Title	A study on nonlinguistic features in singing and speaking voices by brain activity measurement
Author(s)	Nakamura, Tomohiko; Kitamura, Tatsuya; Akagi, Masato
Citation	2009 International Workshop on Nonlinear Circuits and Signal Processing (NCSP'09): 217-220
Issue Date	2009-03-01
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/9964
Rights	This material is posted here with permission of the Research Institute of Signal Processing Japan. Tomohiko Nakamura, Tatsuya Kitamura and Masato Akagi, 2009 International Workshop on Nonlinear Circuits and Signal Processing (NCSP'09), 2009, pp.217-220.
Description	

2009 International Workshop on Nonlinear Circuits and Signal Processing
NCSP'09, Waikiki, Hawaii, March 1-3, 2009

A study on nonlinguistic features in singing and speaking voices by brain activity measurement

Tomohiko Nakamura†, Tatsuya Kitamura‡ and Masato Akagi†

†Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa, 923-1292 Japan
Phone/FAX:+81-761-51-1699/+81-761-51-1149
Email: {tomo-nakamura,akagi} @jaist.ac.jp

‡Konan University
8–9–1 Okamoto,Higashinada, Kobe, Hyogo, 658–8501, Japan
Phone/FAX:+81-078-435-2535/+81-078-435-2540
Email: t-kitamu@konan-u.ac.jp

Abstract

Singing voice has unique acoustical features that are different from those of speaking voice. In this study, we investigate brain activities that elicited by the stimuli concerning singing and speaking voices. We analyze differences of those brain activities to investigate human voice perception. The results of brain activity measurement experiments showed that certain brain activity regions are elicited by the singing voice stimulus. The brain regions include LOrG (lateral orbital gyrus), MOrG (medial orbital gyrus) concerning the emotion system.

1. Introduction

The purpose of this study is to investigate what acoustical features concerning nonlinguistic information elicit brain activities by observing the brain activities when presenting stimuli with different nonlinguistic information to listeners.

Speech has linguistic information and nonlinguistic information. Humans perceive both information from speech. Linguistic information is what speaker said, and nonlinguistic information is the information related to speaker's gender and feelings, etc. Although a large number of studies have been done on investigating perception of linguistic information, few focuses on perception of nonlinguistic information. It is understood that humans can perceive and distinguish singing and speaking voices by using the difference of nonlinguistic information. In this study, we investigate acoustical features that humans capture to perceive speaking voice and singing voice as an example to examine the perception mechanism of nonlinguistic information by measuring the brain activity.

We conduct two brain measurement experiments using stimulus sounds that have the same linguistic information and different nonlinguistic information. In experiment I, we use the stimuli that include real singing and speaking voices, synthesized sounds in which singing voice spectrum shape and vibrato are added to speaking voice in order to investigate brain activities are elicited by singing voice, speaking voice and acoustical features including spectrum shape, vibrate. In experiment II, we use the stimuli that include synthesized sounds which have different acoustical features including F0, spectrum shape, amplitude envelope in order to investigate brain activities are elicited by acoustical features including F0, spectrum shape, amplitude envelope. Moreover, we conducted the psychoacoustic experiments in order to discuss the relation between the results of psychoacoustic experiments and the results of brain activities.

2. Experiment I

2.1. Stimuli

In experiment I, to investigate brain activities elicited by the difference between singing voice and speaking voice, the influence of spectrum shape and vibrate, we used six stimulus sounds. Those stimuli were synthesized using the high-quality analysis-synthesis system STRAIGHT [2]. The six stimuli include: (1) Speak: real (actual) speaking voice; (2) Sing: real (actual) singing voice; (3) Base: synthesized voice with the singer's formant (formant peak at about 3 kHz); (4) VR1: synthesized voice that has vibrate of 0.95-1.05 Hz; (5) VR2: synthesized voice that has vibrate of 5.3-5.9 Hz; (6) VR3: synthesized voice that has vibrate of 18.93-21.0 Hz. Speak and Sing were vocalized professional tenor singer. The other stimuli were synthesized with spectrum shape or vibrate that were important for singing voice [1]. The vibrato rate and band width of VR2 was decided as the parameter of natural singing voice according to Saito's research [1]. The parameters of VR1 and VR3 were decided as the center frequency of 1 Hz, 20 Hz, and the band width of the same ratio. Linguistic information of all stimuli were the same /a/, and the duration of Sing stimuli was 1.79 seconds, others were 1.74 seconds.

2.2. Psychoacoustic experiment

In this study, to investigate singing voice perception, we need to evaluate "Singing-ness". Moreover, to examine whether the stimuli are perceived as human voice, we need to evaluate "Natural-ness". Then, we discuss the relation between the results of psychoacoustic experiments and the results of brain activities.

2.2.1. Method

The subjects were ten normal-hearing Japanese(nine males and one female).

The paired stimuli were presented through binaural headphones at a comfortable loudness level. The number of paired stimuli was 30. Each paired stimulus was randomly presented to each subject three times. Scheffe's paired comparison method [3] was used to evaluate the "Singing-ness" and "Natural-ness" of stimulus (Five-grad evaluation measure: -2, -1, 0, 1, 2). Experimental apparatus are: Personal computer is Panasonic CF-R6, D/A converter and headphone amplifier is YAMAHA DP-U50, headphone is SENNHEISER HDA200.

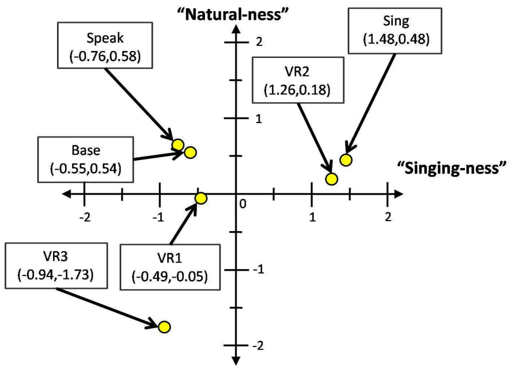

Figure 1: Psychoacoustic experiment results

2.2.2. *Results and Discussions of psychoacoustic experiment*

The results of psychoacoustic experiments are shown in Fig.1. The horizontal axis indicate the degree of "Singing-ness", and the vertical axis indicate the degree of "Natural-ness". Fig.1 indicate that Sing and VR2 are large "Singing-ness". It is showed that the effect of the vibrate rate 5.3-5.9 Hz is large on "Singing-ness". This results are consistent with the results of Saito's study[1]. Fig.1 showed that Speak, Sing and Base are large "Natural-ness".

2.3. Brain activity measurement experiment I

To investigate the brain activities elicited by the singing and speaking voices, and the acoustical features including spectrum shape, fundamental frequency, we conducted brain activity measurement experiment.

2.3.1. *Method*

The subjects were fifteen normal-hearing Japanese(eleven males and four females). They all were right-handed.

We add noise stimulus except the six stimuli. To do not consider the experiment on the singing voice, we instructed subjects to press button when they listen to noise stimulus (oddball task).

In this brain activity measurement experiment, six stimuli and noise stimulus were presented via headphone in functional MRI. Subjects were instructed to close eyes and keep still. Each stimulus was presented fifteen times at optimized order, and noise was presented ten times in one session. Each stimulus was presented every 4 seconds. The experiment was three runs for each subject. For functional brain imaging, a 3.0-T functional MRI was used at ATR BAIC. A total of 30 contiguous axial slices was acquired with a $3.0 \times 3.0 \times 4.0$-mm voxel resolution. A total of 108 scans were taken for each run of the experiment. Each run was approximately 7 min in duration. Images were realigned, unwarped, spatially normalized to a standard space using a template EPI image, and smoothed using an $6 \times 6 \times 6$-mm FWHM Gaussian kernel.

Those obtained brain data were analyzed using SPM5 software (Statistical Parametric Mapping).

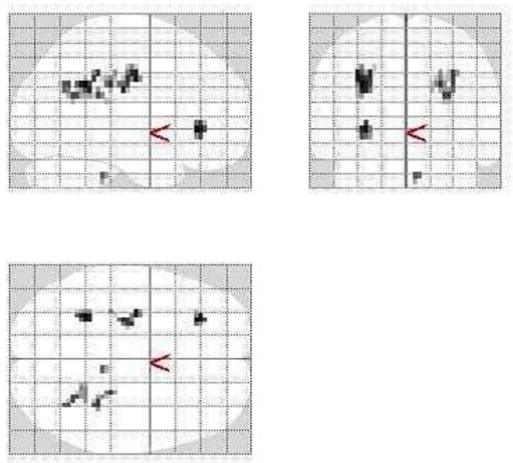

Figure 2: Sing minus Speak in experiment 1

2.3.2. *Results of brain activities in the experiment I*

In the experiment I, trends were investigated using a threshold of P <0.001uncorrected, spatial extent threshold 3 voxels. Sing stimuli more elicited brain activities than Speak stimuli. The results of brain activity by Sing minus brain activity by Speak contrast are shown in Fig.2. These regions include LOrG (lateral orbital gyrus) [-22, 33, 0], SPL (superior parietal lobule) [-30, -45, 32], PrG (precentral gyrus) [-24, -12, 40], AnG (angular gyrus) [27, -57, 36], Cerebellum [6, -33, -32]. However, the effect of brain activities by Speak minus that by Sing showed no activations. Moreover, the results of the synthesized sounds versus the other synthesized sounds showed no large activations. Thus, the effects of different acoustical features (spectrum shape and vibrate) were not clearly.

3. Experiment II

3.1. Stimuli

The purpose of brain activity measurement experiment II was to investigate brain activities elicited by spectrum shape, fundamental frequency and amplitude envelope.

In the experiment II, we used six stimulus sounds. Those stimuli were synthesized from actual singing voice (Sing) and speaking voice (Speak) that were used in experiment I. In consideration of "Natural-ness", we extracted spectrum shape, fundamental frequency and amplitude envelope from those two actual human voices. Next, we synthesized six stimuli based on extracted acoustical features. The six stimuli include: (1) Sp-Sp-Sp: synthesized with F0, spectrum shape, amplitude envelop, from Speak; (2) Sp-Si-Sp: synthesized with spectrum shape from Speak, and F0, amplitude envelop from Sing; (3) Sp-Si-Si: synthesized with F0 from Speak, spectrum shape, amplitude envelop, from Sing; (4) Si-Sp-Sp: synthesized with F0 from Sing, spectrum shape, amplitude envelop, from Speak; (5) Si-Sp-Si: synthesized with spectrum shape from Speak, F0, amplitude envelop, from Sing; (6) Si-Si-Si: synthesized with F0, spectrum shape, amplitude envelop, from Sing. Linguistic information of those stimuli

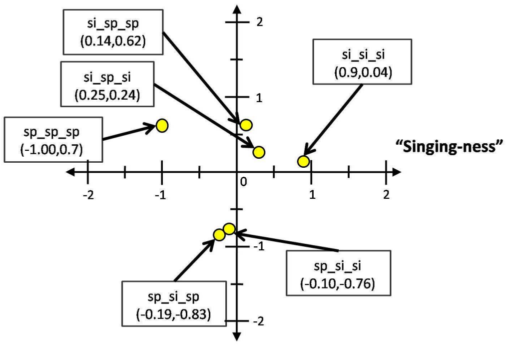

Figure 3: Psychoacoustic experiment results

were the same /a/, and duration of those were also same 1.74 sec, and sound pressure were same.

3.2. Psychoacoustic experiment

3.2.1. Method

This psychoacoustic experiment was conducted in the same way as in the experiment I. Only stimuli are different. The subjects were nine normal-hearing Japanese (eight males and one female).

3.2.2. Results and Discussions of psychoacoustic experiment

The results of psychoacoustic experiment are shown in Fig.3. Fig.3 shows that Si-Si-Si was the lagest "Singing-ness", and Sp-Sp-Sp was the smallest "Singing-ness" in the six stimuli. The evaluation of another stimulus's "Singing-ness" were between evaluation of Si-Si-Si and that of Sp-Sp-Sp. The effect of F0 on singing-voice perception was the largest among all features. In addition, Fig.3 showed that the effect of spectrum shape was the smallest on "Natural-ness".

3.3. Brain activity measurement experiment Ⅱ

To investigate brain activities that were elicited by F0, spectrum shape and amplitude envelope, we conducted brain activity measurement experiment Ⅱ.

3.3.1. Method

The data presented in the experiment were obtained from a total of sixteen healthy subjects. The subjects were twelve males and four females. Fifteen subjects were right-handed and one was left-handed. The experiment design and procedure were the same as experiment I.

3.3.2. Results of experiment Ⅱ

In the experiment Ⅱ, trends were investigated using a threshold of P <0.001 uncorrected, spatial extent threshold 3 voxels. The contrasts of interest included the following: (1) difference between listening to Si-Si-Si and listening to Sp-Sp-Sp; (2) difference between stimuli which are only different in F0; (3) difference between stimuli which are only different in spectrum shape; (4) difference between stimuli which are different in F0 and amplitude envelope.

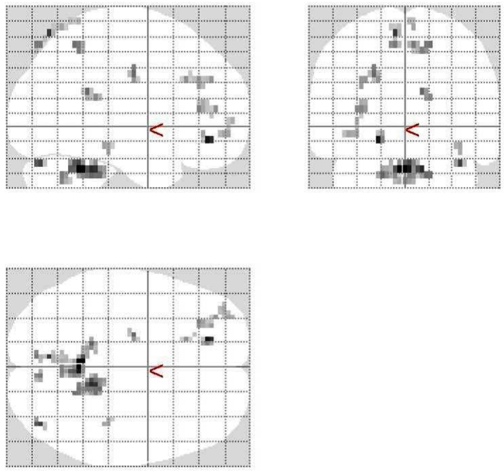

Figure 4: Si-Si-Si minus Sp-Sp-Sp in experiment Ⅱ

The results of the listening to Si-Si-Si minus the listening to Sp-Sp-Sp contrast are shown in Fig.4. These regions include MOrG (medial orbital gyrus) [-18, 42, -8], Cerebellum [-3, -45, -28]. In contrast, the activities regions of Sp-Sp-Sp minus Si-Si-Si include MTG (middle temporal gyrus) [57, -42, 4].

The results of the contrast between stimulus that are only different F0 component are shown in Fig.5. These regions include CG (cingulate gyrus) [-3, 6, 40], [-6,30,28], Ins (insula) [42,-18,4], SMG (supramarginal gyrus) [45, -48, 32], STG (superior temporal gyrus) [45, -45, 16], [57,-39,16], MTG [48,-36, -8], PCun (precuneus) [-3, -60, 12].

The results of the contrast between stimulus that are only different Spectrum shape component are shown in Fig.6. These regions include CG [12, -36, 20], Cd (caudate nucleus) [18, -21, 24], Cerebellum [-33, -57, -40].

The results of the contrast between stimulus that are different F0 and envelope components are regions include IFGOr (inferior frontal gyrus, orbital part) [-39, 30, -12], SG (straight gyrus) [-6, 36, -12].

4. Discussions

The two experimental results showed that certain brain activity regions were elicited by the singing voice stimulus. Overlapping brain activity regions of singing voice minus that of speaking voice in both experiments are orbital part including LOrG and MOrG. These region have been implicated to be involved with emotion[4]. In another study [5], these regions are activated to perceive music. As results, it might be activated those regions, when humans perceive singing voice.

The results of the experiment Ⅱ showed that brain activities were different because of the difference of the acoustical features including F0 and spectrum shape. These brain regions include CG (cingulate gyrus) that belongs to limbic system, Cd (caudate nucleus) that belongs to basal ganglion, Ins (in-

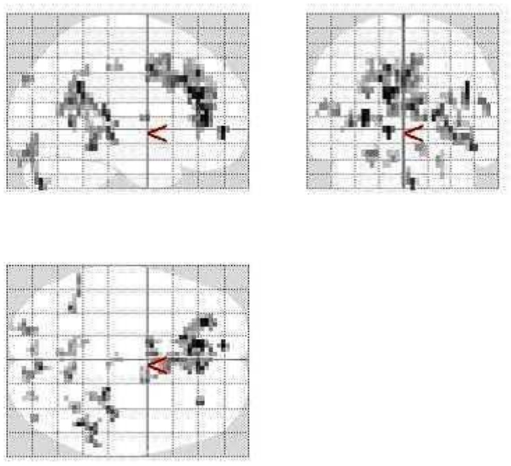

Figure 5: the brain activity of the difference of F0

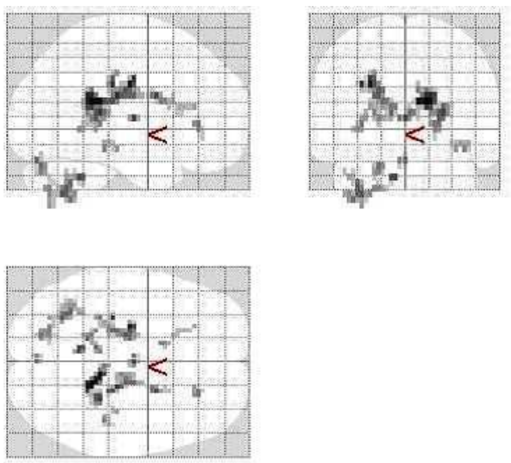

Figure 6: the brain activity of the difference of spectrum shape

sula). Ins (insula) is thought that a route which sends the limbic system the signal concerning feelings. These activity regions are inside regions of brain. As results, we suppose that the processing of nonlinguistic information concerning the singing voice has been activated the regions inside brain regions.

Callan *et al.* also carried out a brain measurement experiment [6] . They used stimuli that were 20 s Japanese songs. Activities in the PrG [-46, 1, 28], SPL [-20, -60, 44], Cerebellum [-40, -48, -28], and OFC (orbitofrontal cortex) [-2, 46, -16] near orbital part for perceiving singing voice are consistent with the findings by Callan *et al.* Those regions (PrG, SPL, Cerebellum) are related to the motor function. It might activate the regions related to vocalization when humans perceive singing voice. The results in Callan's study also activated CG [18, 13, 23], Cd [-8, 5, 18]. As mentioned above, many regions in our results are consistent with the results in Callan's study. In contrast, the active regions only observed in the Callan's study were HG (heschl gyrus) and Hippocampus. It might be the active regions related to linguistic information.

We examined correlations between the results of psychoacoustic experiment and the results of brain activity experiments. The results of experiment I showed that the stimulus with large "Singing-ness" elicited larger brain activities than the stimulus with small "Singing-ness". Moreover, in the experiment I, although Sing and VR2 were also high "Singing-ness", brain activities are quite different. We consider that the difference of "Natural-ness' is a source. There was no region that was more elicited by rising of the evaluation of "Singing-ness".

5. Conclusions

The experimental results showed that certain brain activity regions were elicited by the singing voice stimulus. These brain regions have been implicated to be involved with emotion system. Moreover, it showed that brain activities are different by the different acoustical features. Those brain regions are inside regions of brain. It is necessary for future work to increase the number of subjects so that the activity difference may appear even if the threshold of the analysis is made more severe.

Acknowledgments

This study was supported by the SCOPE (071705001). A part of this work was also supported by a Grant-in-Aid for Science Research (No.20300064).

References

[1] T. Saito, "Construction of Singing-Voice Synthesis System for Clarification of Singing-Voice Perception and Production Mechanism (in Japanese)," JAIST *doctor thesis of information science*, 2006.

[2] H. Kawahara, I. Masuda-Katsuse, and A. Cheveigne, "Restructuring speech representations using a pitch adaptive time-frequency smoothing and an instantaneous-frequency-based on F0 extraction: Possible role of a repetitive structure in sounds," *Speech Communication*, **27**, 187-207, 1999.

[3] S. Ura, "Sensory Evaluation Handbook (in Japanese)," JUSE Press Ltd., 366-384, 1973.

[4] N.R. Horn, M. Dolan, R. Elliott, J.F.W. Deakin, and P.W.R. Woodruff, "Response inhibition and impulsivity: an fMRI study," *Neurosychologia*, **41**, 1959-1966, 2003.

[5] S. Koelsch, T. Fritz, K. Schulze, D. Alsop and G. Schlaug, "Adults and children processing music: An fMRI study," *NeuroImage*, **25**, 1068-1076, 2005.

[6] D.E. Callan, V. Tsytsarev, T. Hanakawa, A.M. Callan, M. Katsuhara, H. Fukuyama, and R. Turner, "Song and speech: Brain regions involved with perception and covert production," *NeuroImage*, **31**, 1327-1342, 2006.