Reprinted from *American Laboratory* April 2002

Functional imaging of the human temporal cortex during auditory sentence processing

Kuniyoshi L. Sakai, Fumitaka Homae, Ruyichiro Hashimoto, and Kei Suzuki

Sensory information processing in the cerebral cortex has been characterized as involving multiple areas that are hierarchically organized and functionally specialized.[1] It is known that speech recognition critically depends on specialized cortical regions, including auditory areas as well as language areas.[2] While cytoarchitectonic studies have identified multiple auditory areas in humans,[3–6] the role of these multiple areas in auditory analysis per se has yet to be elucidated. Using recently developed imaging techniques, differential cortical responses to speech stimuli could reveal functional differentiation in the auditory cortex. One notable drawback is that most functional imaging studies have tested language tasks at the word level, such as lexical decision and word-generation tasks, thereby neglecting the syntactic aspects of the language faculty that may be primarily manifested at the sentence level.

This paper presents the results of recent functional imaging studies intended to identify cortical areas involved in auditory sentence processing. Functional magnetic resonance imaging (fMRI) and optical topography (OT) were used. OT is a newly developed technique that measures temporal changes in hemoglobin oxygenation simultaneously at multiple regions.[7–9] While near-infrared spectroscopy (NIRS) measures spectroscopic reflection and scattering at a single region with a light emitter and a detector,[10,11] OT can acquire a topographical image of cortical activity. The current direction of research in cognitive neuroscience has established the usefulness of these neuroimaging techniques for understanding essential aspects of human cognition, including speech recognition.

Experiment

Subjects

Seven native Japanese-speaking males (aged 20–32) participated in the study. They demonstrated right-handedness (laterality quotients 81–100) according to the Edinburgh inventory.[12] The subjects' consent was obtained according to the Declaration of Helsinki. Approval for the human experiments was obtained from the institutional review board of the University of Tokyo, Graduate School of Arts and Sciences (Tokyo, Japan).

Auditory stimuli

Speech sounds were sentences taken from fairy tales. These were digitized (16 bit, 11,025 Hz) using Oshaberi-mate speech-synthesis software (**Fujitsu,** Tokyo, Japan), which converts Japanese written text into sound waveforms. Sine wave tone and white noise used in the control task were synthesized by Sound Forge sound-editing software (**Sonic Foundry, Inc.,** Madison, WI). In fMRI experiments, auditory stimuli were presented to the subject's ears every 1 sec through separate tubes, and the scanning sounds were confined within the interstimulus interval of 300–600 msec. In OT experiments, stimuli were presented with a stereophonic headphone.

Task paradigm

The behavioral tasks used in the fMRI study were a control (CON) condition in which nonspeech sounds were presented to both ears, and two listening conditions for speech sounds: a diotic (binaural) listening (DIO) condition and a dichotic listening (DIC) condition. All stimuli consisted of targets to be responded to and nontargets to be neglected by the subjects. Under the DIO condition, identical stimuli were presented to both ears simultaneously. Target phrases were presented in the order of an original story with interventions of nontargets, which were made by randomizing the order of syllables of the corresponding target. These jumbled stimuli conformed to the rules of Japanese phonotactics but had no meaning. The subject was asked to respond to a target stimulus by pressing a bulb attached to a pneumatic switch. Under the DIC condition, either a pair of target and nontarget or that of two different nontargets was simultaneously presented to different ears. The subject had to choose one side with a target and pay attention to that target selectively. Therefore, in addition to the recognition of speech sounds required by the DIO condition, the DIC condition also demanded selective attention to target stimuli in either ear.

In the OT study, a dichotic listening paradigm was used for all tasks. Target stimuli and nontarget stimuli were simultaneously presented to different ears every 2 sec, and a target was alternatively presented to either the left or right ear at random intervals. Subjects were asked to track targets and to press a button when a target was shifted to the other side. In a control task, a tone and white noise were presented as targets and nontargets, respectively. In order to confirm that subjects performed tasks by recognizing targets, a tone of different pitch was presented as a nontarget at a lower a

b

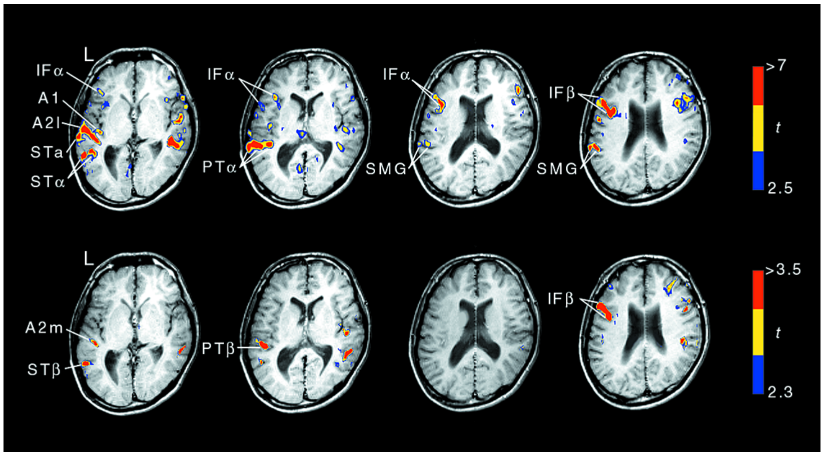

Figure 1 *Representative activity of multiple auditory areas in one subject. a) Series of* t*-maps of (DIO + DIC) vs CON on horizontal brain slices. b) Series of* t*-maps of DIC vs DIO. Color bars denote* t*-values for each comparison. The left side of the brain is shown on the left. The centers of the slices are z = 4, 12, 20, and 28 from the left panel to the right. Multiple activated regions were found in the auditory and language-related areas, mostly in the left hemisphere of this subject.*

rate. These probe stimuli prevented subjects from performing the tasks by tracking nontargets only. Two speech-recognition tasks were used: 1) a repeat task, in which the targets consisted of one repeated sentence; and 2) a story task, in which the targets were successive different sentences of a continuous story. A sentence different from the target for the repeat task and contextually anomalous phrases for the story task were used as probe stimuli. Therefore, these tasks could not be completed appropriately by identifying speech sounds without paying attention to their meanings.

fMRI data acquisition and analysis

fMRI data were acquired using an MRH-1500 1.5 Tesla MRI system (**Hitachi Medical Corp.,** Tokyo, Japan). Horizontal slices were scanned over with a gradient echo echo-planar imaging (EPI) sequence (repetition time 2 sec, field of view 384×384 mm^2, resolution $3 \times 3 \times 8$ mm^3). Analyses of fMRI time-series data were first done on a single-subject basis using in-house software.[13] Time-series data of each voxel were converted to percent signal changes from the initial CON block, and were then averaged for multiple sessions after correction for head movements between scans. The activation under the DIO and DIC conditions was estimated using the following two types of t-tests for each voxel: (DIO + DIC) vs CON ($t > 2.6$), combining DIO and DIC before comparison with CON, and DIC vs DIO ($t > 2.3$) as direct comparison between two experimental conditions. The significance level for each activated region after Bonferroni correction was $P < 0.02$ and $P < 0.04$, respectively.

OT data acquisition and analysis

OT data were acquired using two OT systems with the same calibration (ETG-100 and ETG-A1, **Hitachi Medical Corp.**), one for each hemisphere. Near-infrared laser diodes with two wavelengths (782–793 nm and 823–832 nm) were used as the light sources (maximum intensity 2 mW/mm^2, intensity modulation 1–10 kHz). The reflected lights were detected with avalanche photodiodes located 30 mm from an incident position. Using lock-in amplifiers, the detected signal was separated into individual light sources with each wavelength.[7] Twenty-two points in each hemisphere were simultaneously measured at minimum spatial intervals of 21 mm, and each point was sampled every 500 msec. The measured region in each hemisphere centered on the Sylvian fissure and covered an area of 6×12 cm^2. A correlation coefficient (r) of hemoglobin time points with a boxcar waveform was calculated for each measurement point, and an r-map was created from the r-values. As a threshold for the r-values, 0.73 or –0.73 was chosen for statistical significance ($P = 0.05$, corrected).

Results and discussion

Functional differentiation within the auditory cortex and language areas

Multiple regions were identified that were activated under the DIO and DIC conditions (*Figure 1*).[14] In the t-map of (DIO + DIC) vs CON, activation was observed in multiple auditory and language areas: the primary auditory area (A1), secondary auditory area (A2, both lateral and medial portions), planum temporale (PT), superior temporal gyrus (ST), supramarginal gyrus (SMG), and inferior frontal gyrus (IF). In the t-map of DIC vs DIO, on the other hand, regions in A2 (medial portion), PT, ST, and IF were identified that responded more prominently to the DIC condition (Figure 1*b*). In A2, the medial portion, called A2m (medial A2), clearly showed DIC selectivity, though this selectivity was not apparent in the lateral portion, A2l (lateral A2). Anterior ST (STa) was also identified as an activated region just posterior to the Heschl's gyrus and on the lateral surface of the anterior superior temporal gyrus.

Further, it was confirmed that the spatial relationship between A2m, A21, and STa was consistent among all subjects. These results clearly demonstrate that the DIO and DIC conditions modulate cortical responses differentially among multiple auditory areas. Some areas showed greater responses to the DIC condition than to the DIO condition, while other areas did not show such a difference. These contrasting response patterns suggest functional differentiation among multiple auditory areas. It is notable that both cortical language areas, Wernicke's area (PT and ST) and Broca's area (IF), are parceled into two types of regions with different response patterns. All the regions identified in the left

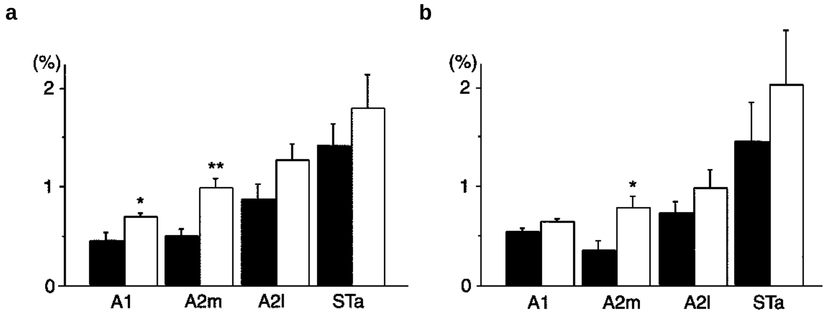

Figure 2 *Differential responses under the DIO and DIC conditions in multiple auditory areas. Histogram comparing the percent signal changes (mean ± SEM (standard error of the mean) of subjects) for each auditory condition vs CON is shown in each region. Filled and open bars denote signal changes under DIO and DIC, respectively. a) Regions in the left hemisphere. b) Regions in the right hemisphere. Signal changes are shown for A1, A2m, A21, and STa, which have close anatomical relationships. They show progressive signal increases under the DIO and DIC conditions. Single (*) and double (**) asterisks denote statistical significance at* $P < 0.05$ *and* $P < 0.005$*, respectively.*

hemisphere were also found in the right hemisphere at similar coordinates of the brain.

It was found that A1, A2m, A2l, and STa showed a progressive increase in responses to both the DIO and DIC conditions (*Figure 2*), suggesting that these regions constitute a functional pathway on the order of A1, A2m, A2l, and STa. In other words, these regions have progressively greater selectivity for the speech stimuli presented. If auditory areas at the early stage process both the story phrases and the nonsense words in the course of figure-background segregation, the DIC condition would elicit larger responses than the DIO condition. In contrast, at the later stage, after filtering out the nonsense words as unattended stimuli, responses to the DIO and DIC conditions would become comparable.

This “filter” process agrees with the decrease in DIC selectivity along the pathway of A2m, A2l, and STa. The auditory pathway proposed here corresponds to previous anatomical findings. Based on the gradient in acetylcholinesterase staining and cytochrome oxidase activity, a hierarchical order of A1, PA (posterior area), LA (lateral area), and STA (superior temporal area) has been proposed.[6] Specifically, the spatial configuration of A1, PA, LA, and STA corresponds to that of A1, A2m, A2l, and STa. It is notable that the anatomical study and the functional study match with respect to the anatomical location of these areas and the order of the proposed hierarchical levels. These findings further suggest that A2 and cortical language areas, such as Wernicke’s and Broca’s areas, play a pivotal role in combining differential mechanisms of speech recognition and in processing language beyond the primary auditory area.

Temporal cortex activation during speech recognition

To address the issue of sentence processing in the cerebral cortex further, the OT technique was adopted for speech recognition experiments.[15] There are several advantages to using OT over other functional mapping techniques. First, it is possible to independently measure the temporal changes in oxyhemoglobin concentration (C_{oxy}) and deoxyhemoglobin concentration (C_{deoxy}). Second, there is no scanning noise to interfere with the experimental auditory stimuli. Third, its signal-to-noise ratio is as high as fMRI, allowing for the observation of cortical activity with a small number of trials. One major disadvantage of OT is that its measurement is restricted to the cortical surface. Nevertheless, OT has the potential to introduce a new dimension to the mapping of human cognitive functions.

Left-dominant activation in the superior temporal cortex (the superior and middle temporal gyri) was observed preferentially during the story task over the repeat task when compared with the control task (*Figure 3*). Wider regions were clearly more activated during the story task than during the repeat task in both hemispheres. This finding is consistent with the results of a previous positron emission tomography (PET) study that showed activation of the left superior temporal cortex when subjects listened to continuous speech in their native language.[16] The direct comparison between the hemodynamics in the story task and that in the repeat task showed focal activation in the left superior temporal cortex (*Figure 4*). Both an increase in C_{oxy} and a decrease in C_{deoxy} were significantly larger in the middle temporal gyrus than in the superior temporal gyrus. The temporal changes in the left superior temporal cortex are shown in *Figure 5*. With a delay of 6 sec, an increase in C_{oxy} was synchronized with each onset of the story task and, after reaching a plateau, C_{oxy} returned to the baseline level at the end of the task. Although a decrease in C_{deoxy} was also synchronized with each period of the story task, C_{deoxy} did not exactly mirror the temporal dynamics of C_{oxy}. Similar, but smaller, changes in both C_{oxy} and C_{deoxy} were observed in the repeat run. These results suggest that the hemodynamics in the midlateral part of the left temporal cortex reflect cognitive factors involved in the processing of sentences. A critical difference between the story and repeat tasks would be the load of speech stimuli to be processed. Recognition of successive different sentences of a story demands more auditory, memory, and language information processing than the recognition of repeated sentences. The temporal cortex activation reported here is also consistent with the role of the primate temporal association area in memory storage and memory retrieval.[17,18]

Based on converging evidence from aphasic and imaging studies, a modular specialization of the cortical language areas that are critically involved in sentence processing has been proposed.[19] This rather a

b

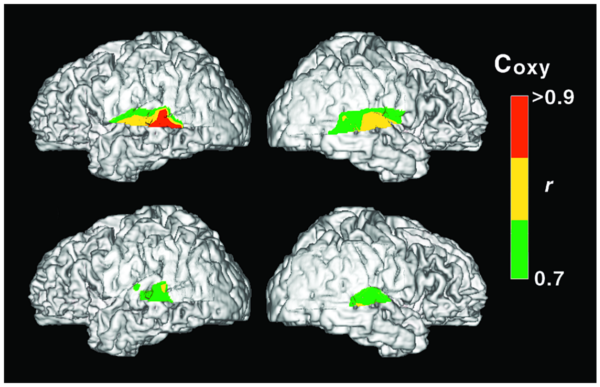

Figure 3 *Superior temporal cortex activation during speech recognition. a) Story task vs control. The left brain is shown on the left. b) Repeat task vs control. OT images in* r*-maps averaged among all subjects were superimposed on lateral images of a brain. The color bar at right denotes correlation coefficients (*r*) of C_{oxy} time points with a boxcar waveform (each period 36 sec, delay 6 sec). Note the prominent activation in the midlateral part of the bilateral temporal cortex.*

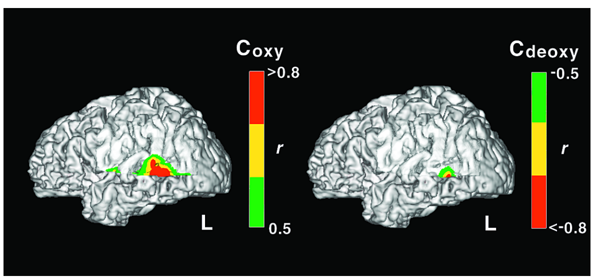

Figure 4 *Focal activation in the left superior temporal cortex during sentence processing. OT images in* r*-maps for a direct comparison between story and repeat tasks. Left:* r*-maps of C_{oxy} temporal changes. Right:* r*-maps of C_{deoxy} temporal changes. Activation is localized in the middle temporal gyrus in both comparisons.*

broad delineation at the current stage of imaging studies reflects the difficulty of dissociating linguistic components from confounding cognitive factors. The language system does not stand alone but interacts with other systems of perception, memory, and consciousness, as well as with the speech output system. Future work combining electrophysiological event-related brain potential and magnetoencephalography, hemodynamic (PET, fMRI, OT), and transcranial magnetic stimulation (TMS) methods will allow the further parceling of language processing in the cerebral cortex and will enable us to address questions about module-specific brain areas and how they actually perform linguistic computations of sentences.

References

1. Felleman DJ, Van Essen DC. Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1991; 1(1):1–47.
2. Geschwind N. Specializations of the human brain. Sci Amer 1979; 241(3):158–68.

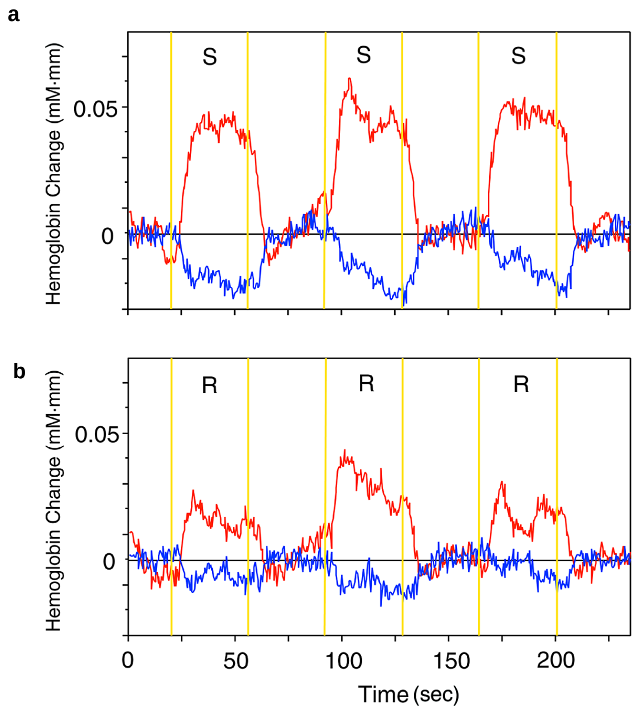

Figure 5 *Hemodynamics in speech-recognition tasks relative to the control task. a) Story and control tasks. b) Repeat and control tasks. There were three periods of either the story task (S) or the repeat task (R) in each run. Red lines show the mostly positive temporal changes of C_{oxy}, whereas the blue lines reveal the mostly negative temporal changes of C_{deoxy}. These temporal changes were calculated from averaged data among subjects.*

3. Brodmann K. Comparative localization of the cerebral cortex in its principles represented on the basis of the cell structure. Leipzeig, Germany: Verlag von Johann ambrosius Barth, 1909.
4. Von Economo C, Horn L. On the convolution, mass, and coritical architecure of the superior temporal area: its individual and side difference. Z Neurol Psychiat 1930; 130:678–757.
5. Galaburda AM, Sanides F. Cytoarchitectonic organization of the human auditory cortex. J Comp Neurol 1980; 190:597–610.
6. Rivier F, Clarke S. Cytochrome oxidase, acetylcholinesterase, and NADPH-diaphorase staining in human supratemporal and insular cortex: evidence for multiple auditory areas. Neuroimage 1997; 6(4):288–304.
7. Maki A, Yamashita Y, Ito Y, Wantanabe E, Mayanayi Y, Koizumi H. Spatial and temporal analysis of human motor activity using noninvasive NIR topography. Med Phys 1995; 22(12):1997–2005.
8. Yamashita Y, Maki A, Koizumi H. Near-infrared topographic measurement system: imaging of absorbers localized in a scattering medium. Rev Sci Instrum 1996; 67(3):730–2.
9. Koizumi H, Yamashita Y, Maki A, et al. Higher-order brain function analysis by trans-cranial dynamic NIRS imaging. J Biomed Opt 1999; 4(4):403–13.
10. Kato T, Kamei A, Takashima S, Ozaki T. Human visual

cortical function during photic stimulation monitoring by means of near-infrared spectroscopy. J Cereb Blood Flow Metab 1993; 13:516–20.

11. Chance B, Zhuang Z, Unah C, Alter C, Lipton L. Cognition-activated low-frequency modulation of light absorption in human brain. Proc Natl Acad Sci USA 1993; 90:3770–4.
12. Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 1971; 9:97–113.
13. Sakai K, Watanabe E, Onodera Y, et al. Functional mapping of the human somatosensory cortex with echo-planar magnetic resonance imaging. Magn Reson Med 1995; 33:736–43.
14. Hashimoto R, Homae F, Nakajima K, Miyashita Y, Sakai KL. Functional differentiation in the human auditory and language areas revealed by a dichotic listening task. Neuroimage 2000; 12(2):147–58.
15. Sato H, Takeuchi T, Sakai KL. Temporal cortex activation during speech recognition: an optical topography study. Cognition 1999; 73(3):B55–66.
16. Mazoyer BM, Tzourio N, Frak V, et al. The cortical representation of speech. J Cognit Neurosci 1993; 5(4):467–79.
17. Sakai K, Miyashita Y. Memory and imagery in the temporal lobe. Curr Opin Neurobiol 1993; 3(2):166–70.
18. Sakai K, Naya Y, Miyashita Y. Neuronal tuning and associative mechanisms in form representation. Learn Mem 1994; 1(2):83–105.
19. Sakai KL, Hashimoto R, Homae F. Sentence processing in the cerebral cortex. Neurosci Res 2001; 39(1):1–10.

The authors are with the Department of Cognitive and Behavioral Science, Graduate School of Arts and Sciences, University of Tokyo, Komaba, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan; tel./fax: (81) 3-5454-6261; e-mail: sakai@mind.c.u-tokyo.ac.jp. Dr. Sakai is also with ***CREST, Japan Science and Technology Corp.,*** *Kawaguchi-shi, Japan.* *The authors would like to thank Dr. Yasushi Miyashita, and Dr. Kyoichi Nakajima, of the Department of Physiology, University of tokyo School of Medicine; Mr. Hiroki Sato and Mr. Tatsuya Takeuchi, in the authors' team, for collaborating on the fMRI and OT experiments; and Ms. Hiromi Matsuda for her administrative assistance. This work was supported by a* ***CREST*** *grant from* ***Japan Science and Technology Corp.*** *and a grant from Toyota Physical and Chemical Research Institute Aichi, Japan, to K.L. Sakai.*