



Early integration of vowel and pitch processing: A mismatch negativity study

Pascale Lidji^{a,b,c,*}, Pierre Jolicœur^a, Régine Kolinsky^{b,c}, Patricia Moreau^a, John F. Connolly^d, Isabelle Peretz^a

^aBRAMS Laboratory, Université de Montréal, Montréal, Canada

^bUNESCOG, Université Libre de Bruxelles, U.L.B., Belgium

^cFonds de la Recherche Scientifique – F.N.R.S., Belgium

^dMcMaster University, Hamilton, Canada

ARTICLE INFO

Article history:

Accepted 9 December 2009

Available online 13 January 2010

Keywords:

MMN

ERPs

Pitch

Vowel

Phoneme

Music

P3a

ABSTRACT

Objective: Several studies have explored the processing specificity of music and speech, but only a few have addressed the processing autonomy of their fundamental components: pitch and phonemes. Here, we examined the additivity of the mismatch negativity (MMN) indexing the early interactions between vowels and pitch when sung.

Methods: Event-related potentials (ERPs) were recorded while participants heard frequent sung vowels and rare stimuli deviating in pitch only, in vowel only, or in both pitch and vowel. The task was to watch a silent movie while ignoring the sounds.

Results: All three types of deviants elicited both an MMN and a P3a ERP component. The observed MMNs were of similar amplitude for the three types of deviants and the P3a was larger for double deviants. The MMNs to deviance in vowel and deviance in pitch were not additive.

Conclusions: The underadditivity of the MMN responses suggests that vowel and pitch differences are processed by interacting neural networks.

Significance: The results indicate that vowel and pitch are processed as integrated units, even at a pre-attentive level. Music-processing specificity thus rests on more complex dimensions of music and speech.

© 2009 International Federation of Clinical Neurophysiology. Published by Elsevier Ireland Ltd. All rights reserved.

1. Introduction

The domain specificity of music and speech has been a matter of debate for years in the growing community of scientists interested in music processing (Patel, 2008; Pinker, 1997; Zatorre and Peretz, 2001). Mounting experimental results from neuropsychology (Peretz, 2006; Steinke et al., 1997) support the notion that specific processes are devoted to music perception and production. In contrast, brain imaging data show, for example, that music activates brain areas classically considered as specialized for language (Koelsch et al., 2002; Levitin and Menon, 2003), suggesting that music and speech cannot be totally independent from each other. The most reasonable view in the present state of the art is to consider that some aspects of music engage domain-specific processes (Peretz and Zatorre, 2005; Peretz, 2006) while other aspects engage processes that are shared between music and language (Patel, 2003; Peretz and Coltheart, 2003; Peretz and Zatorre, 2005; Price et al., 2006).

The aspects of music and speech that are handled by shared processes remain unclear. Pitch height and phonemes, which are

the building blocks of music and speech, respectively, might be good starting points. Among phonemes, vowels are the most similar in structure to musical tones by varying both in timbre and pitch. However, the relative importance of these two dimensions differs. Vowels vary in their spectral structure while tones mostly vary in pitch in a melodic context. However, recent data (Ross et al., 2007) built a new bridge between vowels and tones by showing that the intervals used in the music of most cultures have the same pitch ratio as the formants shaping the vowels of their languages. This suggests that vowels and tones are indeed very similar and may be processed by common mechanisms.

Kolinsky and collaborators (2009) recently found support for the joint processing of vowels and pitch in sung stimuli. In a speeded classification task (Garner, 1974) of bisyllabic nonwords sung on two-note melodic intervals, Kolinsky et al. (2009) observed that vowel and pitch dimensions cannot be attended selectively. When the nonword classification was based on vowel identity, irrelevant variations in pitch interfered with the classification process. A similar interference of irrelevant vowel variations was observed when the classification was based on pitch. Thus, classification performance suggested that vowel and pitch were integral rather than separable dimensions (Garner, 1974). Whereas similar results have been obtained in tone languages (Lee and Nusbaum, 1993; Repp and Lin,

* Corresponding author. Address: UNESCOG, Université Libre de Bruxelles, U.L.B., CP 191, B-1050 Brussels, Belgium. Tel.: +32 2 650 26 40; fax: +32 2 650 22 09.
E-mail address: plidji@ulb.ac.be (P. Lidji).

1990), Kolinsky et al. (2009) provide, to our knowledge, the first demonstration of integration between vowel and pitch processing in a musical context.

However, speeded-classification studies using Garner's interference paradigm do not allow us to specify the processing locus of dimensional interactions. This locus can be as early as pre-attentive sound detection but also as late as response selection, preparation, or even response execution (Holender, 1992). Indeed, as will be developed further on, it has recently been shown that dimensions associated with an integrality pattern in Garner's speeded classification tasks (Caclin et al., 2007) can nonetheless be processed separately in early auditory memory (Caclin et al., 2006). We are thus interested in determining whether pitch and vowel interactions occur very early in auditory processing. One technique that has been often exploited to tap the initial stages of auditory processing is the measurement of the mismatch negativity (MMN) component of event-related potentials (ERPs). The MMN is a negative deflection occurring between 100 and 250 ms after the onset of a stimulus violating an established acoustic regularity. It is generally elicited in non-attentive conditions and typically argued to reflect pre-attentive detection of auditory changes (Näätänen, 1992; Näätänen et al., 2007; Schröger, 1998). The MMN generators are activated when the sensory trace of an acoustically deviant sound is different from the sensory memory trace of the frequent sound, often called the standard (Näätänen, 1995). The MMN mainly originates from generators in the auditory cortex, described as related to pre-attentive change detection. Secondary frontal generators associated to later and probably conscious analysis of the change have also been described (Näätänen and Alho, 1995, but see Deouell, 2007 for a critical review).

The localization of the sources of the MMN has often been used to investigate the processing independence of different sound features (e.g., Giard et al., 1995; Levänen et al., 1993). In a magnetoencephalography (MEG) study, Tervaniemi and colleagues (1999) used the magnetic counterpart of the MMN (MMNm) to compare the strength and location of the brain activations obtained in response to chords and to vowels. The results revealed an MMNm of greater amplitude in response to chords than to vowels in the right hemisphere but no difference in MMNm amplitude in the left hemisphere. Both MMNm loci were posterior to the primary auditory cortex. These data support the idea of relative hemispheric specialization for linguistic and musical dimensions (see also Eulitz et al., 1995; Tervaniemi et al., 2000; Tervaniemi and Hugdahl, 2003; Zatorre et al., 1992). More specifically, these results suggest that the right hemisphere is more strongly involved than the left in fine-grained pitch processing. In addition to this inter-hemispheric specialization, Tervaniemi et al. (1999) found evidence for intra-hemispheric specialization: in each hemisphere, the MMNm source location for the phoneme change was superior to that for chord change. In short, these results suggest that spatially distinct neural populations are involved in the processing of phonemic and musical stimuli. However, the musical (chord) and the phonemic (vowel) changes were presented in separate stimuli and separate blocks in Tervaniemi et al. (1999). Thus, the independence found by the authors may at least in part be due to acoustical differences between chords and vowels. The use of sung stimuli embodying both dimensions would be a convenient solution to this potential problem (Lidji, 2007). In addition to combining pitch and vowel information in the same acoustical stream, hence avoiding comparing acoustically different stimuli (Tervaniemi et al., 1999), the use of these sung stimuli allowed us to compare the present ERP results to the behavioral data of Kolinsky et al. (2009).

In the present study, we used the MMN component of the auditory ERPs as a way to measure early electrophysiological responses to vowel and pitch changes. Furthermore, we designed the exper-

iment so as to determine whether the MMN response to a stimulus deviating from the standard along two stimulus dimensions could be predicted from the simple additivity of the MMN responses to two stimuli, each differing from the standard in a single dimension (e.g., Caclin et al., 2006; Levänen et al., 1993; Schröger, 1996b; Takegata et al., 2001b; Wolff and Schröger, 2001). This test of MMN additivity provides a way to probe the early independence of processing, at the neural level, of vowel and pitch deviations in sung vowels.

This approach has already proven its success in demonstrating that component dimensions of timbre such as spectral centroid and attack time, which had been shown to be interactive with Garner interference tasks (Caclin et al., 2007), leave however separate traces in early sensory memory (Caclin et al., 2006). A recent study (Caclin et al., 2008) in which ERPs were recorded in Garner classification conditions even suggests that these timbre dimensions are initially processed within distinct channels and interact at later processing stages (after 250 ms). This would account for both Garner interference and MMN additivity. Similarly, vowel and pitch may be handled separately in early processing stages and interact later on, as indexed by the dimensional integrality observed in Garner's speeded classification (Kolinsky et al., 2009).

The logic behind the MMN additivity approach used in the present study is as follows: if the changes of two sound dimensions are processed by independent neural generators, then the amplitude of the MMN response to a simultaneous change in both dimensions will be predicted by the sum of the amplitudes of the MMNs to changes in each dimension. This idea has been corroborated by source localization studies (Caclin et al., 2006; Takegata et al., 2001a), which have revealed that such additivity reflects the activity of at least partially distinct MMN neural generators, and thus separate short term memory traces, for each deviant feature. There are several reports of additivity in various auditory dimensions like tone duration and frequency (Levänen et al., 1993), tone duration and intensity (Wolff and Schröger, 2001), or phoneme duration spectral quality and duration (Ylinen et al., 2005).

Conversely, if the sound dimensions have common or interacting neural representations (Paavilainen et al., 2001), a simultaneous change in both features will elicit an ERP response that will differ from the sum of the responses observed for a change in each of the two features measured separately. This difference corresponds generally to an underadditive pattern, that is, an amplitude of the MMN to double deviants smaller than the one predicted by the additive model. The underadditivity can be explained if the same neurons respond to the two kinds of auditory changes. For double changes, they could fire as they did for single changes, or perhaps slightly more intensely if the double change leads to a greater saliency of the deviant, given that the MMN amplitude is proportional to the perceptual distance between standard and deviant stimuli (Näätänen et al., 2007; Savela et al., 2003; Schröger, 1996a). Underadditivity has, for example, been observed for frequency and intensity at frontal scalp sites (Wolff and Schröger, 2001).

The additivity of the MMN has, to our knowledge, never been used to study whether the early pre-attentive brain responses to pitch and vowels are elicited by independent or interactive early neural processes. The aim of the present study was to fill in this gap by examining whether the MMNs to vowel and pitch deviants are additive, a result that would be consistent with independent neural generators for the extraction of vowel identity and pitch, at least in speakers of non-tonal languages. Additivity would suggest that the evidence for interaction found by Kolinsky et al. (2009) arises from relatively late cognitive processes, as already demonstrated for other auditory integral dimensions (Caclin et al., 2006, 2007, 2008). Conversely, a non-additive pattern would suggest that vowel and pitch processing is subserved by shared neural sources even at the early processing levels probed by the MMN component.

2. MMN experiment

2.1. Methods

2.1.1. Participants

Twelve healthy right-handed native speakers of French (five men, mean age 25 years, range: 18–38) volunteered to participate in this study. Most participants had no musical training, except five subjects who had received musical training for 2–5 years but had stopped practicing for at least 6 years at the time of testing. All participants gave informed written consent after the procedures were explained to them. The study was vetted by the Ethics Committee of University of Montreal.

2.1.2. Stimuli and procedure

The stimuli were the synthesized French vowels /ε/ and /ɔ/ (as in the material from Kolinsky et al. (2009), Experiment 5), sung at two different pitches separated by one semitone (C3 = 130 Hz and C3# = 138 Hz) with a duration of 300 ms, including 10 ms rise and fall time. The choice of this frequency range was based on the results of calibration pilot studies which revealed that the MMN amplitude for the one semitone difference matched the amplitude of the MMN to the /ε/-/ɔ/ vowel contrast. The stimuli were synthesized by a source-filter voice synthesizer simulating the vocal tract at a frequency of 130 Hz, and shifted up by one semitone with *Adobe Audition* software to create the 138 Hz stimuli. The loudness of the stimuli was equated with the same software, after measurement with a sonometer. For each vowel, five formants were modeled. The frequency of the first four formants determined the nature of the vowel and F5 was always 1000 Hz higher than F4. In order to obtain a more natural singing feeling, a vibrato deviating by 2% of the fundamental was added. Because the 138 Hz stimuli were generated from the 130 Hz stimuli, the subtle vibrato was identical at both pitch heights (see the pitch contour in Table 1). Spectral analyses (Boersma and Weenink, 2007) showed that the pitch manipulation did not alter the formant structure of the vowels, as shown in Table 1.

The auditory stimuli were presented through e-prime software (Schneider et al., 2002) at a fixed offset to onset interval of 400 ms in a sequence including frequent (standard) and infrequent (deviant) stimuli, that is, an oddball paradigm. The presentation was pseudo-randomized so that any two deviant sounds were separated by at least three standards. The deviants could differ from the standards in three different ways: pitch height only, vowel identity only, or both pitch height and vowel identity. A mixed design was used so that these three types of deviants all occurred in every block (Ylinen et al., 2005). Furthermore, each stimulus served

as the standard, and the other three as deviants, across different blocks of stimulus presentations, as illustrated in Table 2. In sum, there were four blocks each lasting 15 min and the order of presentation was counterbalanced across participants. A total of 300 occurrences of each type of deviant (probability of occurrence = .06) and 4200 standards were presented to each participant, requiring 5100 sound presentations in the session. During the experiment, the participants were sitting in an electrically and acoustically isolated room. They were instructed to ignore the auditory stimulation presented binaurally through headphones at an intensity level of 70 dB SPL, while they watched a silent self-selected subtitled movie.

2.1.3. Electroencephalogram recording and processing

The continuous electroencephalogram (EEG) was recorded (bandpass 0.05–70 Hz, sampling rate 256 Hz, impedance <10 KΩ) via Synamp2 amplifiers (Neuroscan, Compumedics, El Paso, TX) from 64 Ag–AgCl electrodes at the standard 10–10 scalp sites. An electrode at the tip of the nose served as the reference. The electro-oculogram (EOG) was monitored for horizontal and vertical eye movements using two bipolar electrode pairs placed at the outer canthi of the left and right eyes and above and below the left eye. Because of a recording problem at C4 in several participants, this electrode and the homologous C3 were discarded from further statistical analyses.

The data were analyzed offline with Neuroscan 4.3.1 software. The EEG data were corrected for eye movements (Semlitsch et al., 1986), and filtered further with a 0.05–30 Hz band pass filter (24 dB/octave). Artifacts exceeding ±100 μV were rejected. The number of these trials did not exceed 30% in a single block, but this maximum was reached in only one block in a single participant. Prior to epoching, the standard sounds following a deviant sound were discarded from further analyses. These could indeed have evoked an MMN-like response because they were preceded by a physically different sound. Epochs of 800 ms, including a 100 ms

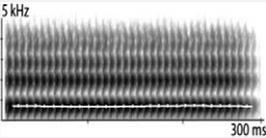
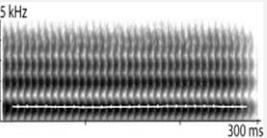
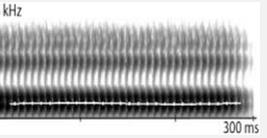
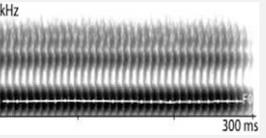
Table 2

Illustration of the experimental design: example of the sequence for one participant.

Stimulus category	Block 1	Block 2	Block 3	Block 4
Standard ($p = .82$)	/ε/ C3 $N = 1050$	/ɔ/ C3	/ε/ C3#	/ɔ/ C3#
Vowel deviant ($p = .06$)	/ɔ/ C3 $N = 75$	/ε/ C3	/ɔ/ C3#	/ε/ C3#
Pitch deviant ($p = .06$)	/ε/ C3# $N = 75$	/ɔ/ C3#	/ε/ C3	/ɔ/ C3
Double deviant ($p = .06$)	/ɔ/ C3# $N = 75$	/ε/ C3#	/ɔ/ C3	/ε/ C3

Table 1

Fundamental frequency (F0), spectrograms, and frequency of the three first formants (F1, F2, F3) for the sung vowels in Hz. F0 contour is marked in white in the spectrograms.

Vowel	/ε/		/ɔ/	
	C3	C3#	C3	C3#
Pitch				
Spectrogram				
F0	130	138	130	138
F1	601	602	513	540
F2	1795	1880	899	950
F3	2615	2760	2475	2626

pre-stimulus interval for baseline correction, were averaged separately for each oddball block (depending on the identity of the standard) and stimulus category (standard, pitch deviant, vowel deviant, double deviant). Epochs of the same stimulus category were also averaged across the four physically different stimuli when playing the role of either standard, vowel deviant, pitch deviant, or double deviant. This averaging allowed us to generalize the results to different vowels and pitches, hence avoiding stimulus-specific effects.

2.1.4. ERPs computation and measurement

The grand average ERPs elicited by the standard and deviant sounds were examined in the non-subtracted waveforms. The fronto-central electrode (Fz) was chosen for the analyses because the MMN amplitude was maximal at this site. We also analyzed the data at the left mastoid (LM) in order to detect a potential mastoid inversion (the data were similar at the right mastoid).

In order to examine whether MMN amplitude and latency differed as a function of stimulus identity ($/\varepsilon/$ at 130 Hz, $/\varepsilon/$ at 138 Hz, $/\sigma/$ at 130 Hz and $/\sigma/$ at 138 Hz), the MMN was first computed for each stimulus and type of deviant (vowel deviant, pitch deviant, and double deviant). For each stimulus, the MMN was delineated individually by subtracting the waveform to this stimulus when used as a standard from the waveform to this same stimulus when used as a pitch deviant, as a vowel deviant, and as a double deviant, using the so-called flip-flop procedure (Sharma et al., 2004). The MMN mean amplitude was quantified by computing the mean voltage within a 40 ms window centered at each individual's peak detected within a time window ranging from 100 to 250 ms after sound onset. The peak latency was defined as the time point of the maximum negativity in the same time window. In a second step, a "stimulus independent MMN" was obtained by subtracting the waveform to the standard, averaged across all four different stimuli when playing this role, from the

similarly computed grand-averaged waveform to each type of deviant. For example, the grand-average waveform for the "vowel deviant" was the average of the waveforms to $/\varepsilon/$ on C3, $/\varepsilon/$ on C#3, $/\sigma/$ on C3 and $/\sigma/$ on C#3 when playing the role of the vowel deviant. Indeed, each stimulus was the vowel deviant in different experiment blocks, depending on which stimulus was the standard (see Table 2). This yielded three difference MMN waves: pitch deviant, vowel deviant, and double deviant. MMN amplitude and latency were quantified in these waves as described for the stimulus-specific MMNs.

To test the MMN additivity hypothesis, a modeled double deviant MMN was computed for each subject as the sum of the pitch deviant and the vowel deviant "stimulus independent" difference waves at Fz and LM. The mean amplitudes of the observed and of the modeled double deviants were measured in an 80 ms window between 120 and 200 ms post-stimulus. This window was chosen because it included the peaks of both the double and the modeled double MMNs for each subject. The mean amplitude of the observed double deviant MMN was compared to that of the modeled double deviant MMN by paired-samples *t*-tests.

The mean amplitude of the P3 in the difference waves was averaged over 40 ms around the peak amplitude, in a window ranging from 200 to 350 ms. The peak latency was located in the same window.

For all analyses of variance, the Greenhouse–Geisser correction for non-sphericity was applied when required and the corrected *p* is reported along with the original degrees of freedom.

2.2. Results

2.2.1. Standard and deviant ERPs

Fig. 1 a and b shows the grand-average waveforms at Fz and LM for the standard sound and each type of deviant, averaged across all four stimuli. The standard waveform at Fz consisted of a P1

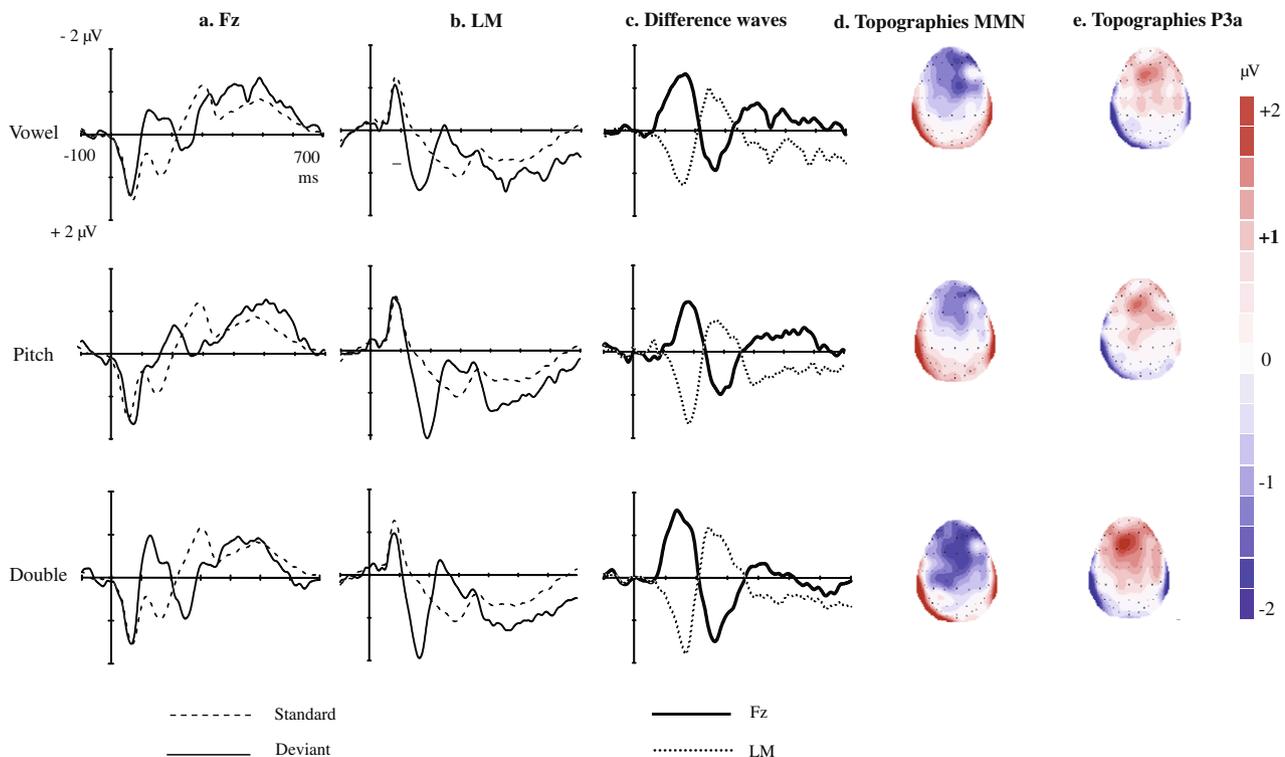


Fig. 1. Grand-average ($N = 12$) ERPs elicited by the standard (dotted line) and deviant (straight line) stimuli recorded at Fz (a) and LM (b) for each category of deviants (vowel, pitch, double), with negative plotted up. (c) The corresponding deviant – standard difference waveforms at Fz (bold line) and LM (dotted line). (d) The average scalp topographies of MMN over a 40-ms window centered at the peak for each condition; and (e) shows the average scalp topographies of the P3a over a 40-ms window centered at the peak. (For the topographies, activity at C4 was computed as an average of the activity recorded at C2, C6, FC4, and CP4).

peaking at 70 ms, an N1 at 118 ms, and a P2 at 162 ms. These components were followed by a negative deflection peaking around 300 ms and a residual sustained negativity. Most of these components inverted polarity at the mastoids. A similar long-lasting negativity was observed in the three grand-average waveforms to deviants. Before this late negativity, one can see a P1, a negative peak in the MMN latency range and a positive deflection peaking between 260 (double) and 288 ms (pitch). The positivity in the P300 latency range can be interpreted as a P3a, given the fronto-central topography (see Fig. 1e) and relatively early latency. This component usually reflects the orientation of attention, even in tasks that do not require an overt response to the deviants (Escera et al., 2000; Friedman et al., 2001; Polich, 2007).

2.2.2. Difference (deviant – standard) waveforms

2.2.2.1. MMN.

To confirm the presence of a significant MMN for each stimulus and each type of deviant (double deviant, vowel deviant and pitch deviant), *t*-tests against zero were performed on the mean amplitude of the difference wave at Fz, computed separately for each stimulus using the flip-flop methods. The difference wave was negative and significantly different from zero for all four stimuli and each deviant type, *t* ranging between -2.9 and -7.9 , with *ps* at least $<.015$.

In order to examine whether the amplitude of the MMN was modulated by the physical identity of the stimulus, we conducted a two-way analysis of variance on MMN mean amplitude and peak latency at Fz and LM. The within-subjects variables considered in this analysis were stimulus identity and type of deviant. The mean amplitude and latency of the MMN did not differ significantly between stimuli, all *F*s < 1 , *ps* $> .20$. Moreover, the interaction between stimulus identity and deviant type was not significant for amplitude, Fz: $F(6, 66) = 1.32$, LM: $F(6, 66) = 1.98$, nor for latency, Fz: $F(6, 66) = 1.82$, LM: $F(6, 66) = 1.67$, all *ps* $> .10$. Thus, stimulus identity did not interact with the type of deviant and did not modulate MMN amplitude and latency. Hence, the ERPs obtained for the four physically different stimuli were averaged in the following

analyses, so as to increase the number of trials for each type of deviation and focus on the type of deviation rather than incidental physical differences across stimuli.

Panel c in Fig. 1 depicts the subtraction of deviant – standard, averaged across the four stimuli, at Fz and LM. The double deviant, $-1.6 \pm 1 \mu\text{V}$, $t(11) = -5.43$, as well as the vowel, $-1.3 \pm 0.6 \mu\text{V}$, $t(11) = -8.29$, and the pitch deviant, $-1.1 \pm 0.8 \mu\text{V}$, $t(11) = -5.22$, elicited a significant MMN at Fz, with a polarity inversion at LM for the double, $1.4 \pm 0.7 \mu\text{V}$, $t(11) = 6.59$, the vowel, $1 \pm 0.6 \mu\text{V}$, $t(11) = 5.61$, and the pitch deviant, $1.4 \pm 0.9 \mu\text{V}$, $t(11) = 5.49$, all *ps* $< .0001$.

The effects of deviant type on the mean amplitude and on the peak latency of the MMN at Fz and LM were analyzed in separate one-way within-subjects ANOVAs. The type of deviant did not significantly modulate the amplitude of the MMN at Fz or LM, $F(2, 22) = 1.34$, and $F(2, 22) = 2.093$, respectively, both *ps* $> .10$. However, it influenced the peak latency of the MMN at both Fz, $F(2, 22) = 17.75$, $p < .0001$ and LM, $F(2, 22) = 33.09$, $p < .0001$ (see Fig. 2a). Post-hoc comparisons with a Bonferroni adjustment showed that the pitch deviants elicited a significantly later MMN peak (average latency, Fz: 189 ± 27 ms; LM: 187 ± 14 ms) than did the vowel (Fz: 154 ± 23 ms; LM: 157 ± 17 ms, $F(1, 11) = 19.97$ and 61.74 , respectively, *p* at least $<.001$), and the double deviants (Fz: 155 ± 21 ms; LM: 160 ± 13 ms, $F(1, 11) = 52.13$ and 84.88 , respectively, $p < .0001$).

In order to test whether the topographies of the brain responses to vowel, pitch, and double deviants differed, a two-way repeated measures analysis of variance was conducted on the MMN mean amplitudes, normalized according to McCarthy and Wood's procedure (McCarthy and Wood, 1985), with condition (3 levels) and electrode (62 levels) as within-subject variables. As expected, the analysis revealed a main effect of electrode, $F(61, 671) = 39.57$, $p < .0001$, but no interaction between electrode and type of deviant, $F < 1$, providing no support for the hypothesis that the scalp distribution of the MMN might be different for vowel, pitch, and double deviants.

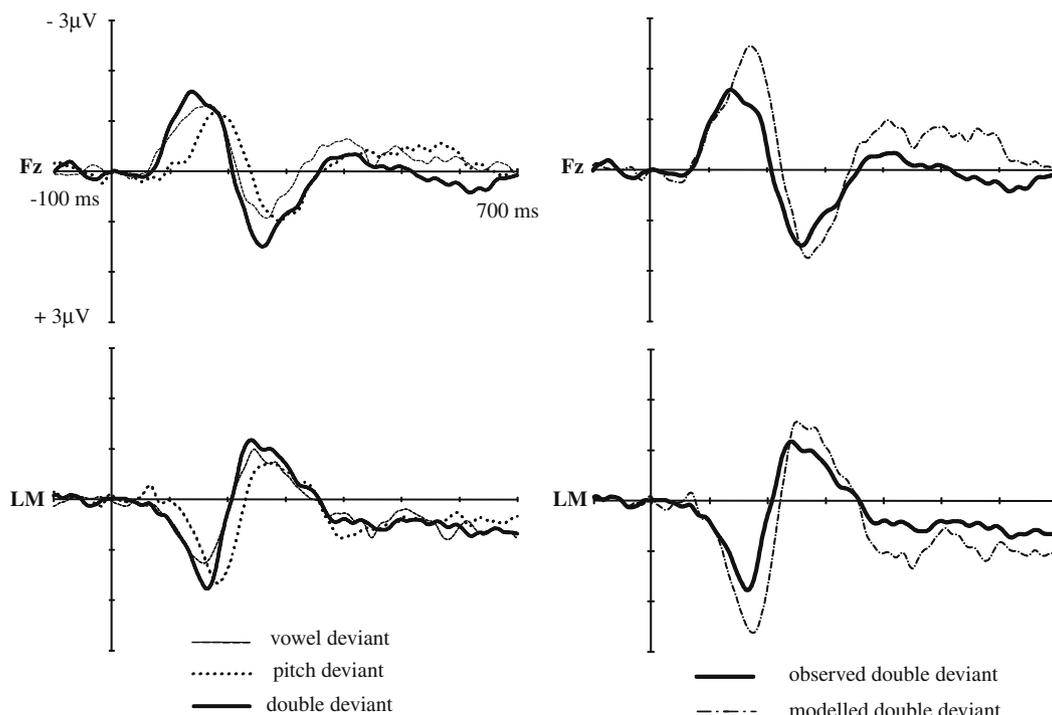


Fig. 2. (a) The MMN difference waves (deviant – standard) for each category of deviant (vowel, pitch, double) at Fz (top) and LM (bottom). (b) The MMN to the observed (bold line) and the modeled (dotted line) double deviant at Fz and LM.

To examine whether there were effects of laterality on the MMN distributions, the anterior scalp region was divided in two sections according to laterality (left vs. right, midline excluded). Only frontal and central electrodes, in which the amplitude of the MMN was the largest, were included in the pooling. The mean, normalized, MMN amplitude of the electrodes were thus pooled as follows: FP2, AF4, F8, F6, F4, F2, FT8, FC6, FC4, FC2, T7, C6, and C2, for the right side of the head, and FP1, AF3, F7, F5, F3, F1, FT7, FC5, FC3, FC1, T7, C5, and C1 for the left side of the head. A two-way analysis of variance on these data, using laterality and type of deviant as within-subject variables revealed, as visible in Fig. 1d, that the MMNs tended to be larger over the right than over the left hemisphere, $F(1, 11) = 3.65$, $p < .08$. The interaction between the type of deviant and the laterality of the MMN was far from significant, $F < 1$, converging with the previous omnibus analysis of the scalp distributions in showing that the topography of the MMN was not significantly different across the three types of deviants.

2.2.2.2. MMN additivity test. The empirical and modeled double deviant difference waves are depicted in Fig. 2b. We quantified the amplitude of the MMN component in both empirical and modeled double deviant difference waves, by computing the mean amplitude over an 80 ms window centered on the peak. The empirical and modeled mean amplitudes were compared in paired-samples *t*-tests. The observed double deviant MMN was significantly smaller than the modeled MMN at both Fz, $t(11) = 2.89$, $p < .02$, and LM, $t(11) = -3.58$, $p < .005$. In other words, the MMNs to vowel changes and to pitch changes were underadditive.

However, given the 35 ms latency difference between the MMNs to pitch and to vowel deviants, in the double deviants, the earlier processing of the vowel deviance might prevent complete processing of the pitch deviance, perhaps explaining the underadditive pattern (see Caclin et al., 2006). If it were the case, participants without, or with smaller, latency differences should exhibit more additivity than participants with larger latency differences. To test this account, we divided participants into two groups based on a median split of the observed latency difference between the MMNs to vowel deviants and to pitch deviants. The “large difference group” had an average latency difference of 59 ms and the “small difference group” had an average latency difference of 11 ms. Group was introduced as a between-subjects variable and the observed vs. modeled double deviant MMN as a within-subject variable. Again, the observed MMN amplitude was smaller than the modeled MMN, $F(1, 10) = 8.59$, $p < .02$. The average amplitude of the MMN did not differ between groups, $F < 1$, $p > .3$. Crucially, the interaction with group was not significant, $F(1, 10) = 1.34$, $p > .2$, hence suggesting that the latency difference between pitch and vowel single deviants is not related to the degree of underadditivity and hence cannot provide a basis on which to explain the underadditivity.

2.2.2.3. P3a. To confirm the presence of a P3a for each type of deviant, one sample *t*-tests against zero were run on the mean amplitude of the P3a in a 40 ms window centered at each individual's peak at Fz, where the amplitude was maximal. The positive component was significantly different from zero for the double, $t(11) = 8.96$, $p < .0001$, and for the pitch deviants, $t(11) = 3.27$, $p < .01$, but only tended towards significance for the vowel deviant, $t(11) = 2.09$, $p < .06$.

The effect of deviant type (vowel, pitch, and double) on the mean amplitude of the P3a was entered in a repeated-measures ANOVA which revealed a significant difference, $F(2, 22) = 6.70$, $p < .05$. Post-hoc tests demonstrated that the P3a amplitude was larger for the double than for the vowel deviant, $p < .01$, but the difference did not reach significance for the double and pitch deviant

comparison, $p < .09$. No difference was found for the pitch and vowel deviant comparison, $p > .10$. The peak latency of the P3a did not differ significantly between the types of deviants, $F(2, 22) = 2.09$, $p > .10$. In summary, the double deviant stimuli produced a larger P3a.

3. Behavioral control experiment

The MMN to pitch changes occurred significantly later than the MMN to vowel and to double changes. To examine whether this latency difference was related to the fact that the pitch changes were more difficult to detect than the other changes, a control behavioral experiment was run.

Fifteen non-musician participants (average age 24, range 18–28, 4 males), 6 of whom had participated in the MMN experiment, had to respond to the deviants as fast as possible by clicking a computer mouse. The response times and accuracy rates were recorded. The stimuli and design were exactly the same as in the MMN experiment. The results revealed a significant difference between the type of deviants on both detection times, $F(2, 28) = 26.32$, $p < .001$, and omissions, $F(2, 28) = 12.55$, $p < .001$. Participants were slower and less accurate to detect the pitch deviants (678 ms and 18% omissions) than both the vowel deviants (634 ms and 9% omissions), and the double deviants (619 ms and 5% omissions), p at least $< .005$. This is consistent with the MMN latency difference and suggests that the pitch change was less salient than the vowel change.

4. Discussion

In this MMN study, rare deviant sung vowels were inserted in a repetitive sequence of more frequent sung vowels. The ERPs to deviants varying in vowel identity only, in pitch only, or in both vowel identity and pitch were compared in order to assess the separability of the pre-attentive sensory memory traces of these sound attributes, as indexed by the MMN. All three types of deviant stimuli elicited a reliable MMN, which was followed by a fronto-centrally distributed positive deflection that can be interpreted as a P3a (Friedman et al., 2001; Polich, 2007). The latency and amplitude of the MMNs were comparable to the ones obtained in studies with similar spoken vowel contrasts (Jacobsen, 2004; Jacobsen et al., 2004; Näätänen et al., 1997; Savela et al., 2003). In the grand-averaged waveforms to standard stimuli, a sustained negativity was also observed (at Fz, seen as a sustained positivity at LM, Fig. 1). This negative deflection in the standard curve may be related to the fact that vowels, like other auditory stimuli (Picton et al., 1978), often evoke a sustained negativity that lasts until the offset of the stimulus (Eulitz et al., 1995; Hewson-Stoate et al., 2006; Jacobsen et al., 2004). In the present case, this sustained negativity seemed to arise around 200 ms and to decrease rapidly at 300 ms, corresponding to the offset of the stimulus. It did, however, not come back to zero and a remaining sustained negativity was present until 600 ms post-stimulus. The latter portion of this sustained negativity (at Fz, positivity at LM) was also present in the waveforms for deviant stimuli.

4.1. MMN

The MMN to pitch deviants occurred later than the MMNs to both vowel and double deviants. It has been frequently documented that the MMN amplitude and latency might be linked to the detectability of the auditory changes (e.g., Tiitinen et al., 1994; Toivainen et al., 1998, but see Horv  rth et al., 2008). This hypothesis was supported by the results of a control behavioral study showing slower and less accurate responses to the pitch

changes than to the vowel changes. In other words, the pitch change was less salient than the vowel change, as reflected in both response times and in MMN latencies. As will be discussed further on, this latency difference does not weaken the interpretation of the additivity results.

The amplitude of the MMN was similar for vowel, pitch, and double deviants. Since the MMN to double deviants was clearly not twice as large as the ones to single deviants, the results were underadditive. This was confirmed by the test of the additive model: the observed double deviant MMN was significantly smaller than the one predicted by summing the MMNs to vowel and pitch deviants at both Fz and LM.

The finding of similar underadditivity patterns at the frontal and the mastoid recording sites differs from prior reports of differential MMN additivity above and below the Sylvian fissure. Paavilainen and colleagues (2003) found the MMNs to pitch and intensity deviants to be additive at the mastoids (relative to the nose) but not at the frontal recording sites. They interpreted this outcome by hypothesizing that the MMN recorded at frontal sites may be partially overlapped by a N2b component (Näätänen et al., 1982) and includes contributions of both temporal and frontal generators (Alho, 1995, but see Deouell, 2007). By contrast, the MMN polarity reversal recorded at the mastoid electrodes provides an estimate of the supratemporal component in isolation. Although the results of Paavilainen et al. (2003) suggest that separate neural populations process *pitch* and *intensity* changes in the auditory cortex, our data indicate that common neural networks handle the pre-attentive memory traces leading to the early response to *pitch* and *vowel* changes. This conclusion is further supported by the fact that the scalp distribution of the MMN did not differ for vowel, pitch, and double deviants.

At first glance, the finding of shared MMN generators for vowel and pitch deviants seems to contradict Tervaniemi and collaborators' (1999) claim that vowel and chord changes are processed by anatomically different sources and that musical-pitch processing shows superiority over phonemic processing in right brain regions. How can we reconcile our underadditive pattern with the generator independence observed by Tervaniemi and colleagues? A first possibility is that the sources of the change-detection mechanisms for pitch and vowel identity are separate and that the MMNs should have been additive in our study. The reasons why it did not occur might be due to unrelated factors, such as saturation and latency differences.

Saturation refers to the possibility that the MMN was already at the maximum possible amplitude for each single deviant, so that a more ample MMN could not be observed for the double deviants. However, several considerations lead us to reject such an account. First, the amplitude of the MMN to the double deviants was $-1.6 \mu\text{V}$, which is below the amplitude of the MMN in experiments showing additivity for changes in vowel identity (Ikeda et al., 2002; Jacobsen et al., 2004; Näätänen et al., 1997; Winkler et al., 1999). More importantly, we performed a calibration experiment in which we used the same vocalic contrast as in the present experiment, but with a larger pitch difference of 9 semitones. With these stimuli, 7 pilot participants produced an average MMN of $-2.9 \mu\text{V}$ for double deviants, hence showing that large MMN amplitudes can be obtained with the same vowels as used in the present study. Based on these calibration results, we reduced the size of the pitch difference to ensure that the amplitude of the MMN would not be saturated in any of the conditions. The observed MMN in the double deviant condition was about half of the MMN we observed in the calibration study, ensuring that there was room for the MMN to be larger in the present experiment. Thus, the absence of additivity is unlikely to have originated from a saturation effect.

Alternatively, one could argue that the latency differences between the MMNs to vowel and pitch single deviants have masked

the additivity in the double deviant condition (Caclin et al., 2006). Such a situation has been documented for the dimensions of frequency and duration. Additivity has been found for MMNs close in latency (around 30 ms, Levänen et al., 1993), whereas the pattern was underadditive with a latency difference of 75 ms (Czigler and Winkler, 1996). Indeed, when the MMN in response to one deviance occurs earlier than the MMN in response to another deviance, the double deviant MMN will follow the latency of the first processed deviance. This could prevent the complete processing of the second dimension in double deviants, hence explaining the underadditive pattern when there are large latency differences. Here, we found little support for this account in terms of processing primacy. First, the average peak latency difference between the MMN to vowel deviants and to pitch deviants was 35 ms in our study, which is rather small compared to the latency differences affecting MMN additivity in the literature (Levänen et al., 1993; Czigler and Winkler, 1996). This average latency difference is about the same as the one observed by Levänen et al. (1993) who observed statistical additivity of their MMN responses. More importantly, we performed a direct test of the impact of MMN latency differences in our data. The additivity pattern did not differ between participants exhibiting a larger (59 ms in average) latency difference between vowel and pitch and the ones showing no or negligible latency difference (11 ms). In both cases, the MMNs to vowel and to pitch deviants were underadditive.

Therefore, the underadditivity of the MMNs to vowel and pitch deviance is not due to artifacts preventing an underlying additivity to occur. Instead, interacting neural populations seem to be involved in the early detection of vowel and pitch changes in *sung stimuli*. The underadditivity observed here might be related to the fact that pitch and vowel information was combined in the same acoustical stream. In contrast, Tervaniemi and colleagues (1999) found separate brain responses when comparing chords and vowels, that is, stimuli differing both acoustically and cognitively. Because MMN amplitude, latency, and source are sensitive to the acoustical features of the auditory stimuli (Näätänen, 1992), the generator difference observed by Tervaniemi et al. (1999) might be related to the physical distinctions between vowels and chords. Our use of sung vowel and pitch differences in the same acoustical context avoided these acoustical confounds.

The observation of underadditivity for vowel and pitch deviants is also consistent with the interaction found by Kolinsky et al. (2009) between vowel and pitch in a Garner's interference paradigm. The present results obtained in an unattended task with the same vowel contrasts further suggest that integration of vowel and pitch information arises early in the processing of sung material and is not due to crosstalk between separate channels at later processing stages (Caclin et al., 2008). Hence, our results support the view that at least some of the building blocks of music (pitch) and language (vowels) can be processed jointly by the same mechanisms at an early processing stage. This could be related to the fact that vowel and pitch are generally associated in both singing, in which vowels support melodic information (Scotto di Carlo, 2007), and speech. Indeed, pitch contour plays a central role in speech prosody (e.g., Patel et al., 1998) and conveys semantic information in tonal languages (e.g., Maddieson, 2005). Music processing specificity probably occurs later, at more complex levels, when pitch is interpreted in a tonal context (Besson et al., 1998; Bonnel et al., 2001). In sum, the processing of isolated pitches is not specific to music, but the tonal encoding of pitch might be (Peretz, 2006).

Yet, the present findings suggesting that pitch and vowel identity are early interactive dimensions are not trivial. Several simple sound features such as duration and frequency (Levänen et al., 1993) or duration and intensity (Wolff and Schröger, 2001) have been found to be processed independently with the same method

as the one used in the present study. Future investigations should aim at measuring MMN while varying pitch and vowel distance in a parametric fashion.

4.2. P3a

The primary goal of this study was to determine whether the processing of vowel and pitch changes is based on common or independent neural substrates. Our study focused on the additivity or non-additivity of the MMNs to vowel and pitch changes, but our stimuli also elicited other ERP components, and in particular a later positivity that was affected by whether the stimulus was a standard or a deviant. All three types of deviants elicited this positive component (though it did not reach significance for the vowel deviants), with a latency and a fronto-central scalp distribution that were broadly consistent with the P3a component (Dien et al., 2004; Escera et al., 2000; Friedman et al., 2001; Polich, 2007). The P3a has been taken to reflect attention switching towards deviants (Friedman et al., 2001). In other words, this component represents an index of the attention-catching power, or saliency, of the deviants. Since the P3a was significantly more ample for the double deviants than for the vowel deviants, and, to a lower extent, than for pitch deviants, it indicates that the double deviant was more salient than were the single deviants.

It is expected that stimuli varying on two features are more salient, and thus elicit a stronger orientation response than stimuli varying on a single feature. However, MMN amplitude and latency are also known to be proportional to the magnitude of the stimulus change (Näätänen et al., 2007). Here, however, the MMN amplitude was similar for the single and double deviants. This difference between the MMN and the P3a patterns supports the idea that the MMN (Näätänen, 1992; Näätänen et al., 2007) and the P3a (Donchin, 1981; Squires et al., 1975) reflect different processes. As mentioned, the MMN is often described as indexing a pre-attentive sensory-memory trace of the incoming stimulus, and its neural generators can slightly differ as a function of the nature of the auditory change (Caclin et al., 2006; Giard et al., 1995; Levänen et al., 1993; Tervaniemi et al., 1999). In contrast, the P3a is a general phenomenon reflecting the orientation of attention towards the deviant stimuli (Polich, 2007; Squires et al., 1975).

5. Conclusion and perspectives

The MMN elicited by vowel and pitch changes did not produce additive effects when manipulated simultaneously in a single stimulus (double deviant), hence suggesting that the early, pre-attentive, processing of these dimensions is integrated. This conclusion is corroborated by the analyses of the voltage scalp distributions of the MMNs across all three types of deviants, which provided no support for the hypothesis that the distributions were produced by different generators. Our result brings a new light on behavioral studies of pitch and vowel processing in songs and show, for the first time, that the integration between these dimensions occurs as early as the MMN latency range and may be independent of conscious attention allocation and of participants' strategies. This finding has important theoretical and clinical applications. At the theoretical level, the early integration of vowel and pitch suggests that the processing specificity of music does not occur at the level of its basic components, such as isolated pitch, but probably later, when tonality is created by the combination of certain musical notes in a specific key (Peretz, 2006). At the clinical level, these results bring a new light on the classical finding that singing or exaggerated intonation seems to improve language production in patients presenting speech disorders such as aphasia or stuttering (e.g., Albert et al., 1973; Healey et al., 1976; Racette

et al., 2006; Schlaug et al., 2008). This beneficial effect of singing on speech production is typically related to the slower rate and increased temporal regularity of singing as compared to speech. The present study brings another explanation into light: singing might emphasize salient pitch changes that can in turn cue vowel production more effectively.

Acknowledgements

Preparation of this paper was supported by a Mini-ARC doctoral grant and a travel fellowship from the Communauté française of Belgium to Pascale Lidji, as well as by a FRFC grant from the Fonds de la Recherche Scientifique (FNRS) to Régine Kolinsky. Isabelle Peretz's and Pierre Jolicoeur's research is supported by grants from the Natural Sciences and Engineering Research Council of Canada, from the Canadian Institutes of Health Research, and from infrastructure support from the Fonds de Recherche en Santé du Québec (FRSQ). The authors wish to thank Cécile Colin and Paul Deltenre for their contribution to the experimental design, Olivier Bélanger and Bernard Bouchard for assisting in the preparation of the experimental material, and Mihaela Felezeu for helping to test the participants.

References

- Albert ML, Sparks RW, Helm NA. Melodic intonation therapy for aphasia. *Arch Neurol* 1973;29:130–1.
- Alho K. Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart, (MMNm) elicited by sound changes. *Ear Hear* 1995;16:38–51.
- Besson M, Faita F, Peretz I, Bonnel AM, Requin J. Singing in the brain: independence of lyrics and tunes. *Psychol Sci* 1998;9(6):494–8.
- Boersma P, Weenink D. Praat: doing phonetics by computer (version 5.0); 2007, 4.3.1.4. Available from: <http://www.praat.org>.
- Bonnel AM, Faita F, Peretz I, Besson M. Divided attention between lyrics and tunes of operatic songs: evidence for independent processing. *Percept Psychophys* 2001;63(7):1201–13.
- Caclin A, Brattico E, Tervaniemi M, Näätänen R, Morlet D, Giard MH, et al. Separate neural processing of timbre dimensions in auditory sensory memory. *J Cogn Neurosci* 2006;18(12):1959–72.
- Caclin A, Giard MH, Smith BK, McAdams S. Interactive processing of timbre dimensions: a Garner interference study. *Brain Res* 2007;1138:159–70.
- Caclin A, McAdams S, Smith BK, Giard MH. Interactive processing of timbre dimensions: an exploration with event-related potentials. *J Cogn Neurosci* 2008;20(1):49–64.
- Czigler I, Winkler I. Preattentive auditory change detection relies on unitary sensory memory representations. *Neuroreport* 1996;7:2413–7.
- Deouell LY. The frontal generator of the mismatch negativity revisited. *J Psychophysiol* 2007;21(3–4):188–203.
- Dien J, Spencer KM, Donchin E. Parsing the late positive complex: mental chronometry and the ERP components that inhabit the neighborhood of the P300. *Psychophysiology* 2004;41:665–78.
- Donchin E. Surprise! ... Surprise? *Psychophysiology* 1981;18(5):493–513.
- Escera C, Alho K, Schröger E, Winkler I. Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiol Neurootol* 2000;5:151–66.
- Eulitz C, Diesch E, Pantev C, Hampson S, Elbert T. Magnetic and electric brain activity evoked by the processing of tone and vowel stimuli. *J Neurosci* 1995;15(4):2748–55.
- Friedman D, Cycowicz YM, Gaeta H. The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neurosci Biobehav Rev* 2001;25:355–73.
- Garner WR. The processing of information and structure. Potomac, MD: Erlbaum; 1974.
- Giard MH, Lavikainen J, Reinikainen K, Perrin F, Bertrand O, Pernier J. Separate representation of stimulus frequency, intensity and duration in auditory sensory memory: an event-related potential and dipole-model analysis. *J Cogn Neurosci* 1995;7:133–43.
- Healey EC, Mallard AR, Adams MR. Factors contributing to the reduction of stuttering during singing. *J Speech Hear Res* 1976;19:475–80.
- Hewson-Stoate N, Schönwiesner M, Krumboltz K. Vowel processing evokes a large sustained response anterior to primary auditory cortex. *Eur J Neurosci* 2006;24:2661–71.
- Horvath J, Czigler I, Jacobsen T, Maess B, Schröger E, Winkler I. MMN or no MMN: no magnitude of deviance effect on the MMN amplitude. *Psychophysiology* 2008;45:60–9.
- Holender D. Expectancy effects, congruity effects, and the interpretation of response latency measurement. In: Alegria J, Holender D, Junça de Morais J, Radeau M,

- editors. Analytic approaches to human cognition. Amsterdam: Elsevier; 1992. p. 351–75.
- Ikeda K, Hashimoto S, Otomo K, Kanno A. Asymmetrical mismatch negativity in humans as determined by phonetic but not physical difference. *Neurosci Lett* 2002;321:133–6.
- Jacobsen T. Mismatch negativity to frequency changes: no evidence from human event-related brain potentials for categorical speech processing of complex tones resembling vowel formant structure. *Neurosci Lett* 2004;362:204–8.
- Jacobsen T, Schröger E, Alter K. Pre-attentive processing of vowel phonemes from variable speech stimuli. *Psychophysiology* 2004;41:654–9.
- Koelsch S, Gunter TC, von Cramon DY, Zysset S, Lohmann G, Friederici AD. Bach speaks: a cortical “language-network” serves the processing of music. *Neuroimage* 2002;17(2):956–66.
- Kolinsky R, Lidji P, Peretz I, Besson M, Morais J. Processing interactions between phonology and melody: vowels sing but consonants speak. *Cognition* 2009;112:1–20.
- Lee L, Nusbaum HC. Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Percept Psychophys* 1993;53(2):157–65.
- Levänen S, Hari R, McEvoy L, Sams M. Responses of the human auditory cortex to changes in one vs. two stimulus features. *Exp Brain Res* 1993;97:177–83.
- Levitin DJ, Menon V. Musical structure is processed in “language” areas of the brain: a possible role for Brodmann Area 47 in temporal coherence. *Neuroimage* 2003;20(4):2142–52.
- Lidji P. Integrality and separability: review and application to the interactions between lyrics and tune in songs. *L'Année Psychol* 2007;107(4):659–94.
- Maddieson I. Tone. In: Haspelmath M, Dryer MS, Gil D, Comrie B, editors. *The World Atlas of language structures*. New York: Oxford University Press; 2005. p. 58–62.
- McCarthy G, Wood CC. Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models. *Electroencephalogr Clin Neurophysiol* 1985;62(3):203–8.
- Näätänen R, Simpson L, Loveless NE. Stimulus deviance and evoked potentials. *Biol Psychol* 1982;14:53–98.
- Näätänen R. Attention and brain function. Hillsdale, NJ: Erlbaum; 1992.
- Näätänen R. The mismatch negativity – a powerful tool for cognitive neuroscience. *Ear Hear* 1995;16(1):6–18.
- Näätänen R, Alho K. Generators of electrical and magnetic mismatch responses in humans. *Brain Topogr* 1995;7(4):315–20.
- Näätänen R, Lehtokoski A, Lennes M, Cheour M, Huotilainen M, Ilvonen A, et al. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 1997;385(6615):432–4.
- 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(12):2544–90.
- Paavilainen P, Mikkonen M, Kilpeläinen M, Lehtinen R, Saarela M, Tapola L. Evidence for the different additivity of the temporal and frontal generators of mismatch negativity: a human auditory event-related potential study. *Neurosci Lett* 2003;349:79–82.
- Paavilainen P, Valppu S, Näätänen R. The additivity of the auditory feature analysis in the human brain as indexed by the mismatch negativity: 1 + 1 approximate to 2 but 1 + 1 + 1 < 3. *Neurosci Letters* 2001;301(3):179–82.
- Patel AD. Language, music, syntax and the brain. *Nat Neurosci* 2003;6(7):674–81.
- Patel AD. Music, language, and the brain. New York: Oxford University Press; 2008.
- Patel AD, Peretz I, Tramo M, Labrecque R. Processing prosodic and musical patterns: a neuropsychological investigation. *Brain Lang* 1998;61:123–44.
- Peretz I, Coltheart M. Modularity of music processing. *Nat Neurosci* 2003;6(7):688–91.
- Peretz I, Zatorre R. Brain organization for music processing. *Annu Rev Psychol* 2005;56:89–114.
- Peretz I. The nature of music from a biological perspective. *Cognition* 2006;100(1):1–32.
- Picton TW, Woods DL, Proulx GB. Human auditory sustained potentials. I. The nature of the response. *Electroencephalogr Clin Neurophysiol* 1978;45:186–97.
- Pinker S. *How the mind works*. New York, London: W.W. Norton & Company; 1997.
- Polich J. Updating P300: an integrative theory of P3a and P3b. *Clin Neurophysiol* 2007;118:2128–48.
- Price C, Thierry G, Griffiths T. Speech-specific auditory processing: where is it? *Trends Cogn Sci* 2006;9(6):271–6.
- Racette A, Bard C, Peretz I. Making non-fluent aphasics speak: sing along! *Brain* 2006;129:2571–84.
- Repp BH, Lin HB. Integration of segmental and tonal information in speech-perception – a cross-linguistic study. *J Phon* 1990;18(4):481–95.
- Ross D, Choi J, Purves D. Musical intervals in speech. *Proc Natl Acad Sci USA* 2007;104(23):9852–7.
- Savela J, Kujala T, Tuomainen J, Ek M, Aaltonen O, Näätänen R. The mismatch negativity and reaction time as indices of the perceptual distance between the corresponding vowels of two related languages. *Cogn Brain Res* 2003;16:250–6.
- Scotto di Carlo N. Effect of multifactorial constraints on opera-singing intelligibility (I). *J Singing* 2007;63:1–13.
- Schlaug G, Marchina S, Norton A. From singing to speaking: why singing may lead to recovery of expressive language function in patients with Broca's aphasia. *Music Percept* 2008;25:315–23.
- Schröger E. A neural mechanism for involuntary attention shifts to changes in auditory stimulation. *J Cogn Neurosci* 1996a;8:527–39.
- Schröger E. Interaural time and level differences: integrated or separated processing? *Hear Res* 1996b;96:191–8.
- Schröger E. Measurement and interpretation of the mismatch negativity. *Behav Res Methods Instrum Comput* 1998;30(1):131–45.
- Schneider W, Eschmann A, Zuccolotto A. *E-prime user's guide*. Pittsburgh: P.A. Psychology Software Tools, Inc.; 2002.
- Semlitsch HV, Anderer P, Schuster P, Presslich O. A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology* 1986;23(6):695–703.
- Sharma M, Purdy SC, Newall P, Wheldall K, Dillon H. Effects of identification technique, extraction method, and stimulus type on mismatch negativity in adults and children. *J Am Acad Audiol* 2004;15:616–32.
- Squires NK, Squires KC, Hillyards SA. Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalogr Clin Neurophysiol* 1975;38:387–410.
- Steinke WR, Cuddy LL, Holden RR. Dissociation of musical tonality and pitch memory from nonmusical cognitive abilities. *Can J Exp Psychol* 1997;51(4):316–35.
- Takegata R, Huotilainen M, Rinne T, Näätänen R, Winkler I. Changes in acoustic features and their conjunctions are processed by separate neuronal populations. *Neuroreport* 2001a;12(3):525–9.
- Takegata R, Syssoeva O, Winkler I, Paavilainen P, Näätänen R. Common neural mechanism for processing onset-to-onset intervals and silent gaps in sound sequences. *Neuroreport* 2001b;12(8):1783–7.
- Tervaniemi M, Kujala A, Alho K, Virtanen J, Ilmoniemi RJ, Näätänen R. Functional specialization of the human auditory cortex in processing phonetic and musical sounds: a magnetoencephalographic (MEG) study. *Neuroimage* 1999;9(3):330–6.
- Tervaniemi M, Medvedev SV, Alho K, Pakhomov SV, Roudas MS, van Zuijen TL, et al. Lateralized automatic auditory processing of phonetic versus musical information: a PET study. *Hum Brain Mapp* 2000;10(2):74–9.
- Tervaniemi M, Hugdahl K. Lateralization of auditory-cortex functions. *Brain Res Rev* 2003;43(3):231–46.
- Tiitinen H, May P, Reinikainen K, Näätänen R. Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature* 1994;372:90–2.
- Toivainen P, Tervaniemi M, Louhivuori J, Saher M, Huotilainen M, Näätänen R. Timbre similarity: convergence of neural, behavioral and computational approaches. *Music Percept* 1998;16(2):223–41.
- Winkler I, Lehtokoski A, Alku P, Vaino M, Czigler I, Csépe V, et al. Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cogn Brain Res* 1999;7:357–69.
- Wolff C, Schröger E. Human pre-attentive auditory change-detection with single, double and triple deviations as revealed by mismatch negativity additivity. *Neurosci Lett* 2001;311:37–40.
- Ylinen S, Huotilainen M, Näätänen R. Phoneme quality and quantity are processed independently in the human brain. *Neuroreport* 2005;16(16):1857–60.
- Zatorre RJ, Evans AC, Meyer E, Gjedde A. Lateralization of phonetic and pitch discrimination in speech processing. *Science* 1992;256:846–9.
- Zatorre R, Peretz I, editors. *The biological foundations of music*. New York: New York Academy of Science; 2001.