JAIST Repository
https://dspace.jaist.ac.jp/

Title	A study on brain activities elicited by synthesized emotional voices controlled with prosodic features
Author(s)	Hamada, Yasuhiro; Kitamura, Tatsuya; Akagi, Masato
Citation	2010 International Workshop on Nonlinear Circuits, Communication and Signal Processing (NCSP'10): 472-475
Issue Date	2010-03-04
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/9968
Rights	This material is posted here with permission of the Research Institute of Signal Processing Japan. Yasuhiro Hamada, Tatsuya Kitamura and Masato Akagi, 2010 International Workshop on Nonlinear Circuits, Communication and Signal Processing (NCSP'10), 2010, pp.472-475.
Description	

JAIST
JAPAN ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

2010 International Workshop on Nonlinear Circuits,
Communication and Signal Processing
NCSP'10, Waikiki, Hawaii, March 3-5, 2010

A study on brain activities elicited by synthesized emotional voices controlled with prosodic features

Yasuhiro Hamada†, Tatsuya Kitamura‡ and Masato Akagi†

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

In this paper, we investigate relationships between 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 of brain activities elicited by different emotions corresponding to prosodic features. Results show emotions affected by prosodic features are suggested to be processed with a different brain region of different hierarchy of feeling.

1. Introduction

Speech contains linguistic information (What are said) and non-linguistic information (emotion, individuality and gender etc.). Many researches have focused on linguistic information, although non-linguistic information is also important for speech communication. Recently, non-linguistic information is attractively attentioned for importance.

Many researches reported that the prosody is strongly related to emotions. Acoustic features forming prosody are mainly F0, power envelope and duration. Hayashi reported that F0 contour conveys much emotional information, from acoustic feature analyses and listening tests using interjectory word /eh/[1].

Brain activity can be measured through recently-developed instruments (e.g., fMRI). Many psychologists and neurologists reported results of brain activity measurements elicited by emotional voices. Wiethoff et al. reported that emotional voices affect right mid superior temporal gyrus rather than natural voice [2]. Bach et al. suggested that the left inferior frontal gyrus plays a specific roll in explicit evaluation of emotional prosody [3]. However, these reports did not consider what acoustic features affect to the brain activity. To understand speech communication by emotion, it is needed to investigate the brain activity related to emotional speech perception influenced by acoustic features. This paper investigates brain activities using synthesized voices controlled acoustic features based on 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 hierarchical structure evolutionary and primitive emotion appeared by the process of the evolution. Primitive emotion is composed of pleasure and unpleasure that is affected by the activity around a hypothalamus. Basic emotion was 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 were acquired by the process in which homo sapiens controls language. A region of social feeling to accomplish social intellect is thought to be acquired with cerebral cortex and limbic system.

3. Stimuli

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

First, we recorded real voices of interjectory word /eh/ uttered by six speakers (five males and one female) in seven contexts(asking again, surprise, affirmation, postponement, doubt, disappointment, hesitation). These contexts are the same as that Hayashi used. Each context is included emotion(asking again, surprise, affirmation, postponement, doubt, disappointment, hesitation). We cut out /eh/ from each

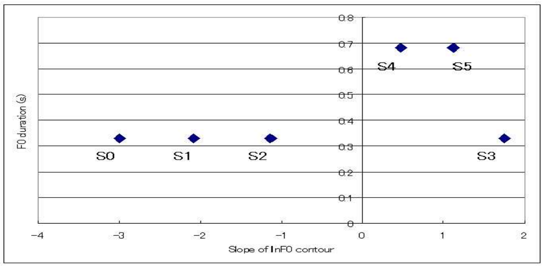

Figure 1: Slope of InF0 contour and F0 duration

Table 1: List of emotional words

1. Love	11. Dislike	21. Surprise
2. Like	12. Negative	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	

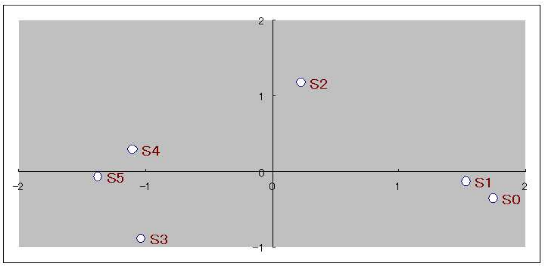

Figure 2: Similarities of stimuli by MDS

recorded voice. Then, we analyzed acoustic features, that are maximum, minimum, average and gradient of fundamental frequency (F0). We synthesize voices according to the extracted values.

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

Next, we conducted a psychoacoustic experiment to choose five synthesized voices from the synthesized voice samples for the 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 highly naturalness. The slopes of the F0 contours and durations of the chosen five synthesized voices and the original one are shown in Figure 1.

4. Psychoacoustic experiment

4.1. Method

We conducted a psychoacoustic experiment. About the five synthesized voices and the original one, 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 relations between the dominant emotional words and hierarchy of feeling proposed by Fukuda[4]. In the psychoacoustic experiment, we used 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 were participated for this experiment.

4.2. Results and Discussion

From the evaluated scores, dominant emotional words for each stimulus were *Affirmation* and *Sympathy*, *Affirmation* and *Calm* (S1), *Disappointment* and *Sadness* (S2), *Asking again* and *Surprise* (S3), *Doubt* and *Negative* (S4) and *Doubt* and *Surprise* (S5).

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

We discuss corresponding relations between results of the psychoacoustic experiment and hierarchy of feeling proposed by Fukuda. *Affirmation*, *Sympathy* and *Calm* are perceived for S0 and S1, which are applied to the classification into social feeling. *Asking again*, *Doubt* and *Negative* are perceived for S3, S4 and S5, which are applied to the classification into social feeling. *Disappointment* and *Sadness* which are basic emotion in many cases thought still discussed are perceived for S2. *Surprise* is perceived for S3 and S5. In many cases, *Surprise* is basic emotion. However, Fukuda define that *Surprise* is the function of attention system such as excitation, arousal and attention evolved as a system different from emotion. From these, we expect that each stimulus affects limbic system that process basic emotion, and cerebral cortex that

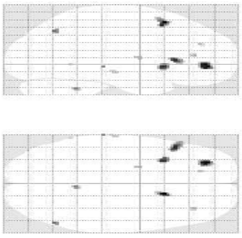

Figure 3: S1 - S0. 'glass brain ' views of statistical parametric maps of fMRI data.(Top-left: lateral view, Top-right: front view, Left-lower: top view)

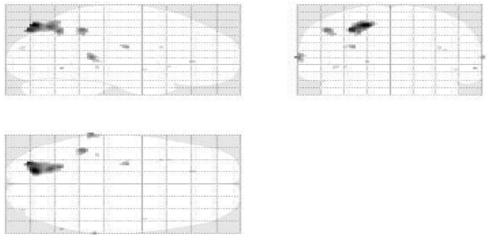

Figure 4: S2 - S0

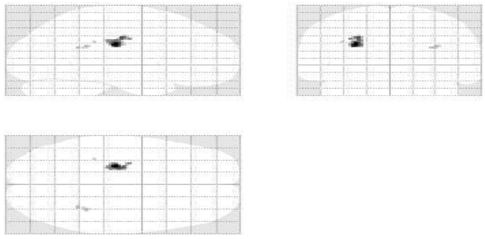

Figure 5: S3 - S0

process social feeling.

5. Brain activity measurement

5.1. Method

The subjects were eighteen normal-hearing Japanese (ten males and eight females). They all were right-handed. For functional brain imaging, a 3.0-T functional MRI was used at ATR BAIC. We instructed to the subjects that this experiment is about emotional voice and push a button when hearing the noise (oddball task). In this experiment, the original and the five chosen voices as stimuli, and one noise was presented via headphone. Subjects were instructed to close eyes and keep still. The stimuli were presented fifteen times at 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 for each session of the experiment. Each session 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 $8 \times 8 \times 8$ mm FWHM Gaussian kernel. Those obtained brain data were analyzed using the Statistical Parametric Mapping software (SPM5).

5.2. Results

We analyzed the differences of brain activities to be listened an original voice with the five synthesized voices. Results were shown in Figures 3 to 7.

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

6. General Discussion

We discuss 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 the stimulus S1(*Affirmation* and *Calm*) minus original voice S0(*Affirmation* and *Sympathy*) elicited different activities on superior frontal gyrus belonging to cerebral cortex. The cerebral cortex is considered to be evaluated with processing of social feeling and intellectual feeling. These results consist with hierarchical hypothesis of feeling. The stimulus S2(*Disappointment*, *Sadness*) minus original voice S0(*Affirmation* and *Sympathy*) elicited different activities on superior parietal lobule belonging to cerebral cortex. These results indicate that *Disappointment* and *Sadness* are related in social feeling and intellectual feeling. The stimulus S3, S4 and S5(*Surprise*, *Doubt*, *Negative* and *Asking again* minus original voice S0(*Affirmation* and *Sympathy*) elicited mainly caudate nucleus or putamen belonging to basal ganglia. The activity on basal ganglia is thought in processing of primitive emotion. Because basal ganglia assumes the adjustment systems of action affected by body homeostasis, emotion of attention, and tension such as *Surprise*, *Doubt*, *Negative* and *Asking again* is supposed to relating these adjustments system though still need more information of emotions.

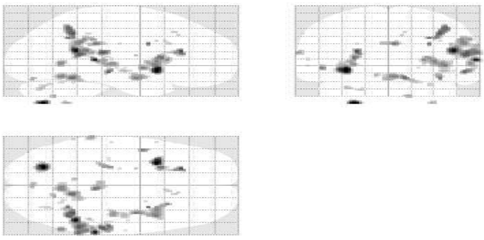

Figure 6: S4 - S0

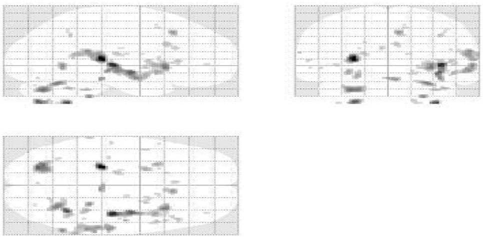

Figure 7: S5 - S0

7. Conclusions

In this paper, we investigated relationships between results of listening test and that of brain activity measurements using synthesized emotional speech with controlled acoustic features. The stimulus included *Disappointment* and *Sadness* elicited mainly superior parietal lobule belonging to cerebral cortex considered to process social feeling. The stimulus contained *Surprise*, *Doubt*, *Negative* and *Asking again* elicited basal ganglia considered to process a basic emotion. These emotions corresponded to prosody are suggested to process in different hierarchy by hierarchical hypothesis of feeling.

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).

Table 2: Statical Analysis of fMRI Activation Patterns during Perception of Emotional Voices

Brain region	Coordinates	Zvalue
S1 - 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 - 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 - 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 - 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] Yasuko 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] Sarah Wiethoff, Dirk Wildgruber, Benjamin Kreifelts, Hubertus Becder, Cornelia Herbert, Wolfgang Grodd, and Thomas Ethofer, "Cerebral processing of emotional prosody-influence of acoustic parameters and arousal," *NeuroImage*, vol. 39, no. 2, pp. 885–893, 2008.

[3] Dominik R.Bach, Didier Grandjean, David Sander, Marcus Herdener, Werner K.Strik, and Erich Seifritz, "The effect of appraisal level on processing of emotional prosody in meaningless speech," *NeuroImage*, vol. 42, no. 2, pp. 919–927, 2008.

[4] Masaji Fukuda, "Hierarchical hypothesis 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] Hideki Kawahara, Ikuyo Masuda-Katsuse, and Alain de Cheveigne, "Restructuring speech representations using a pitch adaptive time-frequency smoothing and an instantaneous-frequency-based f0 extraction: Possible role of a repetitive structure in sounds," *Speech communication*, pp. 187–207, 1999.

[6] Yasuyuki Fukui, "Psychology of feeling (in japanese)," *KAWASHIMA SHOTEN*, 1990.