport*, 16(13), 1519-1523.

Sussman, E. S. (2007) A New View on the MMN and Attention Debate The Role of Context in Processing Auditory Events . *Journal of Psychophysiology*. 21(3-4), 164–175.

Takegata, R., Paavilinen, P., Näätänen, R., Winkler, I. (2001) Preattentive processing of spectral, temporal, and structural characteristics of acoustic regularities: A mismatch negativity study. *Psychophysiology, 38,* 92–98.

Tanaka, M., Okubo, O., Fuchigami, T., Harada, K. (2001) A study of mismatch negativity in newborns. *Pediatrics International*, 43, 281–286

Terhardt, E. (1974). Pitch, consonance, and harmony. *Journal of the Acoustical Society of America*, 55, 1061–1069.

Tervaniemi, M., Winkler, I., Näätänen, R. (1997) Pre-attentive categorization of sounds by timbre as revealed by event-related potentials. *NeuroReport*, 8, 2571–2574.

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

Tervianemi, M. (2001) Musical Sound Processing in the Human Brain Evidence from Electric and Magnetic Recordings. *Annals of the New York Academy of Sciences*, 930, 259-271.

Tervaniemi, M., Rytkönen, M., Schröger, E., Ilmoniemi, R. J., Näätänen, R. (2001) Superior Formation of Cortical Memory Traces for Melodic Patterns in Musicians. *Learning Memory*, 8, 295-300.

Tervaniemi, M., Huotilainen, M. (2003) The Promises of Change-Related Brain Potentials in Cognitive Neuroscience of Music. *Annals of the New York Academy of Sciences* 999, 29-39.

Tervaniemi, M., Brattico, E. (2004) From Sounds to Music. *Journal of Consciousness Studies*, 11, (3–4) 9–27.

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., Castaneda, A., Knoll, M., Uther, M. (2006). Sound processing in amateur musicians and non-musicians: ERP and behavioral indices. *NeuroReport*, 17, 1225-1228.

Tew, S., Fujioka, T., He, C., Trainor, L. J. (2009) Neural Representation of Transposed Melody in Infants at 6 Months of Age. *Annals of the New York Academy of Sciences*, 1169, 287–290.

Tharpe, A., M., Ashmead, D., H. (2001) A Longitudinal Investigation of Infant Auditory Sensitivity. *American Journal of Audiology*, 10, 1059-0889.

Thierry, G., Vihman, M., Roberts, M. (2003) Familiar words capture the attention of 11-month-olds in less than 250ms. *Neuroreport*, 14(18), 2307-2310.

Thiessen, E. D., Hill, A. E., Saffran, J. R. (2005) Infant-Directed Speech Facilitates Word Segmentation. *Infancy*, 7,(1) 53–71.

Toivainen, P., Tervaniemi, M., Louhivuori, J., Saher, M., Huotilainen, M., Näätänen, R. (1998) Timbre Similarity: Convergence of Neural, Behavioral, and Computational Approaches. *Music Perception*, 16(2), 223-241.

Trainor, L. J., Trehub, S. E. (1992) A Comparison of Infants' and Adults' Sensitivity to Western Musical Structure. *Journal Experimental Psychology: Human Perception and Performance*, 18(2), 394-402.

Trainor, L. J. (1997) Effect of frequency ratio on infants' and adults' discrimination of simultaneous intervals. *Journal Experimental Psychology: Human Perception and Performance*, 23(5), 1427-1438.

Trainor, L. J., Samuel, S. S., Desjardins, R. N., Sonnadara, R. R. (2001) Measuring temporal resolution in infants using mismatch negativity. *NeuroReport*, 12(11), 2443-2448.

Trainor, L. J., McDonald, K .L., Alain, C. (2002) Automatic and Controlled Processing of Melodic Contour and Interval Information Measured by Electrical Brain Activity. *Journal of Cognitive Neuroscience*, 14(3), 430- 442.

Trainor, L. J., Tsang, C. D., Cheung, V. H. W. (2002) Preference for consonance in 2- and 4-month-old infants. *Music Perception*, 20, 187-194.

Trainor, L. J., McFadden, M., Hodgson, L., Darragh, L., Barlow, J., Matsos, L., Sonnadara, R. (2003). Changes in auditory cortex and the development of mismatch negativity between 2 and 6 months of age. *International Journal of Psychophysiology*, 51, 5-15.

Trainor, L. J., Wu, L., Tsang, C. D. (2004). Long-term memory for music: Infants remember tempo and timbre. *Developmental Science*, 7, 289-296.

Trehub, S. E., Schneider, B. A., Endman. M. (1980) Developmental Changes in Infants' Sensitivity to Octave-Band Noises. *Journal of Experimental Child Psychology*, 29, 282-293.

Trehub, S. E., Thorpe, L. A., Morrongiello, B. A. (1987) Organizational Processes in Infants' Perception of Auditory Patterns. *Child Development*, 58, 741-749.

Trehub, S. E., Endman, M., Thorpe, L. A. (1990) Infants' perception of timbre: Classification of complex tones by spectral structure. *Journal of Experimental Child Psychology*, 49, 300-313.

Trehub, S.E., Hannon, E. E. (2006) Infant music perception: Domain-general or domain-specific mechanisms? *Cognition,* 100,73-99.

Tsang, C. D., Trainor, L. J. (2002) Spectral slope discrimination in infancy: Sensitivity to socially important timbres. *Infant Behavior and Development*, 25, 183-194

van Dinther. R., Patterson R. D. (2006) Perception of acoustic scale and size in musical instrument sounds. *Journal of the Acoustical Society of America*, 120(4), 2158-2176.

van Leeuwen, T., Been, P., Kuijpers, C., Zwarts, F., Maassen, B., van der Leij, A. (2006) Mismatch response is absent in 2-month-old infants at risk for dyslexia. *Neuroreport*, 17(4), 351-355.

Vanstone, A. D., Cuddy, L. L. (2010) Musical Memory in Alzheimer Disease. *Aging, Neuropsychology, and Cognition,* 17(1), 108-128.

van Zuijen, T., Sussman, E., Winkler, I., Näätänen, R., Tervaniemi, M. (2004) Grouping of sequential sounds--an event-related potential study comparing musicians and nonmusicians. *Journal of Cognitive Neuroscience*, 16, 331-338.

van Zuijen, T., Sussman, E., Winkler, I., Näätänen, R., Tervaniemi, M. (2005). Auditory organization of sound sequences by a temporal or numerical regularity – a mismatch negativity study sounds comparing musicians and non-musicians. *Cognitive Brain Research*, 23, 270–276.

van Zuijen, T .L., Simoens, V. L., Paavilainen, P., Näätänen, R., Tervaniemi, M. (2006) Implicit, Intuitive, and Explicit Knowledge of Abstract Regularities in a Sound Sequence: An Event-related Brain Potential Study. *Journal of Cognitive Neuroscience*, 18(8), 1292–1303.

Vaughan, H. G. Jr., Kurtzberg, D. (1992). Electrophysiologic indices of human brain maturation and cognitive development. In.: Gunnar, M. R., Nelson, C. A. (Eds.) *Minnesota Symposia on Child Psychology*, Erlbaum, Hilsdale, NJ, 1-36.

Vestergaard, M.D., Fyson, N.R.C., Patterson, R.D. (2009). The interaction of vocal characteristics and audibility in the recognition of concurrent syllables. *Journal of the Acoustical Society of America*, 125(2), 1114–1124.

Vuust, P., Pallesen, K. J., Bailey,C., van Zuijen, Gjedde, A., Roepstorff, A., Ostergaard, L. (2005) To musicians, the message is in the meter Pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians. *NeuroImage*, 24, 560–564.

Vuust, P., Ostergaard, L., Pallesen, K .J., Bailey,C., Roepstorff, A. (2009) Predictive coding of music–Brain responses to rhythmic incongruity. *Cortex*, 45, 80–92.

Wallin, N. L., Merker, B., Brown, S. (Editors) (2001) The Origins of Music. MIT Press, Cambridge, MA.

Warrier, C. M., Zatorre, R., J. (2002) Influence of tonal context and timbral variation on perception of pitch . *Perception & Psychophysics*, 64(2), 198-207.

Werner, L. A. (2007) Issues in human auditory development. *Journal of Communication Disorders*, 40, 275-283.

Winkler, I., Paavilainen, P., Alho, K., Reinikainen, K., Sams, M., Näätänen, R. (1990) The effect of small variation of the frequent auditory stimulus on the event-related brain potential to the infrequent stimulus. *Psychophysiology*, 27(2), 228-235.

Winkler, I., Karmos, G., Näätänen, R. (1996) Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain Research*, 742(1-2), 239-252.

Winkler, I., Czigler, I. (1998) Mismatch negativity: deviance detection or the maintenance of the 'standard'. *NeuroReport*, 9(17), 3809-3813.

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P. Lehtokoski, A., Czigler, I., Csépe, V., Ilmoniemi, R. J., Näätänen, R. (1999) Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, 36, 638–642.

Winkler, I., Kushnerenko, E., Horváth, J., Čeponiene, R., Fellman, V., Huotilainen, M., Näätänen, R., Sussman, E. (2003) Newborn infants can organize the auditory world. *Proceedings of the National Academy of Sciences*, 100(20), 11812-11815.

Winkler, I., (2007) Interpreting the Mismatch Negativity. *Journal of Psychophysiology* 21(3–4), 147–163.

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.

Wolff, C., Schröger, E., (2001) Human pre-attentive auditory change-detection with single, double, and triple deviations as revealed by mismatch negativity additivity. *Neuroscience Letters*, 311, 37-40

Wright, A. A., Rivera, J. J., Shyan, M., Hulse, S. H., Neiworth, J. J. (2000) Music perception and octave generalization in rhesus monkeys. *Journal of Experimental Psychology: General*, 129 (3), 291-307.

Yabe, H., Tervaniemi, M., Reinikainen, K., Näätänen, R. (1997) The temporal window of integration in auditory system as revealed by omission MMN. *NeuroReport*, 8, 1971-1974.

Zani, A., Proverbio, A. (2002) The Cognitive Electrophysiology of Mind and Brain. In.: Zani, A., Proverbio, A. (Eds.) *The Cognitive Electrophysiology of Mind and Brain*, Academic Press, San Diego/London.

Zatorre, R. J., Chen, J. L., Penhune, V. B. (2007) When the brain plays music: auditory–motor interactions in music perception and production. *Nature Reviews Neuroscience*, 8, 547-558.