Proceedings of 20th International Congress on Acoustics, ICA 2010

23-27 August 2010, Sydney, Australia

A study on brain activities elicited by emotional voices with various F0 contours

Yasuhiro Hamada (1), Tatsuya Kitamura (2) and Masato Akagi (1)

(1) Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, Ishikawa, Japan
(2) Konan University, 8-9-1 Okamoto, Higashinada, Kobe, Hyogo, Japan

PACS: 43.71.BP PERCEPTION OF VOICE AND TALKER CHARACTERISTICS

ABSTRACT

Many researchers reported that fundamental frequency (F0) contours conveyed much emotional information in listening tests. However, they did not investigate brain activities elicited by emotions produced by the F0 contours. Psychologists and neurologists reported results of measurements of brain activity elicited by emotional voices. However, these reports did not consider how acoustic features affect brain activities. This paper investigates the relationships between the results of the listening test and the measured results of brain activity affected by the F0 contours. The brain activities of subjects were measured while the subjects listened to six stimulative sounds /eh/ with different F0 contours. One of the six stimuli was the original voice resynthesized without modifying any acoustic features (S0), and the others were synthesized by modifying F0 contours (S1 – S5). The slope of the F0 contours in S1 and S2 were downward, and those in S3, S4, and S5 were upward. A psychoacoustic experiment was carried out to investigate what emotions were perceived in each stimulus (S0 – S5). For brain activity measurement, we analyzed the differences of brain activities elicited by the original voice and that of the other five synthesized voices. Results show that S1 and S2 with downward slopes contain social feelings and elicit portions of the cerebral cortex, and S3, S4, and S5 with upward slopes affect the attention system and elicit portions of the basal ganglia. We could perceive the difference of emotions by processing various F0 contours over the hierarchy of brain regions.

INTRODUCTION

Speech conveys not only linguistic but also non-linguistic information (emotion, individuality, and gender etc.). With non-linguistic information, emotion plays an important role in speech communication. For a long time, psychologists and neurologists have researched emotion. Many researchers tried to study emotional perception and production and reported that prosody is strongly related to emotions [1]. In particular, Hayashi reported that F0 contours convey much emotional information in listening tests [2]. However, she did not investigate brain activities elicited by emotions produced by the F0 contours. Recently, brain activity has been studied because of the recently developed techniques (e.g., fMRI) that are used by many researchers. Psychologists and neurologists reported results of measurements of brain activity elicited by emotional voices [3] [4]. However, these reports did not consider how and where F0 contours affect brain activities. In this paper, we investigate the relationships between the results of the listening test and those of brain activity measurement by using synthesized voices controlled with F0 contours.

MATERIAL AND METHODS

Stimuli

The stimuli were six synthesized voices with controlled acoustic features of the interjectory word /eh/. One of the six stimuli was the original voice resynthesized using STRAIGHT [5] without modifying any acoustic features, and the others were synthesized by modifying acoustic features. To measure brain activity with accuracy, stimuli need to be synthesized carefully to be highly natural and to be perceived as different emotions.

First, we recorded real voices of the 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 were the same as those used by Hayashi [1]. The maximum, minimum, average, and slope of lnF0 contours (gradient of regression line) were calculated from each recorded voice. F0 was controlled using the point pitch model within the range of the calculated maximum and minimum values. The correlation between the slope of lnF0 contour and that of power envelope was calculated. We found that there was a strong correlation between the slope of lnF0 contour and that of power envelope (correlation coefficient was 0.88). Thus, we controlled the power envelopes according to the correlation values. Formant frequencies were modified by shifting on the log frequency axis according to the F0 contours. The duration was fixed to be the same as the original voice or twice the length of the original one. Forty-two voice samples were synthesized with attention to naturalness.

Next, listening tests were conducted to choose five synthesized voices from the synthesized voice samples. Subjects evaluated the naturalness and what kind of emotions (asking again, surprise, affirmation, postponement, doubt, disappointment, hesitation) were included. Referring to the results

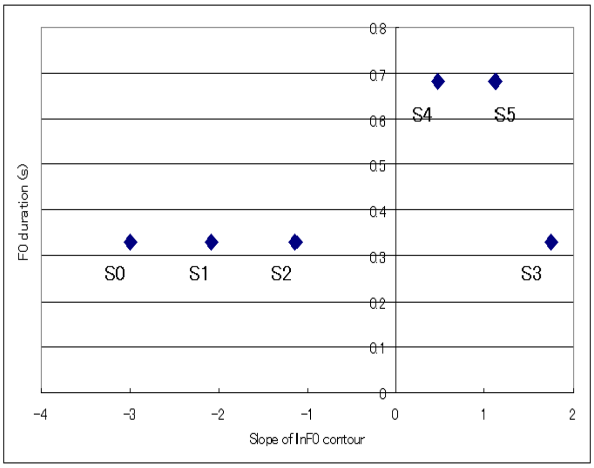

Figure 1. Slope of InF0 contour and F0 duration

of the listening test, we chose five synthesized voices that have large perceptual distances and high naturalness. We named each stimulus S0 (original voice; slope of lnF0 contour was -3.02 and F0 duration was 0.33 ms), S1 (synthesized voices; slope of lnF0 contour was -2.08 and F0 duration was 0.33 ms), S2 (synthesized voices; slope of lnF0 contour was -1.13 and F0 duration was 0.33 ms), S3 (synthesized voices; slope of lnF0 contour was 1.75 and F0 duration was 0.33 ms), S4 (synthesized voices; slope of lnF0 contour was 0.48 and F0 duration was 0.68 ms), S5 (synthesized voices; slope of lnF0 contour was 1.12 and F0 duration was 0.68 ms). The slopes of the F0 contours, durations of the five synthesized voices, and the original one are shown in Figure 1.

Listening test

To investigate what emotion each stimulus had, we conducted listening tests. With regard to the five synthesized voices and the original one, subjects were asked to answer what emotions were included. Eight subjects participated in this experiment. In this experiment, we used twenty-seven emotional words listed in Table 1. The emotional words were selected by referring to previous studies [1] [2]. The stimuli were evaluated using seven rating scales (Left [Not contained], Right [Contained very much]) for each emotional word.

fMRI experiment

Eighteen right-handed normal-hearing Japanese (ten males and eight females; mean age 26.9 years and range 21-53) participated in the fMRI experiment. Whole-brain blood oxygen level dependent (BOLD) fMRI data were acquired using a 3.0-T MR-scanner (Trio, Siemens). In this experiment, the original and the five synthesized voices were stimuli, and one noise was presented via headphones. Subjects were instructed to push a button when they heard the noise (oddball task) and close their eyes and keep still. Each stimulus was presented fifteen times in the optimized order, and the noise was presented ten times in one session. The stimuli were presented every four seconds with sparse imaging method. In the experiment, three sessions were run for each subject. A total of 30 contiguous axial slices were acquired with a 3.0 × 3.0 × 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.

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	

Data analysis was performed using statistical parametric mapping (SPM5). Images were realigned, unwarped, spatially normalized to a standard space using a template EPI image, and smoothed using an 8 × 8 × 8 mm FWHM Gaussian kernel. The statistical analysis relied on a general linear model in which separate regressors were defined for each trial using a stick function convolved with the hemodynamic response function. We analyzed the differences of brain activities when the subject listened to an original voice with the five synthesized voices.

RESULTS

Psychophysical data

From the evaluated scores, dominant emotional words stimuli were *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).

fMRI data

Activities in cerebral cortex

The results show that each stimulus elicited activities on superior temporal gyrus, middle temporal gyrus, supramarginal gyrus, and middle frontal gyrus belonging to the auditory area. In orevious reports, these areas are assumed to process the difference of sound stimuli [6].

The difference of activity on the superior frontal gyrus, left inferior frontal gyrus, and frontomarginal gyrus included on frontal cortex was shown in S1 minus S0 (Figure 2 (b)). In S2 minus S0, superior parietal lobules related to the sensory area and parietal association area were more activated (Figure 2 (c)).

Activities in limbic system

Some researchers reported that amygdala is important for controlling emotional expression especially stronger relating to the stimulus of fear [7] [8]. However, this experimental data did not show important activation in the limbic system.

Activities in basal ganglia

In S3 minus S0, the dominant region was the caudate nucleus (Figure 2 (d) and (e)). In S4 minus S0 and S5 minus S0, the dominant region was putamen in basal ganglia (Figure 2 (f)).

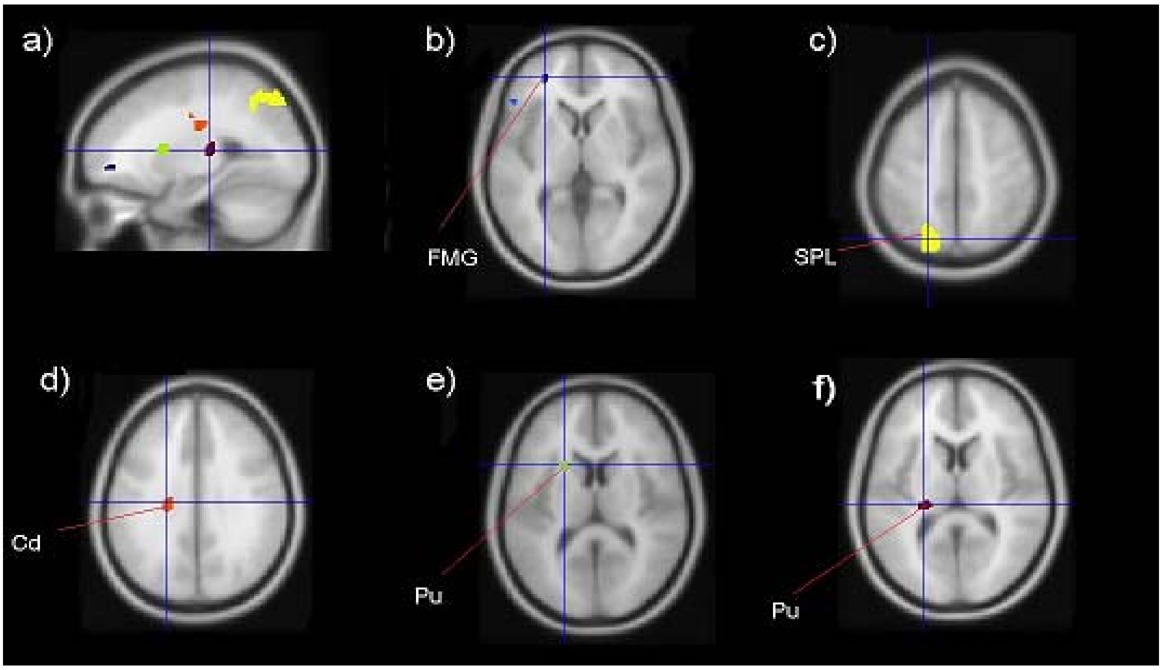

Figure 2: (a) Significant region of brain activities superimposed with (b) - (f), (b) S1 (*affirmation* and *calm*) than S0 (*affirmation* and *sympathy*) for frontomarginal gyrus (-28, 48, 0), (c) S2 (*disappointment* and *sadness*) than S0 for superior parietal lobule (-18, -78, 54), (d) S3 (*surprise* and *asking again*) than S0 for caudate nucleus (-24, -18, 28), (e) S4 (*doubt* and *negation*) than S0 for putamen (-24, 16, 10), and (f) S5 (*doubt* and *surprise*) than S0 for putamen (-24, -26, 8).

Table 2. Significant areas of activation during Perception of Emotional Voices

Brain region	Coordinates	Z value
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
S3 minus 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

DISCUSSION

Hierarchical Hypothesis of Feeling

To discuss relationships between the results of listening tests and those of brain activity measurements, we refer to the hierarchical hypothesis of feeling proposed by Fukuda [9].

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 feelings and intellectual feeling. Emotion has hierarchical structure evolution and primitive emotion that appeared by the process of the 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 evolution. Basic emotion is composed of joy, anger, fear, disgust, and acceptance or love. It is thought that social feelings were acquired by the process in which homo sapiens controls language. In many cases, sadness is classified with basic emotion (basic category) [10] [11]. However, Fukuda proposed that sadness is classified with social feelings because social feelings require more advanced processing than basic emotion and subjective experience is added. Social feelings to accomplish social intellect are thought to be acquired in the cerebral cortex and limbic system. Surprise is classified with basic emotion in some researches as well as sadness. However, this hypothesis defines surprise as a function of the attention system, such as excitation and arousalm and attention evolved as a different system from emotion.

Relationships between results of listening tests and those of brain activity measurement

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 of the listening test show *affirmation*, *sympathy*, and *calm* were perceived for S0 and S1, which were classified into social feelings. *Disappointment* and *sadness* were perceived for S2, which were classified into social feelings. *Doubt* and *negation* were perceived for S4 and S5, which were classified into social feelings. *Asking again* and *surprise* were perceived for S3, S4, and S5, which were classificatied into the attention system.

The results of brain activity measurement showed that the stimulus S1 minus original voice S0 elicited different activities on frontomarginal gyrus and superior frontal gyrus belonging to the cerebral cortex. The cerebral cortex was evaluated with processing of social feelings and intellectual feelings. These results are consistent with the hierarchical hypothesis of feeling. The stimulus S2 minus S0 elicited different activities on the superior parietal lobule and angular gyrus belonging to the cerebral cortex. These results indicate that *disappointment* and *sadness* are related in social feelings and intellectual feelings. The stimuli S3, S4, and S5 minus S0 elicited activities on mainly caudate nucleus or putamen belonging to basal ganglia. The activity on basal ganglia is thought to be processing of primitive emotion. Because the basal ganglia assumes the action adjustment system is affected by body homeostasis, emotion of attention, and tension such as *surprise*, *doubt*, *negation*, and *asking again*, it is supposed to relate these adjustments though it still needs more information of emotions.

Considering the F0 contours, the slope of F0 contours of S1 and S2 are downward, and those of S3, S4, and S5 are upward. Results show that S1 and S2 with downward slopes contained social feelings and elicited activities on portions of the cerebral cortex, and S3, S4, and S5 with upward slopes affected the attention system and elicited activities on portions of the basal ganglia. The various F0 contours affected to the different hierarchy of the brain regions.

CONCLUSION

In this paper, we investigated the relationships between results of the listening test and those of brain activity measurement.

To discuss brain activities affected by F0 contours, we used synthesized voices. From the results of listening tests and those of brain activity measurement, stimuli which the slope of F0 contours were downward contained social feelings and elicited activities on portions of the cerebral cortex that are frontomarginal gyrus or superior parietal lobules. Stimuli which the slope of F0 contours was upward were related to the attention system and elicited activities on portions of the basal ganglia that are the caudate nucleus or putamen. We could perceive the difference of emotions by processing various F0 contours over the hierarchy of brain regions.These results clearly showed the brain activities affected by the F0 contours.

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).

REFERENCES

1 Tsuyoshi Moriyama, Shinya Mori, and Shinji Ozawa, "A synthesis method of emotional speech using subspace constraints in prosody (in Japanese)," *IPSJ journal.* **50**, 1181-1191 (2009)

2 Yasuko Hayashi, "F0 contour and recognition of vocal expression of feelings (in Japanese)," *The technical report of the IEICE* .**98**, pp. 65-72 (1998)

3 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.* **39**, pp. 45–55 (2008)

4 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.* **42**, 919-927 (2008)

5 Hideki Kawahara, "STRAIGHT, exploitation of the other aspect of vocoder: Perceptually isomorphic decomposition of speech sounds," *Acoust. Sci. Tech.* **27** pp. 349-353 (2006)

6 Annett Schirmer and Sonja A. Kotz, "Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing," *Cognitive Sciences.* **10** pp. 24-30 (2006)

7 Luiz Pessoa, Sabine Kastner, and Leslie G. Ungerleider, "Attentional control of the processing of neutral and emotional stimuli," *Cognitive Brain Research.***15** pp. 31-45 (2002)

8 Mary L. Pillips, Leanne M. Williams, Maike Heining, Catherine M. Herba, Tamara Russell, Christopher Andrew, Ed T. Bullmore, and Michael J. Brammer, "Differential neural responses to overt and covert presentations of facial expressions of fear and disgust," *NeuroImage.* **21** pp. 1484-1496 (2004)

9 Masaji Fukuda, "Hierarchical hypothesis of feeling based on evolution of the brain: Positioning of social feeling within evolution," *The Japanese Journal of Research on Emotion.* **16**, pp. 25-35 (2008)

10 David Matsumoto and Paul Ekman, "American Japanese cultural differences in intensity ratings of facial expressions of emotion," *Motivation and Emotion.* **13**, pp. 143-157 (1989)

11 Carroll E. Izard, "Translating emotion theory and research into preventive interventions," *Psychological Bullerin.* **128**, pp. 796-9824 (2002)