Title	A study of brain activities elicited by synthesized emotional voices controlled with prosodic features
Author(s)	Hamada, Yasuhiro; Kitamura, Tatsuya; Akag, Masato
Citation	Journal of Signal Processing, 14(4): 265-268
Issue Date	2010-07
Type	Journal Article
Text version	publisher
URL	http://hdl.handle.net/10119/9949
Rights	Copyright (C) 2010 Research Institute of Signal Processing Japan. Yasuhiro Hamada,Tatsuya Kitamura,Masato Akag, Journal of Signal Processing, 14(4), 2010, 265-268.
Description	

Journal of Signal Processing, Vol. 14, No. 4, pp. 265-268, July 2010

SELECTED PAPER

A Study of Brain Activities Elicited by Synthesized Emotional Voices Controlled with Prosodic Features

Yasuhiro Hamada[1], Tatsuya Kitamura[2] and Masato Akagi[1]

[1]Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan
Phone/FAX:+81-761-51-1391/+81-761-51-1149
E-mail: y-hamada@jaist.ac.jp, akagi@jaist.ac.jp

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

Abstract

In this study, we investigate the relationships between the results of listening tests and those of brain activity measurements using synthesized emotional speech with controlled acoustic features by referring to the hierarchical hypothesis of feeling. We analyze differences in the brain activities elicited by different emotions corresponding to prosodic features. The results suggest that emotions affected by prosodic features are processed in a different brain region with a different hierarchy of feeling.

1. Introduction

Speech contains linguistic information (What is said) and nonlinguistic information (emotion, individuality and gender, etc.). Much research has focused on linguistic information, although nonlinguistic information is also important in speech communication. Recently, nonlinguistic information has attracted attention because of its importance.

Many researchers reported that the prosody is strongly related to emotions. Acoustic features forming prosody are mainly the fundamental frequency (F0), power envelope and duration. Hayashi reported that the F0 contour conveys much emotional information, on the basis of acoustic feature analyses and listening tests using an interjectory word /eh/[1].

Brain activity has been measured through recently developed techniques (e.g., fMRI). Psychologists and neurologists are reporting the results of measurements of brain activity elicited by emotional voices. Wiethoff et al. reported that emotional voices affect the right mid superior temporal gyrus rather than the natural voice [2]. Bach et al. suggested that the left inferior frontal gyrus plays a specific roll in the explicit evaluation of emotional prosody [3]. However, those groups did not consider how acoustic features affect brain activities. To understand speech communication by emotion, it is needed to investigate the influenced of acoustic features on brain activity related to emotional speech perception. We investigate brain activities using synthesized voices controlled with acoustic features, referring to the results of Hayashi's research [1].

2. Hierarchical Hypothesis of Feeling

In this paper, we refer to the hierarchical hypothesis of feeling proposed by Fukuda [4] to arrange the emotion.

In this hypothesis of feeling, affection is classified into emotion and feeling. Furthermore, emotion is classified into primitive emotion and basic emotion. Feeling is classified into social feeling and intellectual feeling. Emotion has a hierarchical structure evolution and primitive emotion appears by the process of evolution. Primitive emotion is composed of pleasure and unpleasure that is affected by the activity around the hypothalamus. Basic emotion is added to the limbic system as the next stage of the evolution. Basic emotion is composed of joy, anger, fear, disgust and acceptance or love. It is thought that social feeling is acquired through the process by which homo sapiens control language. A region of social feeling to acquire social intellect is thought to develop with the cerebral cortex and limbic system.

3. Stimuli

We measured brain activities using synthesized stimuli with controlled acoustic features of the interjectory word /eh/ to be presented in the brain activity measurements, on the basis of the results of Hayashi's study[1]. Considering the experiment time interval, we used six stimuli in brain activity measurement. Stimuli were carefully synthesized to be highly natural and to be perceived as different emotions. One of the six stimuli was an original voice resynthesized using STRAIGHT[5] without modifying any acoustic features, and the others were synthesized by modifying the acoustic features.

First, we recorded real utterances of the interjectory word /eh/ spoken by six speakers (five males and one female) in seven contexts (asking again, surprise, affirmation, postponement, doubt, disappointment, hesitation). These contexts were the same as those used by Hayashi. We isolated /eh/

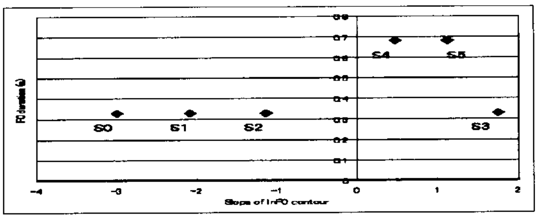

Figure 1: Slope of lnF0 contour and F0 duration

Table 1: List of emotional words

1. Love	11. Dislike	21. Surprise
2. Like	12. Negation	22. Shame
3. Sympathy	13. Doubt	23. Obsequence
4. Pleasure	14. Bitter	24. Asking again
5. Happiness	15. Anger	25. Postponement
6. Joyful	16. Fear	26. Affirmation
7. Neutral	17. Sadness	27. Disappointment
8. Calm	18. Pride	
9. Genial	19. Hesitation	
10. Expectant	20. Impatience	

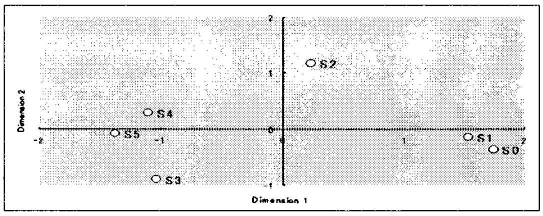

Figure 2: Similarities of stimuli determined by MDS

from each recorded utterance. Then, we calculated the maximum, minimum, average, and gradient of F0. We synthesized voices on the basis of the calculated values.

F0 was controlled using the point pitch model. Because of the correlation between F0 and the power envelope, we controlled power envelopes using the value of F0 contour slopes. The duration was fixed to be the same as that of the original voice or twice as long as that of the original voice. Furthermore, we modified formant frequencies according to F0s. We synthesized forty-two voice samples with attention given to naturalness.

Next, we conducted a psychoacoustic experiment to chose five synthesized voices from among the synthesized voice samples for the use in brain activity measurement. Subjects evaluated naturalness and what kind of emotions (asking again, surprise, affirmation, postponement, doubt, disappointment, hesitation) are included. Referring to the results of the psychoacoustic experiment, we choose five synthesized voices that have large perceptual distances among them and high naturalness. The slopes of the F0 contours and durations of the five chosen synthesized voices and the original voice are shown in Figure 1.

4. Psychoacoustic Experiment

4.1 Method

We conducted a psychoacoustic experiment. Regarding the five synthesized voices and the original voice, subjects were asked to answer what emotions are included. To investigate what characteristic these stimuli have, we define dominant emotional words for each stimulus and show perceptual distances among the stimuli. Furthermore, we discuss corresponding relationships between the dominant emotional words and the hierarchy of feeling proposed by Fukuda[4]. In the psychoacoustic experiment, we used the twenty-seven emotional words listed in Table 1.

The stimuli were evaluated using seven rating scales (Left [Not contained], Right [Contained very much]) for each emotional word. Eight subjects participated in this experiment.

4.2 Results and discussion

From the evaluated scores, the dominant emotional words for each stimulus were found to be *affirmation* and *sympathy* (S0), *affirmation* and *calm* (S1), *disappointment* and *sadness* (S2), *asking again* and *surprise* (S3), *doubt* and *negation* (S4), and *doubt* and *surprise* (S5).

To show perceptual distances among the stimuli, we used multidimensional scaling (MDS). The result of the analysis by MDS is shown in Figure 2. We used the ALSCAL algorithm, and the distance matrix was calculated using the Minkowski distance. In the evaluation of emotional words, Figure 2 might be analogous to the similar dimensions proposed by Schlosberg[6]: The horizontal axis indicates Laxation(+) - Tension(-) and the vertical axis indicates Rejection(+) - Attention(-).

We discuss the corresponding relationships between the results of the psychoacoustic experiment and the hierarchy of feeling proposed by Fukuda. *Affirmation*, *sympathy*, and *calm* were perceived for S0 and S1, which were assigned to the classification of social feeling. *Asking again*, *doubt* and *negation* were perceived for S3, S4, and S5, which were assigned to the classification of social feeling. *Disappointment* and *sadness*, which are basic emotions in many cases thought still discussed were perceived for S2. *Surprise* was perceived for S3 and S5. In many cases, *surprise* is a basic emotion. However, Fukuda defined *surprise* to be a function of the attention system, such as excitation, arousal and attention, evolved as a system different from emotion. On the basis of these results, we expect that each stimulus affects the limbic system, which processes basic emotion, and the cerebral cortex, which processes social feeling.

5. Brain Activity Measurement

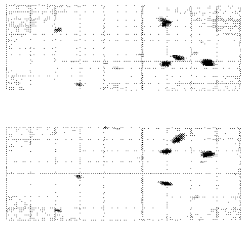

Figure 3: Stronger activation by S1(*affirmation* and *calm*) than by S0 (*affirmation* and *sympathy*). 'Glass brain ' views of statistical parametric maps of fMRI data. (Top left: lateral view; Top right: front view; Left lower: top view)

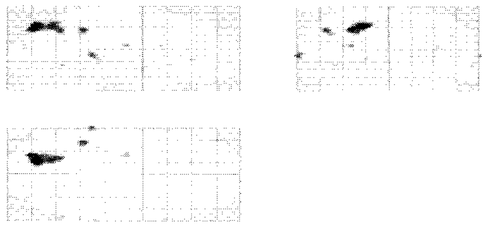

Figure 4: Stronger activation by S2 (*disappointment* and *sadness*) than by S0

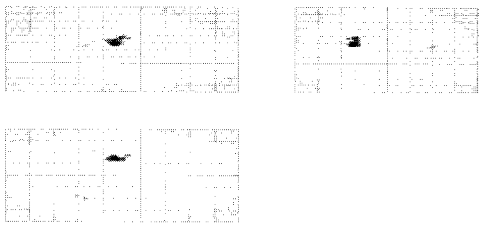

Figure 5: Stronger activation by S3 (*surprise* and *asking again*) than by S0

5.1 Method

The subjects were eighteen normal-hearing Japanese (ten males and eight females). They were all right-handed. For functional brain imaging, a 3.0-T functional MRI was used at ATR BAIC. We informed the subjects that this experiment was about emotional voice and instructed them to push a button upon hearing the noise (oddball task). In this experiment, the original and the five chosen voices were stimuli, and one noise was presented via headphones. Subjects were instructed to close their eyes and keep still. The stimuli were presented fifteen times in an optimized order, and the noise was presented ten times in one session. Each stimulus was presented every 4 seconds. In the experiment, three sessions were run for each subject. A total of 30 contiguous axial slices were acquired with a $3.0 \times 3.0 \times 4.0$ mm voxel resolution. A total of 108 scans were taken in each session of the experiment. Each session was approximately 7 min in duration. Images were realigned, unwarped, and spatially normalized to a standard space using a template EPI image, and then smoothed using an $8 \times 8 \times 8$ mm FWHM Gaussian kernel. The obtained brain data were analyzed using Statistical Parametric Mapping software (SPM5).

5.2 Results

We analyzed the differences in the brain activities during listening to an original voice and the five synthesized voices. Results are shown in Figures 3 to 7 and Table 2.

Results showed that each stimulus elicited activities on the superior temporal gyrus, middle temporal gyrus, supramarginal gyrus and middle frontal gyrus belonging to the auditory area. These areas are said to process the difference in sound stimuli, in previous reports. The difference in the activity on the superior frontal gyrus and left inferior frontal gyrus, included in the orbital area, was shown in S1 minus S0. In S2 minus S0, the superior parietal lobule related to the sensory area and the parietal association area was more activated. In S3 minus S0, S4 minus S0, and S5 minus S0, the dominant regions were the cerebellum, caudate nucleus, and the putamen in basal ganglia.

6. General Discussion

We discuss the relationships between the results of the listening test and those of the brain activity measurements by referring to the hierarchical hypothesis of feeling. Results show that stimulus S1 (*affirmation* and *calm*) minus the original voice S0 (*affirmation* and *sympathy*) elicited different activities on the superior frontal gyrus belonging to the cerebral cortex. The cerebral cortex is considered to be evaluated through the processing of social feeling and intellectual feeling. These results consist of the hierarchical hypothesis of feeling. The stimulus S2 (*disappointment* and *sadness*) minus the original voice S0 (*affirmation* and *sympathy*) elicited different activities on the superior parietal lobule belonging to the cerebral cortex. These results indicate that *disappointment* and *sadness* are related in social feeling and intellectual feeling. The stimuli S3, S4, and S5 (*surprise*, *doubt*, *negation*, and *asking again*) minus the original voice S0 (*affirmation* and *sympathy*) elicited responses mainly in the caudate nucleus or putamen belonging to basal ganglia. The activity on basal ganglia is thought in the processing of primitive emotion. Because basal ganglia includes the adjustment systems of actions affected by body homeostasis, emotion of attention, and tension such as *surprise*, *doubt*, *negation*, and *asking again*, are considered to be related to these adjustment systems, although more information on emotions is still

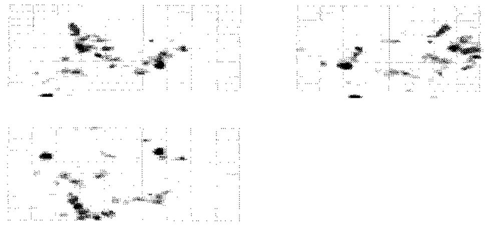

Figure 6: Stronger activation by S4 (*doubt* and *negative*) than by S0

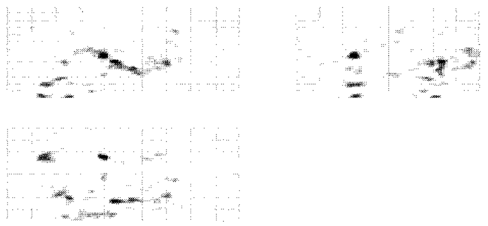

Figure 7: Stronger activation by S5 (*doubt* and *surprise*) than by S0

needed.

7. Conclusions

We investigated the relationships between the results of a listening test and that of brain activity measurements using synthesized emotional speech controlled with prosodic features. The stimulus including *disappointment* and *sadness* elicited activities on mainly in the superior parietal lobule belonging to the cerebral cortex, considered to process social feeling. The stimulus containing *surprise*, *doubt*, *negation*, and *asking again* elicited activities on the basal ganglia, which is considered to process basic emotions. These emotions corresponded to prosody and are suggested to be processes in a different hierarchy by the hierarchical hypothesis of feeling.

Acknowledgments

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

Table 2: Statical analysis of fMRI activation patterns during perception of emotional voices

Brain region	Coordinates	Zvalue
S1 minus S0 (p< 0.01 unc.)		
Superior frontal gyrus	14 18 58	3.13
Frontomarginal gyrus	-28 48 0	3.06
Inferior frontal gyrus	-50 26 6	2.97
Aria orbitoinsularis	-30 16 -2	2.94
S2 minus S0 (p < 0.001 unc.)		
Superior parietal lobule	-18 -78 54	4.51
Angular gyrus	-28 -82 48	4.27
Supramarginal gyrus	-46 -42 48	3.82
Superior temporal gyrus	-68 -36 10	3.76
Superior temporal gyrus	68 -34 10	3.51
Cingulate gyrus	-16 36 4	3.25
S3 - S0 (p < 0.001 unc.)		
Caudate nucleus	-24 -18 28	3.8
Superior frontal gyrus	-22 -14 38	3.54
Supramarginal gyrus	34 -40 26	3.21
Supramarginal gyrus	-34 -34 32	3.14
S4 minus S0 (p < 0.05 FWE.)		
cerebellum	-24 -70 -50	6.07
Area piriformis insulae	-30 12 -6	6.06
Superior temporal gyrus	48 -46 20	5.99
Superior frontal gyrus	-22 18 18	5.71
Putamen	-24 16 10	5.52
Superior parietal lobule	38 -48 44	5.67
Supramarginal gyrus	42 -52 52	5.61
cerebellum	12 -50 -16	5.56
Insular gyrus	42 12 0	5.5
S5 minus S0 (p < 0.05 FWE.)		
Putamen	-24 -26 8	6.32
Insula	38 -18 0	6.07
cerebellum	36 -52 -50	5.89
cerebellum	-24 -70 -32	5.79
Basal operculum	32 18 -2	5.65
Insular gyrus	38 12 0	5.25
Middle temporal gyrus	62 -56 0	5.53
Superior temporal gyrus	60 -38 20	5.49

References

[1] Y. Hayashi: F0 contour and recognition of vocal expression of feelings (in Japanese), The Technical Report of the IEICE, Vol. 98, No. 177, pp.65-72, 1998.

[2] S. Wiethoff, D. Wildgruber, B. Kreifelts, H. Becder, C. Herbert, W. Grodd and T. Ethofer: Cerebral processing of emotional prosoy influence of acoustic parameters and arousal, NeuroImage, Vol. 39, No.2, pp. 885-893, 2008.

[3] D. R. Bach, D. Grandjean, D. Sander, M. Herdener, W. K. Strik and E. Seifritz: The effect of appraisal level on processing of emotional prosody in meaningless speech, NeuroImage, Vol. 42, No. 2, pp. 919-927, 2008.

[4] M. Fukuda: Hierarchical gypothesis of feeling based on evolution of the brain: Positioning of social feeling within evolution (in Japanese), The Japanese Journal of Research on Emotions, Vol. 16, No. 1, pp. 25-35, 2008.

[5] K. Hideki: STRAIGHT, exploitation of the other aspect of vocoder, Acoust. Sci. Tech, Vol. 27, pp. 349-353, 2006.

[6] Y. Fukui: Psychology of Feeling (in Japanese), kawashima shoten, 1990.