NeuroReport 8, 2571–2574 (1997)

Pre-attentive categorization of sounds by timbre as revealed by event-related potentials

Mari Tervaniemi,[1,CA] Istvan Winkler,[1,2] and Risto Näätänen[1]

[1]Cognitive Brain Research Unit, Department of Psychology, PO Box 13 University of Helsinki, FIN-00014 Helsinki, Finland; [2]Department of Psychophysiology, Institute for Psychology, Hungarian Academy of Sciences, Budapest, Hungary

[CA]Corresponding Author

INFREQUENT (10%) pure tones were randomly presented among nine different missing-fundamental tones having the same pitch (10% each) to subjects playing a computer game. MMN (an index of pre-attentive change detection) was elicited by timbre-deviant pure tones with 150 and 500 ms stimulus duration. This suggests that the spectral component of timbre is pre-attentively determined from relatively short (150 ms) acoustic samples. Previous research established that resolving the pitch of the same missing-fundamental tones requires longer (> 150 ms) sounds. Consequently, timbre and pitch are probably determined by separate neural processes. The present results also demonstrate pre-attentive categorization of sounds based on timbre as MMN could only be elicited by the pure tones if their timbre was contrasted with the combined group of the nine standard sounds of qualitatively similar rich timbre.

Key words: Auditory cortex; Auditory sensory memory; Complex tone perception; Mismatch negativity (MMN); N1; Pre-attentive auditory processing; Timbre perception

Introduction

The neural basis of auditory perception can be investigated non-invasively by recording event-related brain potentials (ERPs) during auditory stimulation.[1] ERPs can be conceptualized as a sequence of components each with its own set of neural generators and reflecting different psychological functions.[2] The ERP component termed the mismatch negativity (MMN) reflects a pre-attentive auditory deviance detection process based on auditory sensory memory traces.[3–5] MMN is typically elicited by infrequent discriminable changes in sound parameters such as frequency, intensity, or duration in otherwise homogeneous series of a repetitive standard stimulus. MMN relies on a transient store of auditory sensory information for detecting deviation from invariant features of the acoustic environment (such as repetition of a stimulus).[5,6] This transient auditory store is probably identical to what experimental psychology terms as auditory sensory memory (or echoic memory).[3,7] Because MMN is elicited even when subjects disregard the auditory stimulation (e.g. while playing a visual computer game or performing a difficult visual discrimination task) MMN can be used as a pre-attentive index of auditory sensory memory traces.[8,9]

Most MMN studies have employed pure tones to probe auditory sensory memory traces. For such simple sounds, acoustic parameters and perceived qualities are not easily distinguishable. By employing complex missing-fundamental tones[10] that allowed one to separate the effects of spectral and virtual pitch[11] on the MMN, it was shown recently that the memory traces involved in the MMN process encoded these sounds in terms of their perceived (virtual) pitch rather than in terms of the spectral pitch of their components.[12] Missing-fundamental tones are perceived as having a pitch identical to that of their fundamental frequency, even though the fundamental itself is not present in the frequency spectrum of these complex tones.[10] A subsequent study revealed that resolving the missing-fundamental pitch requires relatively long acoustic samples.[13] This was demonstrated by the elicitation of MMN by 500 ms missing-fundamental tones but not at a stimulus duration of 150 ms. Moreover, the subjects' performance in discriminating the same complex tones indicated that the these sounds could only be discriminated at the longer stimulus duration.[13]

The goal of the present study was to determine whether or not MMN is elicited by infrequent changes in sound timbre with no corresponding change in pitch.[14] Timbre denotes the perceived auditory quality describing sound 'colour'. Timbre can be divided to three components: the temporal component (duration of the constant-amplitude part of the sound, rise and fall times), the spectral

Table 1. Partial (in Hz) used in composing the stimulation of the present study

Partial	Standards with 300 Hz missing-fundamental pitch									Deviant
1	1200	1200	1200	1200	1200	1500	1500	1500	1800	300
2	1500	1500	1500	1800	2100	1800	1800	2100	2100	
3	1800	2100	2400	2100	2400	2100	2400	2400	2400	

Each standard stimulus was constructed of three harmonic partials of equal intensity and aligned phase. Noise (with intensity of –40 dB relative to the summed intensity of the partials) covering the frequency band 200–800 Hz was added to the stimuli to avoid the inclusion of the fundamental frequencies by possible non-linear distortions in the stimulation unit.

component (the amplitude spectrum of simultaneous overtones), and the interaction of these two.[14] The present study investigated the spectral component of timbre by employing tones of identical amplitude envelope but of different harmonic composition. In addition to the primary goal, we wished to establish whether an accurate neural representation of missing-fundamental tone timbre requires samples of the stimuli as long as that needed for resolving their pitch. Elicitation of an MMN would suggest that the timbre of the corresponding tones was resolved and encoded in the memory underlying the MMN process.

Materials and Methods

Ninety percent of the auditory stimuli (standards) were harmonically rich missing-fundamental tones composed of three equally loud harmonic partials (1200–2400 Hz) of the 300 Hz fundamental tone. Nine different complex tones were presented, each with $p = 0.1$ (see Table 1 for frequency compositions). The common 300 Hz fundamental was not present in any of the standard sounds that were, however, perceived as having a (virtual) pitch corresponding to 300 Hz. The harmonically rich complex sounds were occasionally replaced by a 300 Hz pure tone, the deviant ($p = 0.1$). Two separate experiments, the first (short tone) employing 150 ms and the second (long tone) 500 ms tone durations, were conduced. Offset-to-onset ISIs were 350 and 400 ms, respectively.

Ten subjects (age 20–32 years, six females) participated in the short-tone experiment and 12 subjects (age 19–27 years, eight females) in the long-tone experiment. Subjects gave informed consent after the nature of the experiment was explained to them. During the EEG recordings, the stimuli were presented via headphones to the subjects' right ear while they were instructed to play a visual computer game (with no sound) and to ignore the auditory stimulation.

The EEG was recorded (0.1–100 Hz band pass; 250 Hz digitation rate; off-line low-pass filter: 30 Hz) with 11 electrodes attached to the scalp along the midline (Fpz, Fz, Cz, Pz, Oz) and the coronal arcs connecting the left and right mastoids (Lm and Rm, respectively) via Fz (Lm, L2, L1, (Fz), R1, R2, Rm). The common reference electrode was placed on the nose. The EOG was monitored from Fpz (vertical eye movements) and at the outer canthus of the right eye (horizontal eye movements, HEOG). EEG epochs (–50–400 ms with respect to stimulus onset) were averaged separately for standard and deviant stimuli. The ERP to the standard tones was subtracted from that to the deviant to quantify the MMN amplitude. The MMN amplitude was measured as the mean amplitude from the 110–160 ms interval and referred to the mean amplitude of the pre-stimulus period. Statistical significance of the obtained MMNs was tested by ANOVA comparisons of the standard and deviant ERPs recorded at Fz.

Results

Figure 1 shows that infrequent pure tones presented among harmonically rich tones elicited more negative ERPs in the 100–200 ms range from stimulus onset at both stimulus durations (for the short-tone experiment, $F(1,9) = 55.01$, $p < 0.001$; for the long-tone experiment, $F(1,11) = 34.85$, $p < 0.001$). In the short-tone experiment, this negativity peaked an average of 116 ms from deviant stimulus onset, whereas in the long tone experiment the peak ccurred at 144 ms.

Figure 2 illustrates that in both experiments, the amplitude of the negative difference peaked over fronto-central scalp areas and inverted its polarity at mastoid leads. This ties in well with the well-known auditory cortical origin of the MMN[15] supporting the notion that MMN was indeed elicited in both cases.

Discussion

The present data show that timbre deviant tones elicited the MMN with both employed sound duration. This suggests that the auditory memory traces involved in the MMN generating process pre-attentively encode the timbre of sounds on the basis of relatively short (150 ms) acoustic samples, as during the stimulus presentation subjects were engaged in a

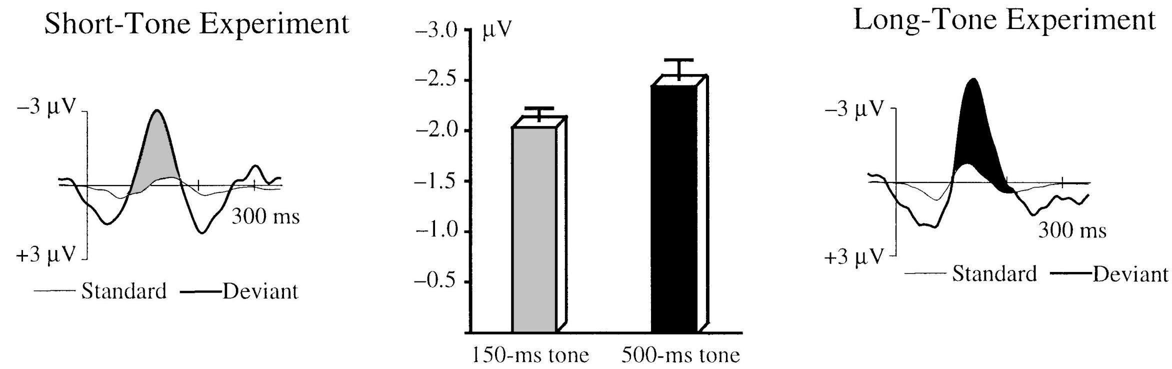

FIG. 1. Grand average frontal (Fz) ERPs elicited by infrequent pure tones (thick line) presented among frequent harmonically rich tones (thin line) with short (150 ms; left) and long (500 ms, right) stimulus duration. The MMN amplitude is marked by shading on the curve and is illustrated in the bar chart panel (the half range of s.e.m. is presented on the top of each bar).

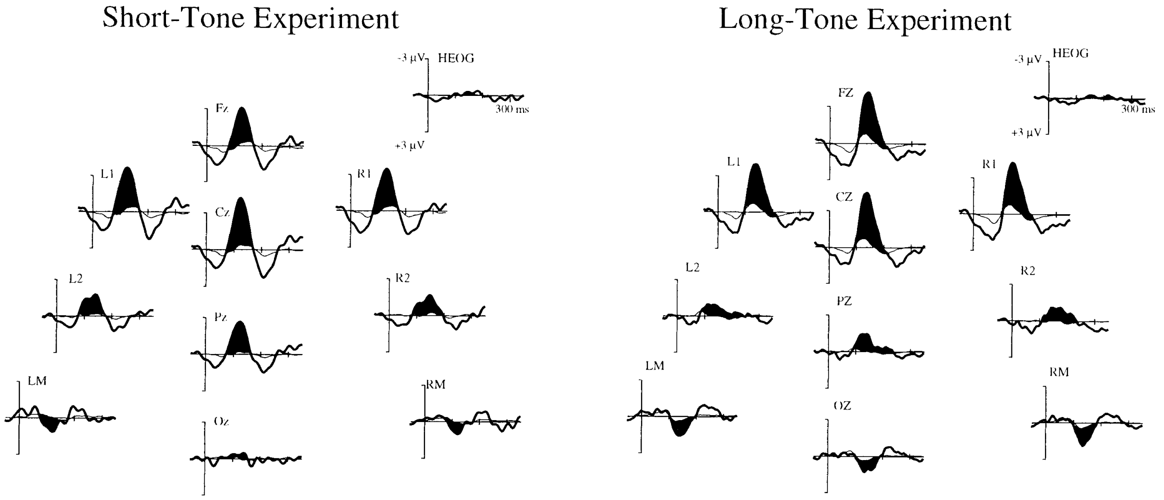

FIG. 2. ERPs elicited by deviant (thick line) and standard tones (thin line) with short (150 ms; left) and long (500 ms; right) stimulus durations from all recorded channels. Shaded areas indicate the MMN.

visual task and ignored the auditory stimuli. Previously Winkler *et al.*[13] found that extracting the pitch of missing-fundamental tones requires tone durations > 150 ms. Therefore, the present results support the notion that pitch and timbre are determined by separate processes in the human auditory system.[16] The precedence of timbre extraction to pitch extraction as shown by this dissociation in MMN elicitation is an ecologically important function of the auditory system: in natural environments, different sound sources typically differ in their characteristic sound spectrum. Consequently, a pre-attentive stream segregation process[17] based on timbre[18,19] facilitates the fast structuring of the acoustic environment.

The relatively early peak latency of the present negative difference and its partly central scalp distribution in the short-tone experiment supports the notion of a contribution from the supratemporal N1 component[2] to the observed deviant standard difference waves. This differential N1 response is probably due to the lack of the 300 Hz spectral component in the standard stimuli, whereas the deviant was a 300 Hz tone. However, the negative deviant standard difference could not have resulted entirely from the difference between the N1 response. First, Pantev *et al.*[20] showed that missing-fundamental tones activate the same tonotopically organized N1-generating neuronal populations as a pure tone of the corresponding fundamental frequency. Therefore, the afferent N1 neurons activated by the present deviant tone could not have responded much more vigorously to the deviant than to the standard tones. Second, Scherg *et al.*[21] demonstrated that with large physical separation between the standard and deviant stimuli, such as in this case, the N1 and MMN responses overlap each other. The earlier peak latency and more central scalp distribution of the deviant standard difference in the short-tone than the long-tone experiment suggests larger N1 contribution to the negative difference obtained in the short-tone experiment. As the onset-to-onset interval was shorter (500 ms) in the short-tone than the long-tone experiment (900 ms) the standard stimulus N1 was reduced (see Fig. 1). However, this N1 difference between the two experiments did not produce a corresponding effect on the deviant standard difference amplitudes suggesting that a large part of the negative difference is unrelated to the N1, i.e., it is a genuine MMN (cf. Ref. 21).

Because the exact timbre of each of the nine standard sounds was different, the MMN elicitation by the deviant tone suggests that the processes underlying MMN grouped the standard sounds together, contrasting them with the qualitatively different deviant timbre. This is because if all nine standard sounds were treated separately, then the pure tone could not have become 'deviant', since it has the same probability as each of the individual standard sounds. As was already reviewed, MMN is only elicited by infrequent changes on the basis of some detected constancy as encoded in standards.[5] There are two possible explanations of how the standard sounds were grouped together regarding their timbre. One explanation suggests that the difference between the individual timbres of the standard sounds was small enough for the auditory system to take it as a 'natural' variability within the same source. Previous results show that the MMN system tolerates some variability in the intensity of the standard ones (an ecologically important feature of this process), provided that the deviant stimulus is sufficiently distinct from the range of standard stimulus variance.[22]

The alternative explanation proposes that, although each harmonically rich standard tone was processed separately (as their spectral composition differed), the auditory system abstracted their timbres into one category, contrasting it with the qualitatively different timbre of the pure-tone deviant. A similar phenomenon has been observed by Tervaniemi *et al.*[23] They presented reading subjects with continuously descending Shepard tones. An MMN was elicited by ascending or repetitive tones when they were infrequently presented among descending tones, despite the variability in the absolute frequency level at which deviant events occurred. The authors concluded that MMN was elicited on the basis of a pre-attentively formed 'abstract' category of descending pitch (for similar results, see Ref. 24). One cannot distinguish between these two alternative explanations on the basis of the present evidence alone. However, the behaviour of the MMN response in the present study is compatible with the assumed role of the MMN-generating processes in maintaining an accurate model of the unattended acoustic environment: the temporary construct of 'rich timbre' enabled the auditory system to detect gross changes despite the equal nominal probabilities.[5]

Conclusion

The present results indicate that the spectral component of timbre is pre-attentively processed on the basis of relatively short sound samples and subsequently encoded into cortical sensory memory traces. In addition, adaptive categorization of sounds based on their timbre quality was demonstrated.

References

1. Johnson R Jr and Baron JC, eds. *Handbook of Neuropsychology* Vol 10. Amsterdam: Elsevier, 1995.
2. Näätänen R and Picton TW. *Psychophysiology* **24**, 375–425 (1987).
3. Näätänen R. *Attention and Brain Function*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1992.
4. Ritter W, Deacon D, Gomes H *et al. Ear Hearing* **16**, 51–66 (1995).
5. Winkler T, Karmos G and Näätänen R. *Brain Res* **742**, 239–252 (1996).
6. Näätänen R, Paavilainen P, Alho K *et al. Neurosci Lett* **98**, 217–221 (1989).
7. Cowan N. *Attention and Memory. An Integrated Framework*. Oxford: Oxford University Press, 1995.
8. Alho K. Woods DL, Algazi A *et al. Electroencephalogr Clin Neurophysiol* **82**, 356–368 (1992).
9. Winkler I, Cowan N, Csépe V *et al. J Cogn Neurosci* **8**, 403–415 (1996).
10. de Boer E. On the 'residue' and auditory pitch perception. In: WD Keidel and WD Neff, eds. *Handbook of Sensory Physiology*. New York: Springer-Verlag 1976: 479–533.
11. Terhardt E. *J Acoustic Soc Am* **55**, 1061–1069 (1974).
12. Winkler I, Tervaniemi M, Huotilainen M *et al. NeuroReport* **6**, 2317–2320 (1995).
13. Winkler I, Tervaniemi M and Näätänen R. *J Acoustic Soc Am* (in press).
14. Grey JM. *J Acoustic Soc Am* **61**, 1270–1277 (1977).
15. Alho K. *Ear Hearing* **16**, 38–50 (1995).
16. Semal C and Demany L. *J Acoustic Soc Am* **89**, 2404–2410 (1991).
17. Bregman AS. *Auditory Scene Analysis*. Cambridge, MA: MIT Press, 1990.
18. Culling JF and Darwin CJ. *Percept Psychophys* **54**, 303–309 (1993).
19. Iverson P. *J Exp Psychol Human Perc Perf* **21**, 751–763 (1995).
20. Pantev C, Hoke M, Lutkenhöner B *et al. Science* **246**, 486–488 (1989).
21. Scherg M, Vajsar J and Picton TW. *J Cogn Neurosci* **1**, 336–355 (1989).
22. Winkler I, Paavilainen P, Alho K *et al. Psychophysiology* **27**, 228–235 (1990).
23. Tervaniemi M, Maury S and Näätänen R. *NeuroReport* **5**, 844–846 (1994).
24. Saarinen J, Paavilainen P, Schröger E *et al. NeuroReport* **3**, 1149–1151 (1992).

Received 12 May 1997; accepted 27 May 1997

ACKNOWLEDGEMENTS: This work was supported by the Academy of Finland, the Finnish Society of Psychology and the National Scientific Research Fund of Hungary (OTKA T0022800).

General Summary
The present results revealed that timbre ('sound colour') is pre-attentively determined by the auditory system on the basis of short acoustic samples. Pre-attentive timbre-based categorization of sounds (demonstrated by these results) is an ecologically important process that facilitates fast segregation of the auditory input containing contributions from multiple sound sources. The contrasting results found from timbre and virtual pitch (in a previous study) suggest that timbre and pitch resolving are separate functions of the auditory system.