Simultaneous storage of two complex temporal sound patterns in auditory sensory memory

Elvira Brattico,[1,CA] Istvan Winkler,[1,2] Risto Näätänen,[1] Petri Paavilainen[1] and Mari Tervaniemi[1,3]

[1]Cognitive Brain Research Unit, Department of Psychology, P.O. Box 13, FIN-00014, University of Helsinki and Helsinki Brain Research Centre, Finland; [2]Institute of Psychology, Hungarian Academy of Sciences, Budapest, Hungary; [3]Institute für Allgemeine Psychologie, Universität Leipzig, Germany

[CA]Corresponding Author: elvirabrattico@supereva.it

Received 1 May 2002; accepted 21 July 2002

We wished to determine whether multiple sound patterns can be simultaneously represented in the temporary auditory buffer (auditory sensory memory), when subjects have no task related to the sounds. To this end we used the mismatch negativity (MMN) event-related potential, an electric brain response elicited when a frequent sound is infrequently replaced by a different sound. The MMN response is based on the presence of the auditory sensory memory trace of the frequent sounds, which exists whether or not these sounds are in the focus of the subject's attention. Subjects watching a movie were presented with sound sequences consisting of two frequent sound patterns, each formed of four different tones and an infrequent pattern consisting of the first two tones of one of the frequent sound pattern and the last two tones of the other frequent sound pattern. The infrequent sound pattern elicited an MMN, indicating that multiple sound patterns are formed at an early, largely automatic stage of auditory processing. *NeuroReport* 13:1747–1751 © 2002 Lippincott Williams & Wilkins.

Key words: Auditory event-related potentials; Auditory sensory memory; Mismatch negativity (MMN); Oddball paradigm; Pattern processing

INTRODUCTION

Natural auditory environments abound in sound events, most of which are perceived as sound groups or patterns [1,2]. Behavioral studies indicate that complex stimuli, such as melodies in music or words in language, are organized as patterns when they enter the focus of attention [3,4]. It has also been proposed that a primitive organization of stimuli can occur even in an automatic manner during the early stages of information processing, with the outcome of this analysis being reflected in auditory sensory memory [2,5]. The term auditory sensory memory [6,7] refers to a brief temporary buffer of acoustic information, which encodes all incoming sounds. The information stored in auditory sensory memory decays within about 10 s, unless it is selected by attention for further analysis, thus entering working memory [8]. Evidence that sound patterns may be formed early to be encoded in auditory sensory memory (and irrespective of their relevance to the ongoing behavioral activity) has been obtained by using the mismatch negativity (MMN) component of the auditory event-related potentials (ERPs) [9–13]. The MMN [14,15] is a brain response typically elicited when a repetitive sound sequence (the standards) is occasionally broken by some different sound (deviant), i.e., when the incoming sound does not match the memory representation of the standard sound.

The memory involved in the MMN generation is of sensory nature: it is not elicited when the deviants are separated from the standards by a time interval > 10 s and this interval cannot be extended or shortened by top-down processes [14–17]. Moreover, MMN is elicited when subjects are not performing any task related to the sounds, preventing the use of active silent rehearsal or other strategies functional to store sounds in working memory [8]. Thus, MMN can be used to study the contents of auditory sensory memory [14,16,17], including the storage of sound patterns having complex spectral or temporal structure [18–20].

Although much effort has been invested to determine the capacity limits of the simultaneous attentive storage of individual sounds [3,6,7], and of multiple sounds buffered into chunks [21], less is known about whether neural representations of multiple sound groups can be maintained pre-attentively at the level of auditory sensory memory (i.e. without the sounds being selected for further processing). However, at least two [22–24], and probably more [25] sounds can be simultaneously represented without chunking in the memory underlying the MMN process. Therefore, we used MMN to test the question whether representations of two complex temporal sound patterns can be stored simultaneously in auditory sensory memory. This question was studied by determining whether infrequent presentations of a tonal pattern (the deviant), combined from parts of the two frequently presented tonal patterns (standards), would elicit the MMN. Since the deviant pattern introduces no new tones, MMN elicitation by task-irrelevant sound pattern deviants would demonstrate that the frequent sounds were encoded in auditory sensory memory as two separate integrated sound patterns rather than as individual tones.

MATERIALS AND METHODS

Ten right-handed volunteers participated in the experiment (two males; 18–25 years, average 20.6 years). They were comfortably sitting in an electrically shielded sound-attenuated room, and were instructed to watch a silent movie (about which they had to fill a questionnaire after the experimental session) and to ignore sounds presented binaurally via headphones.

Three 4-tone patterns, with tonal segments differing from each other only in pitch, were presented in a pseudo-random order. Two of the patterns appeared with a frequency of $p = 0.47$, each (standards), whereas the third, the deviant pattern, which started with the first two tones of one of the standard patterns and ended with the last two tones of the other standard pattern was presented with a frequency of $p = 0.06$ (Fig. 1a). Thus the deviant introduced no new pitch (tone). The standard pattern with the first two tones identical to those of the deviant pattern will be referred to as the first standard and the other standard pattern as the second standard. At least four standards (of either type) preceded each deviant. Six stimulus blocks with 1000 patterns in each were delivered to the subjects.

Each pattern consisted of four 100 ms long pure tones (10 ms rise/fall time; 50 dB above the individually determined hearing threshold) separated by 20 ms silence

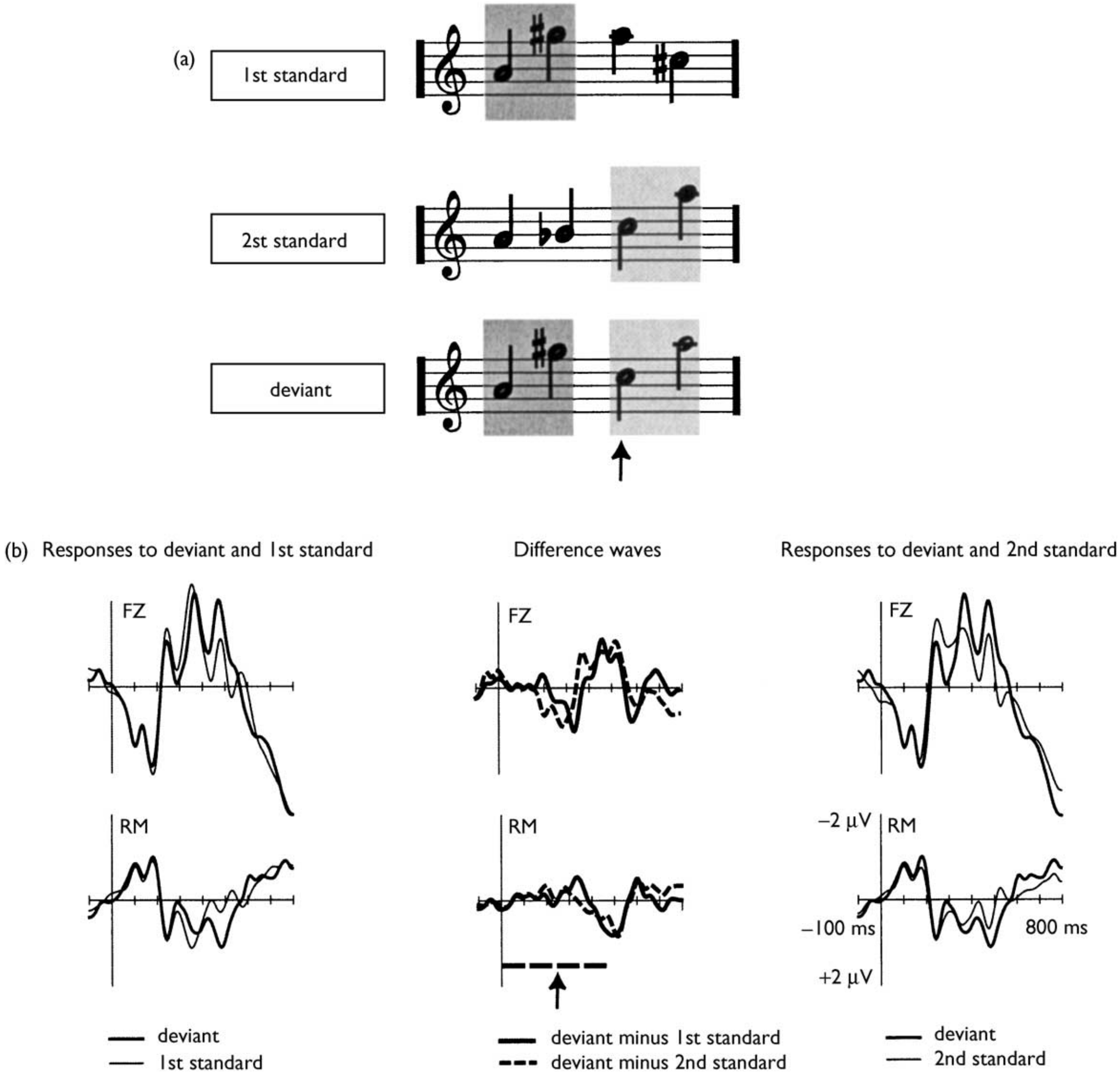

Fig. 1. (**a**) Experimental stimuli illustrated in musical notation. The frequencies of the notes follow the equal temperament tuning as reported by the American Standard Musical Notation. The deviant pattern consisted of the first two tones of the first standard pattern and the last two tones of the second standard pattern. Deviation from the first standard pattern, which started with the third tone in the deviant pattern, is indicated by the arrow. (**b**) Grand-average ERP responses to the deviant and the first standard (left) and the deviant and the second standard pattern (right) recorded from two electrode sites. In the middle column, the two corresponding deviant-minus-standard difference waves are shown. The time course of the sound patterns is illustrated at the bottom of the middle column. The arrow indicates the onset of the deviance.

(Fig. 1a). The total duration of each pattern was thus 460 ms. A total of six different tones were used in the patterns, two were common between the two standards plus the deviant pattern. Deviation onset was 240 ms from the onset of the deviant pattern, at which point the deviant started to differ from the first standard pattern. To match the patterns acoustically, all patterns started at the same pitch and were set within the same frequency range. Consecutive patterns were separated by 500 ms silence, so that they could be easily grouped by subjects into two separated units. The pitches used in the first standard pattern belonged to the A major scale, and the pitches of the second standard pattern belonged to the F major scale. The deviant belonged to a third musical scale, A minor. As shown by previous results [3], this helps the discrimination of the three patterns.

The electroencephalogram (EEG) was recorded from Fpz, Fz, Cz, Pz, from both mastoids, and from L1, L2, R1, and R2 (the one- and two-third locations on the arc connecting Fz with the mastoids on the left and right side, respectively). Horizontal eye movements were monitored from an electrode positioned beside the right eye and vertical eye movements from Fpz. The reference electrode was placed on the tip of the nose. EEG signals were amplified between 0.1 and 30 Hz and digitized with 250 Hz sampling frequency. EEG epochs (100 ms before and 800 ms after the pattern onset) were separately averaged for each stimulus type and then digitally filtered (passband 1–20 Hz). Responses were rejected from averaging when the signal exceeded $\pm 75\,\mu V$ at any electrode. The first 10 trials of each experimental block, standards immediately following a deviant, and standards following a sequence of ≥ 4 consecutive standards of the other type (as the latter two may elicit an MMN) were also rejected from the analysis.

For assessing the MMN response, the responses elicited by the two standards were separately subtracted from the response to the deviant. Peak latencies of the grand-average difference waveforms between the deviant- and the two standard-pattern responses were determined at Fz, Cz and at the mastoid leads from the 350–520 ms post-pattern onset window (110–280 ms with respect to the onset of the deviance). Mean response amplitudes from the 100 ms window centered at the positive (polarity reversed [14]) difference peak at the left mastoid (the electrode location showing the clearest peak) were separately measured for each stimulus category. Amplitude measurements were referred to the mean amplitude in the baseline (from 100 ms before pattern onset to deviation onset, at 240 ms from pattern onset). The significance of the amplitude differences was tested at Fz and at the mastoid leads by one-tailed *t*-tests (for negative difference at Fz and positive at the mastoids), separately between the deviant and the two standard patterns. Latency and amplitude MMN differences were tested at the Fz, Cz and mastoid electrodes by a two-way analysis of variance (ANOVA) with stimulus category and electrode as factors. All results reaching significance ($p < 0.05$) are reported.

RESULTS

Grand-average responses to the standard and deviant patterns and the respective difference waveforms are shown in Fig. 1b at Fz and the right mastoid, the electrode locations showing the largest effect. The response to the deviant pattern significantly differed from the response to the first standard pattern ($t(9) = -3.4$, 4.5, and 4.8; $p < 0.01$ for Fz, the left and right mastoids, respectively) in the MMN latency window (480 ms peak latency, 240 ms after deviance onset). The deviant stimulus response also significantly differed from the response to the second standard pattern ($t(9) = -3.5$, 4.7, and 5.1; $p < 0.01$; for Fz, the left and right mastoids, respectively) in the MMN latency window (500 ms peak latency, 260 ms after deviance onset). The scalp distribution of these difference potentials was compatible with the notion of the MMN component. Though the latency of the present difference response was late, MMN has been shown to occur as late as 250–280 ms after the deviance onset when responses to spectrally and/or temporally complex stimuli were used [19,20]. The non-significant latency difference between the MMN responses associated with the first *vs* the second standard was caused by the difference in the N1 amplitude for the second tonal segment in the first *vs* the second standard pattern.

The question of whether the two frequently presented patterns were simultaneously maintained in auditory sensory memory was further tested by determining whether MMN was elicited by those deviant patterns, which followed mixed sequences of the two standard patterns (i.e. not only by deviants following homogeneous sequences of one of the standards) [22]. Figure 2 shows that significant MMN responses were elicited both by deviants following a homogeneous sequence of the first standard in the preceding four positions ($t(9) = -4.8$; $p < 0.001$ at Fz; peak latency 445 ms from the pattern onset, 205 ms after deviance onset) as well as by deviants following mixed sequences of the two standard patterns in the last four positions ($t(9) = -3.4$; $p < 0.01$ at Fz; peak latency 500 ms from pattern onset, 260 ms after deviance onset). The MMN peak latency measured at the mastoids was significantly longer for deviants following mixed than those following homogeneous standard sequences ($t(9) = 2.9$, 2.3; $p < 0.05$, for the left and right mastoids, respectively); no significant difference for the MMN peak latency was found at Fz and Cz.

DISCUSSION

The present experiment aimed at determining whether representations of two complex sound patterns composed of several individual tones can be simultaneously maintained in auditory sensory memory. The elicitation of MMN by deviant patterns combining the tonal parts of the two standard patterns (thus introducing no new tones) indicates that the frequent sounds were represented in terms of patterns, not as individual tones, in auditory sensory memory.

Representations for both standard patterns were maintained simultaneously as confirmed by the finding of MMN in response to deviants following mixed sequences of the two standard patterns. However, the MMN elicited by deviants following mixed sequences of the standards had a longer peak latency than the one elicited by deviants following a homogeneous sequence of one of the standard patterns. In a previous study [22], tone deviants following homogeneous sequences of one of the standard tones elicited MMNs with larger amplitude than deviants following mixed standard sequences. This and the present results suggest that one sound representation is more easily activated than two. The MMN latency difference, found only at the mastoid leads but not at frontal or central electrode locations, supports the notion that the fronto-centrally recorded MMN (the frontal MMN subcomponent) and the MMN observed at the mastoid leads (the temporal MMN subcomponent) reflect two separate, but linked processes. The neural activity involved in generating the temporal MMN probably reflects the detection of sound change, whereas the frontal MMN is probably involved in calling for a switch of attention to the deviant stimulus [14,26].

The present result that the frequent sounds were represented in terms of patterns indicates that the precise succession of absolute frequency relationships (intervals) between components in each pattern was also stored in sensory memory. Previous studies have shown that the auditory system pre-attentively encodes frequency ratios, even when their absolute frequencies are varied [20,27], as well as the temporal/sequential structure of sound patterns [19,28].

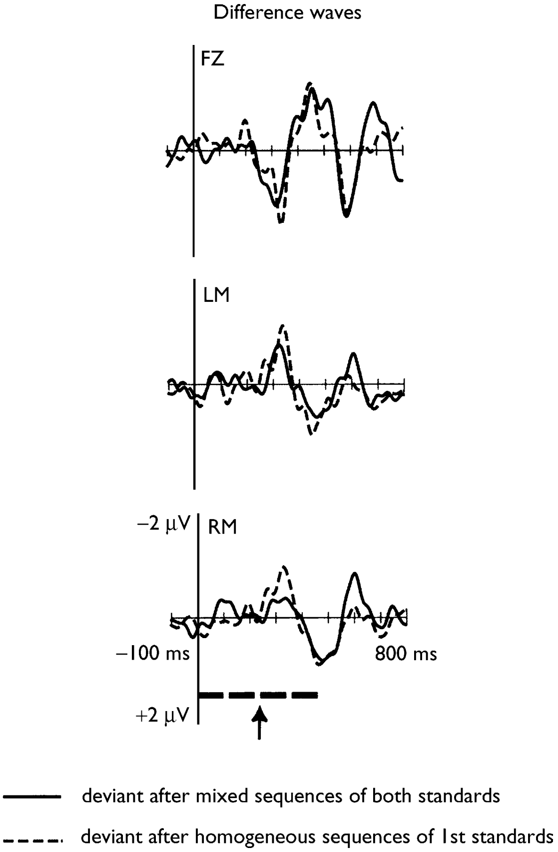

Fig. 2. Deviant-minus-first-standard difference waves as a function of the sequence preceding the deviant. The solid line illustrates the difference response when the deviant followed a mixed sequence of both standards in the preceding four positions. The dashed line shows the difference response when the deviant followed a homogeneous sequence of the first standard in the preceding four positions. The time course of the patterns is illustrated at the lower corner of the figure. The arrow marks the onset of the deviance.

Most theorists claimed that melodies in music, or sentences in language are encoded in working memory and require attentional effort to be memorized [3]. With the present study we were able to show that already at the 250 ms latency and without attentional effort a neural mismatch process is generated in the auditory cortex when the neural sensory memory trace does not correspond to either of two memorized sound patterns. This supports the emerging view [17,28–30] that auditory sensory memory is not a static storage system, dealing with single acoustic events. Rather it contains a composite model of the acoustic environment able to simultaneously maintain two temporally complex sound events, even when these events consist of task-irrelevant sine tones devoid of any meaning. These results suggest that, similarly to the visual modality, in which brief afterimages of multiple objects are essential to integrate them within the visual scene [31], our effortless perception of complex sound groups, such as those appearing in speech and music, is based on early, largely automatic functions of the auditory system that dynamically store separated tones as integrated regularities of the acoustic environment.

CONCLUSION

We showed that at least two sound patterns can be simultaneously represented in the brief acoustic buffer, even when subjects do not attend the sounds. This result supports the view that the auditory system processes sound information in an elaborate way even in the absence of focused attention. For example, it can separately group multiple complex temporal sound patterns and simultaneously maintain their representations. The present results shed light on some of the mechanisms underlying everyday perception, the effortless processing of natural auditory environments.

REFERENCES

1. Watson CS, Wroton HW and Benbassat CA. *J Acoust Soc Am* **57**, 1175–1185 (1975).
2. Bregman AS. *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge (MA): MIT Press; 1990.
3. Deutsch D. In: Deutsch D, ed. *The Psychology of Music*. San Diego: Academic Press; 1999, pp. 349–411.
4. Treisman A. *J Exp Psychol Hum Percept Perf* **8**, 194–214 (1982).
5. Duncan J and Humphreys G. *Psychol Rev* **96**, 433–458 (1989).
6. Cowan N. *Psychol Bull* **96**, 341–370 (1984).
7. Massaro DW. *Experimental Psychology and Information Processing*. Chicago: Rand McNally; 1975.
8. Baddeley A. *Human Memory: Theory and Practice*. Boston: Allyn and Bacon; 1990.
9. Shinozaki N, Yabe H, Sato Y *et al. Neuroreport* **11**, 1597–1601 (2000).
10. Sussman E, Ritter W and Vaughan HGJ. *Neuroreport* **9**, 4167–4170 (1998).
11. Sussman E, Ritter W and Vaughan HGJ. *Psychophysiology* **36**, 22–34 (1999).
12. Winkler I, Schröger E and Cowan N. *J Cogn Neurosci* **13**, 1–13 (2001).
13. Yabe H, Winkler I, Czigler I *et al. Brain Res* **897**, 222–227 (2001).
14. Näätänen R. *Attention and Brain Function*. Hillsdale: Erlbaum; 1992.
15. Picton TW, Alain C, Otten L *et al. Audiol Neuro-Otol* **5**, 111–139 (2000).
16. Ritter W, Deacon D, Gomes H *et al. Ear Hear* **16**, 52–67 (1995).
17. Näätänen R and Winkler I. *Psychol Bull* **125**, 826–859 (1999).
18. Alho K, Tervaniemi M, Huotilainen M *et al. Psychophysiology* **33**, 369–375 (1996).
19. Näätänen R *et al. Neuroreport* **4**, 503–506 (1993).

20. Tervaniemi M *et al. Learn Mem* **8**, 295–300 (2001).
21. Cowan N. *Behav Brain Sci* **24**, 87–185 (2001).
22. Winkler I, Paavilainen P and Näätänen R. *Psychophysiology* **29**, 337–349 (1992).
23. Gomes H, Bernstein R, Ritter W *et al. Psychophysiology* **34**, 712–716 (1997).
24. Takegata R, Paavilainen P, Näätänen R *et al. Neurosci Lett* **266**, 109–112 (1999).
25. Winkler I, Paavilainen P, Alho K *et al. Psychophysiology* **27**, 228–235 (1990).
26. Rinne T, Alho K, Ilmoniemi RJ *et al. Neuroimage* **12**, 14–19 (2000).
27. Paavilainen P, Jaramillo M, Näätänen R *et al. Neurosci Lett* **265**, 179–182 (1999).
28. Näätänen R, Tervaniemi M, Sussman E *et al. Trends Neurosci* **24**, 283–288 (2001).
29. Horváth J, Czigler I, Sussman E *et al. Cogn Brain Res* **12**, 131–144 (2001).
30. Winkler I, Karmos G and Näätänen R. *Brain Res* **742**, 239–252 (1996).
31. Henderson JM and Hollingworth A. *Annu Rev Psychol* **50**, 243–271 (1999).

Acknowledgements: Supported by the Academy of Finland, the Pythagoras Graduate School in Cognitive Musicology, and the Hungarian National Science Fund (OTKA T034112). We thank Titia van Zuijen, Valentina Gumenyuck, Oleg Korzyukov, and Rika Takegata for their helpful comments and suggestions on the previous versions of the manuscript.