

# Automatic Encoding of Polyphonic Melodies in Musicians and Nonmusicians

Takako Fujioka<sup>1,2</sup>, Laurel J. Trainor<sup>1,3</sup>, Bernhard Ross<sup>1</sup>,  
Ryusuke Kakigi<sup>2</sup>, and Christo Pantev<sup>4</sup>

## Abstract

■ In music, multiple musical objects often overlap in time. Western polyphonic music contains multiple simultaneous melodic lines (referred to as “voices”) of equal importance. Previous electrophysiological studies have shown that pitch changes in a single melody are automatically encoded in memory traces, as indexed by mismatch negativity (MMN) and its magnetic counterpart (MMNm), and that this encoding process is enhanced by musical experience. In the present study, we examined whether two simultaneous melodies in polyphonic music are represented as separate entities in the auditory memory trace. Musicians and untrained controls were tested in both magnetoencephalogram and behavioral sessions. Polyphonic stimuli were created by combining two melodies (A and B), each consisting of the same five notes but in a different order. Melody A was in the high voice and Melody B in the low voice in one condition, and this was reversed in

the other condition. On 50% of trials, a deviant final (5th) note was played either in the high or in the low voice, and it either went outside the key of the melody or remained within the key. These four deviations occurred with equal probability of 12.5% each. Clear MMNm was obtained for most changes in both groups, despite the 50% deviance level, with a larger amplitude in musicians than in controls. The response pattern was consistent across groups, with larger MMNm for deviants in the high voice than in the low voice, and larger MMNm for in-key than out-of-key changes, despite better behavioral performance for out-of-key changes. The results suggest that melodic information in each voice in polyphonic music is encoded in the sensory memory trace, that the higher voice is more salient than the lower, and that tonality may be processed primarily at cognitive stages subsequent to MMN generation. ■

## INTRODUCTION

Various elements of music often occur simultaneously. For example, when listening to orchestral music, we recognize melody, rhythm, and harmony, as well as an integrated flow of all of these aspects. Also we are able to listen selectively to different instruments such as violins, flutes, or trumpets. This process is different from the “cocktail party” situation of orienting attention to a single speaker, while ignoring the irrelevant voices and sounds. In contrast, with music we seem to maintain both selective and global listening.

The Western tonal musical system distinguishes two styles of music in terms of the roles of melody and harmony. One style is homophonic music, which combines a main melody with an explicitly accompanying harmonic structure in which the individual melodic lines are not discerned. The other style is polyphonic music, which contains multiple melodic lines (referred to as “voices”) of equal importance, usually separated in

pitch range and often played by different musical instruments. The harmony in polyphonic music is implied by the simultaneous notes in different melodic voices. Listening to one melody in polyphonic music does not exclude following the integrative harmony set up across the voices. In turn, listening to harmonic flow does not exclude tracking melodic information either. Thus, polyphonic music gives us a chance to study listening mechanisms unique to music.

Several psychological studies have investigated which factors affect the strategies of listening to multiple melodic objects. Dowling (1973) investigated the effects of pitch distance between two melodies. He presented two interleaved melodies (Melodies A and B, consisting of  $n$  distinct single notes  $A_i$  and  $B_i$  ( $i = 1 \dots n$ ,  $n = 16$ ), which are played as a sequence of notes  $A_1, B_1, A_2, B_2, \dots, A_n, B_n$ ) and examined recognition of a wrong note embedded in one of the two melody streams. The wrong note was more easily recognized if the melodies were further apart in pitch range. It is also easier to detect wrong notes if the melodies are more harmonically related (Sloboda & Edworthy, 1981). Several other studies have demonstrated that this detection is more robust for the higher than the lower melodic line, even in school-

<sup>1</sup>Baycrest Centre for Geriatric Care, Canada, <sup>2</sup>National Institute for Physiological Sciences, Japan, <sup>3</sup>McMaster University, Canada, <sup>4</sup>University of Münster, Germany

age children (Zenatti, 1969), whereas recognition of lower-pitched melodies can be better achieved by more experienced listeners (e.g., musicians) than naïve subjects (Crawley, Acker-Mills, Pastore, & Weil, 2002; Palmer & Holleran, 1994). Moreover, these studies have shown that listening task differences between attending to one melodic line versus attending to integrated harmony does not change the exceptional dominance of recognition for the higher over lower melody lines, although musicians perform better than nonmusicians in detecting changes in the low stream. Neuroimaging investigations concerning listening to multivoiced music have revealed that many cortical areas are commonly activated in selective listening and global listening conditions, suggesting that different listening strategies mainly cause different attentional loads but not completely different brain areas to be active (Janata, Tillmann, & Bharucha, 2002; Satoh, Takeda, Nagata, Hatazawa, & Kuzuhara, 2001). This corroborates the hypothesis of auditory scene analysis that parsing each line occurs preattentively regardless of musical experience by a preprocessor mechanism before focal attention is placed on the output of such a processor (Bregman, 1990).

To date, researchers have not investigated the preattentive processing of multiple simultaneous melodies. This level of automatic auditory perception, at which multiple sound sources are thought to be formed (Bregman, 1990), can be studied using the mismatch negativity (MMN) response in event-related potentials (ERPs), or its magnetic counter part, mismatch negativity magnetic fields (MMNm) in magnetoencephalogram (MEG). The MMN is a negative wave superimposed onto the mandatory auditory-evoked response, when infrequent deviations of a tone or tonal pattern occur within a sequence of otherwise repeatedly presented standard stimuli (Picton, Alain, Otten, Ritter, & Achim, 2000; Näätänen, 1992), whether or not attention is paid to the stimulus sequence. The MMN can be extracted by subtracting the average waveform of standard responses from that of deviant responses, thus canceling out the obligatory auditory responses common to standards and deviants. MMN is thought to reflect an auditory memory trace process that detects changes in the acoustical environment by comparing the incoming sound with a template of the standard sound previously stored. In general, the larger the size of deviation, the larger and earlier the MMN response that is elicited (Sams, Paavilainen, Alho, & Näätänen, 1985). Thus, the MMN gives objective indexes for auditory discrimination. The source of MMN is located mainly in the auditory cortex, likely accompanied by activation in the frontal lobe (Alain, Woods, & Knight, 1998; Alho, Woods, Algazi, Knight, & Näätänen, 1994; Giard, Perrin, Pernier, & Bouchet, 1990; Hari et al., 1984).

Recent studies have shown that the memory trace underlying the MMN response forms acoustic feature processors extracting invariance in acoustic context at

various levels of complexity. First, stimulus features such as frequency, intensity, duration, and spatial location of a sound appear to be processed in separate memory traces in parallel. If the stimulus sequence contains a single deviant varying in one of those categories, the MMN amplitude decreases accordingly to the probability of the deviants. In particular, when the same deviant stimulus is presented in a row, MMN is significantly reduced in response to the second deviant (Sams, Alho, & Näätänen, 1984). On the contrary, MMN is not attenuated when different features are varied independently within a stimulus block or presented sequentially (Deacon, Nousak, Pilotti, Ritter, & Yang, 1998; Nousak, Deacon, Ritter, & Vaughan, 1996). MMN was obtained to each of five different deviants which each occurred with 10% probability corresponding to a total deviance probability of 50% (Näätänen, Pakarinen, Rinne, & Takegata, 2004). Evidence for parallel feature processing in MMN is strengthened by studies showing that the source locations of MMN to different features are found slightly different from each other (Levänen, Ahonen, Hari, McEvoy, & Sams, 1996; Giard, Lavikainen, et al., 1995). Second, these features seem not only to be processed independently, but are also integrated in memory trace system as a gestalt-like representation at the same time. MMN was obtained in response to an infrequently presented stimulus produced by conjugating features from two separate standard stimuli (Paavilainen, Jaramillo, & Näätänen, 1998; Sussman, Gomes, Nousak, Ritter, & Vaughan, 1998; Gomes, Bernstein, Ritter, Vaughan, & Miller, 1997). The integrative function in the memory trace is also supported by results showing that the amplitude of MMN to a multiple-feature deviant sometimes approximates the summation of the responses to each single-feature deviation separately (Takegata, Paavilainen, Näätänen, & Winkler, 1999, 2001), but not always this summation works in a linear way (Paavilainen, Valppu, & Näätänen, 2001; Wolff & Schröger, 2001). Finally, when acoustic features are combined according to rules or patterns, the invariance of context is also indexed by MMN (Paavilainen, Simola, Jaramillo, Näätänen, & Winkler, 2001; Alain, Achim, & Woods, 1999; Alain, Cortese, & Picton, 1999; Paavilainen, Jaramillo, et al., 1998; Alain, Woods, & Ogawa, 1994). Of interest in the context of the present article is the fact that the representation of auditory features in sensory memory forms part of the neuronal basis of musical processing, and that MMN can be used to examine the presence of memory traces for aspects of complex melodic and harmonic stimuli.

Melodic processing is also indexed by MMN in terms of contour (up-down pattern of pitch) and interval (precise pitch distance) representations, as found in ERP (Trainor, McDonald, & Alain, 2002) and MEG (Fujioka, Trainor, Ross, Kakigi, & Pantev, 2004). MMN was present even in nonmusicians when the interval changes were embedded in melody transpositions from

trial to trial, indicating that melodies are recognized without absolute pitch cues by the relative pitch distances between notes. Furthermore, MMNm was larger in musicians than in nonmusicians, suggesting that long-term musical training enhances neural circuits for automatic melodic encoding.

One feature of musical structure that is of interest in the present study is that of tonality, the perceptual knowledge that certain notes or harmonic chords belong in a particular musical context whereas others do not. In language, the MMN stage of processing is involved in encoding acoustic features and language-specific phonemic categories, but not syntactic and semantic stages of processing (Näätänen, Lehtokoski, et al., 1997). Similarly, the violation of tonal expectancy has not been found to affect MMN, but rather later stages of processing. Conscious detection of tonal violation elicits responses such as P3b (Janata, 1995), late positive component (LPC) (Besson & Fäita, 1995), and P600 (Patel, Gibson, Ratner, Besson, & Holcomb, 1998), but Trainor et al. (2002) found no difference between MMN for in-key and out-of-key changes in the final note of single melodic stimuli. However, the tonal scheme may be more strongly implied by simultaneous notes in polyphonic music than in a single melody line. Indeed, harmonic syntax violation has been found to elicit an early preattentive component, early right anterior negativity [ERAN] (Koelsch, Schröger, & Gunter, 2002; Maess, Koelsch, Gunter, & Friederici, 2001). We therefore tested whether MMN is elicited preferentially for deviants that violate the rules of tonality over those that do not, and compared these results to behavioral tests of deviant detection in the two cases. Specifically, we compared performance on two-semitone change deviant notes that remained within the key of the melodies and one-semitone change deviant notes that went outside the key of the melodies. If MMN was larger for within-key changes, then MMN follows the size of the change regardless of the tonality implications. On the other hand, if MMN was larger for out-of-key changes, then tonality is encoded at the MMN stage of processing for polyphonic music.

The present study was aimed at investigating the neural representation of preattentive processing of simultaneous multiple melodies. Based on the observations introduced above, we hypothesized that (1) multiple melodic lines in polyphonic music are represented in parallel in auditory sensory memory; (2) the high voice might be more strongly encoded than the low voice; (3) tonal expectations for harmony may be represented in sensory memory; and (4) long-term musical training will enhance the encoding of polyphonic melodies. To test these hypotheses we recorded MMNm responses from musicians and nonmusicians using a modified oddball paradigm with 50% standard stimuli and four types of deviants, each occurring with 12.5% probability. Two melodies A and B comprised the poly-

phonic stimuli and were played simultaneously in the same key but in different pitch ranges (high vs. low voice) in two counterbalanced melody–voice combinations (High-A/Low-B and High-B/Low-A) (Figure 1). For each melody–voice combination, the deviations occurred at the final note in either the high or the low melody as either a within-key or as an out-of-key change. Both combinations of the two simultaneous melodies were musically well harmonized and dissonance was avoided. This design allowed us to assess whether the melodies were represented in the auditory memory trace. If the MMNm responses were totally absent for all deviants, we would conclude that the polyphonic music was encoded as a single Gestalt, and that all stimulus changes were processed as a single category that occurred on 50% of trials. This would imply that melodic information was not separately encoded in this level. On the other hand, the presence of MMNm would imply that separate melodic information was encoded at least to some extent. Specifically, according our hypotheses, we expected that MMNm would be: (1) elicited in both high and low voice changes despite 50% total deviations; (2) larger in high than in low voice deviants; (3) greater for out-of-key than within-key changes; and (4) enhanced in musicians compared with nonmusicians. Behavioral discrimination performance was tested after the MEG was recorded in order to compare automatic versus conscious discrimination.

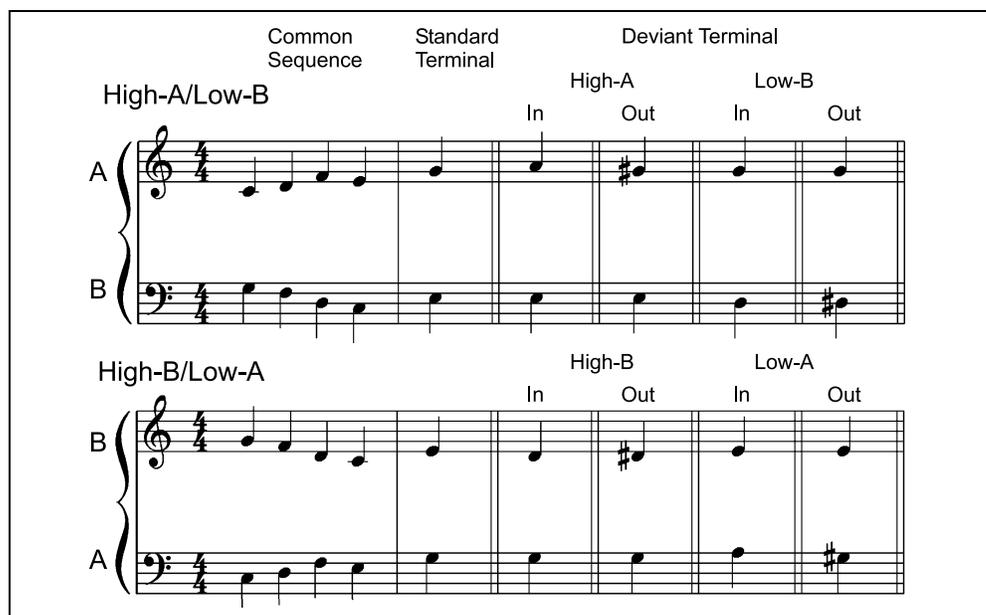
## RESULTS

Clearly pronounced auditory evoked responses to polyphonic melody stimuli were obtained from musicians and nonmusicians. Figure 2A shows an example of superimposed 151-channel magnetic-field waveforms of responses to standard and deviant stimuli, as well as difference waveforms obtained from the right hemisphere in a musician. Figure 2A shows clear P1–N1–P2m component patterns of the slow auditory evoked field peaking at 60, 100, and 180 msec after the onsets of each of the five notes of the stimulus. This triple-phased P1–N1–P2m pattern was largest in response to the first note and decreased in magnitude for the succeeding notes. The difference waveform (Deviants – Standards) showed no substantial peak during the first four notes of the melody (during which there were no deviants), but a pronounced MMNm deflection after the onset of the terminal sound. Individual source waveforms elicited by standard and deviant stimuli are shown in Figure 2B in conjunction with the corresponding difference waveforms.

### MMNm Waveform

The grand-averaged difference source waveforms in each hemisphere to all types of deviation are shown in Figure 3. In musicians, MMNm amplitudes greatly ex-

**Figure 1.** Polyphonic musical stimuli. Two different melodies (A and B) are played in high and low voice corresponding to two lines in musical notation. All melodies consist of five notes. In the High-A/Low-B case, the high voice is melody A and the low voice is melody B, while in the High-B/Low-A case, the melody-voice combination was reversed. The first four notes of a melody form a common sequence, followed by either a standard or a deviant terminal note. Each note has 300 msec duration, and the deviation occurs 1200 msec after the onset of the stimulus (first note). The deviant terminal notes occur either as in-key or as out-of-key changes for one of the melodies, while the other stays with the standard terminal note. Thus, eight types of deviant terminals exist for all melody-voice combinations (High-A/Low-B and High-B/Low-A) varying melody (A vs. B), voice (High vs. Low) and tonality (In vs. Out).



ceeded the noise level for all types of deviants in both left and right hemispheres. Nonmusicians showed smaller MMNm responses than musicians, but MMNm amplitudes were larger for changes in high than in low voices consistently across groups. No pronounced left/right laterality effects were seen in the waveforms in either group.

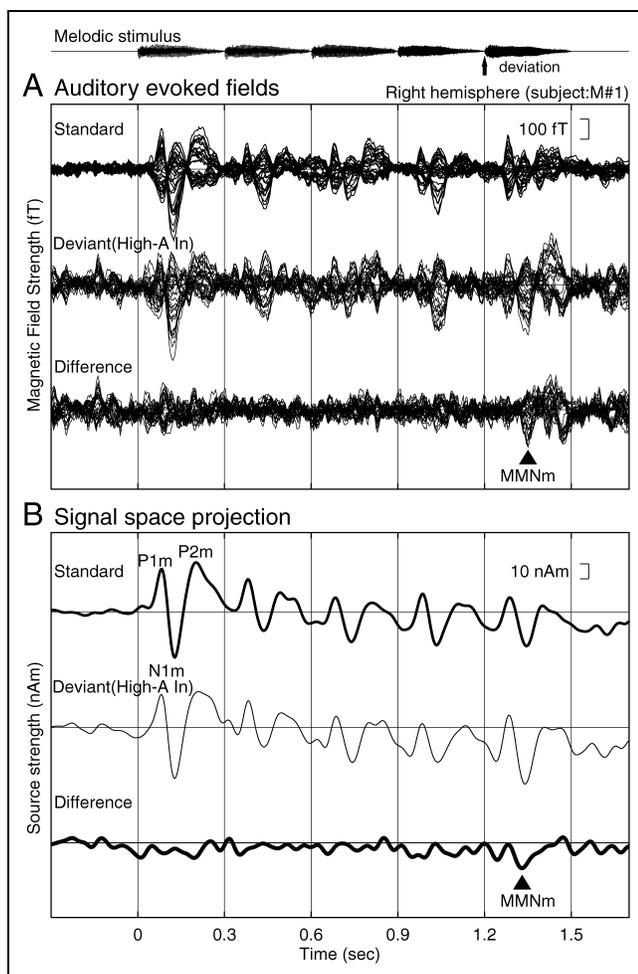
The peak latencies of the grand-averaged waveforms are listed in Table 1. Superimposed residual noise on individual MMNm waveforms resulted in ambiguous peak determination in some cases, and did not allow statistical analysis of peak latency. Nevertheless, Table 1 indicates that both groups exhibited a similar pattern of MMNm latency variation across types of deviants. For higher melody deviations, the latency of MMNm was as short as 130 to 185 msec (mean = 159 msec), whereas for deviation in the lower melody, the MMNm latencies varied between 150 and 230 msec (mean = 181 msec).

### MMNm Amplitudes

The amplitudes of MMNm responses were examined statistically by a repeated-measures analysis of variance (ANOVA) with one between-subject factor (group [musicians, nonmusicians]) and four within-subject factors (melody [A, B], voice [high, low], tonality [in-key, out-of-key], and hemisphere [left, right]). The individual

MMNm amplitudes were derived as a mean value in a time window of 40 msec around the peak latency of the group data. Significant main effects were found for group, voice, and tonality. Musicians showed larger responses than nonmusicians [ $F(1,18) = 6.3, p < .05$ ]. Deviants in the high voice produced larger MMNm than those in the low voice [ $F(1,18) = 9.6, p < .01$ ]. In-key changes (which involved larger pitch differences) produced larger MMNm than out-of-key changes [ $F(1,18) = 11.3, p < .01$ ]. The effects of melody and hemisphere were not significant.

The interaction between group and voice was significant [ $F(1,18) = 4.8, p < .05$ ] with a larger difference between MMNm to deviants in the higher than in the lower voice in musicians ( $p < .05$ ) than in nonmusicians (*ns*) (Figure 4A). There was also a significant triple interaction between tonality, hemisphere, and group [ $F(1,18) = 8.5, p < .01$ ]. This interaction was caused by different laterality between groups for the in-key change, although there was no main effect of hemisphere across groups. The effect of tonality was significant in both musicians ( $p < .05$ ) and nonmusicians ( $p < .05$ ). For the in-key changes, musicians had larger MMNm in the left hemisphere than in the right hemisphere, whereas nonmusicians had larger MMNm in the right hemisphere than in the left hemisphere (Figure 4B). The interaction between melody and voice



**Figure 2.** Waveforms of individual auditory evoked responses obtained from the right hemisphere in a musician. The top row shows the acoustic signal from one of the melody stimulus. (A) The three following rows show superimposed averaged magnetic field waveforms from 30 MEG channels over the right temporal region for the response to the standard, the response to the deviant, and the differences between both. (B) The three lower rows show the corresponding single waveforms of the dipole moment resulting from signal space projection.

was also significant [ $F(1,18) = 6.9, p < .05$ ] (Figure 4C) with greater differences between the high and low voice for the changes in Melody B ( $p < .01$ ) compared to changes in Melody A ( $ns$ ). In addition, the triple interaction across melody, voice, and hemisphere was significant [ $F(1,18) = 4.5, p < .05$ ]. High voice deviants in Melody A elicited larger MMNm in the right hemisphere than in the left hemisphere, whereas the laterality was opposite for high voice deviants in Melody B. However, the hemispheric difference in each condition did not reach significance by a post-hoc test.

### Behavioral Performance

Behavioral discrimination results are presented in Table 2. Across conditions, musicians performed around

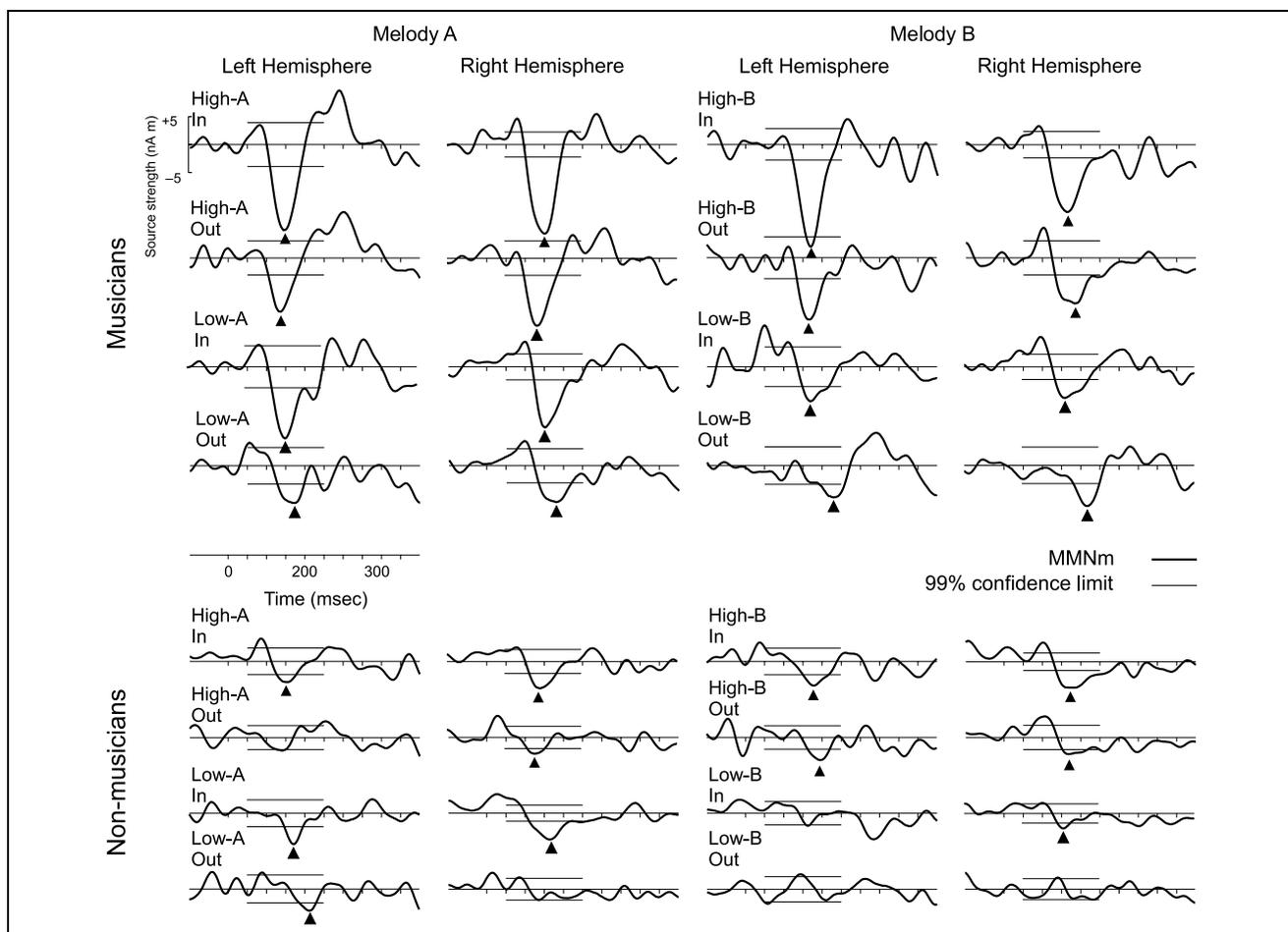
or above 80% correct, whereas nonmusicians were only above chance levels in 4 out of 12 conditions (indicated with the asterisk in Table 2). A four-way repeated-measures ANOVA parallel to that done with MMNm responses was conducted for the scores from the eight polyphonic two-melody conditions using the three within-subject factors of melody [A, B], voice [high, low], and tonality [in-key, out-of-key] and a between-subject factor of group [musicians, nonmusicians]. Group was significant [ $F(1,18) = 47.4, p < .0001$ ], indicating better performance for musicians. Tonality was also significant [ $F(1,18) = 16.9, p < .01$ ], reflecting better discrimination of out-of-key changes than in-key changes. Voice and melody were not significant. There was no interaction between any of the factors.

### Single versus Polyphonic Melody

All behavioral data from all 12 conditions (4 single-melody conditions and 8 two-melody conditions) were compared by another repeated-measures four-way ANOVA using the factors melody [A, B], tonality [in-key, out-of-key], and deviant location [single, high, low], and across group [musicians, nonmusicians]. Note that deviant location was investigated to assess whether there was any difference in recognition performance when the same melody (A or B) was presented alone [single], as the higher of two melodies [high], or as the lower of two melodies [low]. Group was highly significant [ $F(1,18) = 42.9, p < .0001$ ] with better scores in musicians than in nonmusicians (Figure 5). The three within-subject factors were all significant: performance was better on Melody A than Melody B [ $F(1,18) = 7.0, p < .05$ ], on out-of-key than within-key changes [ $F(1,18) = 12.0, p < .01$ ], and deviant location made a difference [ $F(1,18) = 5.7, p < .01$ ]. According to post-hoc tests, performance for both high and low deviant locations (i.e., two-melody conditions) was better than for the single deviant location (high vs. single;  $p < .01$ , low vs. single;  $p < .05$ ). Performance was better in the high compared to low deviant location, but this did not reach significance (mean  $\pm$  SEM; high 74.5%  $\pm$  2.3% low 72.4%  $\pm$  2.3%). The only interaction was between melody and deviant location [ $F(2,36) = 3.6, p < .05$ ], reflecting significantly better scores for Melody A compared to Melody B in the single-melody condition only ( $p < .01$ ).

### DISCUSSION

The present study revealed four main results. First, significant MMNm responses were observed in both musicians and nonmusicians for changes in both high and low voices despite the fact that there were changes on 50% of the trials. Second, MMNm was larger for deviants in the higher voice than in the lower voice.



**Figure 3.** The grand-averaged source waveforms of MMNm from the left and right hemisphere in musicians (top) and nonmusicians (bottom) obtained for each condition varying the factors of melody, voice, and tonality. The time scale refers to the onset of the 5th note. The thick lines represent the median of the MMNm responses across 10 subjects and the thin lines show the upper and lower limits of the 99% confidence interval for the estimated residual noise.

Third, MMNm was larger for in-key than out-of-key deviants despite better behavioral performance on the latter. Fourth, MMNm responses were larger in the musician group than in the nonmusician group, but the pattern of results was generally similar across groups.

### MMNm for Streams of Polyphonic Music

Polyphonic music is complex in that it contains separate melodies that are each pleasing on their own, but that also combine to form a Gestalt that makes sense harmonically. Furthermore, great musicians and composers such as J. S. Bach could improvise music in which both the separate parts and their combination made sense, implying that he could either think at multiple levels at the same time, or could switch attention between them very rapidly.

In fact, there is increasing evidence that memory traces contain multiple mental representations for multiple auditory streams. For example, perceptually segregated auditory streams induced by alternating tonal

patterns were reflected in MMN when the stimuli were separated in frequency (Winkler, Teder-Sälejärvi, Horvath, Näätänen, & Sussman, 2003; Yabe et al., 2001; Shinozaki et al., 2000; Sussman, Ritter, & Vaughan, 1999) or had location differences as an additional cue (Nager, Teder-Sälejärvi, Kunze, & Münte, 2003). Moreover, sensory memory can simultaneously hold two different standard tones (Winkler, Paavilainen, & Näätänen, 1992) or two tonal patterns presented equiprobably (Brattico, Winkler, Näätänen, Paavilainen, & Tervaniemi, 2002).

Multiple memory trace mechanisms were originally suggested for the different acoustic features of a single tone, such as intensity, frequency, duration, or location (Deacon et al., 1998; Nousak et al., 1996; Giard, Lavikainen, et al., 1995; Levänen, Hari, McEvoy, & Sams, 1993). Even in a sequence containing five types of deviants of 10% probability each, leaving only 50% of the trials as “pure” standards, MMN for each type of deviation was observed (Näätänen, Pakarinen, et al., 2004). These experiments support the view that multiple

**Table 1.** Latencies of MMNm Peaks in the Grand-Averaged Waveforms Measured with Respect to the Onset of 5th Note

	<i>Latency(ms)</i>							
	<i>Left Hemisphere</i>				<i>Right Hemisphere</i>			
	<i>High-A</i>		<i>High-B</i>		<i>High-A</i>		<i>High-B</i>	
	<i>in</i>	<i>out</i>	<i>in</i>	<i>out</i>	<i>in</i>	<i>out</i>	<i>in</i>	<i>out</i>
Musicians	144.0	134.4	169.6	166.4	156.8	134.4	166.4	185.6
Nonmusicians	150.4	(140.8)	179.2	195.2	150.4	128.8	172.8	169.6
	<i>Low-A</i>		<i>Low-B</i>		<i>Low-A</i>		<i>Low-B</i>	
	<i>in</i>	<i>out</i>	<i>in</i>	<i>out</i>	<i>in</i>	<i>out</i>	<i>in</i>	<i>out</i>
Musicians	156.8	172.8	169.6	230.4	150.4	179.2	163.2	220.8
Nonmusicians	172.8	(211.2)	(163.2)	(208.0)	166.4	168.0†	160.0	(208.0)†

The symbol † indicates double peaks in the response. In those cases the mean of the two peak latencies is given. The numbers in parentheses indicate that the corresponding MMNm amplitude did not exceed the 99% confidence limits.

memory traces for the various stimulus features can be established rather independently, and that deviation in one stimulus dimension does not interfere with deviation in another. On the other hand, there is also evidence that memory traces can encode different combinations of sound attributes by combining them into a single Gestalt (Sussman, Gomes, et al., 1998; Gomes et al., 1997). In addition, the presence of multiple streams seems to result in dividing memory trace resources or involving other processing, suggested by the delayed or reduced MMN in a multiple compared to single stream context (Nager et al., 2003; Shinozaki et al., 2000). Thus, it appears that the memory trace system as reflected by MMN both separates features of the incoming acoustic context and integrates over those features at the same time.

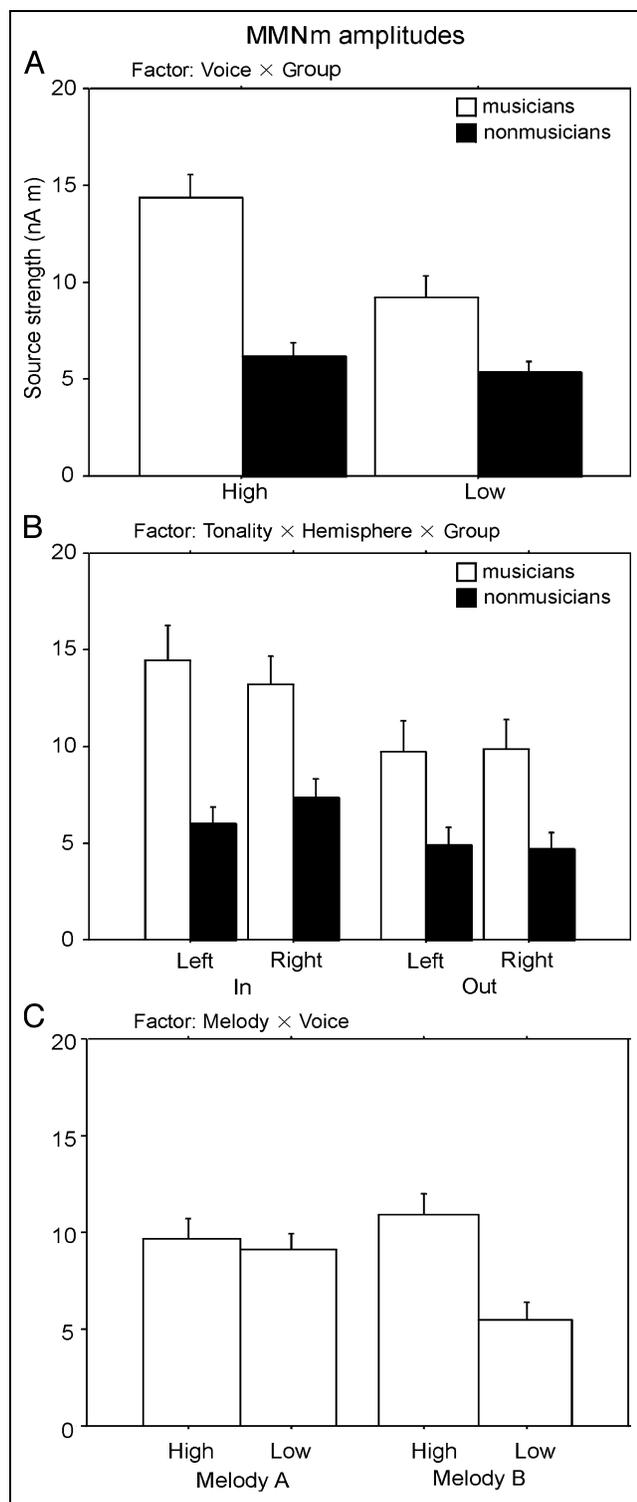
Similar to the Näätänen, Pakarinen, et al. (2004) study on sound features, we found significant MMN when there were 50% standard stimuli, suggesting that high and low melodies were encoded separately at least to some extent. Furthermore, the magnitude of the MMNm for the higher voice change was of similar size as that observed in our previous study, which demonstrated MMNm in response to changes (20% of trials) in the final note of a single five-note melody in both musicians and nonmusicians (Fujioka et al., 2004). The present and previous studies used the same criteria for selecting subject groups and the same MEG recording parameters and equipment.

At the same time, there is evidence in our results for interference between the two melodies suggestive of integrative Gestalt processing. In particular, MMNm was larger for changes in the same melody when it was in the higher voice than when it was in the lower voice. This implies that the two melodies interact to some extent.

We conclude that for polyphonic music, as with simple sound features, both separate melody traces and integrative processes are involved at the level of sensory memory.

### MMNm for In-Key versus Out-of-Key Changes

Both musicians and nonmusicians performed better behaviorally at detecting the out-of-key than the within-key changes, consistent with previous behavioral findings that both musicians and Western-aculturated nonmusicians process melodies according to an implicit knowledge of Western scale structure (Trainor & Trehub, 1994; Krumhansl, 1991). However, in both groups, MMNm was larger for the within-key than the out-of-key changes. Physically, the size of the within-key changes (2 semitones or 1/6 octave) was twice as large as that of the out-of-key changes (1 semitone or 1/12 octave). The fact that MMNm was larger in response to the former than to the latter suggests that MMN is primarily affected by the size of the change rather than its meaning in terms of musical scale and key, consistent with the findings of a previous ERP study (Trainor, et al., 2002). The literature examining MMN responses to frequency changes using pure tones has also consistently found larger responses for larger frequency deviations, regardless of tonality. For example, Scherg, Vajsar, and Picton (1989) showed larger MMN to 2000-Hz deviant pure tones embedded in 1000-Hz standard stimuli, compared with MMN to 1100-Hz deviant tones. Because the 2000-Hz tone is one octave higher in pitch than the 1000-Hz tone, it could be considered as an in-key change. The 1100-Hz change corresponds to between one and two semitones, creating an out-of-key and mistuned tonal sensation. Although the evidence to



**Figure 4.** MMNm amplitudes (mean + standard error of the mean [SEM]) for (A) high and low voice changes and (B) tonality changes and hemisphere in musicians and nonmusicians group, and (C) melody and voice across groups.

date suggests that MMN processes are not sensitive to tonality, further stages of processing are, as reflected in P3b and late positive components (Besson & Faïta, 1995; Janata, 1995).

## High Voice Advantage on MMNm

For both groups, larger and earlier MMNm responses were found for a deviation in the high-pitched melody regardless of which melody was played in high voice, although the effects were more pronounced in the musician than in the nonmusician group. In general, higher pitched instruments (e.g., violin) or voices (e.g., soprano) often play the role of a leading theme in music as written by composers and performed by players. The effect of voice on the MMNm is also in agreement with observations in behavioral studies that higher-pitched melodies are easily recognized in infants (Trehub & Trainor, 1998; Zenatti, 1969) and in both musically trained and untrained subjects (Crawley et al., 2002; Palmer & Holleran, 1994). Because the MMN is thought to increase and become early according to the salience of mental representation, our results indicate that the higher voice is already dominant in sensory memory. The perceptual dominance of higher voices is not likely a result of peripheral masking because the shape of the tuning functions, which facilitate the upward spread of masking (Zwicker & Fastl, 1999; Egan & Hake, 1950), would actually predict the opposite. It is also unlikely that learned top-down selective attention mechanisms, which can prompt perceptual sound grouping and increase the MMN amplitude in response to violation of grouping rules (Sussman, Winkler, Huotilainen, Ritter, & Näätänen, 2002; Sussman, Ritter, & Vaughan, 1998), is the main cause because of the pervasive dominance of higher voices across practiced and naïve listeners and across musical idioms. This enhancement of the MMN response to the high-pitched stream over the low-pitched stream has not been reported when using alternating tone patterns in different pitch ranges (Shinozaki et al., 2000). Because we used simultaneous melodies, this advantage might be specific to polyphonic contexts.

The high voice effect was more pronounced with Melody B than with Melody A across groups, as indicated by the significant interaction between melody and voice. This difference might be due to the direction of pitch deviation. Deviants in Melody A always went higher pitch than the standard terminal note, whereas deviants in Melody B always went lower. Gottselig, Brandeis, Hofer-Tinguely, Borbély, and Achermann (2004) found that when using a single melody, a deviant note that lowered the pitch produced larger MMNm than a deviant that raised the pitch. We found a similar effect for the higher voice, but not for the lower voice, suggesting that the effects of direction of pitch change depend on the voice.

## MMNm in Musicians versus Nonmusicians

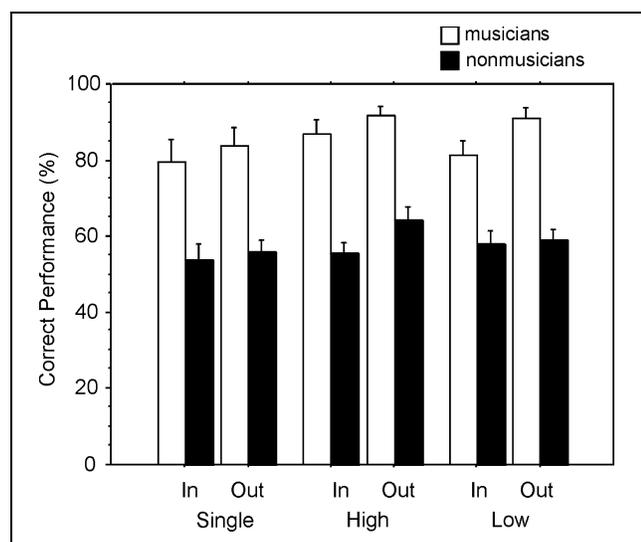
The MMNm responses in musicians were significantly larger than those in nonmusicians. This is in line with the accumulating literature showing superior audi-

**Table 2.** Correct Performance (Mean  $\pm$  Standard Error of the Mean [SEM]) Expressed as Percentage of Correct Answers in the Total Number of Trials of the Two Alternatives Forced Choice Test (2AFC)

	Correct performance (%)			
	Single-A		Single-B	
	<i>in</i>	<i>out</i>	<i>in</i>	<i>out</i>
Musicians	88.0 $\pm$ 4.3***	94.8 $\pm$ 2.7***	70.8 $\pm$ 10.5	72.4 $\pm$ 8.0*
Nonmusicians	58.4 $\pm$ 7.1	62.0 $\pm$ 4.5*	49.2 $\pm$ 4.8	50.0 $\pm$ 3.3
	High-A		High-B	
	<i>in</i>	<i>out</i>	<i>in</i>	<i>out</i>
	Musicians	91.6 $\pm$ 3.4***	94.0 $\pm$ 3.2***	82.0 $\pm$ 7.0**
Nonmusicians	54.8 $\pm$ 2.8	66.4 $\pm$ 3.9*	56.0 $\pm$ 5.6	62.4 $\pm$ 5.6
	Low-A		Low-B	
	<i>in</i>	<i>out</i>	<i>in</i>	<i>out</i>
	Musicians	83.2 $\pm$ 4.9***	89.2 $\pm$ 4.4***	79.6 $\pm$ 5.6**
Nonmusicians	59.6 $\pm$ 3.5*	63.2 $\pm$ 3.8**	56.4 $\pm$ 6.2	54.8 $\pm$ 4.3

The group scores were statistically examined by *t*-tests whether the performance was above chance level (50%). The significance level is indicated by the asterisk (\**p* < .05, \*\**p* < .01, and \*\*\**p* < .0001).

tory preattentive processing in musicians as revealed with the MMN paradigm (Russeler, Altenmüller, Nager, Kohlmetz, & Münte, 2001; Tervaniemi, Rytönen, Schröger, Ilmoniemi, & Näätänen, 2001; Koelsch, Schröger, & Tervaniemi, 1999). Also, enhanced cortical representations in musicians have been observed in the auditory modality (Pantev et al., 1995) and the



**Figure 5.** Behavioral performance in both groups (mean  $\pm$  standard error of the mean [SEM]) indicated separately for single-melody conditions (Single-A and Single-B), and both high and low voices in the two-melody conditions (High-A, High-B, Low-A, and Low-B).

sensory-motor modality (Elbert, Pantev, Wienbruch, Rockstroh, & Taub, 1995), as well as in the cross-modal interaction between them (Schulz, Ross, & Pantev, 2003). These enhanced neural activities in musicians are assumed to result from cortical reorganization due to long-term training.

Significant but minor differences in lateralization between groups were found for tonal violation, although the MMNm responses did not show a general lateralization effect. In musicians, in-key changes showed a left hemispheric advantage, whereas in nonmusicians this was right-lateralized. Considering that the previous single-melody MMN studies did not find any laterality effects in either group (Fujioka et al., 2004; Trainor et al., 2002), the present result with polyphonic music could be related to Gestalt harmonic processing. The response pattern in nonmusicians seems to be consistent with MEG and PET studies showing a right hemispheric advantage in nonmusicians for chords changes to different keys (Tervaniemi, Medvedev, et al., 2000; Tervaniemi, Kujala, et al., 1999). On the other hand, there has been no specific evidence for the left hemispheric advantages in musicians' MMN with respect to tonal violation processing. In general, hemispheric asymmetry in musicians has been demonstrated to be left-lateralized in behavior (Burton, Morton, & Abbess, 1989; Johnson, 1977; Bever & Chiarello, 1974), alpha rhythm in electroencephalogram (EEG) (Hirshkowitz, Earle, & Paley, 1978), in cerebral blood flow velocity (Evers, Dannert, Rödding, Rotter, & Ringelstein, 1999), and

anatomy (Schlaug, Jäncke, Huang, & Steinmetz, 1995), but with contradicting reports (Vollmer-Haase, Finke, Hartje, & Bulla-Hellwig, 1998; Messerli, Pegna, & Sordet, 1995; Gordon, 1970). Neural correlates of various musical aspects appear to be widely distributed in the left and right cerebral and cerebellar hemispheres (Parsons, 2001). Given that our lateralization effect was quite small, and the conflicting reports in the literature, it is difficult to draw definitive conclusions about laterality for sensory memory traces for musical stimuli without further research.

### Behavior and MMNm

In general, MMNm amplitude followed behavioral performance, with musicians showing both larger MMNm and superior behavioral discrimination compared to nonmusicians. Interestingly, MMNm was a more sensitive measure than behavior in that nonmusicians showed significant MMNm under conditions where they were at chance behaviorally. This is consistent with reports from training studies in which MMNm responses develop before behavioral discrimination is achieved (Dalebout & Stack, 1999; Tremblay, Kraus, & McGee, 1998), and suggests that the processes underlying MMN are necessary for conscious discrimination, but not sufficient. It is also evident in our data that there are stages of processing for tonality beyond the level of MMN. First, behavioral performance at detecting changes in one of the melodies was superior when both melodies were played simultaneously compared to when that melody was played alone, but MMNm responses in the two-melody case in the present experiment were similar to MMNm responses in a one-melody case from a previous study in our laboratory using the same subject pool, equipment, and procedures (Fujioka et al., 2004). This suggests that by the time behavioral responses are executed, the auditory system has made use of the richer harmonic tonality information in the two-melody case, but that this is not the case at the MMN stage of processing. Second, MMNm responses were larger for larger pitch changes, but not for key-violating compared to key-consistent changes. The opposite was true for behavior, with superior detection of smaller out-of-key changes compared to larger in-key changes. Again, this suggests that tone patterns are encoded at the level of MMN, but that the tonal implications of these melodies continue to be processed at higher stages of analysis.

### Conclusions

In the present study, automatic auditory processing of two-melody polyphonic compositions was investigated in musicians and nonmusicians by comparing MMNm responses and behavioral discrimination. With respect to the four hypotheses outlined in the introduction, we

conclude: (1) that multiple melodic lines in polyphonic music are represented in parallel in auditory sensory memory; (2) that the memory trace for the higher voice is more salient than that for the lower voice, as reflected by larger and early MMNm; (3) that tonal harmony is not robustly represented in the sensory memory traces, but is processed primarily at a subsequent stage of processing; and (4) that long-term musical training leads to both enhanced sensory memory encoding and superior behavior discrimination.

## METHODS

### Subjects

Ten musicians (5 women) between 20 and 35 years of age and 10 nonmusically trained adults (4 women) between 23 and 34 years of age participated in this study. The musicians had studied more than one instrument and practiced regularly for more than 10 years with formal education including musical schools or private lessons. The nonmusicians had almost no formal musical training, except in their regular school lessons. None of the subjects in either group had absolute pitch perception. All participants were right-handed as assessed by the Edinburgh handedness test and had normal hearing within the range of 250 to 8000 Hz as tested by clinical audiometry. The subjects gave informed consent to participate after they were completely informed about the nature of the study. The Ethics Commission of the Baycrest Centre for Geriatric Care approved all experimental procedures, which are in accordance with the Declaration of Helsinki.

### Stimuli

The scores for the musical stimuli are schematically depicted in Figure 1. Two five-note Melodies A and B were composed using the first five diatonic scale notes in a major scale of the western tonal musical system (i.e., C, D, E, F, and G in C major key). Melody A was a sequence “C-D-F-E-G” whereas Melody B was defined as “G-F-D-C-E.” Then Melodies A and B were combined in parallel but presented in two pitch ranges as the high voice (C5–G5) and the low voice (C4–G4), respectively (American notation). Combinations of the two factors pitch range (high and low voice) and melody were varied resulting in two polyphonic variations of high-A and low-B (High-A/Low-B), and high-B and low-A conditions (High-B/Low-A).

For both melody–voice combinations (High-A/Low-B and High-B/Low-A), the sequences of stimuli were presented as a modified oddball paradigm. Fifty percent of the sequences consisted of combinations of standard A and B melodies. For all deviant trials, either the high or the low voice was altered into a deviant terminal note of either in-key or out-of-key change, whereas the other

voice remained as the standard melody. In-key deviations were created by shifting the original terminal by a whole note (1/6th of an octave = 2 semitones), and out-of-key changes by a semitone shift (1/12th octave) keeping the up-down pitch contour direction of the standard terminal note. Note that the deviant notes in Melody A always went up and those in Melody B always went down from the standard terminal regardless of voice and tonality. Each oddball sequence contained four types of equal probable (12.5%) deviants varying voice and tonality (High-In, High-Out, Low-In, and Low-Out). From one trial to the next, the combined melodies were transposed to one of eight keys sequenced in the order “C-E-C#-F-D-F#-D#-G” to avoid priming effects brought up by repeating notes in successive trials. Although Melodies A and B used the same group of five notes for standards, they had different notes at each time point in the melodies. As well, when Melodies A and B were presented simultaneously, the harmony at each beat was always musically consonant even when one of the two melodies was altered to a deviant.

The sound files were created from digitally recorded piano timbres for each note in sampling rate 44,100 Hz. The duration of each note was 300 msec for a total melody length of 1500 msec. Succeeding melodies were separated by 900 msec silent intervals. There were no sequential deviants occurring within a same voice, although deviants in different voice could occur successively. In total, 1000 trials were presented in an experimental condition, which was divided into two blocks of 500 trials. The total measuring duration for two polyphonic conditions with two blocks each was 80 min.

Individual hearing thresholds were determined for the left and right ears of each subject for a stimulus consisting of G4 and B5 (major 10th interval), which is the median of the eight transposed variations of the standard terminal in the High-A/Low-B sequence. This interval also covers the range of the terminal chords of the High-B/Low-A sequence (major 6th interval). All stimuli were presented at 60 dB above the obtained threshold in each ear.

### MEG Recordings

The magnetic field responses were recorded with a 151-channel whole-cortex magnetometer system (OMEGA, CTF Systems, Port Coquitlam, Canada). The MEG pickup coils of 2 cm in diameter and 3.1 cm intercoil distances are configured as first-order axial gradiometers with 5-cm baseline. The MEG signals were low-pass filtered at 100 Hz and sampled at a rate of 312.5 sec<sup>-1</sup>. For all conditions, the duration of a recording epoch was 2.35 sec including a 0.4-sec prestimulus interval. A trigger-signal corresponding to the onset of the first note of each melody synchronized the stimulus presentation and the data acquisition.

The recordings were performed in seated position in a magnetically shielded room. The subjects were instructed to stay awake and that no specific attention to the stimuli was required. No explanation about the various stimuli was provided. The subjects watched a soundless movie of their own choice, which was projected onto a screen placed in front of the chair. The subject's compliance was verified by video monitoring. The order of experimental conditions was counterbalanced between participants.

### Data Analysis

The recorded magnetic field data were averaged separately for the standard and deviant stimuli for all stimulus types. If the magnetic field amplitude in a channel located above the eyes exceeded 1.0 pT during the latency interval from -0.2 to 1.5 sec, the epoch of data was rejected as artifact contaminated. The average percentage of accepted trials after the artifact rejection was 77% of the recorded epochs without substantial difference between groups and condition (370 epochs for standard and 95 for each deviant in musicians; 380 for standard and 102 for each deviant in nonmusicians).

The analysis technique of signal space projection (SSP) (Tesche et al., 1995) was applied to the MEG data, which collapsed the multichannel magnetic field data into a single time series of magnetic dipole moments. The weighting factor for each MEG channel was the sensitivity of the corresponding sensor to a source at the specified location in the brain. This formed a virtual sensor, which was maximally sensitive to a source at the specified origin and orientation and less sensitive to other sources. This resulted in considerable discrimination against the sensor noise and uncorrelated brain activity from distant brain regions. The SSP is a useful method under the assumption of a single time-varying source at a fixed location. Estimating the coordinates and orientation of the cortical sources is a necessary prerequisite for the method. Although we were interested in source waveforms of the MMNm, we chose the N1m source coordinates for SSP. N1m source localization was performed using an equivalent current dipole (ECD) model based on the averaged evoked response to the onset of the initial tones of the standard polyphonic melodic stimuli within each block. This dipole estimation was conducted using all 151-channel data from MEG. We chose to base the source localization on N1m rather than MMNm for the following reasons. First, N1m responses were larger than MMNm responses and resulted in reliable source estimation in all subjects. Second, it is widely accepted that the MMN is mainly generated in the auditory cortex (Giard, Lavikainen, et al., 1995; Tiitinen et al., 1993; Giard, Perrin, et al., 1990; Scherg et al., 1989; Hari et al., 1984). Although MMN and N1 differ functionally in response to various experimental parameters (e.g., sound intensity, stimula-

tion rate), the dissociation of source location between MMN and N1 components remains controversial. On one hand, there have been reports that the MMN is located slightly (5–10 mm) more anterior or medial to N1 (Tiitinen et al., 1993; Csépe, Pantev, Hoke, Hampson, & Ross, 1992; Sams, Kaukoranta, Hämäläinen, & Näätänen, 1991). On the other hand, recent studies suggest that the two responses may be generated from the same neurons that are sensitive to probability of novel sound (Jääskeläinen et al., 2004; Ulanovsky, Las, & Nelken, 2003). Third, there are also reports demonstrating that the MMN sources for different feature processing are located in slightly different regions within the auditory cortex (Levänen, Ahonen, et al., 1996; Giard, Lavikainen, et al., 1995) and overlapped for the combination of those features (Takegata, Huotilainen, Rinne, Näätänen, & Winkler, 2001). Because we potentially sought more than one MMNm response, there may not exist one location that is optimal for measuring all of our MMN responses. In sum, given the probable localization error that would be introduced by using MMN in conditions in which it was very small and difficult to measure in individual subjects, and given the similarity of MMN and N1 localizations, we reasoned that using N1 to locate the ECD from which we would calculate the SSP would result in the best signal-to-noise ratio.

For each subject, the average of N1m dipole locations and orientations across all stimulus conditions in separate blocks served as an individual estimate for the source in the auditory cortex. Based on these source coordinates, the dipole moment waveforms over the whole stimulus-related epochs were calculated for all stimulus conditions. Another important advantage of the method is that it allows the averaging of source waveforms from repeated measurements in the same subject or between subjects. Grand-averaged source waveforms across both groups of subjects were obtained selectively for the standard and deviant stimuli. Individual difference waveforms were calculated by subtracting the response to the standard from that to the deviant stimuli. MMNm responses were examined after the onset of the fifth note for each type of deviant in the two melody–voice conditions. The baselines of all responses were adjusted to the mean in a 100-msec interval previous to the onset of the deviation. The 99% confidence intervals for the grand-averaged response waveforms and the difference waveforms were estimated from nonparametric bootstrap resampling analysis (Davison & Hinkley, 1997). This method empirically establishes the distribution of the mean from repeated samples of the data itself and allows estimating confidence limits without the assumption of the underlying distribution. This analysis was applied to all data points of the difference waveforms and allowed identifying those time intervals with amplitudes significantly different from zero.

The amplitudes and latencies of the MMNm peaks were measured in 32 grand-averaged difference waveforms corresponding to five experimental parameters (group: musicians, nonmusicians; melody: A, B; voice: high, low; tonality: in-key and out-of-key; hemisphere: left and right). The single subject's MMNm amplitude was defined as the mean value of the waveforms within a 40-msec time interval centered at the mean peak latency. This procedure was necessary because the identification of peak latency and amplitude in the individual data was not always feasible. The MMNm amplitudes were statistically examined by a repeated-measures ANOVA with one between-subjects factor (group) and four within-subjects factors (melody, voice, tonality, and hemisphere). The post-hoc comparison was calculated with Fisher's PLSD tests using the 5% level of significance.

### Behavioral Test

After the MEG recordings, all subjects participated in 12 two-alternative forced-choice (2AFC) tasks, in which 25 trials were presented consisting of two melodic sequences on the basis of the same melodic material as used in the MEG recordings. Four tasks were single-melody conditions (two using Melody A and two Melody B; and two have in-key and two out-of-key changes). The other eight tasks were polyphonic conditions with all possible combination of three factors of melody (A, B), voice (high, low), and tonality (in-key, out-of-key).

The subjects were instructed to listen to a pair of melodic stimulus in each trial, and judge whether both stimuli were similar or different regardless of transposition. The first stimulus of each trial was chosen from the set of standards stimuli, whereas the other stimulus was either another standard stimulus (similar) or a deviant stimulus (different). The melodies were presented in the same order of transposition as in the MEG recordings. The same and different pairs occurred with equal probability. The presentation of stimuli and the recording of the subject's responses were controlled by specially developed software on a desktop computer. The stimuli were presented at an intensity of about 60 dB SL through headphones and the subjects responded by a mouse click on buttons shown on the computer monitor. The silent interval between the first and the second melodies was 900 msec. The next trial started after the subject's response. The subjects were instructed in detail about the tasks and briefly trained by a few trials with feedback until the task was correctly understood before single-melody tasks and polyphonic two-melody tasks. In the actual testing condition, no feedback was provided. All 12 blocks were tested in a randomized order for each subject, taking about 1 hour testing time.

Behavioral data were examined statistically in two separated steps. First, the data for the eight two-melody tasks were examined by the same repeated-measures

four-way ANOVA as used for the MEG results, with one between-subjects factor (group) and three within-subjects factors (melody, voice, tonality). The post-hoc comparisons were done by Fisher's PLSD tests using the 5% level of significance. Second, the data from all conditions were tested by a repeated-measures four-way ANOVA with three within-subjects factors, tonality (in-key, out-of-key), deviant location (single, high, low), and melody (A, B) and a between-subjects factor of groups (musicians, nonmusicians). Fisher's PLSD post-hoc comparisons were conducted where appropriate.

## Acknowledgments

This research was supported by the International Music Foundation, CA, USA. We thank Dr. Terence W. Picton and Dr. Claude Alain for helpful comments throughout the study.

Reprint requests should be sent to Takako Fujioka, PhD, The Rotman Research Institute, 3560 Bathurst Street, Toronto, Ontario, M6A 2E1, Canada, or via e-mail: tfujioka@rotman-baycrest.on.ca.

## REFERENCES

- Alain, C., Achim, A., & Woods, D. L. (1999). Separate memory-related processing for auditory frequency and patterns. *Psychophysiology*, *36*, 737–744.
- Alain, C., Cortese, F., & Picton, T. W. (1999). Event-related brain activity associated with auditory pattern processing. *NeuroReport*, *10*, 2429–2434.
- Alain, C., Woods, D. L., & Knight, R. T. (1998). A distributed cortical network for auditory sensory memory in humans. *Brain Research*, *812*, 23–37.
- Alain, C., Woods, D. L., & Ogawa, K. H. (1994). Brain indices of automatic pattern processing. *NeuroReport*, *6*, 140–144.
- Alho, K., Woods, D. L., Algazi, A., Knight, R. T., & Näätänen, R. (1994). Lesions of frontal cortex diminish the auditory mismatch negativity. *Electroencephalography and Clinical Neurophysiology*, *91*, 353–362.
- Besson, M., & Faita, F. (1995). An event-related potential (ERP) study of musical expectancy: Comparisons of musicians with non-musicians. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1278–1296.
- Bever, T. G., & Chiarello, R. J. (1974). Cerebral dominance in musicians and nonmusicians. *Science*, *185*, 537–539.
- Brattico, E., Winkler, I., Näätänen, R., Paavilainen, P., & Tervaniemi, M. (2002). Simultaneous storage of two complex temporal sound patterns in auditory sensory memory. *NeuroReport*, *13*, 1747–1751.
- Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge: MIT Press.
- Burton, A., Morton, N., & Abbess, S. (1989). Mode of processing and hemisphere differences in the judgement of musical stimuli. *British Journal of Psychology*, *80*, 169–180.
- Crawley, E. J., Acker-Mills, B. E., Pastore, R. E., & Weil, S. (2002). Change detection in multi-voice music: The role of musical structure, musical training, and task demands. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 367–378.
- Csépe, V., Pantev, C., Hoke, M., Hampson, S., & Ross, B. (1992). Evoked magnetic responses of the human auditory cortex to minor pitch changes: Localization of the mismatch field. *Electroencephalography and Clinical Neurophysiology*, *84*, 538–548.
- Dalebout, S. D., & Stack, J. W. (1999). Mismatch negativity to acoustic differences not differentiated behaviorally. *Journal of the American Academy of Audiology*, *10*, 388–399.
- Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge: Cambridge University Press.
- Deacon, D., Noursak, J. M., Pilotti, M., Ritter, W., & Yang, C. M. (1998). Automatic change detection: Does the auditory system use representations of individual stimulus features or gestalts? *Psychophysiology*, *35*, 413–419.
- Dowling, W. J. (1973). The perception of interleaved melodies. *Cognitive Psychology*, *5*, 322–337.
- Egan, J. P., & Hake, H. W. (1950). On the masking pattern of a simple auditory stimulus. *Journal of the Acoustical Society of America*, *22*, 622–630.
- Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., & Taub, E. (1995). Increased cortical representation of the fingers of the left hand in string players. *Science*, *270*, 305–307.
- Evers, S., Dannert, J., Rödding, D., Rotter, G., & Ringelstein, E. B. (1999). The cerebral haemodynamics of music perception. A transcranial Doppler sonography study. *Brain*, *122*, 75–85.
- 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.
- Giard, M. H., Lavikainen, J., Reinikainen, K., Bertrand, O., Pernier, J., & Näätänen, R. (1995). Separate representation of stimulus frequency, intensity, and duration in auditory sensory memory: An event-related potential and dipole-model study. *Journal of Cognitive Neuroscience*, *7*, 133–143.
- Giard, M. H., Perrin, F., Pernier, J., & Bouchet, P. (1990). Brain generators implicated in the processing of auditory stimulus deviance: A topographic event-related potential study. *Psychophysiology*, *27*, 627–640.
- Gomes, H., Bernstein, R., Ritter, W., Vaughan, H. G., Jr., & Miller, J. (1997). Storage of feature conjunctions in transient auditory memory. *Psychophysiology*, *34*, 712–716.
- Gordon, H. W. (1970). Hemispheric asymmetries in the perception of musical chords. *Cortex*, *6*, 387–398.
- Gottselig, J. M., Brandeis, D., Hofer-Tinguely, G., Borbély, A. A., & Achermann, P. (2004). Human central auditory plasticity associated with tone sequence learning. *Learning and Memory*, *11*, 162–171.
- 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*, 127–132.
- Hirshkowitz, M., Earle, J., & Paley, B. (1978). EEG alpha asymmetry in musicians and non-musicians: A study of hemispheric specialization. *Neuropsychologia*, *16*, 125–128.
- Jääskeläinen, I. P., Ahveninen, J., Bonmassar, G., Dale, A. M., Ilmoniemi, R. J., Levänen, S., Lin, F. H., May, P., Melcher, J., Stufflebeam, S., Tiitinen, H., & Belliveau, J. W. (2004). Human posterior auditory cortex gates novel sounds to consciousness. *Proceedings of the National Academy of Sciences, U.S.A.*, *101*, 6809–6814.
- Janata, P. (1995). ERP measures assay the degree of expectancy violation of harmonic contexts in music. *Journal of Cognitive Neuroscience*, *7*, 153–164.
- Janata, P., Tillmann, B., & Bharucha, J. J. (2002). Listening to polyphonic music recruits domain-general attention and working memory circuits. *Cognitive, Affective, & Behavioral Neuroscience*, *2*, 121–140.
- Johnson, P. R. (1977). Dichotically-stimulated ear differences in musicians and nonmusicians. *Cortex*, *13*, 385–389.

- Koelsch, S., Schröger, E., & Gunter, T. C. (2002). Music matters: Preattentive musicality of the human brain. *Psychophysiology*, *39*, 38–48.
- Koelsch, S., Schröger, E., & Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *NeuroReport*, *10*, 1309–1313.
- Krumhansl, C. L. (1991). Music psychology: Tonal structures in perception and memory. *Annual Review of Psychology*, *42*, 277–303.
- Levänen, S., Ahonen, A., Hari, R., McEvoy, L., & Sams, M. (1996). Deviant auditory stimuli activate human left and right auditory cortex differently. *Cerebral Cortex*, *6*, 288–296.
- Levänen, S., Hari, R., McEvoy, L., & Sams, M. (1993). Responses of the human auditory cortex to changes in one vs. two stimulus features. *Experimental Brain Research*, *97*, 177–183.
- Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: An MEG study. *Nature Neuroscience*, *4*, 540–545.
- Messerli, P., Pegna, A., & Sordet, N. (1995). Hemispheric dominance for melody recognition in musicians and non-musicians. *Neuropsychologia*, *33*, 395–405.
- Näätänen, R. (1992). *Attention and brain function*. Hillsdale, NJ: Erlbaum.
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, *385*, 432–434.
- Näätänen, R., Pakarinen, S., Rinne, T., & Takegata, R. (2004). The mismatch negativity (MMN): Towards the optimal paradigm. *Clinical Neurophysiology*, *115*, 140–144.
- Nager, W., Teder-Sälejärvi, W., Kunze, S., & Münte, T. F. (2003). Preattentive evaluation of multiple perceptual streams in human audition. *NeuroReport*, *6*, 871–874.
- Nousak, J. M., Deacon, D., Ritter, W., & Vaughan, H. G. J. (1996). Storage of information in transient auditory memory. *Brain Research, Cognitive Brain Research*, *4*, 305–317.
- Paavilainen, P., Jaramillo, M., & Näätänen, R. (1998). Binaural information can converge in abstract memory traces. *Psychophysiology*, *35*, 483–487.
- Paavilainen, P., Simola, J., Jaramillo, M., Näätänen, R., & Winkler, I. (2001). Preattentive extraction of abstract feature conjunctions from auditory stimulation as reflected by the mismatch negativity (MMN). *Psychophysiology*, *38*, 359–365.
- Paavilainen, 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 approximately 2 but 1+1+1<3. *Neuroscience Letters*, *301*, 179–182.
- Palmer, C., & Holleran, S. (1994). Harmonic, melodic, and frequency height influences in the perception of multivoiced music. *Perception & Psychophysics*, *56*, 301–312.
- Pantev, C., Bertrand, O., Eulitz, C., Verkindt, C., Hampson, S., Schuierer, G., & Elbert, T. (1995). Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. *Electroencephalography and Clinical Neurophysiology*, *94*, 26–40.
- Parsons, L. M. (2001). Exploring the functional neuroanatomy of music performance, perception, and comprehension. *Annals of the New York Academy of Sciences*, *930*, 211–231.
- Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, *10*, 717–733.
- Picton, T. W., Alain, C., Otten, L., Ritter, W., & Achim, A. (2000). Mismatch negativity: Different water in the same river. *Audiology & Neuro-Otology*, *5*, 111–139.
- Russeler, J., Altenmüller, E., Nager, W., Kohlmetz, C., & Münte, T. F. (2001). Event-related brain potentials to sound omissions differ in musicians and non-musicians. *Neuroscience Letters*, *308*, 33–36.
- Sams, M., Alho, K., & Näätänen, R. (1984). Short-term habituation and dishabituation of the mismatch negativity of the ERP. *Psychophysiology*, *21*, 434–441.
- Sams, M., Kaukoranta, E., Hämäläinen, M., & Näätänen, R. (1991). Cortical activity elicited by changes in auditory stimuli: Different sources for the magnetic N100m and mismatch responses. *Psychophysiology*, *28*, 21–29.
- Sams, M., Paavilainen, P., Alho, K., & Näätänen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalography and Clinical Neurophysiology*, *62*, 437–448.
- Satoh, M., Takeda, K., Nagata, K., Hatazawa, J., & Kuzuhara, S. (2001). Activated brain regions in musicians during an ensemble: A PET study. *Brain Research, Cognitive Brain Research*, *12*, 101–108.
- 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.
- Schlaug, G., Jäncke, L., Huang, Y., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science*, *267*, 699–701.
- Schulz, M., Ross, B., & Pantev, C. (2003). Evidence for training-induced crossmodal reorganization of cortical functions in trumpet players. *NeuroReport*, *14*, 157–161.
- Shinozaki, N., Yabe, H., Sato, Y., Sutoh, T., Hiruma, T., Nashida, T., & Kaneko, S. (2000). Mismatch negativity (MMN) reveals sound grouping in the human brain. *NeuroReport*, *5*, 1597–1601.
- Sloboda, J., & Edworthy, J. (1981). Attending to two melodies at once: The effect of key relatedness. *Psychology of Music*, *9*, 39–43.
- Sussman, E., Gomes, H., Nousak, J. M., Ritter, W., & Vaughan, H. G., Jr. (1998). Feature conjunctions and auditory sensory memory. *Brain Research*, *793*, 95–102.
- Sussman, E., Ritter, W., & Vaughan, H. G., Jr. (1998). Attention affects the organization of auditory input associated with the mismatch negativity system. *Brain Research*, *789*, 130–138.
- Sussman, E., Ritter, W., & Vaughan, H. G., Jr. (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology*, *36*, 22–34.
- 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.
- Takegata, R., Huotilainen, M., Rinne, T., Näätänen, R., & Winkler, I. (2001). Changes in acoustic features and their conjunctions are processed by separate neuronal populations. *NeuroReport*, *12*, 525–529.
- Takegata, R., Paavilainen, P., Näätänen, R., & Winkler, I. (1999). Independent processing of changes in auditory single features and feature conjunctions in humans as indexed by the mismatch negativity. *Neuroscience Letters*, *266*, 109–112.
- Takegata, R., Paavilainen, 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.
- Tervaniemi, M., Kujala, A., Alho, K., Virtanen, J., Ilmoniemi, R. J., & Näätänen, R. (1999). Functional specialization of the human auditory cortex in processing phonetic and musical sounds: A magnetoencephalographic (MEG) study. *NeuroImage*, *9*, 330–336.
- Tervaniemi, M., Medvedev, S. V., Alho, K., Pakhomov, S. V., Roudas, M. S., Van Zuijen, T. L., & Näätänen, R. (2000).

- Lateralized automatic auditory processing of phonetic versus musical information: A PET study. *Human Brain Mapping*, *10*, 74–79.
- Tervaniemi, M., Rytönen, M., Schröger, E., Ilmoniemi, R. J., & Näätänen, R. (2001). Superior formation of cortical memory traces for melodic patterns in musicians. *Learning and Memory*, *8*, 295–300.
- Tesche, C. D., Uusitalo, M. A., Ilmoniemi, R. J., Huottilainen, M., Kajola, M., & Salonen, O. (1995). Signal-space projections of MEG data characterize both distributed and well-localized neuronal sources. *Electroencephalography and Clinical Neurophysiology*, *95*, 189–200.
- Tiitinen, H., Alho, K., Huottilainen, M., Ilmoniemi, R. J., Simola, J., & Näätänen, R. (1993). Tonotopic auditory cortex and the magnetoencephalographic (MEG) equivalent of the mismatch negativity. *Psychophysiology*, *30*, 537–540.
- 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*, 430–442.
- Trainor, L. J., & Trehub, S. E. (1994). Key membership and implied harmony in Western tonal music: Developmental perspectives. *Perception & Psychophysics*, *56*, 125–132.
- Trehub, S. E., & Trainor, L. J. (1998). Singing to infants: Lullabies and play songs. In C. Rovee-Collier, L. P. Lipsitt, & H. Hayne (Eds.), *Advances in infancy research* (Vol. 12, pp. 43–77). Norwood, NJ: Ablex.
- Tremblay, K., Kraus, N., & McGee, T. (1998). The time course of auditory perceptual learning: Neurophysiological changes during speech–sound training. *NeuroReport*, *9*, 3557–3560.
- Ulanovsky, N., Las, L., & Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature Neuroscience*, *6*, 391–398.
- Vollmer-Haase, J., Finke, K., Hartje, W., & Bulla-Hellwig, M. (1998). Hemispheric dominance in the processing of J.S. Bach fugues: A transcranial Doppler sonography (TCD) study with musicians. *Neuropsychologia*, *36*, 857–867.
- Winkler, I., Paavilainen, P., & Näätänen, R. (1992). Can echoic memory store two traces simultaneously? A study of event-related brain potentials. *Psychophysiology*, *29*, 337–349.
- Winkler, I., Teder-Sälejärvi, W. A., Horvath, J., Näätänen, R., & Sussman, E. (2003). Human auditory cortex tracks task-irrelevant sound sources. *NeuroReport*, *14*, 2053–2056.
- 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.
- Yabe, H., Winkler, I., Czigler, I., Koyama, S., Kakigi, R., Sutoh, T., Hiruma, T., & Kaneko, S. (2001). Organizing sound sequences in the human brain: The interplay of auditory streaming and temporal integration. *Brain Research*, *897*, 222–227.
- Zenatti, A. (1969). Le développement génétique de la perception musicale. *Monographies Françaises de Psychologie*, *17*. Paris: CNRS.
- Zwicker, E., & Fastl H. (1999). Masking. In: *Psychoacoustics: Facts and models* (2nd ed., pp. 61–110). Berlin: Springer.