Human auditory cortex tracks task-irrelevant sound sources

István Winkler, Wolfgang A. Teder-Sälejärvi, János Horváth, Risto Näätänen and Elyse Sussman

The brain organizes sound into coherent sequences, termed auditory streams. We asked whether task-irrelevant sounds would be detected as separate auditory streams in a natural listening environment that included three simultaneously active sound sources. Participants watched a movie with sound while street-noise and sequences of naturally varying footstep sounds were presented in the background. Occasional deviations in the footstep sequences elicited the mismatch negativity (MMN) event-related potential. The elicitation of MMN showed that the regular features of the footstep sequences had been registered and their violations detected, which could only occur if the footstep sequence had been detected as a separate auditory stream. Our results demonstrate that sounds are organized into auditory streams irrespective of their relevance to ongoing behavior.

Keywords: Auditory scene analysis; Auditory stream segregation; Change detection; Event-related brain potentials (ERP); Mismatch negativity (MMN)

INTRODUCTION

How does auditory cortex form meaningful representations of multiple sources in a real-life setting? Research into this question has only begun in recent years [1]. One possibility is that the auditory system builds and maintains an extensive model representing all of the distinct, simultaneously active sound sources, irrespective of their behavioral relevance. This would allow for faster detection and subsequent selection of salient acoustic events occurring outside the focus of our attention [2,3]. However, continuous updating of the representations for several simultaneously active sound sources would require extensive processing and memory resources. A more economical mode of operation would distinguish only the task-relevant (attended) sound source, thereby creating merely a foreground-background description of the auditory input.

We investigated this issue by recording event-related brain potentials (ERPs) in an ecologically valid setting. ERPs are a non-invasive electrophysiological measures of cortical responses to sensory and cognitive events that include time-sensitive and robust signatures for stimulus change. The current experiment used the mismatch negativity (MMN) response, an electric correlate of a process in the human auditory cortex elicited when a sound deviates from regular aspects of the preceding sound sequence [4,5]. MMN is elicited irrespective of the relevance of the sounds to the ongoing behavior [4]. It has been recently established that MMN can be used to test the segregation of multiple sound streams [6]. This method is based on the finding that adding random sounds into a regular sound sequence, or mixing two sequences with different regular features prevents the brain from detecting the original auditory regularities. When no regular features are found in a sound sequence, no MMN can be elicited by any sound belonging to that sequence. However, if the original and the added sounds are segregated into separate auditory streams, then regular features of the original sound sequence can again be detected and sounds deviating from the detected regularity elicit the MMN [7]. We used this new technique to test whether two sets of task-irrelevant sounds (emitted by separate sources) would be segregated when participants attended sounds coming from a third source. All sounds were naturally produced complex stimuli. Thus, we investigated whether the same principles of stream segregation and change detection applied to complex naturally occurring environmental sounds as those observed for the highly artificial sounds used in most experiments.

If the auditory system created separate auditory streams for all of the simultaneously active sound sources in the environment, then MMN would be elicited by deviants inserted into one of the background sound sequences. If, however, the auditory system only distinguished
attended sound source from the background, then the two sets of irrelevant sounds would be merged into an undifferentiated background and no MMN would be elicited.

MATERIALS AND METHODS
Eight healthy volunteers (three female, 18–25 years of age, mean 20.3 years) participated in the experiment. One subject was recalled for a second session to measure the scalp distribution of the MMN component. Participants gave informed consent after the procedures of the experiment were explained to them. The study was approved by the ethical committee of the Institute for Psychology of the Hungarian Academy of Sciences where the experiment was conducted.

Figure 1a depicts the arrangement in the experimental chamber. Participants watched and listened to a movie. During the experiment, sounds coming from a loudspeaker simulated city street noise entering through an open window. Two additional loudspeakers occasionally delivered the footsteps sequences, which appeared to be crossing the room behind the subject. This setting provided a close approximation of a realistic viewing scenario. Only one of three sound sources (the TV movie) was task relevant, whereas the remaining two (the street noise and the footsteps) were task irrelevant.

Eleven different digitized natural footsteps were presented through two loudspeakers placed symmetrically 1 m apart and 0.5 m behind the participant’s head. The illusion of a person walking across the room was created by linear cross-fading of the footstep sequence of the 11 step sounds. The step sounds were progressively attenuated with a linear function, which sloped from 0 to 1 on the right channel and vice versa on the left sound channel during the course of the test footstep sequence (Fig. 1a, upper horizontal arrow). The control footstep sequence was created by the same procedure except for reversing the order of the individual steps and exchanging the fading functions between the two loudspeakers (Fig. 1a, lower horizontal arrow). This procedure resulted in a series of footsteps apparently crossing the room in the opposite direction compared with the test sequence. The black traces in Fig. 1b show that all steps were slightly different, thereby simulating the genuine spectral variability of a realistic walking pattern. The combined loudness of the two loudspeakers varied between 60 and 72 dB (measured at the participant’s head; Fig. 1c).

A sample of the summed movie sound and street noise is shown by grey lines on Fig. 1b. EEG was measured with a 20-channel electro-cap electrode cap (electrode locations: Fp1, Fp2, F3, Fz, F4, F7, F8, C3, Cz, C4, P3, Pz, P4, T3, T4, T5, T6, O1, O2) [8] and from

![Fig. 1. Stimulation. (a) The experimental setup. The participant watched and listened to a movie presented on a TV screen directly in front of him. Series of footsteps perceived as moving from left to right (test sequence; upper arrow) or right to left (control sequence; lower arrow) were delivered by a pair of loudspeakers placed symmetrically on two sides, slightly behind the participant’s head. Street noise was delivered through a loudspeaker placed directly behind the subject. (b) The test sequence signal (black line) and a sample of the summed movie sound and street noise (grey line) in the time domain. Amplitude is in arbitrary units. (c) Frequency spectrum (relative intensity in dB) of the three types of sounds in the 0–10 kHz frequency interval: footstep sequence (continuous line), movie sound sample (dashed line), and street noise (dotted line).](image-url)
typically appears in the 100–200 ms post-stimulus interval with maximal amplitude over the frontal scalp [4,5].

RESULTS

Figure 2a shows the grand-averaged responses elicited by the deviant and the identical control sound, together with the respective difference waveform. Over the fronto-central scalp, the ERP response elicited by the deviant in the 100–200 ms interval was significantly more negative compared with that elicited by the control sound ($t(7) = 2.39$, $p < 0.03$). The difference waveform peaked 160 ms from the onset of the deviant step.

The scalp distribution of the difference waveform (Fig. 2b; recorded from one recalled subject) closely replicated previous findings of the scalp topography of the MMN potential [9]. Therefore, judged by its latency and scalp distribution, the deviant minus control difference response was identified as a MMN component.

**DISCUSSION**

Recent studies have shown that the segregation of two concurrent auditory sequences precedes and determines the elicitation of the MMN component [6,7,10,11]. As discussed in the Introduction, MMN can only be elicited within sound sequences, whose regular aspects have been detected by the auditory system. In the current study, this could only occur if the footstep sounds were segregated from all other concurrent sounds. Therefore, the elicitation of a MMN response in a background sound sequence, which was delivered concurrently with another irrelevant sequence and one relevant sequence, provided strong evidence that all three sound sources of the current experiment were segregated from each other and their regular features were encoded into separate memory records.

Given that the control sound was identical to the test sound, we can rule out the possibility that the difference between the ERP responses elicited by the deviant and the control sound could have been caused by stimulus-specific neural activity. Furthermore, because the control sound appeared in an earlier position of the footstep train than the deviant sound, neural refractoriness would only be greater for the deviant than for the control sound. Therefore, the amplitude of the obligatory N1 response [12] may have been higher, but not lower, for the control response than for the deviant response. If the control-sound N1 had indeed higher amplitude than the deviant-sound N1, then the deviant minus control difference was significant, possibly underestimating the size of the MMN response. However, since the deviant minus control difference was significant, possibly underestimating the MMN amplitude did not effect the present conclusion.

Only the test sounds could elicited the MMN, because they followed nine regular footsteps sounds that provided enough repetition to establish a regularity. The control sound, which was delivered in the second position of its sequence, appeared at a time when no regularity had yet been established for that sequence [13]. Therefore, the difference between the electrical responses elicited by the identical task-irrelevant test and control sounds could only reflect the elicitation of MMN.

---

**Fig. 2.** Brain responses. (a) Grand-averaged ($n = 8$) electric brain responses elicited by the deviant (continuous grey line) and identical control sounds (dashed grey line) measured from the frontal midline (Fz) electrode (the electrode is marked by black circle on the scalp-distribution map below). The MMN response is marked with a chequered pattern on the corresponding deviant minus control difference waveform (black line). The y (amplitude) axis marks the onset of the deviant/control sound (0 ms). (b) Grey-scale-coded (see scale on the right) distribution of the MMN electric potential distribution over the scalp (measured from 31 locations marked by black dots) at the peak latency (160 ms) of the deviant-minus-control difference response (one representative participant).
The scalp topography of the current difference waveform (Fig. 2b) and the well-known generator location of the MMN response [14,15] suggest that the current MMN was mainly generating in auditory cortex. This means that the neural representations encoding the characteristics of active sound sources provide information for processes operating in the auditory cortex, such as for the MMN-generating change-detector function.

Our results demonstrate that in natural situations, auditory cortical processes maintain separate records of each active sound source, not just distinguishing the currently relevant source. This seems remarkable considering that there are generally no simple acoustic features by which the auditory system can accurately separate the sound sources (see Fig. Ib,c showing this for the current experiment). The processes organizing the auditory input rely largely on temporal regularities of the active sound sources [1]. Therefore, detection of temporal regularities, such as those that characterized our footstep sequences, is a crucial prerequisite for navigating one’s way in the complex auditory world.

In conclusion, our results establish the existence of a multi-source neural model of natural acoustic environments in the human brain. Information stored in these memory records can be accessed by processes operating in auditory cortex. The functional significance of maintaining representations of sound sources that have no current behavioural relevance is that they facilitate rapid refocusing of one’s attention. The delay in establishing a new source representation has been reported to be in the order of 3–5 s [1]. This delay would be much too long in natural situations, in which rapid attention switching is imperative for detecting and reacting to warning signals that occur outside the focus of attention [16]. The memory records encoding the features of auditory streams, which underlie the results of the current study, provide the basis for a natural early warning system in auditory cortex that flags unattended, potentially significant acoustic events.

REFERENCES

Acknowledgements: This research was supported by the Hungarian National Research Fund (OTKA T034112), the National Institutes of Health (ROI DC04263), and the Academy of Finland.