Human cortical representation of virtual auditory space: differences between sound azimuth and elevation

Nobuya Fujiki,1,4,5 Klaus A.J. Riederer,2,3 Veikko Jousmäki,1 Jyrki P. Mäkelä1 and Riitta Hari1,6

1Brain Research Unit, Low Temperature Laboratory, 2Laboratory of Computationl Engineering, 3Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, FIN-02015 HUT, Espoo, Finland 4Department of Otolaryngology – Head and Neck Surgery, Kyoto University Graduate School of Medicine, Sakyo-ku, Kyoto, 606–8507, Japan 5Department of Otolaryngology, Otsu Red Cross Hospital, Otsu, 520–8511, Japan 6Department of Clinical Neurophysiology, Helsinki University Central Hospital, FIN-00290 Helsinki, Finland

Keywords: binaural cue, head-related transfer function (HRTF), magnetoencephalography (MEG), monaural spectral cue, three-dimensional (3-D) sound

Abstract

Sounds convolved with individual head-related transfer functions and presented through headphones can give very natural percepts of the three-dimensional auditory space. We recorded whole-scalp neuromagnetic responses to such stimuli to compare reactivity of the human auditory cortex to sound azimuth and elevation. The results suggest that the human auditory cortex analyses sound azimuth, based on both binaural and monaural localization cues, mainly in the hemisphere contralateral to the sound, whereas elevation in the anterior space and in the lateral auditory space in general, both strongly relying on monaural spectral cues, are analyzed in more detail in the right auditory cortex. The binaural interaural time and interaural intensity difference cues were processed in the auditory cortex around 100–150 ms and the monaural spectral cues later around 200–250 ms.

Introduction

Accurate sound localization is ecologically important for most animal species. Human auditory localization is based on both binaural cues (interaural time and intensity differences, ITDs and IIDs) and on the spectral (monaural) cues (caused by the observer’s own body, especially her/his head and pinnae that spectrally colour the sounds). All these cues of the spatial auditory information can be embodied in head-related transfer functions (HRTFs) that are measured or modelled free-field responses in the ear canal to a sound source in space (Blauert, 1997; Kulkarni & Colburn, 1998). The HRTF depends on the particular person and ear, on the direction of the sound source, and on sound frequency.

Thus convolving the stimuli with the individual HRTF allows all person-dependent monaural and binaural spatial information to be maintained in the reproduced three-dimensional (3-D) sound, leading in very natural sound percepts. In contrast, if the stimuli presented via headphones contain only interaural time and intensity differences, the sound elevation is poorly represented. Moreover, such sounds are typically, and very unnaturally, perceived as originating inside the head.

Although the brain mechanisms of directional hearing have been mainly studied in the peripheral auditory pathways (Viete et al., 1997; Bruckner & Hyson, 1998; Funabiki et al., 1998; Oertel, 1999), several investigations in cats and monkeys (Eisenman, 1974; Thompson & Cortez, 1983; Jenkins & Merzenich, 1984; Ahissar et al., 1992; Middlebrooks et al., 1994; Brugge et al., 1996; Heffner, 1997; Recanzone et al., 2000), as well as in human patients (Walsh, 1957; Sanchez-Longo & Foster, 1958; Shankweiler, 1961; Klingon & Bontecou, 1966; Gazzaniga et al., 1973; Bisiach et al., 1984; Altman et al., 1987; Zatorre et al., 1995; Yamada et al., 1996; Clarke et al., 2000) indicate that the auditory cortex is essential for sound localization.

In noninvasive brain imaging studies on human spatial auditory perception, the sounds are typically, due to technical reasons, presented through headphones instead of free space (loudspeakers). As stated above, such presentation, although sufficient to represent the horizontal plane, cannot represent sound elevation, and is also otherwise unnatural. We have thus applied individual HRTFs to produce a virtual 3-D auditory space; individual HRTFs have not been used in previous brain imaging studies of human sound localization (e.g. Griffiths et al., 1998; Bushara et al., 1999; Palomäki et al., 2000; Warren et al., 2002).

Our goal was to compare reactivity of the human auditory cortex to changes in sound azimuth vs. elevation (horizontal vs. vertical directions in the auditory space) by recording auditory-evoked magnetic fields (Hari, 1990). Human brain mechanisms related to sound elevation analysis have not been explored previously.

Materials and methods

Subjects

We studied eight normal-hearing subjects, (two females, six males; 23–38 years old, mean ± SD 29.4 ± 6.0 years), including three of...
the authors. Seven subjects were right-handed and one female subject left-handed. Informed consent was obtained from all subjects and the study had received prior acceptance by the Ethical Committee of the Helsinki Uusimaa Hospital District.

Stimuli
To produce stimuli with 3-D auditory locations, we first measured individual HRTFs (Riederer, 1998a, 1998b) in an anechoic room (interior 6.6 X 6.6 X 6.1 m$^3$; Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology). The stimuli were designed to allow comparison of cortical metrics for the horizontal, vertical, and front-back directions of the 3-D auditory space.

Figure 1 shows the locations of stimuli, presented in two separate randomized oddball sequences in the anterior and right lateral auditory hemispheres. The interstimulus interval was 500 ms, and in each sequence, the frequent standard sounds had a probability of 0.8 and all four deviant stimuli a probability of 0.05 each. In the anterior space, the standard sounds were exactly in front of the subject, and the four deviants were 30° left, right, up, or down from the centre. In the right lateral hemispace, the standard stimuli were presented exactly to the right of the subject, and the four deviants 30° backwards, frontwards, up, or down from that position.

The 100-ms stimuli consisted of pink noise, convolved with the individual HRTFs that were accurately compensated for the artifacts of the HRTF measurement system and reproduction equipment. The stimuli were reproduced by custom-built tube headphones (Riederer & Niska, 2002; Unides Design Ay., Helsinki, Finland) at the replaceable ear-tips. The sound was transmitted via 3.25-m long plastic tubes to avoid magnetic and electrical artifacts. The frequency response of the system was 100–14 000 Hz (± 10 dB; measured with an artificial ear at the end of the tubes), and it was equalized flat for the stimuli. The intensity of the standard stimulus was adjusted to 40 dB above the individual sensation threshold.

A psychoacoustical test was performed to obtain crude estimates of the subjects’ localization performance: each stimulus was presented repetitively once every 500 ms and the subject was instructed to indicate the sound location with a laser pointer. The point in the room wall was marked and its angles of azimuth and elevation were determined. All stimuli in the anterior and in the right lateral hemispaces were tested twice. Differentiation between locations of deviant and standard stimuli was considered adequate when the reported angle differences in azimuth or elevation were 10° or more in the direction concerned.

As a whole, the azimuth changes in the anterior space were perceived correctly in 97% of stimuli, and the elevation differences in 41%; the correct elevation discrimination increased to 69% when the errors within the up-down confusion were ignored. In the lateral space, azimuth and elevation changes were correctly perceived in 59% and 68% of stimuli, respectively; these values increased to 78% and 81%, respectively, when the errors within the front-back or up-down confusions were ignored.

Neuromagnetic recordings
The subject was seated during the recordings under the helmet-shaped neuromagnetometer. Her/his cortical activity measured by magnetoencephalography (MEG) was recorded with the 204 first-order planar gradiometers of a whole-scalp DC-SQUID (Superconducting Quantum Interference Device) neuromagnetometer (Vectorview™, Neuramag Oy, Helsinki, Finland; Ahonen et al., 1993). The device measures very weak magnetic fields on the brain and is positioned in a magnetically shielded room to avoid external disturbances. For the same reason, no moving magnetic material can be used inside the measurement chamber, and thus normal (magnetic) headphones or loudspeakers were not used.

The recording passband was 0.1–172 Hz, and the signals were digitized at 601 Hz. The vertical electro-oculogram (EOG) was simultaneously recorded between electrodes above and below the left eye, and all traces coinciding with EOG activity exceeding 150 μV were omitted from the on-line averaging. The neuromagnetic responses were averaged time-locked to the sound onsets; 60–100 epochs were averaged for each deviant.

Analysis of MEG responses
The averaged signals were digitally low-pass filtered at 40 Hz. Auditory mismatch fields (MMFs) were calculated by subtracting responses to the standard stimuli from those to the deviants (for a review of the electric and magnetic mismatch responses, see, e.g. Alho, 1995); please note that we call these responses (operationally) MMFs although it has been doubted whether location differences evoke real mismatch fields (for discussion and arguments, see McEvoy et al., 1993).

Two equivalent current dipoles (ECDs) were used to model the sources of MMFs at certain latencies (for a review on dipole modelling, see Hämäläinen et al., 1993). First, if the field pattern was clearly dipolar, a single ECD was found for each hemisphere by a least-squares fit based on the data of a subset of 54 planar gradiometers over the area that included the maximum response at a certain time point. Dipoles were only accepted if the measured and predicted waveforms closely resembled each other at the latencies of interest. Then source waveforms were computed for the two-dipole
model in which the sources were fixed in location and orientation (Hämäläinen et al., 1993). In the head-coordinate system, the x-axis connected the left to the right preauricular point, the y-axis was running towards the nasion perpendicularly to the x-axis, and the z-axis was perpendicular to the xy-plane, with positive direction upwards. The ECDs were superimposed on the subjects’ own magnetic resonance (MR) images.

Results

Figure 2 shows responses of one subject to standard stimuli heard directly from the front, and to deviant sounds heard as elevated 30° upwards. Prominent responses occurred in both temporal areas. The earliest responses to standard stimuli peaked at 75 ms in the left and at 90 ms in the right hemisphere. Similar early deflections were also elicited by the deviant sounds which evoked additional large responses in both temporal areas, with peaks around 135 ms and 240 ms. Source analysis indicated generator areas in the superior temporal cortices for all responses; in the right hemisphere, the late responses to the deviants originated anterior to the early responses (Fig. 2, bottom).

Figure 3 shows mismatch fields of one subject to different deviants illustrated as vector sums:

\[ \sqrt{\left(\frac{\partial B_z}{\partial x}\right)^2 + \left(\frac{\partial B_z}{\partial y}\right)^2} \]

where \(\frac{\partial B_z}{\partial x}\) and \(\frac{\partial B_z}{\partial y}\) are the two orthogonal gradients of the magnetic field component \(B_z\). The vector sums render the analysis less sensitive to slight changes in the direction of source currents. The responses are clearly stronger and better structured in the right than the left hemisphere (bottom vs. top traces). For the anterior space, the first (M1) and second (M2) right-hemisphere MMF deflections peaked 22±32 ms and 10±20 ms earlier to azimuth than elevation changes, respectively; no such differences were observed in the left hemisphere. M1 and M2 were stronger for all stimuli in the right than in the left hemisphere. Even for the right lateral space, the responses of this subject were larger in the right (ipsilateral) hemisphere, and the latencies and amplitudes of M1 or M2 did not differ between azimuth and elevation deviances.

The source latencies and strengths of all subjects were analysed with 3-way ANOVA with hemisphere (left/right), sound direction (contralateral/ipilateral/up/down in the anterior space and front/back/up/down in the lateral space) and response (M1/M2) as factors. In the anterior space (Fig. 4), the peak latency was on average 10 ms shorter in the right than the left hemisphere (\(F_{1,104} = 5.34; P = 0.02\)). The variation of peak latency as a function of sound direction was statistically highly significant (\(F_{3,104} = 16.4, P < 0.0001\)); the latency
was 9–19 ms (mean 17 ms) shorter to sound deviations directed towards contralateral than ipsilateral space with respect to the hemisphere ($P = 0.01$). The latency was 10±51 ms shorter to changes in azimuth (filled and open blue bars in Fig. 4) than in elevation (red bars) (contralateral vs. up, mean 36 ms, $P < 0.0001$; contralateral vs. down, mean 33 ms, $P < 0.0001$; ipsilateral vs. up, mean 19 ms, $P = 0.02$). Source strength showed a similar, statistically nonsignificant tendency, with stronger mean values in the right than left hemisphere in all conditions and stronger mean values to azimuth than elevation changes.

For sounds in the lateral space, the variation of M1 and M2 source latencies as a function of sound direction was statistically significant ($F_{3,93} = 3.5, P = 0.02$). The latency was 11±25 ms shorter to elevation than backward changes (up vs. back, mean difference 17 ms, $P = 0.005$; down vs. back, mean difference 18 ms, $P = 0.004$). The mean latencies and source strengths did not differ between the hemispheres.

At individual level, M2 was significantly stronger in the right than the left hemisphere in four subjects (the $t$-values from $t$-tests varied from –2.53 to –6.35 and $P$-values from 0.04 to 0.0004 in different subjects; d.f. = 7), significantly weaker in two ($t$-values 3.38 and 5.96, d.f. = 7, $P$-values 0.01 and 0.0006), and showed no hemispheric differences in the remaining two subjects. M1 was statistically significantly stronger ($t = –3.76$, d.f. = 7, $P = 0.007$) in the right than left hemisphere in one subject, and showed no significant hemispheric differences in the others.

Figure 5 shows the M1 and M2 source locations to all stimuli. The sources were distributed to a considerably wider brain area in the right than the left hemisphere and differed only in the right hemisphere between conditions: The right M1 source was posterior and superior to the right M2 source for sounds presented both in the anterior sound space (4 mm; $F_{1,26} = 5.27$; $p = 0.03$, and 3 mm; $F_{1,26} = 3.24$; $P = 0.08$, respectively; 2-way ANOVA) and in the lateral sound space (6 mm; $F_{1,18} = 6.23$; $P = 0.02$; 5 mm; $F_{1,18} = 10.1$; $P = 0.005$). Further, the M1 activation elicited by azimuth changes rightward was 6 mm posterior and 6 mm superior to the M1 activations for downward changes ($P < 0.05$).
Discussion

Three-dimensional sounds, including both binaural and monaural localization cues, elicited prominent responses around 140 ms and 240 ms in the supratemporal auditory cortices of both hemispheres. These responses were consistently earlier to azimuth than elevation changes in the anterior space, and they covered a wider area and were often stronger in the right than the left hemisphere. Azimuth changes in the anterior space elicited earlier and stronger responses in the hemisphere contralateral to the sound location. The weakest responses, with the longest latencies, were elicited by sound changes backwards. In the right hemisphere, the sound changes appeared to activate a wider cortical area, so that, e.g. the M1 and M2 sources differed by approximately half a centimeter in the anteroposterior direction.

In the studies of Kaiser et al. (2000a, 2000b), binaural complex sounds, containing ITDs, elicited MMFs around 110–140 ms, corresponding to the occurrence of our M1 response, whereas differences in sound spectra, but not in sound location, elicited MMFs around 180 ms, close to the occurrence of our M2 response. Altogether these findings suggest that binaural ITD or IID cues, which help us greatly to perceive the sound azimuth, are processed in the auditory cortex rather early, around 100–150 ms, and that the monaural spectral cues, used in both azimuth and elevation estimation, are analyzed later around 200–250 ms.

ITD and IID vary systematically along the azimuth, but they are non-existent in the median (front-back) plane and constant across the so-called ‘cones of confusion’ (see e.g. Blauert, 1999). Front-back separation and elevation discrimination are still disputed, although they are primarily cued by spectral changes in the pinna cavities, i.e. monaural cues. The more accurate recognition of azimuth than other directions (Middlebrooks, 1992) is in line with the shortest peak latencies and largest amplitudes of both M1 and M2 responses to azimuth deviances in the anterior auditory space. Accordingly, the proportion of spatially sensitive neurons in the monkey auditory cortex is larger for stimulus azimuth than elevation, and most neurons respond best to contralateral sound locations (Recanzone et al., 2000).

Our findings on latency differences suggest that the azimuth deviances in the anterior space, relying on both binaural and monaural localization cues, are processed mainly in the hemisphere contralateral to the sound localization. Ungan et al. (2001) suggested that IIDs are analyzed most prominently in the hemisphere contralateral to the stimulus location whereas ITDs are analyzed in both hemispheres. Our stimuli contained both ITD and IID cues and therefore naturally precluded detailed analysis of such differences. Nevertheless, in line with this suggestion, our M1 response was earlier and stronger in the hemisphere contralateral than ipsilateral to the virtual source.

In free field, subjects localize more accurately sounds emanating from the left than the right hemifield (Burke et al., 1994). Moreover, subjects are more accurate in monaural elevation localization with the left than right ear (Butler, 1994). These findings can be attributed to higher fidelity of spectral cue analysis in the right than left hemisphere. In line with this, patients with right, but not left, temporal lobe lesions are impaired in detecting increasing vs. falling periodicity pitch (Zatorre, 1988). This hemispheric specialization seems to be reflected also in auditory evoked fields to frequency modulations, which are stronger in the right than the left auditory cortex (Pardo et al., 1999).

Previous MEG studies of cortical responses to lateralized sounds elicited by ITDs (McEvoy et al., 1993; Sams et al., 1993) and IIDs (Mäkelä & McEvoy, 1996) reported no differences between the two hemispheres. The more detailed analysis of the 3-D sounds in the right hemisphere, suggested by the more structured organization of sources of responses, and the earlier latencies of the right-hemisphere activation could be related to the content of individual HRTF stimuli. Also in line with the idea that the monaural spectral cues affect the activity more in the right than the left auditory cortex, the amplitude of the auditory 100-ms responses to sounds with different azimuths, created by nonindividual HRTFs, varied significantly as a function of azimuth in the right but not in the left hemisphere (Palomäki et al., 2000).

The source analysis of our data suggested that the monaural spectral cues are processed in a more anterior site of the right superior temporal cortex than the binaural localization cues. The right anterior temporal cortex of humans could correspond to the multidirectional anterior ectosylvian area of cat auditory cortex where single auditory neurons may code sound source locations over 360°, with temporally different firing patterns separating different spatial locations (Middlebrooks et al., 1994).

The hemispheric dominance of M1 and M2 responses varied among individuals, in line with several previous reports that have suggested variable hemispheric dominance in directional hearing (Altman et al., 1987; Burke et al., 1994; Butler, 1994; Zatorre et al., 1995; Yamada et al., 1996; Woldorff et al., 1999; Clarke et al., 2000; Kaiser et al., 2000b). In animal experiments, lesions of the auditory cortex reduce the ability to localize sounds and to attend to them within the contralateral auditory space (Jenkins & Merzenich, 1984). However, in humans hemispherectomy worsens to some extent the accuracy of sound localization in the horizontal plane bilaterally, more prominently in the contralateral auditory hemifield, but the subjects still display some localization accuracy in both hemifields (Zatorre et al., 1995). Patients with hemispheric lesions due to cerebrovascular accidents in adult age are less accurate in free-field sound localization than healthy control subjects but this impaired performance is not limited to the contralateral auditory hemispace (Haeske-Dewick et al., 1996). Thus the auditory cortices of both hemispheres seem to play a role in processing of sound direction in humans. Plastic reorganization after lesion is a less probable explanation of the bilateral localization ability especially in adult patients after cerebrovascular accident.

We have previously reported on a patient whose right-hemisphere auditory responses were totally abolished due to stroke but the left-hemisphere responses were normal. Binaural directional stimuli elicited normal MEG responses in the healthy hemisphere, and the patient was able to describe adequately both right-to-left and left-to-right azimuthal illusory sound movement with ITD stimuli, suggesting that one hemisphere can analyze sound location with binaural localization cues in both auditory hemifields (Mäkelä & Hari, 1992). In line with this suggestion, we observed M1 and M2 responses in both hemispheres.

In previous studies with positron emission tomography (PET) (Bushara et al., 1999; Weeks et al., 1999) and functional magnetic resonance imaging (fMRI) (Alain et al., 2001), carried out with nonindividual HRTFs, activity has been observed in the superior or middle temporal cortices and in the inferior parietal lobules during location discrimination task. Warren et al. (2002) used both PET and fMRI and found the activation bilaterally in planum temporale and in the parietotemporal operculum in response to sound movement contrasted to activation elicited by externalized sound stimuli. The 100-ms auditory response, N100m has been attributed, at least in part, to neuronal activity in planum temporale, and could well reflect activity of these same brain regions.
Stimuli producing illusion of sound movement also activate, although somewhat inconsistently, the human frontal and superior parietal cortices (for references, see Warren et al., 2002). Furthermore, in monkeys the prefrontal cortex is important in sound localization (Romanski et al., 1999). The auditory activation of frontal and superior parietal areas could be related to spatial attention or movement preparation (Warren et al., 2002). The responses in the present study were elicited by location differences of standard and deviant sounds and probably reflect more low-level processing, explaining why we did not see prominent activation outside the auditory cortices. It is also possible that such activities generated mainly by currents in the gyri, invisible for MEG measurements due to their radial orientation, or that the neural activity in these regions is not synchronous enough to produce detectable MEG signals.

In conclusion, our results suggest that azimuth and elevation changes in the three-dimensional auditory space are processed bilaterally but differentially in the human left and right auditory cortices. Binaural localization cues are processed earlier than monaural spectral cues and show contralateral hemispheric dominance with respect to sound location. Thus the azimuthal sound locations, especially in front of the subject, are easily perceived. The earlier and more structured source regions of responses in the right hemisphere imply that the sounds that include both binaural cues and monaural spectral cues are processed more extensively in the right auditory cortex.

Acknowledgements

This work was supported by the Academy of Finland, Sigrid Jusélius Foundation, the Japan Society for the Promotion of Science, and by the Graduate School of Electronics, Telecommunication and Automation, Ministry of Education, Finland. We thank Dr Jyri Huopaniemi (Nokia Research Center) for advice for stimulus generation in the early phases of this work and Mr Jaakko Järvinen for help in the measurements.

Abbreviations

ECD, equivalent current dipole; EOG, electro-oculogram; fMRI, functional magnetic resonance imaging; HRTF, head-related transfer function; ITD, interaural intensity difference; ITD, interaural time difference; MEG, magnetoencephalography; MMNm, mismatch field; MR, magnetic resonance; PET, positron emission tomography; SQUID, superconducting quantum interference device; 3-D, three-dimensional.

References


