Neural correlates of sound externalization

Akiko Callan [a,*], Daniel E. Callan [b], Hiroshi Ando [a]

[a] NICT Multisensory Cognition and Computation Laboratory, Kyoto 619–0288, Japan
[b] ATR Neural Information Analysis Laboratories, Kyoto 619–0288, Japan

ABSTRACT

When we listen to sounds through headphones without utilizing special transforms, sound sources seem to be located inside our heads. The sound sources are said to be lateralized to one side or the other to varying degree. This internal lateralization is different than sound source localization in the natural environment in which the sound is localized distal to the head. We used fMRI to investigate difference in neural responses between lateralization and localization. Individualized binaural recordings were used as externalized auditory stimuli and stereo recordings were used as internalized auditory stimuli. Brain activity was measured while 14 participants performed an active auditory localization task and while 12 participants performed a stimulus type identification task. Irrespective of the task condition, we observed enhanced activity in the bilateral posterior temporal gyri (pSTG) for the externalized stimuli relative to the internalized stimuli. Region of interest analysis indicated that both left and right pSTG were more sensitive to sound sources in contra- than ipsilateral hemifields. Moreover, greater back than front activity was also found in the left pSTG. Compared to impoverished spatial auditory stimuli, realistic spatial auditory stimuli enhance neural responses in the pSTG. This may be why we could observe contralateral hemifield preference in bilateral pSTG that many previous studies have failed to observe. Overall, the results indicate the importance of using ecologically valid stimuli for investigating neural processes in human cortex.

© 2012 Elsevier Inc. All rights reserved.

Introduction

When sounds reach our ears, the acoustic waveform vibrates our eardrums and through a series of steps is transformed into neural impulses and propagated to multiple brain regions. Location of the sound sources is represented in time and level differences between left and right ears (interaural time difference: ITD, interaural level difference: ILD), spectral characteristics arising from the direction-dependent diffraction and reflection properties of the head, pinna, and torso (head-related transfer functions: HRTFs), and echoes of the sounds under reverberant environments. Thus, the brain processes complex auditory signals, and as a result of the processing, we hear the characteristics of sound including pitch and loudness, determine direction and distance of sound sources, and sense information about the environment around us. While it is important to be able to identify what a sound is, it is also important to be able to detect sound source location and to predict direction of a moving sound source in order to focus attention to events for survival in many species, especially when visual cues are not available.

Sound sources are localized by interaural (i.e. binaural) and spectral cues. Interaural cues, such as ITD and ILD, are primary cues for localization in the horizontal plane. Spectral cues, such as HRTFs, are important for estimation of elevation, localization in the median plane, and resolution of front-back confusion. Moreover, spectral cues are necessary in order to experience the sound as localized outside the head instead of originating inside the head (1996; Plenge, 1974). In addition to those cues, we also perceive reverberation of a room in everyday settings. Room reverberation distorts spatial cues for sound localization but is an absolute cue for auditory distance (Bronkhorst and Houtgast, 1999; Mershon and Bowers, 1979). Room reverberation also provides information on room characteristics, and improves the subjective realism and externalization achieved in a headphone simulation (Shinn-Cunningham et al., 2005).

Spatial processing in the human auditory cortex has been investigated by functional magnetic resonance imaging (fMRI) studies using either internalized (inside-the-head located) auditory stimuli (e.g., Baumgart et al., 1999; Krumbholz et al., 2005) or externalized (outside-the-head located) auditory stimuli (e.g., Deouell et al., 2007; Warren et al., 2002). Both types of studies suggested the involvement of the posterior superior temporal gyrus (pSTG) in spatial processing. However, no fMRI studies have compared neural activity between internalized and externalized auditory stimuli in one study. Neural responses to natural complex stimuli are very likely to be different from neural responses to artificial simple stimuli. In fact, electroencephalography (EEG) and magnetoencephalography (MEG) studies that compared internalized stimuli with externalized stimuli report much stronger responses (Getzmann and Lewald, 2010a) and a larger dynamic range of the N1m responses (Palomaki et al., 2005) for the externalized stimuli. Furthermore, Palomaki et al. (2005) found that neural responses to individualized HRTF stimuli including room reverberation predicted behavioral performance better than neural responses to non-individualized HRTF stimuli.

* Corresponding author at: NICT Multisensory Cognition and Computation Laboratory, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619–0288, Japan. Fax: +81 774 95 2647.
E-mail address: acallan@nict.go.jp (A. Callan).

How these neurons represent auditory space remains unknown though considerable neuroimaging studies indicate the existence of spatially selective neurons in the pSTG. Auditory space can be separated into left-right, front-back and up-down dimensions. Among them, the left-right dimension has been extensively studied. The auditory cortices receive both contralateral and ipsilateral projections from each ear but the cortical neurons respond more strongly to sounds presented monaurally to the contralateral ear (e.g. Schonwiesner et al., 2007; Woldorff et al., 1999). Thus, contralateral representation of auditory space in the pSTG is expected but it is not clear whether binaurally presented spatial sounds are processed contralaterally. Several neuroimaging studies (Krumbholz et al., 2005; Palomaki et al., 2005; Salminen et al., 2010) have reported contralateral responses to binaurally presented stimuli but, on the other hand, several other studies (Brunetti et al., 2005; Woldorff et al., 1999; Zimmer and Macaluso, 2005; Zimmer et al., 2006) have reported no contralateralized processing. The type of auditory stimulus (i.e. externalized realistic spatial sounds vs. internalized spatially impoverished sounds), the state of the auditory stimulus (i.e. stationary vs. moving), and sensitivity of the measurements (i.e. N1m amplitudes vs. BOLD responses) were considered to be reasons for these contradictory results. A few neuroimaging studies (Getzmann and Lewald, 2010b; Pavani et al., 2002) comparing horizontal and vertical sound movements suggest that azimuth and elevation components of dynamic auditory spatial information are processed in common cortical substrates. As for front-back dimension, as far as we know, no neuroimaging has investigated the cortical representation.

The purpose of this study is to investigate differences in how externalized sounds and internalized sounds are processed in the human brain. In psychoacoustics, outside-the-head localization is referred to as "localization" and inside-the-head localization is referred to as "lateralization." Therefore, in other words, we are investigating differences in neural processing of localization and lateralization. For this purpose, we compared three types of sound recording methods; binaural, stereo and mono. All three types of recordings were done in the same reverberant room so that all auditory stimuli included the same room reverberation. Both binaural and stereo recording methods used two microphones so that they both include ITD and ILD cues. The mono recording method that used one microphone did not include ITD and ILD cues. In the binaural recording, microphones were positioned at the entrance of the participant's ear canals so that the recorded sounds included individualized HRTFs that allow participants to experience externalized auditory space using headphones. Based on previous EEG and MEG studies (Getzmann and Lewald, 2010a; Palomaki et al., 2005), we hypothesized that externalized auditory stimuli activate the cortical substrate involved in auditory spatial processing more than internalized auditory stimuli. In addition, we investigated neural correlates of auditory localization in the left-right and front-back dimensions. Because the low sensitivity of functional imaging methods was considered as one of the reasons for the lack of contralaterality in several previous studies (Werner-Reiss and Groh, 2008), we performed a region of interest analysis in the bilateral STG for this investigation. The elevation (up-down dimension) of the sound source is computed predominantly from monaural spectral cues and is assumed to be processed in the pSTG (Getzmann and Lewald, 2010b; Pavani et al., 2002). Since front-back location is also computed from monaural spectral cues, we expected modulation of brain activity in the pSTG for front/back localization as well.

Materials and methods

Participants

Fourteen adults (7 male; 22–40 years of age, mean 27.5) participated in Experiment 1. About 1 month later, 12 of them (6 male, 22–40 years of age, mean 28.3) participated in Experiment 2. All participants had no neurological or psychiatric history, and had pure-tone thresholds within normal range (≤20 dB HL) for octave frequencies between 250 and 8000 Hz. Thresholds were obtained with 5 dB steps and levels lower than 15 dB were not tested. Five subjects had 20 dB thresholds at either or both 250 and 500 Hz only in their right ears. Other than that, their thresholds are at or better than 15 dB. We did not check participants' precise interaural threshold differences. The interaural threshold differences may influence localization processing. However, in people with normal hearing, plasticity in spatial hearing has been demonstrated by the ventriloquism aftereffect (e.g. Recanzone, 1998) and unilateral earplugging experiments (e.g. Kumpik et al., 2010). Therefore, within normal hearing range, we assume that effects of interaural threshold differences are not significant. Moreover, individual differences were controlled by employing a random effects model for group analysis. All participants gave written informed consent for experimental procedures approved by the institutional review board at the National Institute of Information and Communications Technology.

Stimuli and procedure

The auditory stimuli were recorded in three ways: binaural, stereo and mono. The sound used for recording was uniformly distributed noise (band-pass filtered white noise; cutoff frequency, 0.6 kHz and 22 kHz; duration 100 ms; rise and fall times, 20 ms) (Zimmer and Macaluso, 2005). The noise was presented through a loudspeaker (Eclipse TD508II, Fujitsu Ten Ltd.) from one of 12 horizontal directions (0°, 30°, 60°, 90°, 120°, 150°, 180°, −150°, −120°, −90°, −60°, −30°, angle 0° was in front of the subject, angles >0° were in the right hemifield) with a distance of 2 m from the participant. In order to avoid variance caused by different acoustical characteristics of different speakers, only one speaker was used.

Recording methods were as follows. For the binaural recording (BR), the participant sat down in a chair. In-ear binaural microphones (SP-TFB-2, The Sound Professionals Inc.) were positioned at the entrance of the participant's ear canals, and stimuli were recorded through the microphones. For the stereo recording (SR), the same microphones were placed in similar positions to where participants' ears were located (the height from the floor was 115 cm and the distance between left and right microphones was 15.6 cm). For the mono recording (MR), the left microphone was placed in the center of the positions used for stereo recording. The recordings took place in a slightly reverberant room. For each direction, the noise was recorded 18 times and 9 of the recordings were used for the fMRI experiment. Altogether, 108 binaurally recorded sounds (12 directions × 9 times) were prepared for each participant. One hundred eight stereo and 108 mono sounds were made once and used for all participants. The duration of each of the sound stimuli was 250 ms. In order to include all reverberation, we kept a 150 ms interval in addition to the 100 ms duration of the sound source. Sound levels for each recording type were matched using coefficients that were calculated to match the root mean square energy of the sounds in front of the subjects (angle 0°).

For each trial, participants heard trains of three sounds (sound onset asynchrony 300 ms; duration of each train 850 ms; inter-train interval 2150 ms) and perceived auditory apparent motion (about 18.8 km/h). We divided 12 directions into 4 quadrants. Three sounds that belonged to the same quadrant were presented as a train either in ascending or in descending order (e.g., 0°, 30°, 60° or 60°, 30°, 0°). Auditory stimuli were delivered via MR-compatible headphones (Hitachi Advanced Systems' ceramic transducer headphones; frequency range 30–40,000 Hz, approximately 20 dB SPL passive attenuation). There were three experimental conditions defined by the recording method (binaural, stereo, mono) as well as a null condition in which the subject did nothing. We employed a rapid event-related paradigm and the conditions were presented randomly.