Superior temporal and inferior frontal cortices are activated by infrequent sound duration decrements: An fMRI study

Teemu Rinne, a,b,c,d,e,* Alexander Degerman, a and Kimmo Alho a

a Department of Psychology, University of Helsinki, Finland
b BioMag Laboratory, Medical Engineering Centre, Helsinki University Central Hospital, Finland
c Cognitive Brain Research Unit, Department of Psychology, University of Helsinki, Finland
d Helsinki Imaging Center, University of Helsinki, Finland
e Human Cognitive Neurophysiology Laboratory, UC Davis and VANCHCS, Martinez, CA 94553-4612, USA

Received 25 June 2004; revised 1 December 2004; accepted 14 January 2005
Available online 3 March 2005

Functional magnetic resonance imaging (fMRI) was used to examine the processing of infrequent changes occurring in an unattended sound sequence. In event-related brain potentials (ERPs), such sound changes typically elicit several responses, including an enhanced N1, the mismatch negativity (MMN), and the P3a. In the present study, subjects were presented with a repeating sound of 75 ms in duration, which was occasionally replaced, in separate blocks, by a 15-ms, 25-ms, or 35-ms sound (large, medium, and small change, respectively). In the baseline block, only the frequent 75-ms sound was presented. During the scanning, the subjects were instructed to ignore the sounds while watching a silent wildlife documentary. We assumed that in this condition, the MMN mechanism would contribute more to the observed activation than the other change-related processes. We expected sound changes to elicit fMRI activation bilaterally in the supratemporal cortices, where the electric MMN is mainly generated, and that the magnitude of this activation would increase with the magnitude of sound duration change. Unexpectedly, however, we found that only blocks with medium duration changes (25 ms) showed significant activation in the supratemporal cortex. In addition, as reported in some previous EEG and fMRI studies, contrasts between different levels of sound duration change revealed additional activation in the inferior frontal cortex bilaterally. This activation tended to be greater for the small and medium changes than for the large ones.

© 2005 Elsevier Inc. All rights reserved.

Keywords: fMRI; Superior temporal cortex; Inferior frontal cortex; Auditory change detection; MMN

Introduction

The present study used functional magnetic resonance imaging (fMRI) to examine processing of infrequent sound changes occurring in a repetitive train of unattended sounds. Most of the previous knowledge about brain mechanisms involved in processing of such sound changes is based on electric and magnetic brain responses recorded with electroencephalography (EEG) and magnetoencephalography (MEG), respectively. This research has largely focused on certain components of the event-related brain potential (ERP) such as the N1, mismatch negativity (MMN), and P3a. The N1 (peak latency about 100 ms from sound onset) is generated by a fast change in acoustic energy (e.g., sound onset) and its amplitude is determined by the physical properties of the sound. A large N1 is typically elicited by sounds presented at a slow rate or in the beginning of a stimulus train (Hari et al., 1982; Näätänen and Picton, 1987). In contrast, sounds presented at a fast presentation rate will typically elicit only a small N1. However, if the pitch of a sound in a fast-rate repetitive sequence is occasionally changed significantly (e.g., from 1000 Hz to 2000 Hz), then a slightly enhanced N1 is elicited by these infrequent sound changes (Scherg et al., 1989). It is assumed that this enhancement is due to activation of new non-refractory neural populations. In addition to an enhanced N1, infrequent pitch changes also elicit a subsequent MMN response. The MMN is generated by a mechanism that detects deviations from regular aspects of the ongoing auditory stimulation by comparing the incoming sensory information to a representation formed by the repetitive features in preceding auditory inputs (Näätänen and Winkler, 1999). Such a regular feature could be, for example, a repeating single sound, a repeating tone pattern, or even an invariant higher level relationship between sounds (Näätänen et al., 2001). Thus, it is assumed that N1 and MMN reflect the activation of fundamentally different processes: while the N1 is generated by a fast change in the stimulus energy level (stimulus onset) and its amplitude is determined by the physical properties of the sound, the MMN is elicited by changes
in (higher-level) representations formed on the basis of repeating auditory information (Naätänen and Winkler, 1999). It has been suggested that both the N1 and MMN mechanisms may trigger involuntary switching of attention to the change in the unattended auditory environment (Escera et al., 2000; Naätänen, 1990). The switch of attention itself is generally associated with the P3a response (Escera et al., 2000). Large and salient changes in unattended sounds are more likely to elicit a P3a than small changes.

Although N1, MMN, and P3a occur in close temporal succession within the first 300 ms from change onset and may even partly overlap in time, dissociation of these components from each other is typically possible due to the millisecond-scale temporal resolution of EEG. However, spatial analysis of the brain sources of these components is often more difficult. Previous studies using EEG and MEG source analysis techniques suggest that N1, MMN, and P3a have at least partially different sources in studies using EEG and MEG source analysis techniques suggest that N1, MMN, and P3a have at least partially different sources in the supratemporal cortex (STC) and that additional areas in frontal and parietal lobes are likely to be involved in their generation (Alho, 1995; Alho et al., 1998; Escera et al., 1998). In contrast to EEG and MEG, spatial information is directly encoded into the fMRI signal and, therefore, it seems worthwhile to use fMRI to map the generator sources of N1, MMN, and P3a with high spatial accuracy. However, the use of fMRI to investigate the mechanisms underlying these auditory ERP components is challenging for several reasons. First, the temporal resolution of fMRI does not allow one to separate the N1, MMN, and P3a generators from each other in time. Second, the mapping between ERPs and the fMRI signal is not fully understood. Further, the loud noise (>100 dB SPL) associated with fMRI recording (Ravicz et al., 2000) activates the auditory system (Bilecen et al., 1998), may impair the perception of other sounds, and may modulate the activation of N1, MMN, and P3a generators (Novitski et al., 2001).

Most previous fMRI studies focusing on auditory change detection have used relatively large frequency changes to elicit hemodynamic responses (Doeller et al., 2003; Liebenthal et al., 2003; Opitz et al., 1999, 2002; Sabri et al., 2004). The use of a pronounced difference between frequent and infrequent sounds is understandable as the scanner noise might impede the perception of the minute details of sounds. However, as noted above, large frequency changes generate several different ERP components and, therefore, the fMRI signal elicited by such changes may reflect activation of several different functional mechanisms related to stimulus detection, change detection, and involuntary attention.

It could be argued that sound duration decrements could be used in an fMRI study to dissociate N1 and MMN type of activation from each other: such changes do not elicit the N1 enhancement in ERPs as stimulus energy is decreased rather than increased (Naätänen et al., 1989). In a previous study, Mathiak et al. (2002; see also, Kircher et al., 2004) manipulated the fMRI imaging sequence so that the gradient noise could be used to produce frequent (quite complex) stimuli and infrequent sound duration decrements (acquisition of 4 instead of 8 echoes; in addition, infrequent amplitude decrements were produced by lowering the strength of the gradients). These duration decrements elicited activation in the right STC. The authors also recorded magnetoencephalographic (MEG) responses to the same sounds and found that the dipole-modeled sources of MMNm (the magnetic counterpart of the electric MMN) were located in STC within the area of fMRI activation. These results demonstrate that duration decrements can be used in an fMRI study to elicit change-related activation. It should be noted, however, that the spatial match between fMRI activation and MEG source dipole in STC does not unequivocally prove that fMRI and MEG results are caused by the same functional processes.

A parametric design could be one way to ascertain whether the ERP and fMRI results are due to activation of the same brain mechanism: if ERP and fMRI signals are modulated similarly by the magnitude of sound change, it is likely that the methods are measuring the activation of the same functional unit (Doeller et al., 2003; Horovitz et al., 2002; Opitz et al., 2002). In a recent study, Doeller et al. (2003) used three levels of frequency (500 Hz vs. 667, 833, or 1000 Hz) changes in an event-related fMRI paradigm. Their results showed that the frequency changes were associated with activation in the right STC and that this activation increased with the magnitude of the frequency change. In a separate session, the authors recorded ERPs to the same sounds. The infrequent frequency changes elicited a change-related response consisting of several components including at least MMN and P3a. In addition, it is likely that N1 was enhanced by the frequency changes although the authors do not discuss this possibility. In general, the amplitude of the change-related components in their data seemed to increase with the magnitude of the frequency change (though the authors report variations from this monotonic pattern). Thus, it is probable that not only the MMN but also the N1 and P3a generator mechanisms contributed to their fMRI results. Furthermore, the activation of attention-related systems is very likely as their subjects did not have any specific task to engage their attention during fMRI scanning.

In addition to STC activation, Doeller et al. (2003) reported fMRI activation in the bilateral inferior frontal cortex (IFC). This activation was not seen in statistical parametric maps (SPMs) comparing infrequent vs. frequent sounds but was revealed by comparisons between different levels of frequency changes. Frontal activation associated with processing of infrequent sound changes has also been reported by other recent fMRI (Opitz et al., 2002; Schall et al., 2003) and positron emission tomography (PET; Dittmann-Balcar et al., 2001; Muller et al., 2002) studies. Previously, frontal lobe contribution to the electric MMN has been suggested based on EEG (Baldeweg et al., 2002; Giard et al., 1990; Rinne et al., 2000) and intra-cranial recordings (Liias et al., 2001). However, the functional role of the frontal lobes in processing of infrequent sound changes and generation of MMN remains poorly understood.

In the present study, we used a three-level parametric design with duration decrements. We presented our subjects with a repeating 75-ms sound (onset-to-onset interval 300 ms) which was occasionally replaced, in separate blocks, by a 15-ms, 25-ms, or 35-ms sound (large, medium, and small sound change, respectively). During the baseline condition, only the 75-ms sound was presented. fMRI scanning was structured so that 1 s of imaging, associated with scanning noise, alternated with a 2.9-s silent period. Most of the frequent sounds and all the infrequent sound duration decrements were presented during these silent periods. During the scanning, subjects were instructed to ignore the sounds while watching a silent wildlife documentary. We assumed that in this paradigm, the MMN mechanism would be activated and would contribute to the observed fMRI signal relatively more than the other possible change-related processes such as those underlying N1 (when stimulus energy is decreased, an enhanced N1 should not be elicited) and P3a (the subjects concentrated on watching a video while ignoring the meaningless sounds). We expected that
Sound duration changes would elicit fMRI activation in the auditory areas of STC and that the magnitude of this activation would increase with the magnitude of sound duration change. In addition, we expected that these duration decrements would activate areas in the IFC and that this activation might be modulated by the magnitude of duration changes.

**Methods**

**Subjects**

Fourteen healthy right-handed adults volunteered as subjects (age 20–28 years, 6 males). Informed consent was obtained from each subject prior to the experiment. The protocol was approved by the Ethical Committee of the Helsinki University Central Hospital. Two subjects were excluded from the analysis due to mislocalization of the fMRI slices.

**Stimuli**

The stimuli were complex tones, which consisted of three sinusoidal partials (500, 1000, and 1500 Hz). The second and third components were 3 and 6 dB lower in intensity, respectively, than the first component. This tone structure was chosen to facilitate MMN elicitation: the electric MMN amplitude is larger with spectrally rich tones than with sinusoidal sounds (Tervaniemi et al., 1993, 2000). There were four different tones of 75, 35, 25, and 15 ms in duration (5 ms rise and fall times). The 75-ms tone was the frequent sound and the 35-ms, 25-ms, and 15-ms tones were used to produce small, medium, and large infrequent sound changes, respectively. The sound sequence was ordered in a blocked design. There were four different blocks of 39 s in duration (Fig. 1, top). In the first block, only the 75-ms frequent sound was repeated with constant onset-to-onset intervals of 300 ms (130 presentations). The second block was otherwise similar but 10 (8%) of the sounds were replaced with a 35-ms infrequent sound. The third and fourth blocks were built up similarly by using either 25-ms or 15-ms sounds, respectively. The onset-to-onset interval between two successive infrequent sounds alternated between 2.4 and 5.4 s so that the sound changes were always presented during the silent breaks in the imaging sequence and they were not (completely) time locked to the fMRI data acquisition (Fig. 1, bottom). The block with only the frequent 75-ms sounds was presented first followed by the three blocks with infrequent sounds in random order. This four-block sequence was repeated 12 times. The sound sequences were delivered using Presentation software (www.neuro-bs.com).

**Imaging procedure**

Magnetic resonance imaging was performed using a 1.5-T Siemens Sonata scanner (Erlangen, Germany) with a standard head coil. First, a T1-weighted high-resolution anatomical image was acquired for each subject (3D MP-RAGE, Siemens). The anatomical image was used to set the functional volume of 12 slices (thickness 4 mm, no gap between slices) to cover bilateral supratemporal and inferior frontal cortices. A gradient-echo echoplanar sequence was used for functional imaging (TR = 3.9 s, TE = 40 ms, flip angle = 90°, matrix 64 x 64, voxel size 4 mm x 4 mm x...
A total of 480 functional volumes were recorded. As the recording of one volume lasted about 1.2 s, the imaging sequence contained 2.7-s periods when no data were collected and, thus, the gradient noise was absent (Fig. 1, bottom). Note that clustered volume acquisition (Edmister et al., 1999) with a long TR was used to create periods of relative silence for sound presentation, which is in contrast to procedures of previous studies using significantly longer TRs (>10 s) to separate the activation to auditory stimulus from that to the scanner noise (Hall et al., 1999). The functional session lasted for 31 min. After the completion of functional recording, one EPI volume covering the whole head was collected. This additional volume, which included the slices of the 12-slice functional volume, was used in co-registration of the functional data to the anatomical image.

fMRI data analysis

Image analysis was carried out using the Functional Magnetic Resonance Imaging of the Brain Centre (FMRIB) software library (FSL, version 3.1, www.fmrib.ox.ac.uk/fsl). The first 10 volumes were excluded from data analysis. Data were motion-corrected (Jenkinson et al., 2002), spatially smoothed with a Gaussian kernel of 8 mm full-width half-maximum (FWHM), and subjected to high-pass temporal filtering (cutoff 120 s). Statistical analysis was carried out using the FMRIB Improved Linear Model (FILM) with local autocorrelation correction (Woolrich et al., 2001). In the first level, the experimental design was set up in an event-related manner so that the presentation time of each infrequent (15-ms, 25-ms, or 35-ms) sound was entered separately to the model. The hemodynamic response was modeled with a gamma-function and its temporal derivative. The model was temporally filtered (similarly to the data). Finally, several contrasts were specified to create Z statistic images testing separately presentation of each infrequent sound vs. frequent sounds and for linear and quadratic relationships between activation and magnitude of sound duration decrement.

For group analyses, the individual level Z statistic images for all subjects were transformed into standard space using FMRIB’s Linear Image Registration Tool (FLIRT; Woolrich et al., 2004). As the first-level analysis showed that the overall level of activation in response to sound changes was relatively low, the second-level analysis was restricted to bilateral spherical regions of interest (ROI) with a radius of 20 mm in the STC and IFC. These ROIs were defined according to coordinates reported by a previous study (Doeller et al., 2003). These ROIs were large enough to allow anatomical variation between subjects and to cover the locations of activated areas in the STC and IFC reported by previous fMRI and PET studies of auditory change detection. Group-level activation within these ROIs was considered significant when 150 or more contiguous voxels survived an uncorrected threshold \( Z = 1.64 \).

EEG recording

In order to verify that the short-duration sounds presented together with the fMRI noise elicit the MMN response, the subjects of the fMRI study (\( N = 10 \)) participated in an additional ERP experiment. The stimuli were identical to the fMRI experiment. Tape-recorded fMRI noise (gradient and coolant pump) was mixed with the sounds and played back during the ERP recording. MMN peak amplitudes were determined from the infrequent tone-frequent tone ERP subtraction.

Results

The contrast between the condition with medium sound changes (25-ms vs. 75-ms sounds) and the condition with only the frequent sounds (75 ms) revealed activation in the right STC (Fig. 2). No significant activation in similar contrasts was found for the small or large sound duration changes. These findings were surprising since a monotonic function between the magnitude of sound duration change and the STC activation was expected. A contrast testing such a relationship between different levels of sound duration change (large > medium > small) yielded no activation in the left or right STC.

In addition to the STC, activation was detected in the IFC bilaterally. This IFC activation was revealed by a contrast testing inverse linear relationship between activation and the magnitude of sound change. That is, in the IFC, activation

Fig. 2. Summary of the fMRI results (\( N = 12 \)) overlaid onto the MNI standard head. Two different comparisons are illustrated: contrast between medium sound changes and frequent sounds revealed activation in the right STC. IFC activation was revealed bilaterally by a contrast testing for inverse linear relation (small > medium > large) between activation and the magnitude of sound change. Note that because the overall level of activation in response to sound changes was relatively low, the analysis was restricted to bilateral spherical ROIs with a radius of 20 mm in the STC and IFC. Activation was considered significant (color-coded areas) when 150 or more contiguous voxels survived an uncorrected threshold \( Z = 1.64 \).
and large sound changes but the differences were not significant. MMN peak amplitude appeared to be larger to medium than to small electrode at right mastoid) is shown. Note that the change-related response ERP subtraction at a frontal midline electrode (~Fz; referenced to an with the tape-recorded fMRI scanner noise. Infrequent tone-frequent tone ERPs recorded in a separate session showed clear MMN responses to the infrequent sound changes (Fig. 3) presented with tape-recorded fMRI scanner noise. There were no significant differences between the MMN peak amplitudes to small, medium, or large sound changes.

Discussion

In the present study, auditory areas of the right STC were activated in response to medium sound duration changes (25-ms vs. 75-ms sounds). The right hemisphere dominance of this change-related activation is consistent with previous fMRI studies reporting activation to infrequent pitch, intensity, and duration changes, and with previous EEG and MEG studies on MMN/ MMNm to such changes (Giard et al., 1995; Levänen et al., 1996; Paavilainen et al., 1991). However, the present data did not support the hypothesis that STC activation increases with the magnitude of sound change. The STC activation was stronger for medium than for small and large sound changes. This unexpected finding might be explained in several ways: first, it is possible that the short-duration (15-ms) sounds constituting the large sound change differed from other sounds in some way so that the sound-change-related STC activation could not be detected. For example, it might be that due to perceptual difficulties in the fMRI environment, the 15-ms tones were too short to elicit a significant difference in the fMRI response. The lack of significant activation to the large sound duration changes (15-ms vs. 75-ms sounds) would explain the observed non-monotonic function between the magnitude of sound duration change and the STC activation if, as expected, the STC activation to the small duration change (35-ms vs. 75-ms sounds) was weaker than that to the medium duration change (25-ms vs. 75-ms sounds). Second, it is possible that the medium duration change is a more optimal stimulus for the auditory change detection mechanism than the small and large duration changes. This is tentatively suggested by previous ERP studies showing that a duration decrement similar to the medium change in the present study elicits an MMN with a higher amplitude and reliability than a smaller duration, intensity, or frequency change (Tervaniemi et al., 1999).

In addition to the STC activation, areas in the bilateral IFC were found to respond to infrequent sound duration changes. This activation was revealed by contrasts testing for an inverse linear relation between activation and magnitude of sound change. That is, IFC activation to the small duration changes was stronger than that to the large changes. This finding is consistent with previous studies recording fMRI responses to different levels of frequency change (Doeller et al., 2003; Opitz et al., 2002). Alternatively, it could be argued that this result is not caused by sound changes per se but by activation of an auditory area in the frontal cortex (Romanski and Goldman-Rakic, 2002) responding to sound duration or energy (i.e., larger activation to 35-ms than to 15-ms tones).

Previous fMRI studies focusing on auditory change detection have often used relatively large frequency changes to elicit the change-related hemodynamic response (Doeller et al., 2003; Liebenthal et al., 2003; Opitz et al., 1999, 2002; Sabri et al., 2004). Although large frequency changes undoubtedly elicit pronounced signals in fMRI, such changes do not optimally differentiate in fMRI the separate auditory cortex mechanisms suggested by a large body of ERP literature. The present study was designed to maximize MMN type of activation in fMRI while minimizing the contribution of other change-related processes: first, the frequent sounds were presented with a fast presentation rate (at 300-ms onset-to-onset intervals). In ERPs, such fast presentation rates are typically associated with low amplitude obligatory ERP components (e.g., P1, N1, and P2, Näätänen and Picton, 1987). Second, duration decrements were used to elicit change-related activation. As stimulus energy is reduced when sound duration is decreased, the infrequent duration decrements should not activate mechanisms reacting to increasing stimulus energy. Third, during the experiments, the subjects were instructed to watch a video and to ignore the sounds. Although this is not a perfect control for attention, a video-watching task reduces the chance that the subjects covertly attend to the sounds and that attention-related systems are activated. Data from a separate session (Fig. 3), in which the electric responses were recorded to the same sounds used in the fMRI experiment, show that the change-related ERP elicited by the duration decrements consisted predominantly of MMN and, therefore, it may be assumed that the present fMRI results (Fig. 2) reflect mainly the activation of the MMN mechanism. Thus, the present results support the view that a network of brain areas in STC and IFC underlies the MMN.

In spite of the fact that MMN has been extensively studied since late 1970s, the role of IFC in auditory change detection and analysis still remains to be clarified. Originally, it was suggested that the frontal lobe contribution to MMN mechanism signifies the initiation of switching of attention to the sound detected by the STC change-detection process (Näätänen, 1990; Näätänen and Michie, 1979). Alternatively, it has been suggested that, instead of attention switching, the IFC activation is related to a contrast enhancement mechanism which would be activated when the STC system gets in difficulty in discriminating stimuli (Doeller et al., 2003; Opitz et al., 2002). In addition to these hypotheses, it is also possible that the IFC activation detected by fMRI might be related to an inhibitory system that allows the subjects to ignore the sound changes when no change-related response is required.
Acknowledgments

Supported by the Academy of Finland (grants #102316, #103170, and #80819). We thank Drs. G. Christopher Stecker and David L. Woods for helpful comments on the manuscript.

References

Alho, K., 1995. Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. Ear Hear. 16, 38–51.


