

## ORIGINAL ARTICLE

# An fMRI Study of the Ventriloquism Effect

Akiko Callan, Daniel Callan, and Hiroshi Ando

Center for Information and Neural Networks (CiNet), National Institute of Information and Communications Technology, Osaka University, Suita, Osaka 565-0871, Japan

Address correspondence to Akiko Callan, Center for Information and Neural Networks (CiNet), National Institute of Information and Communications Technology, Osaka University, 1-4 Yamadaoka, Suita, Osaka 565-0871, Japan. Email: acallan@nict.go.jp

## Abstract

In spatial perception, visual information has higher acuity than auditory information and we often misperceive sound-source locations when spatially disparate visual stimuli are presented simultaneously. Ventriloquists make good use of this auditory illusion. In this study, we investigated neural substrates of the ventriloquism effect to understand the neural mechanism of multimodal integration. This study was performed in 2 steps. First, we investigated how sound locations were represented in the auditory cortex. Secondly, we investigated how simultaneous presentation of spatially disparate visual stimuli affects neural processing of sound locations. Based on the population rate code hypothesis that assumes monotonic sensitivity to sound azimuth across populations of broadly tuned neurons, we expected a monotonic increase of blood oxygenation level-dependent (BOLD) signals for more contralateral sounds. Consistent with this hypothesis, we found that BOLD signals in the posterior superior temporal gyrus increased monotonically as a function of sound azimuth. We also observed attenuation of the monotonic azimuthal sensitivity by spatially disparate visual stimuli. The alteration of the neural pattern was considered to reflect the neural mechanism of the ventriloquism effect. Our findings indicate that conflicting audiovisual spatial information of an event is associated with an attenuation of neural processing of auditory spatial localization.

**Key words:** auditory, fMRI, spatial processing, multimodal integration, ventriloquism

## Introduction

In ventriloquism, the audience perceives speech sounds as coming from a direction other than their true direction. The speech sounds are localized at the location of the puppet's obviously moving mouth instead of the speaker's unmoving mouth. This ventriloquism effect is one example of "visual capture." In general, vision dominates (or captures) perception when spatially disparate visual and auditory stimuli are simultaneously presented. In the visual system, the spatial features of the stimulus are encoded by the location of stimulation on the retina. Location information is inherent at the peripheral level. In contrast, spatial perception in audition is much more challenging because the spatial features have to be extracted from sensory receptors organized according to sound frequency rather than spatial configuration, and because auditory cues are easily distorted by echoes

and reverberation in the environment. In humans, the spatial resolution of the visual system is superior to that of the auditory system. The ventriloquism effect can be explained as a phenomenon in which the sensory modality with the higher acuity dominates over and captures the other sensory modality with lower acuity (Warren et al. 1981; Alais and Burr 2004). Although the phenomena of the ventriloquism effect have been studied extensively [for a review, see Witten and Knudsen (2005)], its neural basis is still unresolved.

To investigate the neural basis of the ventriloquism effect, it is necessary to understand how source location is represented in the human auditory cortex. A topographical place code and a population rate code are 2 major hypotheses for the coding of auditory localization. The topographical place code assumes that auditory space is represented by the activation of neurons that correspond to particular locations in space (Jeffress 1948).

Conversely, the population rate code assumes that sound-source locations are represented by patterns of activity across populations of broadly tuned neurons (Middlebrooks et al. 1998). Spatially tuned neurons have been observed in mammalian superior colliculus (SC; Palmer and King 1982; Middlebrooks and Knudsen 1984) and in avian inferior colliculus (Knudsen and Konishi 1978; Takahashi et al. 1984). However, neurophysiological studies have reported broadly tuned, but not narrowly tuned, neurons in the auditory cortex in cats (Rajan et al. 1990; Brugge et al. 1996; Middlebrooks et al. 1998) and in monkeys (Werner-Reiss and Groh 2008). Thus, in mammals, most location-sensitive auditory cortical neurons are broadly tuned and respond to stimuli located throughout the contralateral space. These results suggest that the population rate code (not the topographical place code) is instantiated in the human auditory cortex.

Recent electrophysiological [EEG and magnetoencephalography (MEG)] studies using an adaptation paradigm have provided support for the population rate coding of sound azimuth localization in the human auditory cortex (Salminen et al. 2009; Magezi and Krumbholz 2010). The EEG experiment showed that the adaptation effect was stronger when an interaural time difference (ITD) changed toward the midline than when it changed away from the midline, and that the majority of spatially sensitive neurons in each hemisphere are tuned to the contralateral hemifield (Magezi and Krumbholz 2010). The MEG experiment showed that all adaptors were effective when the probe and the adaptor were within the same hemifield, but not effective when the adaptor was at the midline or in the opposite hemifield (Salminen et al. 2009). Both results indicate that the human auditory cortex represents sound-source locations with 2 populations of broadly tuned neurons. One population is sensitive to the contralateral hemifield and the other population is sensitive to the ipsilateral hemifield, especially to sounds located at lateral extremes (far to the left or to the right side of the perceiver). Thus, the existence of population rate coding in the human auditory cortex is supported by changes in the response level of N1 (the negative deflection at ~100 ms after sound onset). However, which part of the auditory cortex is responsible for the N1 changes and how neural activity levels represent the sound azimuth are still unknown. Compared with EEG and MEG, functional magnetic resonance imaging (fMRI) has higher spatial resolution and is expected to be able to clarify which cortical regions are associated with the population rate code. Nevertheless, there are no fMRI studies that have reported the relationship between the sound azimuth and neural activity. A single-unit recording study in monkeys reported a monotonic azimuthal sensitivity as well as a contralateral sensitivity at the population level (Werner-Reiss and Groh 2008). Despite these findings, many fMRI studies fail to show contralateral preference in the human auditory cortex [for a review, see Werner-Reiss and Groh (2008)]. The lack of sensitivity of the fMRI measurements as well as the impoverished nature of the stimuli used might be one of the reasons why many studies have failed to find a contralateral preference in the human auditory cortex. To investigate neural mechanisms for auditory spatial perception, we need to enhance neural response levels by presenting optimal stimuli and increase the sensitivity of the fMRI measurements.

Natural complex auditory stimuli that include both interaural and spectral temporal cues [i.e., head-related transfer functions (HRTFs)] are considered to elicit more salient neural responses in the posterior superior temporal gyrus (pSTG) than artificial simple auditory stimuli that only include interaural cues (Palmaki et al. 2005; Getzmann and Lewald 2010; Callan et al. 2013). Using realistic (externalized) auditory spatial stimuli and

applying region-of-interest (ROI) analyses, our previous fMRI study was able to show that both left and right pSTG were more sensitive to sound sources in contra- than in ipsilateral hemifields (Callan et al. 2013). As for stimulus content, we chose speech stimuli that are more complex and natural rather than the simple tones and lights used in most previous multimodal integration studies (Vander Wyk et al. 2010). Physiological investigation revealed that most neurons in the auditory cortex respond maximally to sounds located far to the left or to the right, but change most abruptly across the frontal midline (Stecker et al. 2005). Stecker et al. (2005) proposed that, in rate coding, the optimal spatial acuity across the frontal midline observed in behavioral studies is achieved by using the slopes of activity rather than the peaks of activity. Compared with sounds located far to the left or to the right, sounds located around the frontal midline can induce more abrupt neural response changes. Therefore, if we compare auditory stimuli located around the frontal midline, we may be able to detect a blood oxygenation level-dependent (BOLD) signal change that is monotonically related to the sound azimuth even with fMRI.

In this fMRI study, we first determined regions involved in sound azimuth representation and then investigated how conflicting visual stimuli affect neural activity in these regions. To enhance the detectability of BOLD signal changes, we used the following: (1) realistic spatial sounds that include both interaural and spectrotemporal cues; (2) complex natural speech as stimuli; (3) sound locations around the frontal midline where neural responses show greatest modulation; and (4) ROI analyses. We hypothesized that the regions associated with sound azimuth presentation in the auditory cortex (i.e., pSTG) will show increased BOLD signals monotonically as the sound-source changes from the frontal midline to more contralateral locations. We further hypothesized that the ventriloquism effect results from conflicting visual information suppressing the monotonic signal increase associated with sound azimuth representation.

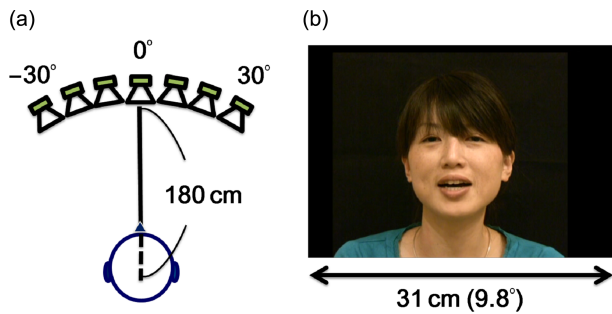
## Materials and Methods

### Participants

Sixteen adults (8 males; 22–44 years of age, mean 28.25) participated in this experiment. All participants had no neurological or psychiatric history, had pure-tone thresholds within the normal range ( $\leq 20$  dB HL) for octave frequencies between 250 and 8000 Hz, and gave written informed consent for experimental procedures approved by the institutional review board at the National Institute of Information and Communications Technology.

### Stimuli

The stimuli were audio and video recordings of a female-native Japanese speaker articulating Japanese greeting words at a natural speech rate. Five Japanese greeting words were used: “ohayou gozaimasu” (good morning), “konnichiwa” (good afternoon), “konbanwa” (good evening), “arigatou” (thank you), and “sayonara” (good bye). Each of the greeting words were repeated 5 times. In each video recording, the speaker’s face, shoulders, and neck were visible (see Fig. 1 for a sample frame). Video recordings were made in an anechoic chamber using a digital video camera (30 images/s frame rate, 48 kHz in a 16-bit audio sample rate; IVISHV10, JVC Corp., Japan) using a miniature condenser microphone (20–20,000 Hz frequency response; DPA4060) attached to the person’s collar as an audio input. The videos were edited using the software (Edius version 4; Canopus)



**Figure 1.** Experimental set-up. (a) Recording set-up of auditory stimuli. The speech sounds were presented through a loudspeaker from 1 of the 7 horizontal directions ( $-30^\circ$ ,  $-20^\circ$ ,  $-10^\circ$ ,  $0^\circ$ ,  $10^\circ$ ,  $20^\circ$ , and  $30^\circ$ , left to right, angle  $0^\circ$  was in front of the participant) with a distance of 180 cm from the participants. (b) An example frame of the visual stimuli.

and the duration of each video clip ranged from 0.9 to 1.2 s (“ohayou gozaimasu” 1.2 s, “konnichiwa” 1 s, “konbanwa” 1 s, “arigatou” 0.9 s, and “sayonara” 0.9 s). Then, we extracted the speech sounds from the videos. The extracted speech sounds were used for auditory stimuli recordings.

For the recordings, we used binaural (BR) and stereo-recording (SR) methods. The binaural-recorded sounds can elicit externalized (outside-the-head) perceptions of auditory space through headphones. In contrast, the stereo-recorded sounds can provide left or right information about the sound sources, but cannot elicit externalized perception; therefore, the sounds are localized inside but not outside of the head. For the BR, participants sat down on a chair. In ear, binaural microphones (20–20 000 Hz frequency response; 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 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 the left and right microphones was 15.6 cm). The speech sounds were presented through a loudspeaker (55–20 000 Hz frequency response; Eclipse TD508II; Fujitsu Ten Ltd) from 1 of the 7 horizontal directions ( $-30^\circ$ ,  $-20^\circ$ ,  $-10^\circ$ ,  $0^\circ$ ,  $10^\circ$ ,  $20^\circ$ , and  $30^\circ$ , left to right, angle  $0^\circ$  was in front of the participant) with a distance of 180 cm from the participants (Fig. 1a) and digitally recorded with a 16-bit resolution at a sampling rate of 48 kHz. To avoid variance caused by different acoustical characteristics of different speakers, only one speaker was used. For each direction, each greeting word was presented 5 times. BR stimuli were prepared for each participant. On the other hand, SR stimuli were prepared only once and the same SR stimuli were used for all participants. All auditory stimuli included 5.3 ms no-speech-sound-periods in the beginnings that reflect the distance between the speaker and the listener (180 cm) given the speed of sound (340 m/s). The durations of auditory stimuli were matched with the corresponding video stimuli. Sound levels for each recording type were matched using coefficients that were calculated to match the root mean square (RMS) energy of the sounds in front of the participants (angle  $0^\circ$ ).

## Procedure

For each trial, participants listened to one of the auditory stimuli and responded whether the sound source was located either left or right by pressing 1 of the 2 buttons. To minimize interruption by MRI scanning noise, we presented auditory stimuli during interscan intervals. Thus, participants never heard the scanning

noise while listening to the stimuli. The left or right forced choice task is the same task used for the ventriloquism aftereffect with macaque monkeys (Woods and Recanzone 2004). The left-right forced choice task was used instead of localizing perceived stimulus locations for 2 reasons. The first reason is because the left-right response is considered to be simpler and faster than the localization response, so that the task minimizes differences in the hemodynamic response resulting from long response times (Poldrack 2000). The second reason is because behavioral studies (Jack and Thurlow 1973; Thurlow and Jack 1973) of the ventriloquism effect determined that the localization task was not appropriate to measure the ventriloquism effect because the degree of visual capture obtained when subjects are instructed to localize a sound is quite small.

Stimulus presentation and timing was controlled by using the Presentation<sup>®</sup> software ([www.neurobs.com](http://www.neurobs.com)). Timing uncertainties were generally smaller than 0.2 ms and the various conditions did not lead to differences in the performance of the program. Auditory stimuli were presented with or without corresponding visual stimuli. When both auditory and visual stimuli were presented, their presentation timings were synchronized. Auditory stimuli were represented in 1 of the 7 locations, but visual stimuli were always presented in front of the participants. Auditory stimuli were delivered via MR-compatible headphones (Hitachi Advanced Systems’ ceramic transducer headphones; frequency range 30–40 000 Hz,  $\sim 20$  dB SPL passive attenuation). Mean intensities of auditory stimuli were “ohayou gozaimasu” 66 dB, “konnichiwa” 66 dB, “konbanwa” 67 dB, “arigatou” 65 dB, and “sayonara” 67 dB SPL. All visual stimuli were projected on a screen and viewed through a mirror mounted on the MRI head coil at the viewing distance of 180 cm (total display size  $9.8^\circ \times 7.4^\circ$  of visual angle; Fig. 1b).

We applied a block design paradigm with 4 experimental conditions (auditory-only BR, auditory-only SR, audiovisual BR, and audiovisual SR) and a baseline rest condition. Each participant performed five 10 min imaging runs. During each run, a 2-min cycle that includes 4 experimental blocks alternating with a rest block was presented 5 times. An example of one cycle is “auditory-only BR + rest + auditory-only SR + rest + audiovisual BR + rest + audiovisual SR + rest.” Each experimental block included 7 stimuli (i.e., 7 trials) and lasted 21 s. The rest block lasted 9 s and participants were instructed to simply look at the fixation on the screen. The order of the experimental conditions was randomized for each run. The effect of stimulus length was controlled by presenting each greeting word the same number of times for all conditions. The locations of the auditory stimuli were presented randomly throughout the experiment. To control participants’ gaze direction, a fixation marker was presented in the center of the screen throughout the experiment except when video stimuli were presented. Participants were instructed to try not to move their eyes and to always look at the center of the screen.

## MRI Data Acquisition and Preprocessing

For structural and functional brain imaging, a 3-T scanner (Siemens MAGNETOM Trio, A Tim system) was used at the ATR Brain Activity Imaging Center. Functional  $T_2^*$ -weighted images were acquired using a gradient echo planar imaging sequence (TR = 3000 ms, matrix size =  $64 \times 64$  pixels, field of view =  $192 \times 192$  mm, slice thickness = 3 mm with a 1-mm gap, 30 slices). For each functional imaging run, 205 volumes were obtained. The first 2 scans from each run were discarded to allow for  $T_1$  equilibration effects. The acquisition time was 1800 ms so there was a

1200-ms quiet period between scans. We presented auditory stimuli during these interscan intervals, so that the presentation of auditory stimuli was not interrupted by MRI scanning noise.

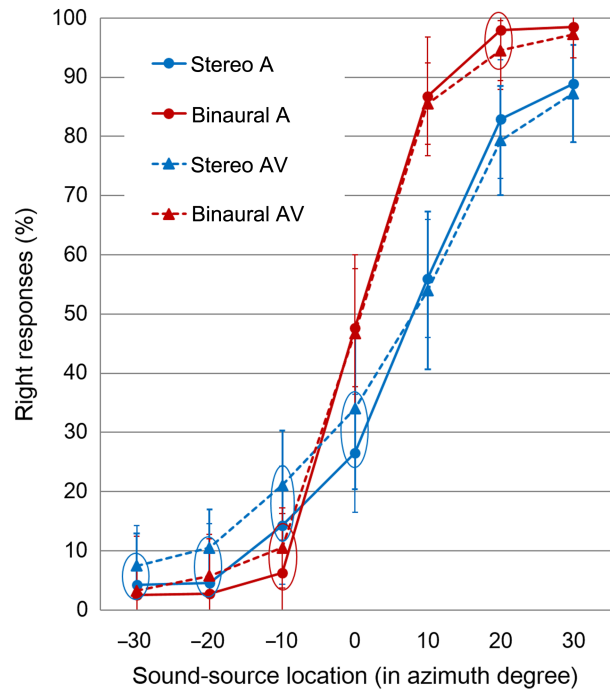
Images were preprocessed using programs within SPM8 (Wellcome Department of Cognitive Neurology, London, UK). Images were realigned and spatially normalized using a template defined by the Montreal Neurological Institute (MNI), and were smoothed using a twice voxel size ( $6 \times 6 \times 8$  mm) FWHM Gaussian kernel. Before the acquisition of functional images,  $T_2$ -weighted anatomical images were acquired in the same plane as the functional images (matrix size =  $256 \times 256$  pixels). The  $T_2$ -weighted images were coregistered with the mean of the functional images and used to calculate the parameters for the spatial normalization of the functional images.

### fMRI Data Analysis

For both experiments, preprocessed fMRI data were analyzed statistically on a voxel-by-voxel basis using SPM8 (128 s high-pass filter, serial correlations corrected by an autoregressive AR (1) model). The task-related neural activity was modeled with a series of events convolved with a canonical hemodynamic response function. Six movement parameters derived from the realignment were also included in the model. The 7 locations ( $-30^\circ$ ,  $-20^\circ$ ,  $-10^\circ$ ,  $0^\circ$ ,  $10^\circ$ ,  $20^\circ$ , and  $30^\circ$ ) of the 4 presentation types were used as experimental conditions. Left (or right) lateralized activity was tested by a parametric contrast [3, 2, 1,  $-1.5$ ,  $-1.5$ ,  $-1.5$ ] (or [ $-1.5$ ,  $-1.5$ ,  $-1.5$ ,  $-1.5$ , 1, 2, 3]) for the auditory-only BR condition and auditory-only SR condition separately.

Then, a second-level random-effect analysis was performed to yield statistical parametric maps of the resulting  $t$ -values for each contrast at the group level. Significant activation clusters were determined using a height threshold of  $P < 0.001$  uncorrected and an extent threshold of  $P < 0.05$  corrected for family-wise error (FWE). Additionally, small-volume correction (SVC) for multiple comparisons ( $P < 0.05$ , FWE corrected) was performed using bilateral superior temporal gyrus templates (3875 voxels) in the Automated Anatomical Labeling atlas (Tzourio-Mazoyer et al. 2002). The use of SVC allows researchers to conduct principled correction using Gaussian Random Field Theory within a predefined ROI. To conduct ROI analyses, we used an SPM toolbox called "Marsbar" (<http://marsbar.sourceforge.net/>). ROI analyses were performed in the following steps: (1) ROIs were defined as spheres with a 5-mm radius centered on the peak voxels located in the superior temporal gyrus for the BR and SR conditions separately; (2) we obtained an averaged fMRI signal of all voxels within the spheres and calculated mean beta values for all experimental conditions; (3) for each presentation type, we obtained a slope of left (or right) lateralized activity; and (4) significant differences were investigated by comparing between the auditory-only and audiovisual conditions with paired  $t$ -tests.

In addition, using behavioral data as regressors, we examined which brain regions were activated more by participants who were more affected by visual stimulus presentation. For this analysis, first, we evaluated laterality of sound-source perception by subtracting 50% from percentages of "right" responses at each sound location. Secondly, for each participant, we averaged absolute values of them and made laterality indexes for the auditory-only and audiovisual conditions separately. Then, we subtracted the audiovisual index from the auditory-only index. Bigger differences indicated stronger ventriloquism effects, because the audiovisual average was expected to be smaller than the auditory-only average if the centrally presented visual stimuli captured auditory perception of the laterally presented auditory



**Figure 2.** Percentage of "right" responses. Means and standard errors of the means (SEM) are plotted. On both SR and BR conditions, slopes were steeper for auditory-only conditions than for audiovisual conditions. At each sound location, a significant difference ( $P < 0.05$ ) between the auditory-only and audiovisual conditions was indicated by a circle. Stereo A, SR auditory-only; binaural A, BR auditory-only; stereo AV, SR audiovisual; binaural AV, BR audiovisual.

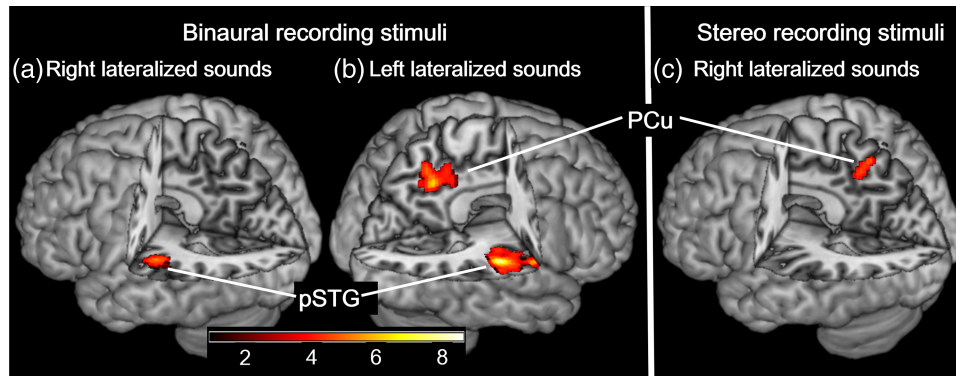
stimuli. Finally, we tested a linear regression between those behavioral data and BOLD responses (the audiovisual > baseline contrast images) for the BR and SR stimuli separately. For visualization of results, the activation maps are superimposed on a high-resolution anatomical MRI brain template (ch2better.nii.gz) using the MRIcron software (Rorden et al. 2007).

## Results

### Behavioral Data

For each experimental condition, percentages of "right" responses were calculated for each participant and the mean percentages were plotted in Figure 2. Sound sources located closer to the midline were naturally more difficult to lateralize so that psychometric functions could be generated. In this analysis, 100% "right" responses were expected for the most right stimuli (i.e.,  $30^\circ$ ), 50% "right" responses were expected for the center stimuli (i.e.,  $0^\circ$ ), and 0% "right" responses were expected for the most left stimuli (i.e.,  $-30^\circ$ ). Although a response curve with the BR stimuli was consistent with the expected psychometric function, a response curve with the SR stimuli was not. The mean percentage of right responses for  $0^\circ$  was 26.5% instead of 50%. The result indicated that the  $0^\circ$  SR stimuli represented not the center but the left.

In SPSS (IBM SPSS version 22), we performed a within-subjects three-way analysis of variance with presentation type (auditory-only and audiovisual), recording method (BR and SR), and sound-source location ( $-30^\circ$ ,  $-20^\circ$ ,  $-10^\circ$ ,  $0^\circ$ ,  $10^\circ$ ,  $20^\circ$ , and  $30^\circ$ ) as factors. There were significant main effects of recording method ( $F_{1,15} = 12.748$ ,  $P < 0.01$ ) and of location ( $F_{6,90} = 171.524$ ,  $P < 0.01$ ). On average, participants perceived the SR stimuli as more left-lateralized



**Figure 3.** Brain regions activated to a greater extent for more lateral sound locations. Significant activation was rendered on a high-resolution anatomical MRI brain template and a color bar indicates *t*-value. (a) The left pSTG was activated to a greater extent for more right lateral ( $30^\circ > 20^\circ > 10^\circ > 0^\circ$ ) BR stimuli. (b) The right pSTG and the right precuneus (PCu) were activated to a greater extent for more left lateral ( $-30^\circ > -20^\circ > -10^\circ > 0^\circ$ ) BR stimuli. (c) The left PCu was activated to a greater extent for more right lateral ( $30^\circ > 20^\circ > 10^\circ > 0^\circ$ ) stereo-recording stimuli.

than the BR stimuli. Furthermore, the percentage of right responses increased to a greater extent for more right lateral locations. All 3 two-way interactions were significant. The first interaction was between the presentation type and the recording method. It indicated that the difference between the audio-only and audiovisual conditions was bigger for the SR stimuli than for the BR stimuli ( $F_{1,15} = 5.523$ ,  $P < 0.05$ ). The second interaction was between the recording method and the sound-source location. It indicated that the sigmoid function for the left-to-right locations for the BR stimuli was sharper than that for the SR stimuli ( $F_{6,90} = 22.189$ ,  $P < 0.01$ ). The third interaction was between the presentation type and the sound-source location. It indicated that the sigmoid function for the auditory-only condition was sharper than that for the audiovisual condition ( $F_{6,90} = 2.674$ ,  $P < 0.05$ ). Bonferroni post hoc tests indicated significant ( $P < 0.05$ ) differences between the auditory-only and audiovisual conditions at  $-30^\circ$ ,  $-10^\circ$ , and  $20^\circ$ . The three-way interaction was not significant.

In addition, as planned comparisons, we performed paired *t*-tests to investigate in which locations visual stimulus presentation significantly affected auditory perception for the BR and SR stimuli separately. Significant ( $P < 0.05$ ) differences between auditory-only and audiovisual conditions were observed at  $-30^\circ$ ,  $-20^\circ$ ,  $-10^\circ$ , and  $0^\circ$  with the SR stimuli and  $-10^\circ$  and  $20^\circ$  with the BR stimuli (Fig. 2). We also tested whether visual stimulus presentation affected left and right sound source locations differently (i.e., whether differences between auditory-only and audiovisual conditions were larger at left or right sound source locations). For this analysis, we compared  $-30^\circ$  with  $30^\circ$ ,  $-20^\circ$  with  $20^\circ$ , and  $-10^\circ$  with  $10^\circ$  for the BR and SR conditions separately. There were no significant differences in these comparisons.

## fMRI Data

### The Auditory-Only Conditions

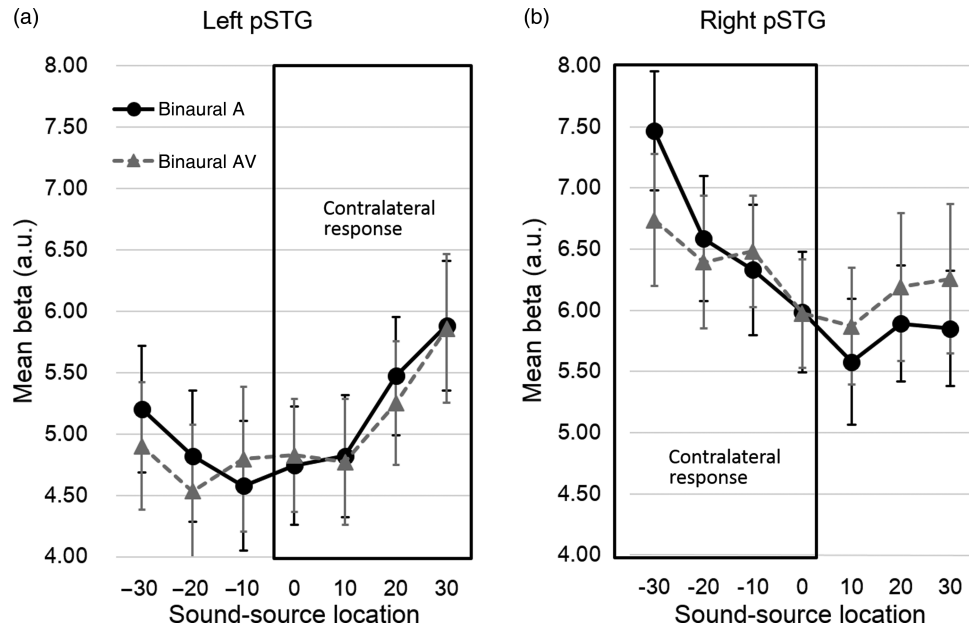
Under auditory-only conditions, we determined brain regions activated to a greater extent for more lateral sound locations. The results of the whole-brain analyses (height threshold of  $P < 0.001$  uncorrected and an extent threshold of  $P < 0.05$  FWE corrected) for the BR stimuli revealed that linearly increasing activity for the left lateral sounds existed in the right pSTG (MNI coordinates at the peak voxel  $x y z = 52, -24, 10$ , 859 voxels) and in the right precuneus ( $x y z = 8, -50, 42$ , 315 voxels), and that linearly increasing activity for the right lateral sounds existed in the left pSTG ( $x y z = -56, -32, 12$ , 153 voxels; Fig. 3). Significant increased

activity for the contralateral sounds in the left and right pSTG were also found by the additional SVC ( $P < 0.05$  FWE corrected) using the bilateral superior temporal gyrus templates. The whole-brain analyses for the SR stimuli found linearly increasing activity for the right lateral sounds in the left precuneus ( $x y z = -18, -46, 46$ , 147 voxels). The whole-brain analyses and the SVC failed to show linearly increasing activity in the STG with the SR stimuli.

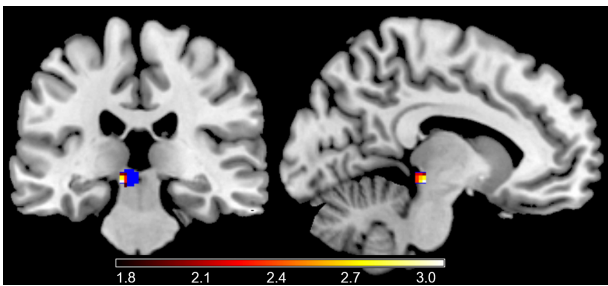
### ROI Analyses Comparing the Auditory-Only Condition with the Audiovisual Condition

We performed ROI analyses using peak voxel coordinates in the STG for the BR and SR conditions separately. For the BR condition, the center coordinates for the ROI spheres were  $x y z = -56, -32, 12$  and  $x y z = 52, -24, 10$ . Mean beta values in the left and right regions are plotted in Figure 4. A significant decrease in slope for the audiovisual condition compared with the auditory-only condition for the BR stimuli was found in the right pSTG ( $t_{15} = 1.79$ ,  $P < 0.05$ ; Fig. 4b). In the right pSTG, visual stimuli presented in the midline reduced neural responses for left-lateralized sounds compared with neural responses when those sounds were presented without visual stimuli. For the SR condition, the center coordinates for the ROI spheres were  $x y z = -56, -46, 12$  and  $x y z = 44, -28, 16$ . The ROI analyses with the SR stimuli failed to show significant differences between the auditory-only and the audiovisual conditions.

Krumbholz et al. (2005) proposed a hierarchical organization of auditory spatial processing in which analysis of simple interaural cues begins at the brainstem and more complex signal sensitivity (in their case, moving sound) emerges in the pSTG. From this proposal, it follows that the ventriloquism effect for the simple SR stimuli was not found in the pSTG because the neural change is processed below the level of the primary auditory cortex, in the brainstem. In the brainstem, the SC is a site for multi-sensory integration (Stein and Stanford 2008). To check whether the SC showed linearly increasing activity, we performed additional SVC analyses using spheres with a 5-mm radius centered on the left ( $x y z = -6, -28, -6$ ) and right ( $x y z = 6, -28, -6$ ) SC peak voxels reported in a previous study (Linzenbold and Himmelbach 2012). Linearly increasing activity for the left lateral SR stimuli was not found, but linearly increasing activity for the right lateral SR stimuli was found in the left SC ( $x y z = -10, -26, -8$ ,  $P < 0.05$  FWE corrected; Fig. 5). Using contrast estimates of the peak voxel for each participant, we compared the slope of the auditory-only



**Figure 4.** Activation of pSTG correlated with sound azimuth. Means and SEMs of beta (parameter estimates) are plotted. Both left and right pSTG responded significantly greater for contralateral BR stimuli. (a) The left pSTG responded more for right lateral sounds. (b) The right pSTG responded more for left lateral sounds and the slope was significantly steeper for the auditory-only condition than for the audiovisual condition. Binaural A, BR auditory-only; binaural AV, BR audiovisual.



**Figure 5.** The left superior colliculus activated to a greater extent for more right lateral sound locations. Activation ( $P < 0.05$  for display purpose) is plotted on a high-resolution anatomical MRI brain template and a color bar indicates  $t$ -value. The blue region corresponds to the left region sphere for the SVC analysis.

condition with the slope of the audiovisual condition. In contrast to findings in the right pSTG, a significant difference ( $P < 0.05$ ) was found for the SR stimuli but not for the BR stimuli.

#### Linear Regression Between the BOLD Responses and the Behavioral Data

We performed linear regression analyses between the BOLD responses during the audiovisual trials and the behavioral data. With the BR stimuli, we found a significant positive correlation (the height threshold  $P < 0.001$  and the extent threshold  $P < 0.05$  FWE corrected) in the bilateral middle temporal gyri ( $x y z = 56, -12, -22$  and  $x y z = -58, -14, -14$ ), the left middle occipital gyrus ( $x y z = -38, -70, -2$ ), the left middle frontal gyrus ( $x y z = -34, 18, 30$ ), the right pSTG including Heschl's gyrus ( $x y z = 38, -26, 12$ ), the right parahippocampal gyrus ( $x y z = 34, -30, -18$ ), and the bilateral hippocampus ( $x y z = 32, -22, -20$  and  $x y z = -32, -8, -26$ ). The scatter plots show the covariation between brain activity (contrast estimates from the peak voxel in the right middle temporal and in the left middle occipital clusters) and the behavioral data (Fig. 6). As can be seen in Figure 6, there is one outlier in

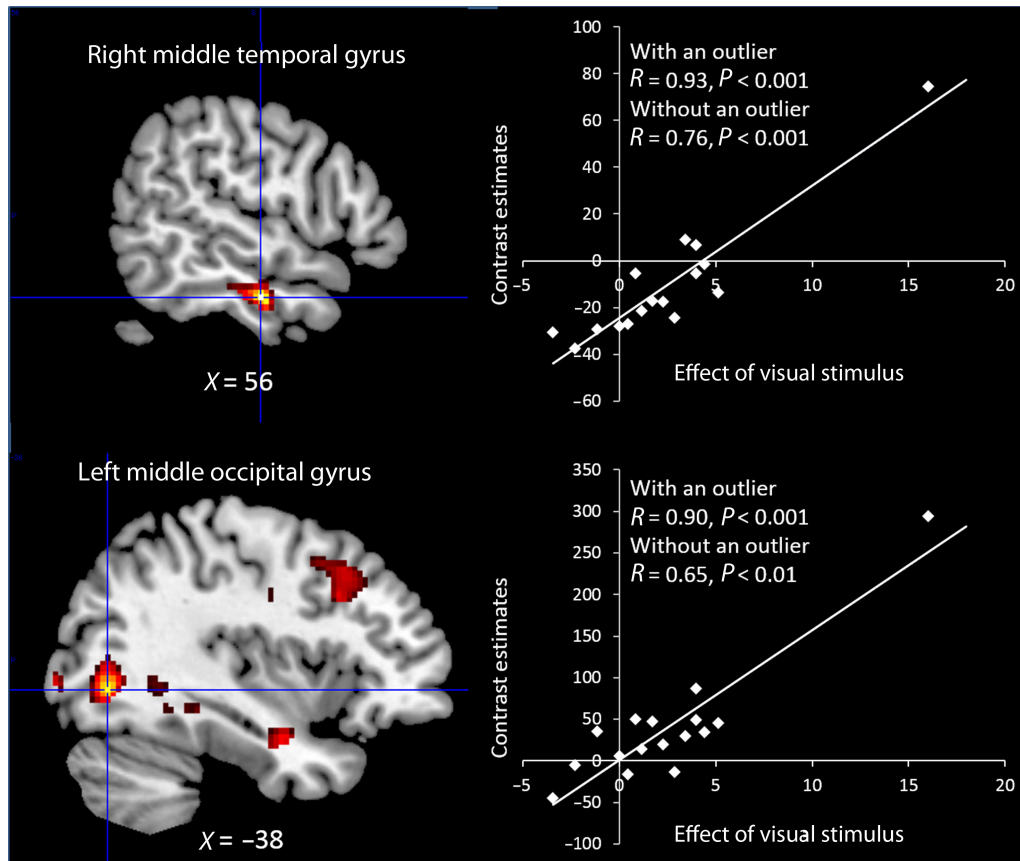
each region. We calculated correlation without the outliers and correlations were still significant ( $R = 0.76$ ,  $P < 0.001$  in the right middle temporal gyrus, and  $R = 0.65$ ,  $P < 0.01$  in the left middle occipital gyrus). During the BR audiovisual trials, participants who were more influenced by the visual stimuli activated those regions more. With the SR stimuli, we could not find any significant correlation.

## Discussion

This is the first study to show sound azimuth-related BOLD signal changes in the human auditory cortex. The observed signal change pattern was very similar to the averaged responses of 119 neurons in the monkey auditory cortex (Werner-Reiss and Groh 2008). Neural responses did not vary much for ipsilateral sound locations and increased monotonically for more contralateral sound locations. In the audiovisual condition, centrally presented visual stimuli attenuated the monotonic change related to the sound azimuth. This attenuation of the auditory monotonic response function may be the underlying neural mechanism responsible for the ventriloquism effect.

### Behavioral Results

The significant difference between the auditory-only and audiovisual conditions confirmed that the ventriloquism effect was induced in this experimental setting. Participants' performance for left and right identification deteriorated when they were presented with visual stimuli. The behavioral effect was stronger for the SR stimuli than the BR stimuli. We suspect that the different acuity levels of auditory stimuli caused this difference. As can be seen in Figure 2, a sigmoid function of the BR stimuli was sharper than one of the SR stimuli, indicating that participants identify whether the sound source was located either left or right more clearly for the BR stimuli compared with the SR stimuli. The mean percentage of "right" responses for the most left ( $-30^\circ$ ) sound was 2.5% for the BR stimuli and 4.25% for the SR



**Figure 6.** The result of the linear regression analysis between the neural responses and the behavioral data. Left: the activation maps are superimposed on a high-resolution anatomical MRI brain template. Crosshairs corresponds to peak voxels from which we extracted contrast estimates in the scatter plots. Right: scatter plots illustrating the relationship between the neural responses during the audiovisual trials and the behavioral index of the ventriloquism effect. The large behavioral index indicates strong influence of visual stimuli on auditory localization. Each dot represents a separate participant.

stimuli. The mean percentage for the most right ( $30^\circ$ ) sound was 98.5% for the BR stimuli and 88.9% for the SR stimuli. Considering that the sensory modality with lower acuity is captured by the sensory modality with higher acuity (Alais and Burr 2004), it is reasonable to assume that the SR stimuli with lower acuity than the BR stimuli were more easily captured by conflicting visual information.

With the left or right forced choice task, we expected that the  $0^\circ$  auditory stimuli would yield near-chance level (50%) right responses. The BR stimulus result was consistent with the expectation but the SR stimulus result was not. Mean percentage for the SR  $0^\circ$  stimuli was 26.5%, indicating that they were on average localized in the left side but not in the midline. To figure out the reason of this left-lateralized responses, we inspected ITD and interaural level difference (ILD) cues of the BR and SR stimuli. We computed mean ITD values by using interaural cross-correlation and mean ILD values by the ratio of the left channel RMS to the right channel RMS. For the ILD computation, the first 10 ms of the speech signals were used to prevent the reverberant sound field from affecting the estimate. The numerical values of the ITDs for the azimuth angles  $-30^\circ$ ,  $-20^\circ$ ,  $-10^\circ$ ,  $0^\circ$ ,  $10^\circ$ ,  $20^\circ$ , and  $30^\circ$  were  $-0.27$ ,  $-0.20$ ,  $-0.12$ ,  $0.02$ ,  $0.08$ ,  $0.21$ , and  $0.35$  ms, respectively, for the BR stimuli and  $-0.19$ ,  $-0.14$ ,  $-0.07$ ,  $0$ ,  $0.07$ ,  $0.14$ , and  $0.19$  ms, respectively, for the SR stimuli. The minus ITD indicates that the left ear sound precedes the right ear sound. On the other hand, for the azimuth angles  $-30^\circ$ ,  $-20^\circ$ ,  $-10^\circ$ ,  $0^\circ$ ,  $10^\circ$ ,  $20^\circ$ , and  $30^\circ$ , the numerical values of the ILDs were 4.80, 4.08, 2.67, 0.09,  $-1.77$ ,

$-3.33$ , and  $-4.03$  dB, respectively, for the BR stimuli and 1.50, 1.29, 0.77, 0.49,  $-0.73$ ,  $-1.04$ , and  $-1.25$  dB, respectively, for the SR stimuli. The positive ILD indicates that the left ear sound is louder than the right ear sound. The mean  $0^\circ$  ITDs for both the BR and SR stimuli were close to 0 ms, indicating a midline position. The mean  $0^\circ$  ILD for the BR was 0.09 (close to 0) dB, but the mean ILD for the SR was 0.49 dB, indicating a left position. The positive ILD value of the SR  $0^\circ$  stimuli is considered a reason why they were localized to the left side on average.

### Sound Azimuth-Related BOLD Signal Changes with the Auditory-Only Condition

Using the externalized auditory (i.e., BR) stimuli, we found linearly increasing activity for more contralateral sounds in the pSTG. The similar neural response changes in the pSTG were not found for the SR stimuli. A parsimonious explanation of this result is that the ITD and ILD cues of the SR stimuli were less salient than ones of the BR stimuli because of the lack of the head-shadow effect. In fact, as we can tell from the mean ITD and ILD values of our stimuli, they were less salient for the SR stimuli than the BR stimuli. Nevertheless, we do not think that this is the only reason why linearly increasing activity for more left lateral sounds in the right pSTG was not found for the SR stimuli. Using N1m responses recorded by MEG, Palomaki et al. (2005) compared individualized BR, non-individualized (generic) HRTF, ITD, ILD, and combined ITD and ILD (ITD + ILD) stimuli.

In their study, the ITD and ILD stimuli were constructed from averaged ITD and ILD values of BR stimuli, so that the ITD and ILD values of their ITD + ILD stimuli were as salient as the values of their BR stimuli. The N1m amplitude response dynamics were analyzed by subtracting the contralateral N1m amplitudes from their ipsilateral counter parts. The results of this analysis revealed significant effects of stimulus type in both the left and the right hemisphere, but the dynamic range in the right hemisphere was larger than that in the left hemisphere (Palomaki et al. 2005). Moreover, the response dynamics in the right hemisphere reflected the amount of spatial cues in the stimuli (i.e., BR > HRTF > ITD + ILD > ITD > ILD; Palomaki et al. 2005). Their study also reported that the correlation between the N1m amplitude and behavioral performance was much higher for the BR stimuli than for the generic HRTF stimuli (Palomaki et al. 2005). These results indicate that more realistic auditory spatial stimuli yield a larger dynamic range of brain activity in the right auditory cortex, and that the larger dynamic range is associated with better localization accuracy.

The differential neural activity in the pSTG between the BR and SR stimuli supports the hypothesis that realistic complex auditory stimuli elicit greater neural responses than artificial simple auditory stimuli in the auditory cortex (Palomaki et al. 2005; Getzmann and Lewald 2010; Callan et al. 2013). The involvement of the pSTG in processing complex sounds has been suggested by neurophysiological studies (Rauschecker et al. 1995; Rauschecker, 1998). It may also explain why previous fMRI studies using ITD and ILD cues failed to observe contralateral preferences in the auditory cortex (Woldorff et al. 1999; Brunetti et al. 2005; Zimmer and Macaluso 2005; Zimmer et al. 2006). The right pSTG was activated to a greater extent by the more left-lateralized sounds and the left pSTG was activated to a greater extent by the more right-lateralized sounds. The sound azimuth-related BOLD signal change in the pSTG was revealed by the ROI analysis (Fig. 4). Our results provide strong support for location coding by opponent neural populations (i.e., population rate coding) in the human auditory cortex (Stecker et al. 2005). Since this experiment only investigated sound azimuth from  $-30$  to  $30^\circ$ , we cannot state how BOLD signal changes for more lateralized sounds. However, based on neurophysiological work with monkeys (Werner-Reiss and Groh 2008), we assume that neural responses are greatest for sounds at the lateral peaks, but show the least modulation for sounds around the lateral peaks. Neural responses in the monkey auditory cortex showed similar activation levels between  $60^\circ$  and  $90^\circ$  (Werner-Reiss and Groh 2008). To understand the whole response function in the right pSTG, an fMRI experiment to cover the whole  $360^\circ$  horizontal azimuth needs to be conducted.

In addition to the pSTG, sound azimuth-related BOLD signal changes were observed in the right precuneus with the BR stimuli and in the left precuneus with the SR stimuli. The right precuneus was activated to a greater extent for more left lateral sounds, and the left precuneus was activated to a greater extent for more right lateral sounds. Krumbholz et al. (2009) reported that the precuneus was involved in spatial attention shifts not only in the visual modality but also in the auditory modality, and responded more strongly to attention shifts toward the contralateral than to the ipsilateral hemisphere. We suspect that the precuneus activation we found in this study may reflect spatial attention shifts caused by presenting auditory stimuli from various locations.

### Effects of Visual Stimuli on Auditory Spatial Perception

The main purpose of this study is to determine how conflicting visual stimuli affect auditory spatial perception. The auditory-

only minus audiovisual contrast was used for this purpose. A significant difference was only observed in the right pSTG region with the BR stimuli (Fig. 4b). The results indicate that neural responses to the left lateral sounds were attenuated by observing visual stimuli presented at the frontal midline. The attenuation of the monotonic azimuthal sensitivity by spatially disparate visual stimuli was considered to reflect the neural mechanism of the ventriloquism effect such that visual spatial information perceptually captures auditory spatial information. The attenuation does not mean that monotonically increasing neural activity in the pSTG disappeared completely. If participants perceived the left-lateralized sounds were located at the center every time visual stimulus was presented, one would expect to see the complete diminishment of the monotonic neural response change. However, as can be seen in Figure 2, the mean percentage of the left lateral sounds on the audiovisual condition was more than that of the left lateral sounds on the auditory-only condition but still <50%. It means that the overall responses for the audiovisual stimuli were less left-lateralized than the auditory-only stimuli but still left-lateralized. The behavioral results agree with the neuroimaging results, showing that the slope for the audiovisual sounds was significantly smaller (i.e., less left lateralized) than that for the auditory-only condition but still significantly positive (i.e., left lateralized).

Neuroimaging and lesion studies show that the auditory cortices in both hemispheres are involved in auditory spatial perception; however, the right hemisphere plays the dominant role in sound localization in humans (Palomaki et al. 2005). Although we found sound azimuth-related BOLD change in both left and right pSTG, the right pSTG cluster size (859 voxels) was much bigger than the left pSTG cluster size (153 voxels; Fig. 3). Moreover, the effect of visual stimuli on auditory spatial processing was only observed in the right pSTG. These results also indicate the right auditory cortex dominance in sound localization. Cytoarchitectonic studies in the human brain have shown a greater volume of white matter in left than in right Heschl's gyrus and the posterior temporal lobe indicating faster transmission and greater temporal resolution available in the left auditory cortex (Zatorre et al. 2002). A functional imaging study comparing spectral and temporal processing in auditory cortex found that temporal resolution is better in the left auditory cortical areas and spectral resolution is better in the right auditory cortical areas (Zatorre and Belin 2001). It is well known that the ITDs and ILDs are 2 main mechanisms to localize a sound source in the horizontal plane (Blauert 1997). Spectral cues are considered to be supplemental for horizontal localization judgment. However, greater involvement of the right pSTG in spatial processing indicates that spectral information processing is also important to determine the spatial location of the sound source in the horizontal dimension.

One limitation of this present study is that we could not compare between the neural responses to stimuli that are actually physically displaced in space versus stimuli that are displaced only as a result of the ventriloquism illusion. In this study, we could not specify trials in which they experienced the ventriloquism illusion based on their responses, because we employed the left-right forced choice task instead of the localization task. Therefore, even if visual stimulus presentation affected their perception and they perceived a  $3^\circ$  sound as a  $20^\circ$  sound, their response remains the same ("right"). Because of this reason, we could not distinguish the illusion trials from the no-illusion trials. However, our results are in line with Bonath et al.'s (2007) study that reported results of such a comparison. In their experiment, participants were asked to indicate the location of the



perceived tone (central, left, or right). In the case when a central sound was presented with a left or right visual stimulus, a trial where the central sound was perceived as a left or right sound was treated as an illusion trial and a trial where the sound was perceived as a central sound was treated as a no-illusion trial. By subtracting illusion trials from no-illusion trials, they found reduced activity in the planum temporale (PT) of the hemisphere ipsilateral to the visual stimulus (i.e., shifted auditory percept). They also compared left or right sounds with central sounds without visual stimuli and found enhanced activity in the PT of the hemisphere contralateral to the auditory stimuli. Their results indicate that the lateralized auditory percept is represented by asymmetrical neural responses in the PT and the laterally shifted auditory percept, in their study, resulted from a relatively enlarged response in the contralateral PT by reducing activity in the ipsilateral PT. The loci (i.e., the pSTG including the PT) found in this study are consistent with the loci found in their study. Moreover, the neural activity pattern change found in this study is analogous to their study. In this study, when visual stimuli were presented at the center (instead of audio in the center as in Bonath et al. 2007), the neural responses to lateral sounds that are normally asymmetric became more symmetric (in accordance to the central visual stimuli) by reducing activity in the contralateral pSTG.

In the behavioral results, the ventriloquism effect was bigger for the SR stimuli than the BR stimuli. Nevertheless, we failed to find neural response changes caused by visual stimulus presentation for the SR stimuli. The lack of significant findings could be simply due to less salient ITD and ILD cues of the SR stimuli. As we discussed earlier, we think this is unlikely because of the more dynamic response change in the pSTG for the BR stimuli than the matched ITD + ILD (Palomaki et al. 2005). The pSTG, or PT, more specifically, has been considered as a computational hub that segregates the components of the acoustic world and matches these components with learned spectrotemporal representations (Griffiths and Warren 2002). In our previous fMRI study, we found more enhanced activation by the BR stimuli compared with the SR stimuli in the pSTG (Callan et al. 2013). The MEG study showed that artificial ITD and ILD stimuli can activate the auditory cortex, but realistic auditory stimulus with spectrotemporal cues can yield a larger dynamic range of brain activity (Palomaki et al. 2005). In line with these findings, we found linearly increasing activity for more lateral sounds in the pSTG with the BR, but not with the SR stimuli. We assume that direction-specific modulation in the pSTG produced by the SR stimuli was not large enough to show the significant effects of visual stimuli.

Krumbholz et al. (2005) proposed a hierarchical organization of auditory spatial processing in which analysis of simple interaural cues begins at the brainstem and more complex signal sensitivity (in their case, moving sound) emerges in the pSTG. Based on the proposal, we assumed that the ventriloquism effect for the simple SR stimuli was processed below the level of the primary auditory cortex, in the brainstem. Because the SC in the brainstem is a site for multisensory integration (Stein and Stanford 2008), we performed additional SVC analyses at the SC. In contrast to finding in the right pSTG, linearly increasing activity in the left SC for the right lateral sounds was only found for the SR stimuli, but not for the BR stimuli. Moreover, a significant difference between the slope of the auditory-only condition and the slope of the audiovisual was also found only for the SR stimuli. The result supported our assumption that the ventriloquism effect for the SR stimuli was processed below the level of the primary auditory cortex, in the SC.

## Correlation Between the Neural Responses and the Behavioral Data

We found that azimuthal representation in the right pSTG is attenuated by a spatially conflicting visual stimulus. However, we still do not know how visual processing affects auditory processing. Results of regression analyses between the BOLD responses and the behavioral data provide insights into the underlying neural mechanism. In these analyses, we found significant positive correlation in the bilateral middle temporal gyri, the left middle occipital gyrus, the left middle frontal gyrus, and the right pSTG including Heschl's gyrus, the parahippocampal gyrus, and the hippocampus. Most of those areas have been reported to be involved with spatial processing. The middle occipital gyrus showed a preference for visual spatial over nonspatial processing (Renier et al. 2010), the middle frontal gyrus showed engagement in the storage of spatial information (Leung et al. 2002), the hippocampus is involved in spatial memory (Burgess et al. 2002), and the parahippocampal gyrus plays a critical role in spatial navigation (Epstein 2008). Stronger facilitation of those areas by participants who were more influenced by spatially conflicting visual stimuli indicates that they processed visual spatial information more extensively than participants who were less influenced by spatially conflicting visual stimuli.

## Plasticity in Human Sound Localization

The neural response changes in the pSTG observed in this study indicate the plastic nature of spatial processing in the pSTG. The plastic nature of auditory spatial perception has been demonstrated by the ventriloquism aftereffect. The ventriloquism aftereffect is a phenomenon characterized by an enduring shift in the perception of acoustic space in the absence of the visual stimulus after prolonged (20–30 min) exposure to a consistent audio-visual spatial disparity (Recanzone 1998). More recent studies reported that even milliseconds of single exposure to an auditory-visual discrepancy could cause recalibration of perceived auditory space (Wozny and Shams 2011), and that a few minutes of exposure produced consolidated and long-lasting aftereffects (Frissen et al. 2012).

Since audiovisual spatial disparity was randomly changed in this study, we cannot investigate whether long-lasting neural response changes associated with the ventriloquism aftereffect occur or not. However, visually driven neural changes in the pSTG provide strong evidence that the cortical plasticity is mediated by the pSTG. A behavioral study investigating whether ventriloquism aftereffects generalize across sound frequencies found that the aftereffects could be generalized across a four-octave range of test frequencies (Frissen et al. 2005). In their study, the ITD-dominant low-frequency (400 Hz) adapter produced the aftereffect to the ILD-dominant high-frequency (6400 Hz) test tone, and the results supported a hypothesis that the locus of recalibration was at least beyond the level of the peripheral ITD and ILD mechanisms (Frissen et al. 2005). The ITD and ILD processing pathways converge in the inferior colliculus (for review, see Konishi 2003). Therefore, our findings, showing recalibration in the pSTG for the BR stimuli and in the SC for the SR stimuli, are in line with the hypothesis provided by the behavioral study of the ventriloquism aftereffect.

## Ventriloquism Effect in Relation to the Auditory Dorsal Pathway

Like the visual system, it has been proposed that the auditory system have 2 functionally distinct and anatomically segregated processing streams at the cortical level (2-stream hypothesis;

Rauschecker and Tian 2000; Alain et al. 2001). The 2 streams are dorsally or ventrally located and are considered to process location (“where”) or identity (“what”) information of objects, respectively. In the auditory system, the dorsal pathway emanates from the pSTG and the ventral pathway emanates from the anterior STG. Our results that show the involvement of the pSTG in auditory localization are consistent with the dorsal “where” auditory system. However, it is unclear how the system is associated with attenuation of neural activity in the pSTG by spatially conflicting visual stimuli?

In recent years, the involvement of the dorsal pathway in perception–action processing has been discussed. Arnott and Alain (2011) suggest that auditory spatial processing in the dorsal pathway may be understood as a form of action processing in which the visual system may be guided to a particular location of interest. In their account, auditory spatial processing serves to inform and orient eyes toward the location of interest. They also discuss that the dorsal stream functions as a kind of filter that directs attention to particular regions in space so that the ventral stream processes particular objects. In this study, participants looked at the center of the screen in both the auditory-only and audiovisual conditions. This is to ensure that differential neural activity between the auditory-only and audiovisual conditions was not caused by simple eye positions. Attenuation of neural activity in the pSTG may indicate that visual and auditory dorsal pathways interact with each other, and that the location of interest provided by vision can modify auditory localization processing.

## Conclusion

Using realistic (externalized) auditory spatial stimuli, our fMRI study was able to show that bilateral pSTG was more sensitive to sound sources in contra- than in ipsilateral hemifields. ROI analyses in the pSTG revealed monotonically increasing neural responses for more contralateral locations. The results support population rate coding, but not topographical place coding, in the human auditory cortex. The comparison between the auditory-only conditions and the audiovisual conditions successfully demonstrated the neural basis of the ventriloquism effect as arising from attenuation of the monotonically increasing functions of sound azimuth processing in the pSTG by the capture of spatially discordant visual stimuli.

## Funding

Funding to pay the Open Access publication charges for this article was provided by the National Institute of Information and Communications Technology.

## Notes

We express our gratitude to Drs Hiroaki Kato and Ryouichi Nishimura for helpful advice in early phases of this work. *Conflict of Interest:* None declared.

## References

- Alain C, Arnott SR, Hevenor S, Graham S, Grady CL. 2001. “What” and “where” in the human auditory system. *Proc Natl Acad Sci USA*. 98:12301–12306.
- Alais D, Burr D. 2004. The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol*. 14:257–262.
- Arnott SR, Alain C. 2011. The auditory dorsal pathway: orienting vision. *Neurosci Biobehav Rev*. 35:2162–2173.
- Blauert J. 1997. *Spatial hearing*, Revised Edition. Cambridge, MA: MIT Press.
- Bonath B, Noesselt T, Martinez A, Mishra J, Schwiecker K, Heinze HJ, Hillyard SA. 2007. Neural basis of the ventriloquist illusion. *Curr Biol*. 17:1697–1703.
- Brugge JF, Reale RA, Hind JE. 1996. The structure of spatial receptive fields of neurons in primary auditory cortex of the cat. *J Neurosci*. 16:4420–4437.
- Brunetti M, Belardinelli P, Caulo M, Del Gratta C, Della Penna S, Ferretti A, Lucci G, Moretti A, Pizzella V, Tartaro A, et al. 2005. Human brain activation during passive listening to sounds from different locations: an fMRI and MEG study. *Hum Brain Mapp*. 26:251–261.
- Burgess N, Maguire EA, O’Keefe J. 2002. The human hippocampus and spatial and episodic memory. *Neuron*. 35:625–641.
- Callan A, Callan DE, Ando H. 2013. Neural correlates of sound externalization. *Neuroimage*. 66C:22–27.
- Epstein RA. 2008. Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends Cogn Sci*. 12:388–396.
- Frissen I, Vroomen J, de Gelder B. 2012. The aftereffects of ventriloquism: the time course of the visual recalibration of auditory localization. *Seeing Perceiving*. 25:1–14.
- Frissen I, Vroomen J, de Gelder B, Bertelson P. 2005. The aftereffects of ventriloquism: generalization across sound-frequencies. *Acta Psychologica*. 118:93–100.
- Getzmann S, Lewald J. 2010. Effects of natural versus artificial spatial cues on electrophysiological correlates of auditory motion. *Hear Res*. 259:44–54.
- Griffiths TD, Warren JD. 2002. The planum temporale as a computational hub. *Trends Neurosci*. 25:348–353.
- Jack CE, Thurlow WR. 1973. Effects of degree of visual association and angle of displacement on the “ventriloquism” effect. *Percept Mot Skills*. 37:967–979.
- Jeffress LA. 1948. A place theory of sound localization. *J Comp Physiol Psychol*. 41:35–39.
- Knudsen EI, Konishi M. 1978. A neural map of auditory space in the owl. *Science*. 200:795–797.
- Konishi M. 2003. Coding of auditory space. *Annu Rev Neurosci*. 26:31–55.
- Krumbholz K, Nobis EA, Weatheritt RJ, Fink GR. 2009. Executive control of spatial attention shifts in the auditory compared to the visual modality. *Hum Brain Mapp*. 30:1457–1469.
- Krumbholz K, Schonwiesner M, Rubsamen R, Zilles K, Fink GR, von Cramon DY. 2005. Hierarchical processing of sound location and motion in the human brainstem and planum temporale. *Eur J Neurosci*. 21:230–238.
- Leung HC, Gore JC, Goldman-Rakic PS. 2002. Sustained mnemonic response in the human middle frontal gyrus during on-line storage of spatial memoranda. *J Cogn Neurosci*. 14:659–671.
- Linzenbold W, Himmelbach M. 2012. Signals from the deep: reach-related activity in the human superior colliculus. *J Neurosci*. 32:13881–13888.
- Magezi DA, Krumbholz K. 2010. Evidence for opponent-channel coding of interaural time differences in human auditory cortex. *J Neurophysiol*. 104:1997–2007.
- Middlebrooks JC, Knudsen EI. 1984. A neural code for auditory space in the cat’s superior colliculus. *J Neurosci*. 4:2621–2634.
- Middlebrooks JC, Xu L, Eddins AC, Green DM. 1998. Codes for sound-source location in nontopographic auditory cortex. *J Neurophysiol*. 80:863–881.
- Palmer AR, King AJ. 1982. The representation of auditory space in the mammalian superior colliculus. *Nature*. 299:248–249.

- Palomaki KJ, Tiitinen H, Makinen V, May PJ, Alku P. 2005. Spatial processing in human auditory cortex: the effects of 3D, ITD, and ILD stimulation techniques. *Brain Res Cogn Brain Res*. 24:364–379.
- Poldrack RA. 2000. Imaging brain plasticity: conceptual and methodological issues—a theoretical review. *Neuroimage*. 12:1–13.
- Rajan R, Aitkin LM, Irvine DRF, McKay J. 1990. Azimuthal sensitivity of neurons in primary auditory-cortex of cats. 1. Types of sensitivity and the effects of variations in stimulus parameters. *J Neurophysiol*. 64:872–887.
- Rauschecker JP. 1998. Cortical processing of complex sounds. *Curr Opin Neurobiol*. 8:516–521.
- Rauschecker JP, Tian B. 2000. Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proc Natl Acad Sci USA*. 97:11800–11806.
- Rauschecker JP, Tian B, Hauser M. 1995. Processing of complex sounds in the macaque nonprimary auditory cortex. *Science*. 268:111–114.
- Recanzone GH. 1998. Rapidly induced auditory plasticity: the ventriloquism aftereffect. *Proc Natl Acad Sci USA*. 95:869–875.
- Renier LA, Anurova I, De Volder AG, Carlson S, VanMeter J, Rauschecker JP. 2010. Preserved functional specialization for spatial processing in the middle occipital gyrus of the early blind. *Neuron*. 68:138–148.
- Rorden C, Karnath HO, Bonilha L. 2007. Improving lesion-symptom mapping. *J Cogn Neurosci*. 19:1081–1088.
- Salminen NH, May PJ, Alku P, Tiitinen H. 2009. A population rate code of auditory space in the human cortex. *PLoS ONE*. 4:e7600.
- Stecker GC, Harrington IA, Middlebrooks JC. 2005. Location coding by opponent neural populations in the auditory cortex. *PLoS Biol*. 3:520–528.
- Stein BE, Stanford TR. 2008. Multisensory integration: current issues from the perspective of the single neuron. *Nat Rev Neurosci*. 9:255–266.
- Takahashi T, Moiseff A, Konishi M. 1984. Time and intensity cues are processed independently in the auditory system of the owl. *J Neurosci*. 4:1781–1786.
- Thurlow WR, Jack CE. 1973. Certain determinants of the “ventriloquism effect”. *Percept Mot Skills*. 36:1171–1184.
- Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*. 15:273–289.
- Vander Wyk BC, Ramsay GJ, Hudac CM, Jones W, Lin D, Klin A, Lee SM, Pelphrey KA. 2010. Cortical integration of audio-visual speech and non-speech stimuli. *Brain Cogn*. 74:97–106.
- Warren DH, Welch RB, McCarthy TJ. 1981. The role of visual-auditory “compellingness” in the ventriloquism effect: implications for transitivity among the spatial senses. *Percept Psychophys*. 30:557–564.
- Werner-Reiss U, Groh JM. 2008. A rate code for sound azimuth in monkey auditory cortex: implications for human neuroimaging studies. *J Neurosci*. 28:3747–3758.
- Witten IB, Knudsen EI. 2005. Why seeing is believing: merging auditory and visual worlds. *Neuron*. 48:489–496.
- Woldorff MG, Tempelmann C, Fell J, Tegeler C, Gaschler-Markefski B, Hinrichs H, Heinz HJ, Scheich H. 1999. Lateralized auditory spatial perception and the contralaterality of cortical processing as studied with functional magnetic resonance imaging and magnetoencephalography. *Hum Brain Mapp*. 7:49–66.
- Woods TM, Recanzone GH. 2004. Visually induced plasticity of auditory spatial perception in macaques. *Curr Biol*. 14:1559–1564.
- Wozny DR, Shams L. 2011. Recalibration of auditory space following milliseconds of cross-modal discrepancy. *J Neurosci*. 31:4607–4612.
- Zatorre RJ, Belin P. 2001. Spectral and temporal processing in human auditory cortex. *Cereb Cortex*. 11:946–953.
- Zatorre RJ, Belin P, Penhune VB. 2002. Structure and function of auditory cortex: music and speech. *Trends Cogn Sci*. 6:37–46.
- Zimmer U, Lewald J, Erb M, Karnath HO. 2006. Processing of auditory spatial cues in human cortex: an fMRI study. *Neuropsychologia*. 44:454–461.
- Zimmer U, Macaluso E. 2005. High binaural coherence determines successful sound localization and increased activity in posterior auditory areas. *Neuron*. 47:893–905.