ELSEVIER

Neuroscience Letters 290 (2000) 66–70

Neuroscience Letters

www.elsevier.com/locate/neulet

Effects of spectral complexity and sound duration on automatic complex-sound pitch processing in humans – a mismatch negativity study

M. Tervaniemi[a,*], E. Schröger[b], M. Saher[a], R. Näätänen[a, c]

[a]Department of Psychology, Cognitive Brain Research Unit, University of Helsinki, P.O. Box 13, FIN-00014 Helsinki, Finland
[b]Department of General Psychology, University of Leipzig, Germany
[c]BioMag Laboratory, Medical Engineering Centre, Helsinki University Central Hospital, Helsinki, Finland

Received 25 February 2000; received in revised form 19 June 2000; accepted 21 June 2000

Abstract

The pitch of a spectrally rich sound is known to be more easily perceived than that of a sinusoidal tone. The present study compared the importance of spectral complexity and sound duration in facilitated pitch discrimination. The mismatch negativity (MMN), which reflects automatic neural discrimination, was recorded to a 2.5% pitch change in pure tones with only one sinusoidal frequency component (500 Hz) and in spectrally rich tones with three (500–1500 Hz) and five (500–2500 Hz) harmonic partials. During the recordings, subjects concentrated on watching a silent movie. In separate blocks, stimuli were of 100 and 250 ms in duration. The MMN amplitude was enhanced with both spectrally rich sounds when compared with pure tones. The prolonged sound duration did not significantly enhance the MMN. This suggests that increased spectral rather than temporal information facilitates pitch processing of spectrally rich sounds.

Keywords: Frequency discrimination; Complex sounds; Human; Auditory event-related potentials; Mismatch negativity; P3a; Automatic processing

Accurate pitch perception is a prerequisite for adequate music and speech processing for instance, in the case of melodic, harmonic, and prosodic processing. Neural dynamics underlying pitch processing have recently been extensively investigated by recording the mismatch negativity (MMN) component of the event-related potential (ERP). The MMN is elicited when the neural code of the incoming stimulus with 'deviant' parameter(s) is discrepant with the cortical representation of the repetitive 'standard' stimulus [9,11]. The MMN is elicited by any discriminable sound change, for instance, in sinusoidal tones (for a review, see [9]), phonemes [1,2,7,10,21], chords [3,6,21], and also abstract sound features [13].

The MMN latency and amplitude reflect the magnitude of the physical difference between the deviant and standard stimulus [15]. Furthermore, the MMN amplitude and latency reflect perceptual accuracy, as determined by musicality tests (for a review, see [18]) and by the hit rates and the reaction times in a behavioral experiment employing the same stimuli as in the MMN recordings, (e.g. [1,7,23]). This correlation between the MMN parameters and behavioral responses imply that pre-attentive neural functions determine the accuracy of the subsequent attentive processes [12].

The frequency discrimination of tones with a rich spectral structure is facilitated when compared with that of simple tones. The frequency-MMN amplitude was enhanced and latency shortened with synthesized piano tones in comparison to sinusoidal tones with matched magnitudes of frequency change [19]. Correspondingly, the MMN was elicited by a frequency change as small as 3% presented in one out of ten harmonic partials of a complex sound [22]. In behavioral studies, this frequency change causes a 1-Hz change of the perceived pitch in 155-Hz fundamental frequency, equaling a pitch change of 0.65% [4], whereas in MMN studies with pure-tone stimulation, somewhat larger frequency differences are necessary for reliable MMN elicitation [15,23]. Very recently it was shown that the frequency-MMN amplitude was larger and latency shorter in spectrally rich tones consisting of three harmonic partials

---

[*] Corresponding author. Tel.: +358-9-1912-3408; fax: +358-9-1912-2924.
E-mail address: tervanie@helsinki.fi (M. Tervaniemi).

0304-3940/00/$ - see front matter © 2000 Elsevier Science Ireland Ltd. All rights reserved.
PII: S0304-3940(00)01290-8

than in pure tones [20]. Consistently with this, the subjects' behavioral responses were more accurate for spectrally rich tones than for sinusoidal tones.

In addition to spectral complexity, the pitch processing is influenced by sound duration. The encoding of virtual pitch is facilitated in long compared to short tones (150 vs. 500-ms sounds [24] whereas the encoding of spectral pitch is not further facilitated if sound duration exceeds 30 ms [14].

Facilitated pitch discrimination with spectrally rich sounds can be explained by information increase: spectrally rich tones carry more information than sinusoidal tones, in both spectral and temporal terms [8]. The present study was conducted to systematically compare the relative importance of spectral complexity and sound duration in automatic pitch processing. To this end, the MMN to a frequency change in sinusoidal and spectrally rich sounds of two durations was investigated.

Ten healthy subjects (five males; aged 21–29 years; nine right-handed) with normal hearing were employed. During the recordings, subjects were instructed to watch a silenced movie of their own choice and not to pay attention to sound stimuli.

The sound sequences included standard tones ($P = 0.9$) and two different kinds of deviant tones (+2.5% and −2.5% frequency change, $P = 0.05$ each). In pure tones with one partial, the standard was of 500 Hz in frequency and the deviant of 513 or 488 Hz. In three-partial sounds, the standard stimulus was a 500 + 1000 + 1500 Hz sound and the deviant stimulus was a 513 + 1026 + 1539 or 488 + 976 + 1464 Hz sound. In five-partial sounds, the standard was a 500 + 1000 + 1500 + 2000 + 2500 Hz sound and the deviant was a 513 + 1026 + 1539 + 2052 + 2565 Hz or 488 + 976 + 1464 + 1952 + 2440 Hz sound. The stimulus sequences were pseudorandom so that each deviant tone was preceded by at least one standard tone.

The tone duration was, in separate blocks, 100 and 250 ms (including 5 ms rise and fall times). Stimuli were binaurally presented via headphones at an intensity of 50 dB above the individually determined hearing threshold. The constant stimulus onset asynchrony (SOA) was 500 ms. The sequences consisted of 1000 sounds. Each of the six conditions (100 and 250 ms, each with one, three and five partials) included two sequences.

The electroencephalograph (EEG) recordings were performed in an acoustically and electrically shielded room. The EEG was recorded (passband 0.1–30 Hz, sampling frequency 250 Hz) with Ag–AgCl electrodes from seven scalp locations: Fz, Cz and Pz (10–20 system), left and right mastoids (LM and RM, respectively), and two electrodes placed at 1/3 of the arc connecting Fz to the mastoids (L1 on the left and R1 on the right side, respectively). In addition, horizontal eye movements were monitored with an electrode attached to the right outer canthus and vertical eye movements with the Fpz electrode. The EEG and electrooculogram (EOG) were referenced to the nose during the recordings.

The EEG was epoched (−100–400 ms including 100 ms pre-stimulus baseline) and averaged separately for each deviant stimulus and for standards immediately preceding deviants, for standards following deviants, and for the remaining standards. All epochs including voltage changes exceeding 100 μV were automatically rejected. Frequencies higher than 30 Hz were filtered out.

Difference waveforms (deviant-tone ERP minus standard-tone ERP) were calculated. To improve the signal-to-noise ratio, difference waves for frequency changes +2.5 and −2.5% were pooled together. The MMN was quantified by first determining the MMN peak latency from the Fz grand-average difference waves separately for each deviant as the most negative peak between 100–300 ms. The MMN amplitude was thereafter measured from the individual difference waves as the average amplitude calculated over the 20-ms time window centered at this peak latency.

The statistical significance of MMN was tested with two-tailed *t*-tests by comparing the mean MMN amplitude at Fz and mastoid leads to zero. Two-way analysis of variance (ANOVA) were used to determine whether the sound structure (levels: one-, three-, and five-partials) and the sound duration (levels: 100 and 250 ms) affected the MMN amplitude and latency at the Fz electrode.

The P3a was quantified by first determining the peak latency from the grand-average difference waves separately for each deviant as the most negative peak between 200 and 400 ms. The P3a amplitude was thereafter measured from individual difference waves as the average amplitude calculated over the 40-ms time window centered at this peak latency. The statistical significance of P3a was tested by comparing the mean P3a amplitude at Cz to zero.

Fig. 1 illustrates the grand-average ERPs elicited by standard and deviant tones. The 100-ms sounds elicited prominent P1 and N1 responses whereas, the 250-ms sounds elicited a P1 without a subsequent N1. This was due to the relatively fast stimulation rate (2 sounds/s) which led the N1 to saturate with the longer, 250-ms sounds.

Fig. 2 illustrates the deviant-minus-standard difference waves. In all conditions, deviant tones elicited a negative response with a fronto-central maximum and a slightly larger amplitude over the right than left hemisphere. It peaked at 190–235 ms and was significant at Fz in all conditions ($t(9)$ between 4.1 and 8.4, $P < 0.01$ at the Fz electrode). Its polarity inversal in the mastoid leads was significant in all conditions but the one with 100-ms sounds with one partial (for the remaining conditions: $t(9)$ between 3.5 and 6.9, $P < 0.007$ at the LM electrode). The topography of the response is consistent with its being an MMN.

There was a significant main effect of the sound structure on the MMN amplitude ($F(2, 18) = 16.1$, $P < 0.0001$). Subsequent Newman–Keuls paired comparisons indicated that the MMN amplitude was significantly larger for the three- and five-partial sounds than for the one-partial sounds ($P < 0.0005$ and $P < 0.0003$, respectively). The MMN latency was not affected by the sound structure ($F(2, 18) = 0.68$, $P < 0.52$).

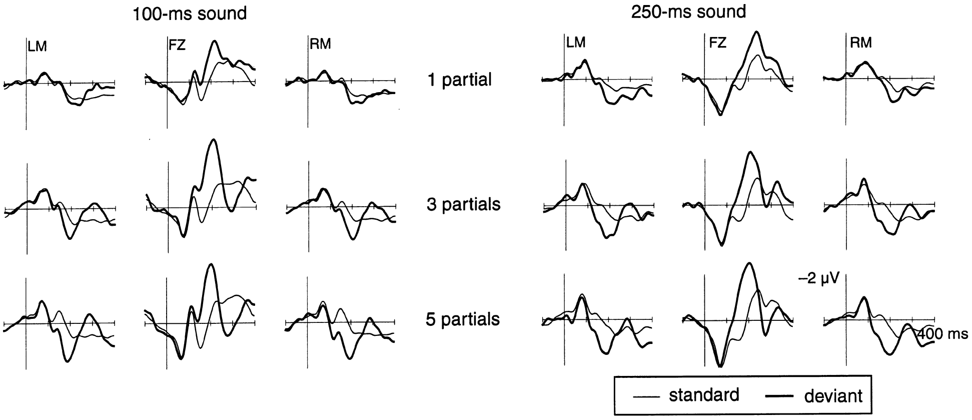

Fig. 1. The grand-average ERPs elicited by standard (thin line) and frequency-deviant (thick lines) one-, three-, and five-partial sounds with sound duration of 100 ms (left panel) and 250 ms (right panel).

The sound duration did not significantly affect the MMN amplitude or latency ($F(1,9) = 2.0$, $P < 0.19$; $F(1,9) = 0.31$, $P < 0.59$, respectively). Neither was there a significant interaction between the sound structure and duration in MMN amplitude or latency ($F(2,18) = 2.3$, $P < 0.13$; $F(2,18) = 0.84$, $P < 0.45$, respectively).

The MMN was followed by P3a with short three-partial sounds and with short and long five-partial sounds ($t(9)$ between 2.65 and 3.45, $P < 0.05$ at the Cz electrode). Its amplitude at Cz differed between conditions (1-way repeated measures ANOVA $F(1,5) = 3.58$, $P < 0.009$). Subsequent Newman–Keuls test showed that this resulted from P3a being significantly larger for both five-partial sounds than for long one-partial sounds ($P < 0.05$).

Previously, it was shown that the presence of spectral harmonics facilitates neuronal and behavioral pitch perception [20]. The present study compared the relative importance of increased spectral complexity and sound duration in this facilitation. According to the present results, the increased spectral rather than temporal information facilitates pitch processing with spectrally rich sounds. This was indexed by the MMN amplitude enhancement caused by enriched sound structure but not by prolonged sound duration. This result is further confirmed by the missing statistical interaction between the sound structure and duration. The absence of a duration effect on pre-attentive pitch discrimination seems to contradict the finding according to which the encoding of virtual pitch is facilitated in long compared to short tones [24]. Yet, this result is consistent with the finding that prolonging the duration of sinusoidal tones up to 300 ms does not improve pitch discrimination (as indexed by MMN and behavioral performance) provided that a minimal duration of about 30 ms is exceeded [14]. However, according to the temporal theory of pitch perception [8] not only an effect of the spectral structure but also of duration should have been expected. Since no such effect occurred it seems that temporal information seems to be less important than spectral information for encoding pitch in the present experimental condition.

The pitch-processing accuracy (as indexed by the MMN) was not significantly increased by further increasing the number of the partials from three to five. This may be caused by a saturation effect, in other words, the three lowest partials carried sufficient information for effectively

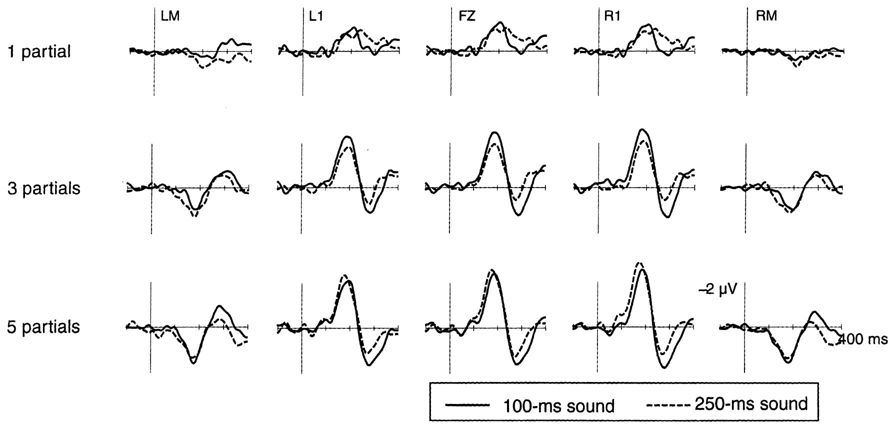

Fig. 2. The deviant-minus-standard difference waves elicited by frequency-deviant sounds with one-, three-, and five-partials (as indicated on the left) when sound duration was 100 ms (solid line) and 250 ms (dashed line).

determining the complex-tone pitch. Alternatively, since the partials four and five fell above the dominance region of the complex-tone pitch [8], they did not further facilitate the pitch processing.

Interestingly, MMN was small in one-partial sounds with no subsequent P3a, while the MMN was larger and followed by a distinct P3a with spectrally rich sounds. Since the P3a indexes an involuntary attention shift to deviant or novel stimuli (for a review, see, [5]) the present results suggest that tiny pitch deviations are more likely to cause distraction in spectrally rich sounds than in sinusoidal tones. Moreover, the present pattern of results confirms the hypothesis that the MMN signal may lead to an attention switch when a particular threshold is exceeded [9,16].

As sounds were not task-relevant, (i.e. subjects were watching a movie during the recordings), the pattern of results demonstrates that the effect of spectral richness already occurs (at least partly) at pre-attentive levels of processing that do not require the intentional allocation of attention. This has a possible application in constructing and testing acoustic devices, which inform the user about potentially relevant events during his/her parallel task. In addition, the present results may have an impact for the MMN paradigms used in psychopharmacological and clinical research in which a reasonable signal-to-noise ratio is required in minimal time. Thus, the MMN signal can be optimized by rather increasing the sound complexity than by prolonging the sound duration (at least when sound duration exceeds 100 ms), the experimental session remaining shorter. Moreover, with spectrally rich sounds, a sizable MMN signal can be obtained with small frequency deviations. Thus, the overlap of enhanced N1 component adding variability into MMN quantification and interpretation can be minimized [17].

In sum, the present data indicate that, at least when sound duration exceeds 100 ms, facilitated pitch discrimination with spectrally rich sounds when compared with sinusoidal tones is not further increased if the sound duration is prolonged. This suggests that increased spectral rather than temporal information facilitates pitch processing of spectrally rich sounds.

The study was supported by the Academy of Finland and by the Deutscher Akademischer Austauschdienst. The authors thank Dipl.-Psych. Christian Wolff for his technical assistance.

[1] Aaltonen, O., Eerola, O., Lang, A.H., Uusipaikka, E. and Tuomainen, J., Automatic discrimination of phonetically relevant and irrelevant vowel parameters as reflected by mismatch negativity, J. Acoust. Soc. Am., 96 (1994) 1489–1493.

[2] Alho, K., Connolly, J.F., Cheour, M., Lehtokoski, A., Huotilainen, M., Virtanen, J., Aulanko, R. and Ilmoniemi, R.J., Hemispheric lateralization in preattentive processing of speech sounds, Neurosci. Lett., 258 (1998) 9–12.

[3] Alho, K., Tervaniemi, M., Huotilainen, M., Lavikainen, J., Tiitinen, H., Ilmoniemi, R.J., Knuutila, J. and Näätänen, R., Processing of complex sounds in the human auditory cortex as revealed by magnetic brain responses, Psychophysiology, 33 (1996) 369–375.

[4] Darwin, C. and Ciocca, V., Grouping in pitch perception: effects of onset asynchrony and ear of presentation of a mistuned component, J. Acoust. Soc. Am., 91 (1992) 3381–3390.

[5] Escera, C., Alho, K., Schröger, E. and Winkler, I., Involuntary attention and distractibility as evaluated with event-related brain potentials, Audiol. Neuro-otol., 5 (2000) 151–166.

[6] Koelsch, S., Schröger, E. and Tervaniemi, M., Superior attentive and pre-attentive auditory processing in musicians, NeuroReport, 10 (1999) 1309–1313.

[7] Kraus, N., McGee, T.J., Carrell, T.D., Zecker, S.G., Nicol, T.G. and Koch, D.B., Auditory neurophysiologic responses and discrimination deficits in children with learning problems, Science, 273 (1996) 971–973.

[8] Moore, B.C.J., Introduction to the Psychology of Hearing, Academic Press, London, 1989.

[9] Näätänen, R., Attention and Brain Function, Erlbaum, Hillsdale, NJ, 1992.

[10] 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. and Alho, K., Language specific phoneme representations revealed by the electric and magnetic brain responses, Nature, 385 (1997) 432–434.

[11] Näätänen, R., Paavilainen, P., Tiitinen, H., Jiang, D. and Alho, K., Attention and mismatch negativity, Psychophysiology, 30 (1993) 436–450.

[12] Novak, G.P., Ritter, W., Vaughan Jr., H.G. and Witznitzer, M.L., Differentiation of negative event-related potentials in an auditory discrimination task, Electroenceph. clin. Neurophysiol., 75 (1990) 255–275.

[13] Paavilainen, P., Jaramillo, M. and Näätänen, R., Binaural information can converge in abstract memory traces, Psychophysiology, 35 (1998) 483–487.

[14] Paavilainen, P., Jiang, D., Lavikainen, J. and Näätänen, R., Stimulus duration and the sensory memory trace: an event-related potential study, Biol. Psychol., 35 (1993) 139–152.

[15] Sams, M., Paavilainen, P., Alho, K. and Näätänen, R., Auditory frequency discrimination and event-related potentials, Electroenceph. clin. Neurophysiol., 62 (1985) 437–448.

[16] Schröger, E., On the detection of auditory deviations: a pre-attentive activation model, Psychophysiology, 34 (1997) 245–257.

[17] Sinkkonen, J. and Tervaniemi, M., Towards optimal recording and analysis of the mismatch negativity, Audiol. Neuro-otol., 5 (2000) 225–234.

[18] Tervaniemi, M., Pre-attentive processing of musical information in the human brain, J. New Music Res., 28 (1999) 237–245.

[19] Tervaniemi, M., Alho, K., Paavilainen, P., Sams, M. and Näätänen, R., Absolute pitch and event-related brain potentials, Music Perception, 10 (1993) 305–316.

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

[21] Tervaniemi, M., Kujala, A., Alho, K., Virtanen, J., Ilmoniemi, R.J. and Näätänen, R., Functional specialization of the human auditory cortex in processing phonetic and musical sounds, NeuroImage, 9 (1999) 330–336.

[22] Tervaniemi, M., Schröger, E. and Näätänen, R., Pre-atten-

tive processing of spectrally complex sounds with asynchronous onsets: an event-related potentials study in human subjects, Neurosci. Lett., 227 (1997) 197–200.

[23] Tiitinen, H., May, P., Reinikainen, K. and Näätänen, R., Attentive novelty detection in humans is governed by pre-attentive sensory memory, Nature, 372 (1994) 90–92.

[24] Winkler, I., Tervaniemi, M. and Näätänen, R., Two separate codes for missing fundamental pitch in the auditory cortex, J. Acoust. Soc. Am., 102 (1997) 1072–1082.