RAPID REPORT

NEURAL CORRELATES OF AUDITORY FEEDBACK CONTROL IN HUMAN

A. TOYOMURA,[a,b]* S. KOYAMA,[b] T. MIYAMAOTO,[c]
A. TERAO,[d] T. OMORI,[e] H. MUROHASHI[f] AND
S. KURIKI[b]

[a]*Research Institute of Science and Technology for Society, Japan Science and Technology Agency, Japan*

[b]*Research Institute for Electronic Science, Hokkaido University, Kita 12 Nishi 6, Kita, Sapporo, 060-0812, Japan*

[c]*Brain Function Research Laboratory, Hokkaido University Graduate School of Medicine, Japan*

[d]*Graduate School of Letters, Hokkaido University, Japan*

[e]*Tamagawa University Research Institute, Tamagawa University, Japan*

[f]*Graduate School of Education, Hokkaido University, Japan*

Abstract—Auditory feedback plays an important role in natural speech production. We conducted a functional magnetic resonance imaging (fMRI) experiment using a transformed auditory feedback (TAF) method to delineate the neural mechanism for auditory feedback control of pitch. Twelve right-handed subjects were required to vocalize /a/ for 5 s, while hearing their own voice through headphones. In the TAF condition, the pitch of the feedback voice was randomly shifted either up or down from the original pitch two or three times in each trial. The subjects were required to hold the pitch of the feedback voice constant by changing the pitch of original voice. In non-TAF condition, the pitch of the feedback voice was not modulated and the subjects just vocalized /a/ continuously. The contrast between TAF and non-TAF conditions revealed significant activations; the supramarginal gyrus, the prefrontal area, the anterior insula, the superior temporal area and the intraparietal sulcus in the right hemisphere, but only the premotor area in the left hemisphere. This result suggests that auditory feedback control of pitch is mainly supported by the right hemispheric network. © 2007 IBRO. Published by Elsevier Ltd. All rights reserved.

Key words: auditory feedback control, pitch control, vocalization, transformed auditory feedback method, functional magnetic resonance imaging.

Perception of one's own speech plays an important role in fluent speech production. Pitch and sound pressure are either involuntarily or voluntarily controlled through auditory feedback while talking. Auditory feedback is also in-dispensable for language acquisition of children or second language learning of adults. In order to quantitatively evaluate the role of the auditory feedback while talking, a transformed auditory feedback (TAF) method has been developed (Kawahara et al., 1996; Burnett et al., 1998; Houde and Jordan, 1998; Toyomura and Omori, 2005). With this method, subjects are required to vocalize while hearing an artificially TAF voice through headphones. Compensations within 500 ms to hold pitch level against fluctuation were clarified for the first time using this method (Kawahara et al., 1996; Burnett et al., 1998). However, a detailed elucidation of the neural basis for auditory feedback control of pitch has not been reported.

Non-invasive brain imaging studies have recently revealed vocalization-related brain regions, for example the superior temporal area, the insula, and the cerebellum, by using functional magnetic resonance imaging (fMRI) (Riecker et al., 2000; Fu et al., 2006), magnetoencephalography (MEG) (Kuriki et al., 1999; Houde et al., 2002; Gunji et al., 2003) and positron emission tomography (PET) (Wise et al., 1999; Schulz et al., 2005). With regard to auditory feedback, some studies have investigated brain activation during delayed auditory feedback (DAF) tasks (Hirano et al., 1997; Hashimoto and Sakai, 2003) or a TAF task when subjects were singing (Zarate and Zatorre, 2005). For example Hashimoto and Sakai (2003) found that the activation in the superior temporal area was correlated with the degree of DAF effects.

In the present study, we conducted an fMRI experiment to delineate the neural mechanism for auditory feedback control of pitch when vocalizing a vowel, /a/. Subjects were asked to perform a TAF task in an MRI scanning room; the pitch of feedback voice was altered during continuous vocalizing of an /a/ sound while subjects were required to hold the pitch of the feedback voice constant (TAF condition). The subjects had to change the pitch of their own voice to keep the altered pitch at the original level when the pitch was suddenly modulated. In the non-TAF condition, the pitch of the feedback voice was not modulated and the subjects just vocalized /a/ continuously. In both TAF and non-TAF conditions, the overt oral articulation mechanism was expected to be equivalent. Under the TAF condition, the pitch control based upon auditory feedback was expected to work selectively and more strongly than under the non-TAF condition. Thus, a comparison between these conditions could be expected to extract the brain regions involved in the auditory feedback control of pitch.
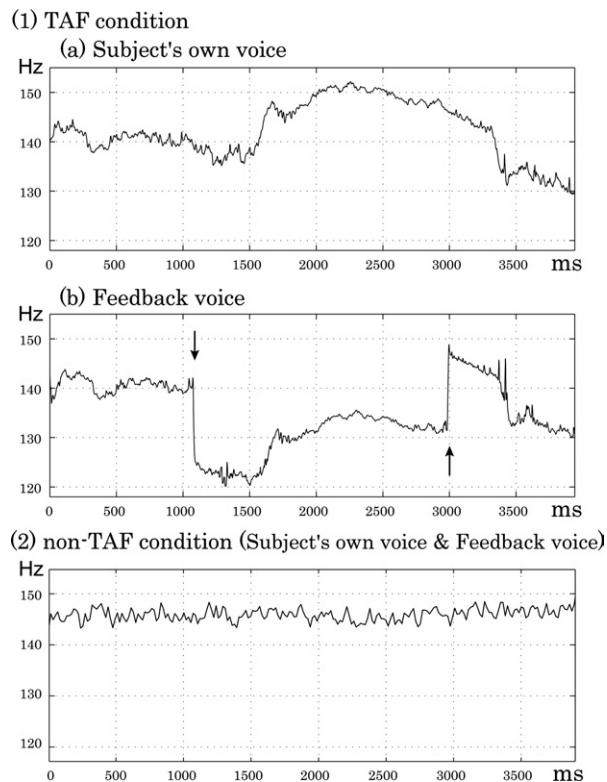
*Correspondence to: A. Toyomura, Research Institute for Electronic Science, Hokkaido University, Kita 12 Nishi 6, Kita, Sapporo, 060-0812, Japan. Tel: +81-11-706-3360; fax: +81-11-706-3663.
E-mail address: toyomu@es.hokudai.ac.jp (A. Toyomura).*

(1) TAF condition

(a) Subject's own voice

(b) Feedback voice

(2) non-TAF condition (Subject's own voice & Feedback voice)

Fig. 1. Examples of pitch trajectory (F0) when a subject performed under (1) TAF and (2) non-TAF conditions. Pitch trajectory of (a) subject's own voice, (b) feedback voice ((1) TAF condition). Arrows indicate feedback perturbations. In the non-TAF condition, the pitch trajectory of a subject's own voice and the feedback voice are the same.

EXPERIMENTAL PROCEDURES

Subjects

Twelve right-handed subjects (2 female; age range: 19–37; mean: 27.1) participated. None had a history of speech, voice disorders or had trained as a singer. Written informed consent was obtained from all subjects and the study was approved by the Committee of Medical Ethics, Graduate School of Medicine, Hokkaido University.

Experimental design

We used a sparse image acquisition protocol (Hall et al., 1999) to avoid the scanning noise of the MRI system (a Signa Echo-Speed 1.5 T system (GE, Schenectady, NY, USA)) during vocalizing and imaging artifacts due to vocalizing. During a scanning mechanical noise-free period (6 s), the subjects wearing an MRI-compatible headphones and microphone set (Hitachi Advanced Systems Corporation, Tokyo, Japan) were required to vocalize /a/ for 5 s according to a visual instruction presented on a screen between each of two consecutive scanning periods. There were two conditions. In the TAF condition, the pitch of the feedback voice was randomly shifted up or down 2 semitones or less from the original pitch during the period. The subjects were required to hold the pitch of the feedback voice constant. For example, when the pitch of the feedback voice shifted upward, the subjects had to shift their pitch down accordingly. In the non-TAF condition, the feedback voice was fed back to subjects without any pitch-shift. The subjects were merely asked to vocalize /a/ during the 5-s period. Fig. 1 shows examples of the pitch trajectories of subjects' voices and

feedback voices in the TAF and non-TAF conditions. The visual instruction was presented at the center of the screen against a white background. Each subject performed the TAF condition when a two-headed black arrow was presented, while they performed the non-TAF condition when a black cross was presented. The auditory feedback pitch was modulated by using an effector (RFX-2000, Zoom Corporation, Tokyo, Japan). To avoid any influence from scanning noises during the scanning period, a low pass filter was applied. To reduce any effects of bone-conduction, pink noise at 70 dB SPL was generated by an analyzer (PAA-2, Phonic Corporation, Tampa, FL, USA) and mixed into the feedback voice using a sound mixer (MG10/2, Yamaha Corporation, Hamamatsu, Japan). Prior to the scanning, the subjects practiced the tasks until they became proficient. There were three runs for each subject; one run consisted of eight blocks of five trials. The first five trials were always TAF trials and the next five trials were non-TAF trials. Overall, the subject vocalized 120 times (60 times for each condition) in the experiment. The pitch shift occurred two or three times in a single trial.

Image acquisition and processing

Nine parallel axial slices (thickness 5 mm, gap 2.5 mm) were acquired using echo planar imaging with a sparse image acquisition protocol (64×64 matrix, field of view 24×24 cm, TR 9s, TE 40 ms, TA 2.67s, flip angle 90°). T1-weighted images (corresponding to nine axial slices, thickness 5 mm, gap 2.5 mm, 256×256 matrix, field of view 24×24 cm, TR 500 ms, TE 14 ms) served as an anatomical reference for the functional images. Data were analyzed using SPM99 (Wellcome Department of Cognitive Neurology, London, UK). The data were realigned, coregistered with structural data, spatially normalized to the Montreal Neurological Institute standard stereotactic space, and smoothed with an isotropic gaussian kernel of 8 mm full width at half maximum. Task specific effects were estimated with a general linear model. For random effects analyses, a contrast image between tasks was generated for each subject and used for inter-subject comparisons. The statistical threshold was set at $P<0.00001$ for the voxel level, uncorrected for multiple comparisons.

RESULTS

Contrasting the TAF and non-TAF conditions, larger areas were activated in the right than the left hemisphere. Fig. 2 and Table 1 show activation maps and the coordinates of activations. In the right hemisphere, significant activations were observed in the supramarginal gyrus, the prefrontal area (Brodmann's area, BA9), the anterior insula, the superior temporal area, and the intraparietal sulcus ($P<0.00001$, uncorrected). In the left hemisphere, only the premotor area showed significant activation. Even at a higher threshold of $P<0.05$, corrected at the voxel level, significant activations were still observed in the same regions, except for in the right intraparietal sulcus.

DISCUSSION

In this study we compared brain activities during vocalization with (TAF condition) and without (non-TAF conditions) pitch control based upon auditory feedback. The activation area in common between our experiment and fMRI experiments under DAF (Hirano et al., 1997; Hashimoto and Sakai, 2003) and distorted speech feedback (McGuire et al., 1996; Fu et al., 2006) is the right superior temporal area. This area is also a part of voice sensitive area (Belin
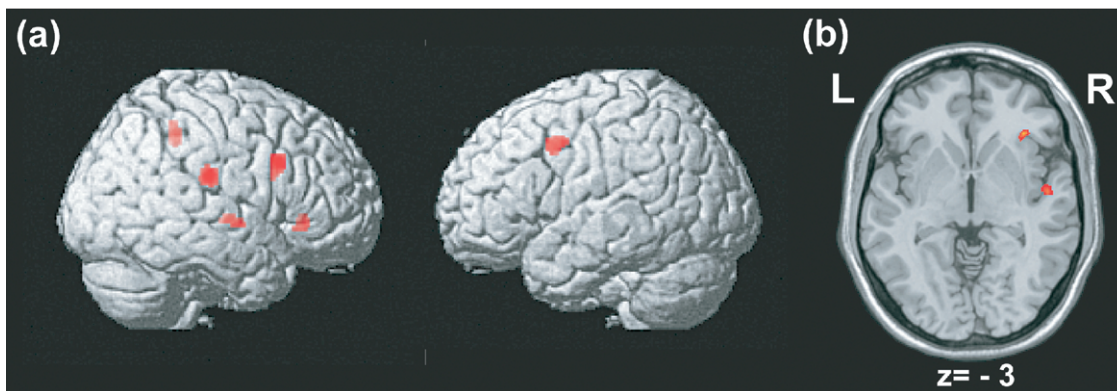
Fig. 2. Activations under the TAF condition compared with the non-TAF condition at a threshold of $P<0.00001$, $k>30$, uncorrected at the voxel level for multiple comparisons. (a) Right and left lateral view. (b) Axial view ($z=-3$) shows activations at the right anterior insula and the right superior temporal area. L, left; R, right. Note that all activated areas except the right intraparietal sulcus were activated when a threshold of $P<0.05$ (corrected), was applied.

et al., 2000; Kriegstein and Giraud, 2004; Belin 2006 for a review).

The differences between TAF and non-TAF conditions were mainly observed in the right hemisphere. This was probably due to the experimental manipulation being in pitch. Although, we are not able to specify to what extent these differences are attributable to acoustical ones (input) between the conditions, we speculated that the activation in the right intraparietal sulcus depends on the acoustical differences between the conditions. This area was activated neither during the DAF task (Hirano et al., 1997; Hashimoto and Sakai, 2003) nor under distorted speech feedback (McGuire et al., 1996; Fu et al., 2006). The region is known as one for attention processing (Bremmer et al., 2001). Levanen et al. (1996) showed that the right parietal region is related to auditory changes detected using MEG. They suggested that a stronger involvement of the right than the left hemisphere in pitch change detection measured by mismatch field (Levanen et al., 1996). If these facts are considered, the right intraparietal sulcus was activated in the present study because occasional pitch changes recruited a bottom-up attentional system whereas DAF and distorted speech were applied without a break.

McGuire et al. (1996) introduced continuous distortion of a subject's speech feedback (by pitch shift) while the subjects read aloud. The effect of distortion was observed

Table 1. Coordinates of activations for contrast[a]

Region	Side	x	y	z	Z score
Supramarginal gyrus	R	62	−22	22	5.68
Prefrontal area	R	56	14	32	5.34
Anterior insula	R	36	30	−2	5.29
Premotor area	L	−52	8	38	4.92
Superior temporal area	R	52	−10	−2	4.89
Intraparictal sulcus	R	32	−42	44	4.73

[a] Stereotactic coordinates (x, y, z) in the Montreal Neurological Institute space are shown for each voxel with a local maximum of Z values in the contrasts indicated ($P<0.00001$, $k>30$, uncorrected for the voxel level).

in the lateral superior temporal area with a greater activation in the right than the left hemisphere. A similar pattern of activation was seen when subjects read aloud, but the words they heard were spoken by someone else (McGuire et al., 1996, PET study). Recently, Fu et al. (2006) using fMRI reported that the right superior area ($\{x, y, z\}=\{53, -13, 9\}$) was more activated for undistorted self voice than distorted self voice. Further the right superior area ($\{61, -26, -2\},\{53, 4, 4\},\{61, -33, 4\}$) was more activated when subjects were able to recognize their distorted voice as their own than when they were not (Fu et al., 2006). Bilateral superior temporal areas without significant hemispheric difference were also activated during a DAF condition (Hirano et al., 1997; Hashimoto and Sakai, 2003). Considering these findings, the activation in the right superior temporal area in the present study is closely related to pitch change perception of one's own voice.

Hashimoto and Sakai (2003) also reported the activation of bilateral supramarginal gyrus with a DAF effect. Guenther (2006) discussed the role of supramarginal gyrus in somatosensory feedback control, using a speech production model (DIVA model) which is similar to a feedforward and feedback control model in the motor system (Kawato, 1999). However, this area is not a voice sensitive one (Belin et al., 2000; Kriegstein and Giraud, 2004; Belin 2006 for a review). The right supramarginal gyrus thus appears to be involved in feedback control of pitch rather than in voice perception itself.

A significant activation in the right prefrontal area (BA9) was observed in the present study. It has been reported that this area is activated during neither a DAF task (Hirano et al., 1997; Hashimoto and Sakai, 2003) nor under distorted speech feedback (McGuire et al., 1996; Fu et al., 2006). However, this area is activated during pitch discrimination for speech sounds (Zatorre et al., 1992). Zatorre et al. (1992) presented subjects with a pair of consonant–vowel–consonant successively in each trial. The second syllable had a higher fundamental frequency in the half of the pair and a lower frequency in the other. When the subjects were asked to judge whether the pitch of the

second stimulus was higher than the first, significant stronger activations in the right prefrontal cortex (BA45/46, BA9) were observed compared with when the subjects were asked to just alternate key presses. They suggested that the right prefrontal cortex is related to auditory working memory (Colombo et al., 1990; Zatorre and Samson, 1991; Romanski and Goldman-Rakic, 2002). Since we instructed our subjects to hold the pitch of the feedback voice constant in the TAF condition, the subjects had to retain the original pitch level. Although BA9 in the right hemisphere was activated in both studies, the BA45/46 in the right prefrontal cortex showed significant activations only in the study of Zatorre et al. (1992) but not in ours. Zatorre et al. (1992) presented the stimuli discretely (mean inter-stimulus interval 300 ms), thus the duration of the retention was longer in their study than ours. Hence activation of BA45/46 might be recruited only when a longer retention time is required.

The right anterior insula was activated in the present study. In the previous study, the left insula is related to coordination of speech articulation (Dronkers, 1996, lesion-based study; Kuriki et al., 1999, MEG study; Wise et al., 1999, PET study; Gunji et al., 2003, MEG study), while the right anterior insula is recruited when singing (Riecker et al., 2000; Zarate and Zatorre, 2005). Zarate and Zatorre (2005) showed that the right anterior insula, the anterior cingulate cortex, the superior temporal sulcus, the inferior frontal lobe, the supplementary motor area and the putamen exhibited greater activation when subjects compensated for the pitch shift while they were singing. Among these activation areas, the right insula, the putamen and the superior temporal sulcus activated only when subjects were singers. Although the subjects in our experiment were non-singers, the right insula exhibited significant activation. Other areas were not activated in the present study. There are several possible reasons for this discrepancy. First, pitch control during singing and that during simple vocalizing could be different. Second, the threshold of statistical significance might be different between the two studies. Third, the number of trials might cause the difference in adaptation and signal-to-noise ratio. The subjects in the present study vocalized as many as 120 times. Since Zarate and Zatorre (2005) did not report the statistical significance level and the number of trials or a detailed experimental method, we are not able to specify the reason(s) for the discrepancy here.

The left premotor area was also activated in the present study. The left premotor area has been shown to be related to speech perception and production (Wilson et al., 2004). However, it should be noted that the coordinates for speech production and perception were located more posterior than that for the present study (see their supplementary Table 1). On the other hand, Yokoyama et al. (2006) reported robust and extended activation of the left premotor activation while Japanese subjects read visually presented sentences. They were late bilinguals for English and both Japanese and English sentences were shown to activate this area more strongly than three nouns presented in a row (control stimuli). Their coordinate for the premotor area ({−46, 8, 30}) is very similar to that in the present study. This activation might be thus related to mental manipulation such as pitch compensation rather than speech perception and production.

CONCLUSION

In summary, the supramarginal gyrus, the prefrontal area, the anterior insula, the superior temporal area and the intraparietal sulcus in the right hemisphere and the premotor area in the left hemisphere were shown to be involved in auditory feedback control of pitch. The right hemisphere dominance is in line with previous neuroimaging research on voice perception (Kriegstein and Giraud, 2004; Belin, 2006).

REFERENCES

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. Nature 403:309–312.

Belin P (2006) Voice processing in human and non-human primates. Philos Trans R Soc Lond B Biol Sci 361:2091–2107.

Bremmer F, Schlack A, Duhamel JR, Graf W, Fink GR (2001) Space coding in primate posterior parietal cortex. Neuroimage 14: S46–S51.

Burnett TA, Freedland MB, Larson CR, Hain TC (1998) Voice F0 responses to manipulations in pitch feedback. J Acoust Soc Am 103:3153–3161.

Colombo M, D'Amato MR, Rodman HR, Gross CG (1990) Auditory association cortex lesions impair auditory short-term memory in monkeys. Science 247:336–338.

Dronkers NF (1996) A new brain region for coordinating speech articulation. Nature 384:159–161.

Fu CHY, Vythelingum GN, Brammer MJ, Williams SCR, Amaro EJ, Andrew CM, Yaguez L, van Haren NEM, Matsumoto K, McGuire PK (2006) An fMRI study of verbal self-monitoring: Neural correlates of auditory verbal feedback. Cereb Cortex 16:969–977.

Guenther FH (2006) Cortical interactions underlying the production of speech sounds. J Commun Disord 39:350–365.

Gunji A, Kakigi R, Hoshiyama M (2003) Cortical activities relating to modulation of sound frequency: How to vocalize? Brain Res Cogn Brain Res 17:495–506.

Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW (1999) "Sparse" temporal sampling in auditory fMRI. Hum Brain Mapp 7:213–223.

Hashimoto Y, Sakai KL (2003) Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: an fMRI study. Hum Brain Mapp 20:22–28.

Hirano S, Kojima H, Naito Y, Honjo I, Kamoto Y, Okazawa H, Ishizu K, Yonekura Y, Nagahama Y, Fukuyama H, Konishi J (1997) Cortical processing mechanism for vocalization with auditory verbal feedback. Neuroreport 8:2379–2382.

Houde JF, Jordan MI (1998) Sensorimotor adaptation in speech production. Science 279:1213–1216.

Houde JF, Nagarajan SS, Sekihara K, Merzenich MM (2002) Modulation of auditory cortex during speech: An MEG study. J Cogn Neurosci 14:1125–1138.

Kawahara H, Kato H, Williams JC (1996) Effects of auditory feedback on F0 trajectory generation. Proc 4th Int Conf Spoken Language Processing 287–290. http://www.icsa-speech.org/archive.

Kawato M (1999) Internal models for motor control and trajectory planning. Curr Opin Neurobiol 9:718–727.

Kriegstein KV, Giraud A-L (2004) Distinct functional substrates along the right superior temporal sulcus for the processing of voices. Neuroimage 22:948–955.

Kuriki S, Mori T, Hirata Y (1999) Motor planning center for speech articulation in the normal human brain. Neuroreport 10:765–769.

Levanen S, Ahonen A, Hari R, McEvoy L, Sams M (1996) Deviant auditory stimuli activate human. left and right auditory cortex differently. Cereb Cortex 6:288–296.

McGuire PK, Silbersweig DA, Frith CD (1996) Functional neuroanatomy of verbal self-monitoring. Brain 119:907–919.

Riecker A, Ackermann H, Wildgruber D, Dogil G, Grodd W (2000) Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. Neuroreport 26: 1997–2000.

Romanski LM, Goldman-Rakic PS (2002) An auditory domain in primate prefrontal cortex. Nat Neurosci 5:15–16.

Schulz GM, Varga M, Jeffires K, Ludlow CL, Braun AR (2005) Functional neuroanatomy of human vocalization: An H215O PET study. Cereb Cortex 15:1835–1847.

Toyomura A, Omori T (2005) Auditory feedback control during a sentence-reading task: Effect of other's voice. Acoust Sci Tech 26:358–361.

Wilson SM, Saygin AP, Sereno MI, Iacoboni M (2004) Listening to speech activates motor areas involved in speech production. Nat Neurosci 7:701–702.

Wise RJS, Greene J, Buchel C, Scott SK (1999) Brain regions involved in articulation. Lancet 353:1057–1061.

Yokoyama S, Miyamoto T, Riera J, Kim J, Akitsuki Y, Iwata K, Yoshimoto K, Horie K, Sato S, Kawashima R (2006) Cortical mechanisms involved in the processing of verbs: an fMRI study. J Cogn Neurosci 18:1304–1313.

Zarate JM, Zatorre RJ (2005) Neural substrates governing audiovocal integration for vocal pitch regulation in singing. Ann N Y Acad Sci 1060:404–408.

Zatorre RJ, Samson S (1991) Role of the right temporal neocortex in retention of pitch in auditory short-term memory. Brain J Neurol 114:2403–2417.

Zatorre RJ, Evans AC, Meyer E, Gjedde A (1992) Lateralization of phonetic and pitch discrimination in speech processing. Science 256:846–849.