NeuroReport 8, 2809–2812 (1997)

Vocal identification of speaker and emotion activates differerent brain regions

Satoshi Imaizumi,[CA] Koichi Mori, Shigeru Kiritani, Ryuta Kawashima,[1] Motoaki Sugiura,[1] Hiroshi Fukuda,[1] Kengo Itoh,[2] Takashi Kato,[2] Akinori Nakamura,[2] Kentaro Hatano,[2] Shozo Kojima[3] and Katsuki Nakamura[3]

Department of Speech and Cognitive Sciences, Graduate School of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyoku, Tokyo 113; [1]Department of Nuclear Medicine and Radiology, IDAC, Tohoku University, 4-1 Seiryocho, Aobaku, Sendai 980-77; [2]Department of Biofunctional Research, National Institute for Longevity Sciences, 36-3 Gengo, Moriokacho, Obu, Aichi 474; [3]Primate Research Institute, Kyoto University, Inuyama, Aichi 484, Japan

[CA]Corresponding Author

REGIONAL cerebral blood flow was measured in six healthy volunteers by positron emission tomography during identification of speaker and emotion from spoken words. The speaker identification task activated several audio-visual multimodal areas, particularly the temporal poles in both hemispheres, which may be involved in connecting vocal attributes with the visual representations of speakers. The emotion identification task activated regions in the cerebellum and the frontal lobe, suggesting a functional relationship between those regions involved in emotion. The results suggest that different anatomical structures contribute to the vocal identification of speaker and emotion.

Key words: Cerebellum; Emotion; Frontal lobe; PET; Prosody; rCBF; Speaker identification; Speech; Temporal pole

Introduction

Speech sounds convey not only linguistic information but also non-linguistic content such as speaker identity and his/her emotional state. Both speaker identity and emotional states have been widely considered to be conveyed by the prosodic cues of speech, such as voice quality, vocal pitch contour and tempi. Although right hemisphere dominance in prosody perception has been widely accepted,[1–3] little is known about the mechanisms by which the brain extracts these two different prosodic aspects. In the present study we considered whether different anatomical structures of the human brain are involved in speaker and emotion identification from spoken words.

Materials and Methods

Speech stimuli: Eight professional actresses/actors (four females and four males) produced three examples each of eight non-affective words with four different emotional tones: surprise, disgust, happy and angry. Through a preliminary perceptual test, two actresses aged 30 and 57 (S1 and S2), and two actors aged 30 and 56 (S3 and S4) were selected as test speakers who achieved about 80% correct on speaker and emotion identification. In total, 384 utterances were used after digitization through an audio interface with a 20 kHz sampling rate and a 16 bit resolution.

Subjects: Six right-handed male volunteers (age 18–25 years) participated in the study. None had any signs or history of medical or neurological disease and all had normal magnetic resonance images (MRI) of the brain. Handedness was assessed by the H.N. Handedness Inventory. Written informed consent was obtained from each subject in accordance with the guidelines approved by the National Institute for Longevity Sciences and the Helsinki Declaration of Human Rights, 1975. Prior to the positron emission tomography (PET) experiments, each subject had a catheter placed into his left brachial vein for tracer administration. A high resolution MRI scan (1.5 T) was performed on a separate occasion.

Task procedures: All subjects performed two tasks, a speaker identification (SI) task and an emotion identification (EI) task. For the SI task, the subjects were instructed to guess the speaker of utterances presented in a random sequence and to press the button assigned to each speaker. For the EI task, they were instructed to guess the emotion that the speaker was trying to express through vocalization and to report by pressing the button assigned to each specific emotion. The subjects were instructed to respond by their right hand using four fingers as soon as possible after stimulus offset. The finger-emotion or the finger-speaker assignment was fixed as follows: the index finger was assigned to surprise or S4, the middle to disgust or to S2, the third to happy or S3, and the small to angry or S1. The sound stimuli were presented to both ears at the most clearly perceived level of each subject (about 60 dBSL) through a pair of stereo earphones.

PET measurements: Each subject was placed comfortably in a supine position on the scanner. The room was darkened (0.7 lux) and kept quiet for the duration of the study. Before the start of the PET measurement, subjects were trained to perform two tasks. The rCBF was measured with a PET scanner (ECAT EXACT HR (Siemens/CTI, Knoxville, Tenn., USA); the in-plane resolution was the full-width of 3.1 mm at half-maximum[4]) using a 3-D collection mode and a bolus injection of approximately 15 mCi (555 MBq) $H_2{}^{15}O$, which was given before each PET scan. Reconstruction and filtering gave a final image resolution of 4.0 × 4.0 × 3.4 mm. Prior to the PET measurements, a transmission scan was performed using three external $^{68}Ge/^{68}Ga$ rod sources. This scan directly measured the attenuation coefficients, and these data were used to obtain corrected emission images. All tasks were started 20 s prior to the bolus injection. Each PET measurement commenced immediately after the bolus injection and continued for 120 s. An electroculogram (EOG) was also recorded for each subject during the PET measurements.

All PET images were smoothed with a 3-D Gaussian filter 8 mm wide, then normalized for global cerebral counts of 100 counts/voxel.

Anatomical standardization and statistical evaluation of the PET data: The standard anatomical structures of the computerized brain atlas of Roland *et al.*[5] were fitted interactively to each subject's MRI using both linear and non-linear parameters. These parameters were subsequently used to transform each subject's PET images and MRIs into the standard atlas form.[6] After anatomical standardization of the PET images, we made subtraction pictures of each task minus the other task, voxel by voxel, for each subject. Mean and variance pictures, as well as descriptive Student's pictures, were then calculated. Voxels with t-values > 4.03 ($p < 0.005$) were considered to represent regions of significantly changed rCBF. Finally, each activation was superimposed onto the average reformatted MRI of the same six subjects. Anatomical localization of areas of the activation in each subtraction condition was performed in relation to the mean reformatted MRI.

Results

The SI task and the EI task activated different anatomical structures. As shown in Table 1, the SI task significantly activated fields in the visual-related multimodal areas, the temporal poles (Fig. 1f) and the parieto-occipital junction (Fig. 1c) in both hemispheres, and the lateral-occipital gyrus (Fig. 1c) in the left. Activations were found more in the right hemisphere, i.e. in the inferior parietal gyrus (Fig. 1b), and in a large field ranging across the superior frontal gyrus and the superior frontal sulcus (Fig. 1a), which showed the highest t-values for the

Table 1. Anatomical regions specifically activated in SI task: Talairach coordinates, t-values of peak activations, sizes of activated regions and corresponding panels in Figure 1

Anatomical structure	Talairach coordinates			*t*-value	Size	Panel in Fig. 1
	X	Y	Z			
Superior frontal gyrus	–19	16	60	16.4	1096	a
Inferior parietal gyrus	–55	–54	44	5.1	272	b
Parietal-occipital junction	–51	–72	21	8.6	712	c
Lateral-occipital gyrus	14	–96	20	5.6	240	c
Frontal-parietal operculum	–48	–8	16	9.5	320	c
Parietal-occipital junction	43	–14	15	4.5	128	c
Parahippocampal gyrus	16	–40	8	5.1	80	d
Caudate head	–8	16	7	7.4	608	d
Lingual gyrus	–10	–80	–8	7.7	56	e
Temporal pole	–38	–1	–11	7.9	488	f
Temporal pole	50	22	–14	7.2	272	f

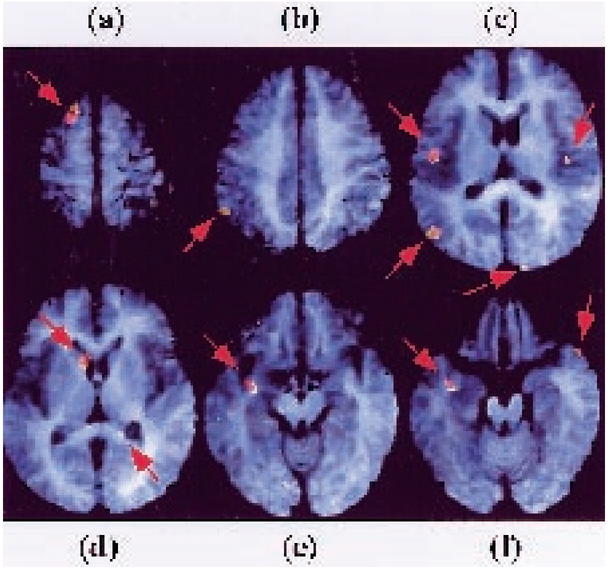

FIG. 1. A series of axial sections of the mean reformatted MRI from the brains of the six subjects. The bright areas on the MRI show the fields of activation which were significantly active in SI task minus EI task. The panel numbers refer to Table 1.

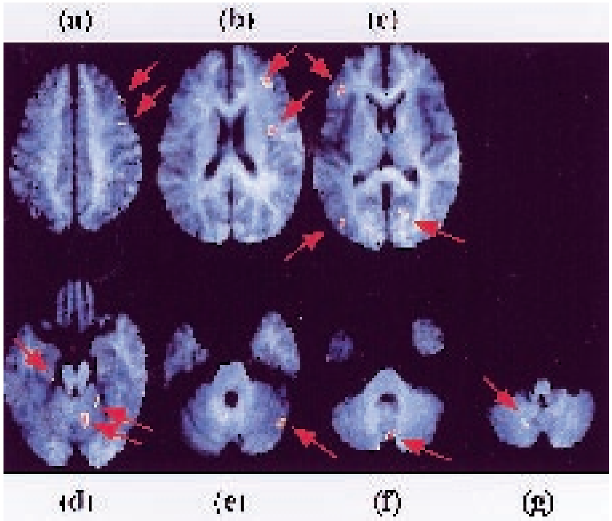

FIG. 2. A series of axial sections of the mean reformatted MRI from the brains of the six subjects. The bright areas on the MRI show the fields of activation which were significantly active in EI task minus SI task. The panel numbers refer to Table 2.

SI task. The left parahippocampal gyrus (Fig. 1d), the right caudate head (Fig. 1d), and the right lingual gyrus (Fig. 1e) were also activated.

The EI task (Table 2) significantly activated fields in the cerebellum (Fig. 2d,e,f,g). There were three activated fields in the left middle frontal gyrus (Fig. 2a,b) and one in the right inferior frontal gyrus (Fig. 2c). Other fields activated were one in the left insula cortex, one in the right parahippocampal gyrus (Fig. 2d) and one in the lateral occipital gyrus (Fig. 2c). The highest activity in the right hemisphere was found in the inferior frontal gyrus, whereas the highest activity found in the left hemisphere was in the insula (Fig. 2b).

Discussion

No concrete theory has been proposed regarding which anatomical structures in the human brain are responsible for speaker identification from voice. One of the most important results obtained in the present study is the significant activation in the temporal poles of both hemispheres (Fig. 1f). The temporal poles have been known as higher-order visual cortical areas that are related to object recognition and memory.[7] In humans, damage to this part of the cortex results in deficits in the recognition of familiar objects and faces. The PET measurements during visual face identification tasks revealed that

Table 2. Anatomical regions specifically activated in EI task: Talairach coordinates, *t*-values of peak activations, sizes of activated regions and corresponding panels in Figure 2

Anatomical structure	Talairach coordinates			*t*-value	Size	Panel in Fig. 2
	X	Y	Z			
Middle frontal gyrus	42	6	41	4.7	32	a
Middle frontal gyrus	49	26	39	5.9	248	a
Insula	41	–2	23	12.2	960	b
Middle frontal gyrus	34	42	20	5.9	472	b
Lateral-occipital gyrus	–40	–78	12	5.2	160	c
Inferior frontal gyrus	–42	39	6	19.4	984	c
Lingual gyrus	6	–63	4	11.9	3080	c
Parahippocampal gyrus	–24	–28	–16	6	40	d
Cerebellum	18	–52	–20	4.5	280	d
Cerebellum	46	–62	–22	9.7	728	d
Cerebellum vermis	4	–73	–35	7.1	440	e
Cerebellum	–17	–64	–40	4.6	104	f
Cerebellum	–13	–48	–50	4.5	128	g

the anterior region of the temporal lobes of both hemispheres, including the temporal poles, were activated.[8] The present results suggest, for the first time, that the temporal poles also contribute to speaker identification from spoken words. The temporal poles may be involved in connecting vocal attributes with visual representations of speakers, and particularly facial representations. The significant activation in the right lingual gyrus, which has been reported as being activated in visual gender identification tasks,[8] may concern gender categorization, which is necessary to identify a speaker.

The left parahippocampal gyrus was significantly activated in the SI task (Fig. 1d). In contrast, the right parahippocampal gyrus was activated in the EI task (Fig. 2d). This result suggests that there is a task-dependent role shared between the right and left parahippocampal gyri, i.e. the left for speaker identification and the right for emotion identification. The parahippocampal gyri may be involved in re-establishing the connections between vocal representations and the pertinent biographical information that was activated and stored during earlier encounters with these speakers or emotions.

A striking result of the EI task was the activities found in the cerebellum (Fig. 2d,e,f,g). At least three reports have suggested a substantial relationship between the cerebellum and emotion, i.e. affective symptoms and cerebellar pathology,[9] a functional relationship between the cerebellum and parts of the forebrain involved in emotion,[10] and potential involvement of the feedback loop between the cerebellum, septal region and hippocampus in epilepsy and emotional disorders.[11] New research increasingly indicates that the cerebellum coordinates and integrates a wide range of processes not confined to the motor domain. Affective stimuli frequently induce subjective feelings which are associated with autonomous changes in gesture, facial expression and other physiological indices such as blood pressure and body temperature. Identification of emotion is not simply a memory activation process, but is rather an autonomic activation of behavior. Therefore, it seems natural to assume that the cerebellum is involved in the identification of emotion.

For the EI task, significant activities were found in the left and right prefrontal cortex, i.e. one in the right inferior frontal gyrus (Fig. 2c) and three in the left middle frontal gyrus (Fig. 2a,b). This suggests that both left and right prefrontal cortices contribute to emotion identification. Several previous reports based on elctroencephalographic measurements have suggested that prefrontal cortical asymmetries are linked to differences in the basic dimensions of emotion,[12] i.e. negative emotions for the right and positive emotions for the left.[13] The present results suggest that the left middle frontal gyrus and the right inferior frontal gyrus may generate such prefrontal cortical asymmetries.

No significant acitivies were found in the amygdala, which has been reported to play a crucial role in the processing of the emotional salience of faces with a specificity of response to fearful expressions.[14–16] This might be the case, because no fearful vocal expressions were used in the present study.

It seems reasonable to conclude that the different anatomical structures of the human brain contribute to vocal identification of speaker and emotion, both of which mainly rely on the prosodic cues of speech. One should bear in mind that these PET activations can be resulted from deactivation or active inhibition.[17]

Conclusions

The rCBF was measured in six healthy volunteers with PET to determine whether different anatomical structures of the human brain are involved in the vocal identification of speaker and emotion. During the speaker identification task several visual-related areas, particularly the temporal poles in both hemispheres, were activated. During the emotion identification task the cerebellum, the left middle frontal gyrus and the right inferior frontal gyrus were activated. Although right hemisphere dominance has been widely accepted for processing prosody which conveys a speaker's identity and emotional state, different anatomical structures contribute to these two tasks when examined in detail.

References

1. Zatorre RJ, Evans AC, Meyer E and Gjedde A. *Science* **256**, 846–849 (1992).
2. Kawashima R, Itoh M, Hatazawa J *et al. Neurosci Lett* **161**, 69–72 (1993).
3. George M, Parekh PI, Rosinsky N *et al. Arch Neurol* **53**, 665–70 (1996).
4. Wienhard K, Dahlbom M, Eriksson L *et al. J Comp Assisted Tomogr* **18**, 110–118 (1994).
5. Roland PE. Graufelds CJ, Wahlin J *et al. Hum. Brain Map.* **2**, 1–12 (1994).
6. Roland PE and Zilles K. *Trends Neurosci.* **17**, 458–467 (1994).
7. Nakamura K and Kubota K. *Behav Brain Res* **77**, 53–77 (1996).
8. Sergent J and Signoret J. *Phil Trans R Soc Lond B* **335**, 55–62 (1992).
9. Gutzmann H and Kuhl K. *Arch Gerontol Geriatr* **6**, 61–71 (1987).
10. Snider S. *Neurosci Biobehav Rev* **6**, 47–53 (1982).
11. Heath R, Dempesy C, Fontana C and Fitzjarrell A. *Biol Psychiatry* **15**, 541–556 (1980).
12. Tomarken A, Davidson R, Wheeler R and Doss R. *J Personality Soc Psychol* **62**, 676–687 (1992).
13. Sobotka S, Davidson R and Senulis JA. *J Electroencephalogr Clin Neurophysiol* **83**, 236–247 (1992).
14. Young AW, Aggleton JP, Hellawell DJ *et al. Brain* **118**, 15–24 (1995).
15. Adolpha R, Tranel D, Damasio H and Damasio A. *J Neurosci* **15**, 5879–5891 (1995).
16. Scott S, Young AW, AJ Calder AJ *et al. Nature* **35**, 254–257 (1997).
17. Kawashima R, O'Sullivan BT, Roland PE. *Proc Natl Acad Sci USA* **92**, 5969–5972 (1995).

**Received 20 May 1997;
accepted 10 June 1997**