

Cerebral networks for spontaneous and synchronized singing and speaking

Yoko Saito^{a,b,c}, Kenji Ishii^b, Kazuo Yagi^c, Itaru F. Tatsumi^d and Hidehiro Mizusawa^a

^aDepartment of Neurology and Neurological Science, Tokyo Medical and Dental University, ^bPositron Medical Center, Tokyo Metropolitan Institute of Gerontology, ^cDepartment of Radiological Science, Tokyo Metropolitan University of Health Sciences and ^dLanguage, Cognition and Brain Science Research Group, Tokyo Metropolitan Institute of Gerontology, Tokyo, Japan

Correspondence and requests for reprints to Kenji Ishii, MD, Positron Medical Center, Tokyo Metropolitan Institute of Gerontology, I-1 Nakacho, Itabashi-ku, Tokyo 173-0022, Japan

Tel: +81 3 3964 3241 ext. 3503; fax: +81 3 3964 2188; e-mail: ishii@pet.tmig.or.jp

Received 8 September 2006; accepted 26 September 2006

Singing in unison is usually easier than singing alone, but the neural mechanism underlying these two contrasting modes of singing remains unknown. We investigated neural correlates of singing by a functional magnetic resonance imaging study focusing on the capacities of spontaneity and synchronization and compared them with those of speaking. The left inferior frontal gyrus appears important for self-generation of text in singing and speaking without auditory

input, whereas the left posterior planum temporale plays a key role in synchronizing both text and melody, in combination with the bilateral inferior parietal lobule for singing along, and with the left angular gyrus for speaking in chorus. These findings indicate that text and melody are not processed symmetrically or parallel in singing a well-learned song. *NeuroReport* 17:1893–1897 © 2006 Lippincott Williams & Wilkins.

Keywords: functional magnetic resonance imaging, inferior frontal gyrus, planum temporale, singing, speaking, spontaneity, synchronization

Introduction

When singing, we occasionally sing alone; often, however, we sing along with others. This singing may or may not be accompanied by instrumental music. Similarly, although we usually speak alone, we occasionally speak together in a group, for example, during group chanting. Neural networks for singing should involve control mechanisms of two contrasting modes: self-production and synchronized production. It is essential for the self-production of a song to retrieve and represent the internally stored lyrics, rhythm, and melody without the help of external sensory information. On the other hand, it is important for the synchronized production of a song to monitor the external auditory information and to concurrently render synchronous control of articulation. In terms of simple voluntary movements such as finger tapping, there have been several reports that have focused on the differences in the central mechanisms for self-paced movements and externally cued movements [1]. Recently, investigations have been conducted on more complex tasks of finger tapping to external auditory rhythms, such as synchronization compared with syncopation [2], and finger tapping following various rhythms [3]. This work aroused an interest in auditory-motor correlations rather than the spontaneous control of voluntary movements. No previous studies have been found focusing on the capacities of spontaneity and synchronization for singing. It may be useful to investigate the neural mechanism underlying singing from the standpoint of spontaneity and synchronization for the planning of therapy for apraxia of speech, aphasia, and dementia [4]. The purpose of this study was to investigate the neural mechanisms underlying the two different modes of singing: singing alone and

singing along. In this paper, we use the term 'spontaneous' singing to refer to the self-production of a well-learned song without the help of external sensory information, and it does not imply any improvisational behavior.

Some recent works have investigated the neural networks from the viewpoint of simple pitch and melody generation [5,6]. Most of the previous studies that investigated singing, however, showed contrasting mechanisms for language and music [7,8]. These studies have shown a symmetric view of language and music: the inferior frontal gyrus, anterior insula, and superior temporal gyrus in the right hemisphere play a predominant role in singing, whereas the homologous regions in the left hemisphere are mainly involved in speaking. On the basis of a conventional model of cerebral lateralization, some lesion studies have indicated that language and music are dealt with parallel networks [9,10]. On the other hand, behavioral experiments and recent neuroimaging studies have suggested that processes for language and music are not completely separate [11–15]. The complete viewpoint of the inter-relationship between text and melody processing in singing remains unknown. To elucidate the neural correlates underlying spontaneous and synchronized processes in singing, we contrasted them with those in speaking in order to address the issue of integration and segregation of music and language in singing.

Materials and methods

Participants

We studied 20 normal right-handed Japanese nonmusicians (19–24 years old; 10 men and 10 women) without any neurological or hearing impairment, after obtaining their

written informed consent. None had experienced any specific training for singing or playing an instrument, except for ordinary schoolwork. Most of them regularly listened to music, but sang only occasionally in day-to-day life. This study was approved by the Ethics Committee of the Tokyo Metropolitan University of Health Sciences.

Experimental paradigm and stimuli

The task design consisted of the following conditions: (i) overtly singing ‘Umi’ (The Sea) – a very popular song among the Japanese – alone (Sg_solo); (ii) overtly singing Umi in synchronization with the auditory presentation of the song, including the melody and lyrics (Sg_syn); (iii) listening to the singing undertaken in the previous task (L_Sg); (iv) overtly reciting alone the lyrics of Umi without the melody (Sp_solo); (v) overtly reciting Umi along with an auditory presentation of the lyrics (Sp_syn); (vi) listening to the reciting undertaken in the previous task (L_Sp); and (vii) rest condition. The processes involved in each task are summarized in Table 1. For the participants, it was easy to not only sing the song, but also recite the lyrics because the lyrics had a seven-and-five syllable meter, a characteristic style observed in traditional Japanese poems. The auditory stimuli (sampling rate, 44.1 kHz) of singing and speaking were recorded by a female singer, without adding any effect. The averaged rate, fundamental frequency (F0), and intensity of the auditory stimuli were recorded (mean ± SD) to be 1.60 ± 0.14 moras/s, 221 ± 43 Hz, and 69.9 ± 8.1 dB SPL (SPL: sound pressure level), respectively, for singing, and 2.5 ± 0.13 moras/s, 215 ± 43 Hz, and 69.9 ± 6.5 dB SPL, respectively, for speaking. The rate of presentation alone was significantly different between singing and speaking (*P* < 0.01, paired *t*-test). In the sound-proof room, the participants underwent prescan training. During training,

they practiced singing synchronously with an auditory presentation (Sg_syn task), and this was done until they sang the entire song correctly by themselves without making any mistakes in the lyrics, melody, rhythm and tempo, as assessed by the examiner. Prescan training for the speaking version was also done in the same way.

Data acquisition

Magnetic resonance imaging (MRI) data acquisition was performed with a 1.5-T scanner (Signa; General Electric, Milwaukee, Wisconsin) by using a gradient-echo echoplanar imaging-sequence (TR, 5000 ms; TE, 35.1 ms; flip angle, 60°; matrix size, 128 × 128; field of view, 24 × 24 cm²; and 21 slices with 5 mm thickness). The functional MRI procedure involved six runs that were labeled Sg_solo, Sg_syn, L_Sg, Sp_solo, Sp_syn, and L_Sp and the order was counter-balanced across the participants. One run consisted of six blocks in which the rest block (60 s) alternated with the task block (60 s), and the six runs were performed at 5-min intervals. The participants gently bit disposable wooden bite-bars with their molars during scanning to minimize head and jaw motion artifacts associated with articulation, but the articulation was intelligible. During the scan, the performance of the participant was monitored for text and melody components by experimenters. After each scan, we asked the participants whether they could hear the auditory input and their own song/speech clearly during the scan and if they could perform the task as instructed.

Data analysis

Analysis of the functional MRI data was performed using SPM99 (Wellcome Department of Cognitive Neurology, University College London, London, UK) implemented on MATLAB (Mathworks, Natick, Massachusetts, USA). After

Table 1 Components involved in each task

Components		Articulation; self-monitoring		External perception		Spontaneous production		Synchronized production	
		Text	Melody	Text	Melody	Text	Melody	Text	Melody
Task									
Spontaneous singing: Sg_solo		○	○	-	-	○	○	-	-
Synchronized singing: Sg_syn		○	○	○	○	-	-	○	○
Listening to the song: L_Sg		-	-	○	○	-	-	-	-
Spontaneous speaking: Sp_solo		○	-	-	-	○	-	-	-
Synchronized speaking: Sp_syn		○	-	○	-	-	-	○	-
Listening to the speech: L_Sp		-	-	○	-	-	-	-	-
Paired <i>t</i> -test									
A ₁	Sg_solo > Sg_syn								
A ₂	Sg_syn > Sg_solo								
B ₁	Sp_solo > Sp_syn								
B ₂	Sp_syn > Sp_solo								
C ₁	Sg_solo > Sp_solo								
C ₂	Sg_syn > Sp_syn								
D ₁	Sp_solo > Sg_solo								
D ₂	Sp_syn > Sg_syn								

Circles indicate that the task includes the component, whereas dashes indicate the task that does not include the component. Light gray boxes show components related to spontaneous processes, and black boxes show components related to synchronized processes. Sg_solo, spontaneous singing; Sg_syn, synchronized singing; Sp_solo, spontaneous speaking; Sp_syn, synchronized speaking.

realignment, normalization, and smoothing (full-width at half-maximum, 6 mm), the activated voxels were identified using the general linear model approach [16]. In each participant, a single summary 'contrast' image of a given condition was produced by convolving the images with a standard hemodynamic response function. Next, one-sample *t*-test based on the random effects model was applied contrasting passive listening conditions compared with rest: (L_Sg - rest) and (L_Sp - rest). Then, paired *t*-test was applied contrasting one task condition with another task condition, to determine the brain regions that were specifically activated by each component of the task: (Sg_solo - rest) vs. (Sg_syn - rest), (Sp_solo - rest) vs. (Sp_syn - rest), (Sg_solo - rest) vs. (Sp_solo - rest), and (Sg_syn - rest) vs. (Sp_syn - rest) (Table 1). The specific activations of each paired *t*-test was inclusively masked with the minuend contrast to avoid the detection of pseudo activation caused by deactivation (decrease in signal intensity relative to the rest condition) in the subtrahend contrast [17]. For instance, contrast of (Sg_solo - rest) > (Sg_syn - rest) were inclusively masked by (Sg_solo - rest). The threshold for the pass criterion of the inclusive mask was set to *P* < 0.001 (uncorrected). For all statistical comparisons, we used a height threshold of *P* < 0.05 (corrected) at the cluster statistics level and a spatial extent of *K* = 30 voxels. Montreal Neurologic Institute coordinates were converted to the original coordinate system of Talairach and Tounoux (1988) by Matthew Brett's transformations.

Results

At the postscan interview, all participants reported that they could hear their own song/speech and the auditory input of singing/speaking clearly during the scan. In the passive listening tasks, the L_Sg activated the bilateral Heschl's gyrus, bilateral upper bank of the superior temporal sulcus, bilateral posterior part of the planum temporale, left planum polare, right dorsal premotor cortex, and right supplementary motor area, whereas the L_Sp activated the bilateral Heschl's gyrus, bilateral superior temporal sulcus, left posterior planum temporale, and right planum polare.

Specific activations for spontaneous processes and for synchronized processes in singing and speaking are summarized in Table 2a and displayed in Fig. 1. Specific activations for singing processes and for speaking processes in spontaneous and synchronized conditions are summarized in Table 2b. No significant activated area that was involved in speaking but not in singing (D₁ and D₂) existed.

Discussion

The spontaneous tasks selectively activated the left inferior frontal gyrus in contrast to the synchronized tasks: the left anterolateral part of the inferior frontal gyrus in singing and the left pars triangularis of the inferior frontal gyrus in speaking. On the basis of our task design, it can be interpreted that the activation in the left pars triangularis of the inferior frontal gyrus in speaking is specific to the spontaneous generation of text, but it is also possible that

Table 2a Activations for spontaneity and synchronization in each singing and speaking

Brain regions		Singing (Fig. 1a)				Speaking (Fig. 1b)			
		x	y	z	Z-value	x	y	z	Z-value
Spontaneous process (green)		Sg_solo > Sg_syn				Sp_solo > Sp_syn			
Anterolateral part of the IFG	L	-32	37	-5	4.72	-42	25	1	4.61
Pars triangularis of the IFG	L								
Synchronized process (red)		Sg_syn > Sg_solo				Sp_syn > Sp_solo			
Posterior planum temporale	L	-44	-36	12	5.12	-48	-33	2	5.26
	R	40	-33	10	4.58				
Anterior edge of the IPL	L	-54	-27	21	4.73				
	R	50	-20	24	4.68				
Angular gyrus	L					-56	-49	25	4.93

Sg_solo, spontaneous singing; Sg_syn, synchronized singing; Sp_solo, spontaneous speaking; Sp_syn, synchronized speaking; IFG, inferior frontal gyrus; IPL, inferior parietal lobule.

Table 2b Activations for singing and speaking in each spontaneity and synchronization

Brain regions		Spontaneous process				Synchronized process			
		x	y	z	Z-value	x	y	z	Z-value
Singing process		Sg_solo > Sp_solo				Sg_syn > Sp_syn			
Dorsal premotor cortex	R	50	12	36	4.79	50	2	40	5.01
IFG/Anterior insula	R	45	29	-5	4.62	40	23	-5	4.64
Anterior cingulate cortex	R	2	23	36	4.02	2	17	34	4.55
Planum temporale	R	52	-22	6	5.16	60	-29	11	4.97
Planum polare	R					48	-4	10	4.42
Middle insula	R					46	2	-3	4.83
Anterior edge of the IPL	L					-50	-22	26	4.34
Speaking process		Sp_solo > Sg_solo NS				Sp_syn > Sg_syn NS			

Talairach coordinates and the Z-values of peak activation; activated regions detected by paired *t*-test: *P* < 0.05 corrected for multiple comparisons, using an inclusive mask: *P* < 0.001, uncorrected.

Sg_solo, spontaneous singing; Sg_syn, synchronized singing; Sp_solo, spontaneous speaking; Sp_syn, synchronized speaking; IFG, inferior frontal gyrus; IPL, inferior parietal lobule; L, left; R, right.

Comparison of spontaneity and synchronization

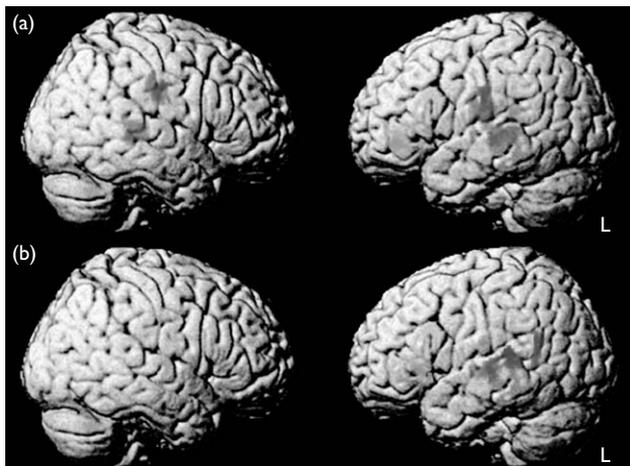


Fig. 1 Activations specific to the spontaneous process (green) and those specific to the synchronized process (red). (a) Singing; (b) speaking ($P < 0.05$, corrected).

the activation of the left anterolateral part of inferior frontal gyrus in singing is related not only to the spontaneous generation of text but also to that of melody [13]. Strong evidence indicates that the major role played by the left inferior frontal gyrus is language generation in various aspects such as phonology, semantics, and syntax [18]. Damage to the left inferior frontal gyrus could cause a condition called dynamic aphasia, in which the patient faces a difficulty in producing spontaneous speech without any difficulty in naming, repetition, and comprehension. Robinson *et al.* [19] attributed this condition to the inability to select an appropriate verbal response from a large number of competing candidates. Warren *et al.* [20] reported the case of a patient suffering from dynamic aphasia owing to left inferior frontal gyrus atrophy who was able to generate melody without lyrics. Therefore, activity in the left inferior frontal gyrus may be related to the spontaneous generation of text rather than melody. Text generation in singing, however, is more complex than that in speaking, considering the need for supplying words for the tune and rhythm. This may explain the difference in the location of the left inferior frontal gyrus activation in singing and speaking. Another possible interpretation of the activation of left inferior frontal gyrus is associated with the increasing motor preparation [15] or cognitive control in spontaneous tasks.

On the other hand, activations specific for the spontaneous generation of melody, which imply the actions complementing the left inferior frontal gyrus activation, were not observed. The right inferior frontal gyrus and right anterior insula have been reported to be involved in melody generation [8,13]. In our study, these areas were found to be active specifically for singing but not for speaking under both spontaneous and synchronized conditions. These results suggest that the roles of the right inferior frontal gyrus and right anterior insula are not exclusive for the spontaneous condition. In addition, the right dorsal premotor cortex was observed as a singing-specific area regardless of spontaneous, synchronized, or listening conditions, implying that the right dorsal premotor cortex is involved in preparing the motor components of music production, such as melody and rhythm [6]. These findings

suggest that melody and text generation in spontaneous singing of a well-learned song are not treated in parallel; melody might subsequently be generated along with text generation. To investigate the specific regions involved in the spontaneous generation of melody, the spontaneity and synchronization of pure melody, avoiding the influence of text, need to be examined in the future.

As the synchronized tasks relative to the spontaneous tasks contain two major components of external perception and synchronized production, activations to be detected by paired *t*-tests need to be interpreted in the light of the activation by passive listening tasks. Among the activations observed that are specific to the synchronized tasks, the posterior planum temporale was also activated by listening tasks. The left posterior planum temporale has been reported as common areas between covert singing and speaking and passive listening to songs and speech [14]. Moreover, the left posterior planum temporale has been reported to be involved in the processing and primary storage of various complex sound patterns [21]. Considering these points, the role of the left posterior planum temporale is not a passive perceptual function; it appears to play a more active role of bridging the external auditory information with further judgment and motor planning. The cortical region surrounding the left posterior Sylvian fissure in the posterior parietal-temporal boundary has been reported to be implicated in verbal working memory [22], auditory-motor integration for speech [23], and auditory-motor integration for both speech and music [24]. Our results suggest that the left posterior planum temporale plays an essential role in the synchronized processing of both melody and text, in combination with more specific regions: in combination with the bilateral anterior edge of the inferior parietal lobule and right posterior planum temporale, for synchronized singing, and with the left angular gyrus, for synchronized speaking. Activations in the left anterior edge of the inferior parietal lobule, right posterior planum temporale, right planum polare, and right middle insula were detected in synchronized singing as compared with synchronized speaking, suggesting that these areas are essential for processing melody rather than text under synchronized conditions [24,25].

Conclusion

Our results elucidate the neural networks of spontaneous and synchronized singing. The left inferior frontal gyrus may be essential for the spontaneous generation of text: the anterolateral part of the left inferior frontal gyrus in singing and pars triangularis of the left inferior frontal gyrus in speaking. In contrast, the right inferior frontal gyrus and right anterior insula are involved in melody generation regardless of spontaneous or synchronized conditions. On the other hand, the left posterior planum temporale plays the role of an auditory hub for synchronization of text and melody, in combination with the bilateral anterior edge of the inferior parietal lobule in singing and the left angular gyrus in speaking. Our findings suggest that text and melody in a well-learned song are not dealt with by symmetric or parallel networks. The neural mechanisms of spontaneous and synchronized singing will provide us with the basics to understand the neural mechanisms for a higher level of spontaneity (e.g. propositional speech and improvisation music), other forms of music (e.g. instrumental

melody and percussion), and their further application in education and rehabilitation.

Acknowledgement

We would like to thank Takashi Kodama, MSc, for the technical support.

References

- Rao SM, Harrington DL, Haaland KY, Bobholz JA, Cox RW, Binder JR. Distributed neural systems underlying the timing of movements. *J Neurosci* 1997; **17**:5528–5535.
- Jantzen KJ, Steinberg FL, Kelso JA. Brain networks underlying human timing behavior are influenced by prior context. *Proc Natl Acad Sci U S A* 2004; **101**:6815–6820.
- Chen JL, Zatorre RJ, Penhune VB. Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms. *Neuroimage* 2006; **32**:1771–1781.
- Gotell E, Brown S, Ekman SL. Influence of caregiver singing and background music on posture, movement, and sensory awareness in dementia care. *Int Psychogeriatr* 2003; **15**:411–430.
- Perry DW, Zatorre RJ, Petrides M, Alivisatos B, Meyer E, Evans AC. Localization of cerebral activity during simple singing. *Neuroreport* 1999; **10**:3453–3458.
- Langheim FJ, Callicott JH, Mattay VS, Duyn JH, Weinberger DR. Cortical systems associated with covert music rehearsal. *Neuroimage* 2002; **16**:901–908.
- Riecker A, Ackermann H, Wildgruber D, Dogil G, Grodd W. Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. *Neuroreport* 2000; **11**:1997–2000.
- Jeffries KJ, Fritz JB, Braun AR. Words in melody: an H₂¹⁵O PET study of brain activation during singing and speaking. *Neuroreport* 2003; **14**:749–754.
- Peretz I, Kolinsky R, Tramo M, Labrecque R, Hublet C, Demeurisse G, Belleville S. Functional dissociations following bilateral lesions of auditory cortex. *Brain* 1994; **117**:1283–1301.
- Hebert S, Racette A, Gagnon L, Peretz I. Revisiting the dissociation between singing and speaking in expressive aphasia. *Brain* 2003; **126**:1838–1850.
- Serafine ML, Crowder RG, Repp BH. Integration of melody and text in memory for songs. *Cognition* 1984; **16**:285–303.
- Koelsch S, Gunter TC, v Cramon DY, Zysset S, Lohmann G, Friederici AD. Bach speaks: a cortical 'language-network' serves the processing of music. *Neuroimage* 2002; **17**:956–966.
- Brown S, Martinez MJ, Parsons LM. Music and language side by side in the brain: a PET study of the generation of melodies and sentences. *Eur J Neurosci* 2006; **23**:2791–2803.
- Callan DE, Tsytsarev V, Hanakawa T, Callan AM, Katsuhara M, Fukuyama H, Turner R. Song and speech: brain regions involved with perception and covert production. *Neuroimage* 2006; **31**:1327–1342.
- Ozdemir E, Norton A, Schlaug G. Shared and distinct neural correlates of singing and speaking. *Neuroimage* 2006; **33**:628–635.
- Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith CD, Frackowiak RS. Statistical parametric mapping in functional imaging: a general linear approach. *Human Brain Mapp* 1995; **2**:189–210.
- Makuuchi M, Kaminaga T, Sugishita M. Brain activation during ideomotor praxis: imitation and movements executed by verbal command. *J Neurol Neurosurg Psychiatry* 2005; **76**:25–33.
- Vigneau M, Beaucousin V, Herve PY, Duffau H, Crivello F, Houde O, et al. Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage* 2006; **30**:1414–1432.
- Robinson G, Blair J, Cipolotti L. Dynamic aphasia: an inability to select between competing verbal responses? *Brain* 1998; **121**:77–89.
- Warren JD, Warren JE, Fox NC, Warrington EK. Nothing to say, something to sing: primary progressive dynamic aphasia. *Neurocase* 2003; **9**:140–155.
- Griffiths TD, Warren JD. The planum temporale as a computational hub. *Trends Neurosci* 2002; **25**:348–353.
- Jonides J, Schumacher EH, Smith EE, Koeppe RA, Awh E, Reuter-Lorenz PA, et al. The role of parietal cortex in verbal working memory. *J Neurosci* 1998; **18**:5026–5034.
- Wise RJ, Scott SK, Blank SC, Mummery CJ, Murphy K, Warburton EA. Separate neural subsystems within 'Wernicke's area'. *Brain* 2001; **124**:83–95.
- Hickok G, Buchsbaum B, Humphries C, Muftuler T. Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *J Cogn Neurosci* 2003; **15**:673–682.
- Brown S, Martinez MJ, Hodges DA, Fox PT, Parsons LM. The song system of the human brain. *Cogn Brain Res* 2004; **20**:363–375.