We investigated the brain mechanisms enabling one automatically discriminate phoneme categories irrespective of the large inter-speaker variability in the acoustic features of the voices. For this purpose, subjects were presented with 450 different speech stimuli, each uttered by a different speaker, belonging to three vowel categories, while a 306-channel magnetoencephalogram (MEG) was obtained to record the magnetic counterpart of the mismatch negativity (MMNm), elicited only when sensory memory traces for repetitive sounds are formed in the auditory cortex. Despite this wide acoustic variation, category changes elicited prominent MMNm responses, which were considerably stronger in the left than in the right hemisphere in the right-handed subjects. These results implicate the presence of long-term memory traces for vowels, which can recognize the vowel-specific invariant code enabling correct vowel percept even in the presence of realistic acoustic variation.

Key words: Categorical perception; Magnetoencephalography (MEG); Mismatch negativity (MMN); MMNm

INTRODUCTION

Speech stimuli are perceived as belonging to certain language-specific categories first demonstrated by Liberman et al. [1], even in the presence of large acoustical variation, which can be associated, for example, with speaker, condition, or word-context change. In search for the neurophysiological basis of categorical speech-sound perception, it was found that the language-specific memory traces of speech sounds play a central role [2–6]. These memory traces can be probed by the mismatch negativity (MMN) component of the auditory event-related potential (ERP), which is elicited by occasional changes in unattended sounds [7]. Language-specific memory traces appear to be located in the left auditory cortex [5,6,9] and accommodate irrelevant acoustic variations in experiments employing synthetic speech stimuli [2,6,10]. The MMN and its magnetic counterpart (MMNm) may also be elicited by an interruption of abstract features or rules in the otherwise continuously changing auditory stimulation [11–13]. The categorical perception of speech and other sounds varying in acoustic content can be regarded as a particular type of abstract invariant acoustic feature extraction [14]. Some experiments [3,10] using synthetic speech stimuli have shown that permanent traces of phoneme categories can accommodate acoustical variation, whether that variation was due to inter-speaker differences in pitch or timbre, or word context (allophonic variation). The present study aimed at extending the previous results to the real-life situation where we have to perceive speech of different speakers. The brain activity was measured with a whole-head MEG device, which allowed one to compare the activity in the two hemispheres. This study was approved by the Ethical committee of the Department of Psychology, the University of Helsinki.

MATERIALS AND METHODS

Nine healthy (eight right-handed and one left-handed), normal-hearing native speakers of Russian (age 23–33, mean 29 years, five males) served as subjects. The stimulus sequences comprised 450 examples of /i/, /a/, or /u/ phonemes (150 exemplars per each category), each spoken by a different male native speaker. Stimuli were recorded via a fixed telephone network channel, digitized at 8000 Hz and 16 bits, edited to have a duration of 250 ± 20 ms, and normalized in amplitude. Three phoneme categories were chosen so that their distances from each other were the largest in the two-formant coordinate space for Russian
The subjects reported having no difficulty in classifying the stimuli. Presenting speech sounds produced by different speakers but belonging to the same phoneme category for several times, we intended to form a memory trace for the specific vowel category in the subject’s auditory sensory memory system (Fig. 1). As a consequence of the acoustic variation of the speech sounds, their sensory memory trace is formed according to phoneme categories thus activating the permanent phoneme recognition traces. After repeating the same category 3–6 times, we broke the expectation for the specific phoneme category by presenting a stimulus from another phonetic category. We called this experimental procedure roving-standard paradigm [16] in order to stress the fact that a new standard category is formed after each deviant. The stimuli immediately following the change of a category were regarded as deviants, while the stimuli following ≥ 3 repetitions of the same category were regarded as standards.

The stimuli were binaurally delivered through plastic tubes with Presentation software (Version 0.43, Neurobehavioral Systems Inc., USA) at a constant the inter-stimulus onset-to-onset interval (stimulus-onset asynchrony, SOA) of 900 ms and intensity 55 dB above the individual hearing threshold estimated using the same stimuli.

Subjects sat in a magnetically shielded chamber (Euroshield Ltd., Finland) watching a silent video of their own choice. They were instructed to pay no attention to the auditory stimuli. The sounds were presented until ≥ 300 artifact-free MEG responses had been collected for the deviant stimulus (approximate block duration of 35–40 min).

The MEG responses were recorded with a whole-head magnetometer (a 306-channel Vectorview system, 4-D Neuroimaging Oy, Finland). Eye movements and blinks were recorded with vertical and horizontal bipolar electrooculograms (EOG). MEG epochs (sampling rate 603 Hz, passband 0.03–90 Hz) starting 100 ms before and ending 500 ms after stimulus onset were separately averaged for standards and deviants: epochs contaminated by extracerebral artefacts (EOG variation > 150 μV or MEG variation at any channel > 3000 fT/cm during the epoch) were automatically omitted. Averaged responses were digitally off-line filtered with a passband of 1–20 Hz and baseline-corrected off-line in relation to the mean amplitude at a 50 ms pre-stimulus interval.

MMNm responses were obtained by subtracting the response to standard stimuli from that to deviants (Fig. 2a). The MMNm responses were evaluated using 54–63 channels (containing triplets of two gradiometers and one magnetometer) above each temporal area separately for each subject’s left and right hemispheres. The MMNm sources were estimated by means of sequential equivalent current dipole (ECD) fitting in the latency range of 100–250 ms from stimulus onset using a spherical head model [17,18]. Similarly, source generators of the N1m elicited by standard stimuli were estimated at the latency range of 80–150 ms.

The MMNm generator sources over each hemisphere were modelled in five right-handed subjects (Fig. 2b, Fig. 3) and one left-handed subject. In the remaining three subjects, the MMNm sources in the right hemisphere could not be modelled because of the weakness of the response. In these subjects, a mirroring procedure was used to estimate the
strength and direction of the source in the right temporal cortex at a location symmetrical to that of the left-hemisphere MMNm ECDs. The dipole moments of the MMNm generators in the two hemispheres were compared with each other by means of a two-tailed t-test. (The data of the left-handed subject were excluded from the calculation of the mean values and from the statistical comparison.)

RESULTS
The ECDs of the MMNm sources explained on the average 81% \((n = 8)\) of magnetic signals over the left temporal cortex and 75% \((n = 5)\) over the right one. The mean values of the MMNm dipole moment were significantly larger in the left \((12 \text{ nAmp})\) than right \((8 \text{ nAmp})\) hemisphere in the right-handed subjects \((t(7) = 6.29, p < 0.0005)\). In the left-handed subject, the MMNm dipole was stronger in the right than left hemisphere. The MMNm source dipole moments for each of the nine subjects are presented in Fig. 4. In contrast, the N1m source dipole moments did not significantly differ between the hemispheres \((23 \text{ vs } 21 \text{ nAmp for the left and right hemisphere, respectively}; t(7) = 0.59, p < 0.56)\). The MMNm latencies and locations could not be subjected to statistical analysis because the ECDs could not be calculated in the right auditory cortex in three of the right-handed subjects. However, the MMNm activity tended to start earlier in the left than right hemisphere \((123 \text{ and } 142 \text{ ms, respectively}, n = 5)\).

DISCUSSION
The main finding of the present study was that even when natural speech stimuli constantly changing and varying in acoustic content are used, an MMNm to the phoneme-category change is elicited. This result confirms and extends previous findings involving the phonological MMN recorded using synthesized stimuli [2,3,6]. Thus far, the MMN and MMNm studies of speech perception have been constrained by using relatively few stimuli in order to control for the acoustical variation.

In the present study, the experimental paradigm, where speech sounds were uttered by hundreds of speakers, and recorded outside the laboratory (thus taking in the noise of the background), offered an opportunity to test an ecologically valid situation. In fact, if we are not usually exposed to speech sounds uttered by continuously different speakers, we everyday effortlessly listen to and extract meaningful sounds (phonemes, words, phrases), from the countless and noise-masked variations of our auditory environment. Thus, the present MMNm results help us understanding the neural bases of this human ability.

The MMN or MMNm may be elicited not only to the replacement of single physical features of the sounds (like duration, frequency and etc.), but also to an interruption of abstract features or rules in the otherwise continuously
changing auditory stimulation [12,13]. The auditory sensory memory system, as reflected by MMNm, seems to be able to extract phonetically invariant information (another type of abstract feature) out of extensive acoustic variation. This conclusion is based on the main finding of the present study. The dipole moments indicated the dominance of the MMNm response in the left hemisphere, whereas the N1m response showed no such lateralization in the right-handed subjects. The MMNm asymmetry observed in the absence of the N1 hemispheric differences, together with previous finding [9] for consonant-vowel syllable, suggests the existence of long-term memory traces, i.e. preformed templates, for categorical processing of speech stimuli in the left hemisphere. A tendency for an earlier MMNm response over the left than over the right auditory cortex also suggested strong hemispheric lateralization for the categorical perception of vowels, consistent with that for language-specific recognition traces [2,5,19].

CONCLUSION

In the present study, MMNm indexed the ability of the central auditory system to automatically extract the invariant relevant information for vowel recognition even from acoustically continuously changing speech sounds uttered by hundreds of speakers. This process seems to be strongly lateralized to the left auditory cortex when natural speech sounds are used.

REFERENCES

<table>
<thead>
<tr>
<th>Manuscript Page/line</th>
<th>Details required</th>
<th>Author’s response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No Author Query.</td>
<td></td>
</tr>
</tbody>
</table>