Development of the Brain Mechanism for Understanding Speakers' Intents from Speech

Satoshi Imaizumi[1], Yuki Noguch[1]i, Midori Homma[1], Kazuko Yamasaki[1]
Masaharu Maruishi[2], Hiroyuki Muranaka[2]

[1]Prefectural University of Hiroshima, Japan
[2]Hiroshima Prefectural Rehabilitation Center, Hiroshima, Japan
imaizumi@pu-hiroshima.ac.jp

Abstract

To clarify how the brain understands the speaker's mind for verbal acts, fMRI images obtained from 24 subjects and behavioral data obtained from 339 subjects were analyzed when they judged the linguistic meanings or emotional manners of spoken phrases. The target phrases had linguistically positive or negative meanings and were uttered warmheartedly or coldheartedly by a woman speaker. The results of the fMRI analyses suggest that neural resources responsible for the speakers' mind reading are distributed over the superior temporal sulci, inferior frontal regions, medial frontal regions and posterior cerebellum. The correct judgment of the speaker intentions significantly increased with age for the phrases with inconsistent linguistic and emotional valences. Female children showed faster development than male children. The neural mechanism to interpret speaker's real intensions from spoken phrases develops slowly during the school age.

1. Introduction

Phrases with positive linguistic meanings may convey negative meanings when uttered with coldhearted emotion. For instance, "It's wonderful" uttered with evident sarcasm may inform "It's not wonderful at all." Contrary, phrases with negative linguistic meanings may convey positive meanings when uttered with warmhearted emotion. A speech act "How stupid you are" may be interpreted as "How nice you are" when uttered with warmhearted emotion with a friendly feeling. Listeners have to correctly interpret speaker's emotion and integrate it with linguistic meaning of the speech act to understand hidden but true intentions of the speaker.

Humans powerfully and flexibly interpret the speaking behavior of other people based on an understanding of their minds. The awareness that other people have beliefs and desires different from our own and that their speaking behavior can be explained by these beliefs and desires can be referred to as "theory of mind" [1-5] or "mentalizing" [5]. Most conventional theory of mind tests based on a false belief estimate that "mentalizing" ability matures by early stage of life as age 4.

The ability to understand real intent of speakers seems to play an important role to establish smooth relationships with others also in speech communication. Little has been known, however, about how linguistic and emotional information is integrated in the brain to understand speaker's intent, and how this brain mechanism develops and matures in what stage of life. Using spoken phrases with positive or negative linguistic meanings uttered warmheartedly and coldheartedly by a female, we analyzed brain activities based on a functional MRI measurement when subjects judged linguistic and emotional meanings of the phrases, and discussed possible neural mechanisms for mentalizing speaking acts. We also analyzed development of the mentalizing ability testing the young people aged 6 up to 28.

2. Method

All experiments in this study were conducted in accordance with Declaration of Human Rights, Helsinki 1975 and the research ethics regulations by the authors' affiliate. Written informed consent was obtained from each subject after explaining the purpose and the outline of the method for this research and the advantages and disadvantages expected for the subjects.

2.1. Speech materials

Two sets of speech samples were prepared: one for testing young children, and the other for testing adults. For testing adults, frequently used 40 phrases with positive linguistic meanings, such as "I love you," and 40 phrases with negative linguistic meanings, such as "I hate you," were uttered warmheartedly or coldheartedly by a female speaker of Tokyo dialect. The warmhearted utterances were made with strong pleasure, while the coldhearted ones were made with strong hatred. These utterances were classified into four types, "praise" as the linguistically and emotionally positive phrases, "sarcasm" as the linguistically positive and emotionally negative phrases, "banter" as the linguistically negative and emotionally positive phrases, and "blame" as the linguistically and emotionally negative phrases. These utterances were digitized and analyzed to extract prosodic characteristics such as F0 contour and utterance length, and then used for listening tests under functional brain imaging.

For testing young children, frequently used 20 phrases with positive linguistic meanings, such as "How nice you are!" and 20 phrases with negative linguistic meanings, such as "How bad you are!" were uttered warmheartedly or coldheartedly by a female speaker. These utterances were digitized and analyzed to extract prosodic characteristics, and then used for listening tests.

2.2. Acoustic analyses

As shown in Figure 1, fundamental frequency at various characteristic timing points and total length of utterances were measured and analyzed by ANOVA with two factors of emotion Manner (Warmhearted vs. Coldhearted) as a within-phrase measure and linguistic Meaning (Positive vs. Negative)

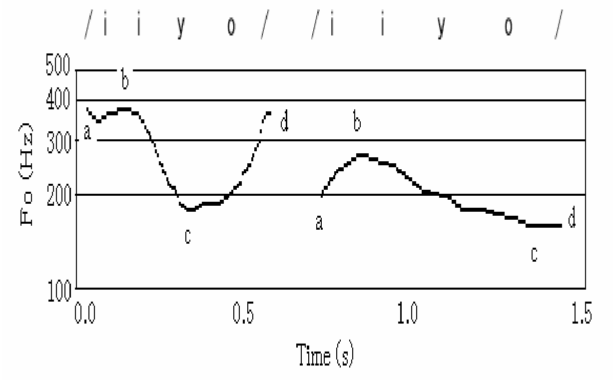

Figure 1: *F0 contour of two utterances of /iiyo/ (meaning "okay") with warmhearted emotion (left) and with coldhearted emotion (right). A: initial F0, b: maximum F0, c: minimum F0, d: final F0, F0 range= b-c, length=time difference between d and a.*

as a between-subject measure.

2.3. Listening tasks for MRI measurement

The listening subjects for the brain imaging were 24 right-handed healthy adults (12 males and 12 females aged 24.7 in average). Two listening tasks were imposed, the language task and the emotion task. For the language task, the listening subjects were instructed to judge whether the linguistic meaning of a presented speech sample is positive or negative as fast as possible. For the emotion task, they were instructed to judge whether the emotional prosody of a presented speech samples is warmhearted or coldhearted as quickly and correctly as possible.

The response time and the percent correct were analyzed by ANOVA with three factors of emotional Manner (Warmhearted vs. Coldhearted), linguistic Meaning (Positive vs. Negative), and Sex (Female vs. Male).

For the control task used in brain imaging, tones with a rising pitch or a falling pitch were prepared. The length and the presentation level of the tones were adjusted as same as those of the speech samples. The control task was to judge whether the pitch of a tone raises or falls.

Table I: *Subjects for developmental analyses*

Class	Age	Male	Female	Total	Mean Age
Low	7	14	15	29	7.7
	8	20	14	34	
	9	19	16	35	
Middle	10	19	18	37	10.5
	11	18	17	35	
	12	20	19	39	
High	13	18	18	36	13.4
	14	17	20	37	
	15	17	15	32	
Adult	20-28	10	15	25	24.1
total		173	166	339	

With the subject in the fMRI system (Siemens, Magnetic Symphony 1.5T), one of the listening tasks and the control task were performed alternately at 30-second intervals, 4 times each, and this block design, which took 4 minutes to complete, was performed under the 2 listening tasks. The speech and tone stimuli were presented at the most comfortable level of individual subjects once per 3 seconds. Pressing two buttons by the right index and middle fingers, the judgment and the response time were measured.

The scanning conditions were: The interscan interval (TR), 6 seconds; acquisition time (TA), 4.4 seconds; flip angle, 90°. The image data obtained were analyzed using a statistical parametric mapping (SPM2, the Wellcome Department of Imaging Neuroscience, http: //www.fil.ion.ucl.ac.uk /spm/spm2.html)[6] .

2.4. Developmental analyses

In order to develop a screening test for early detection of communication problems in children, the mind-reading ability from speech was measured for 339 listening subjects (173 male and 166 female) using spoken phrases directed to children based on the two tasks, the language task and the emotion task. Table I shows the age and number of subjects for the developmental analyses.

3. RESULTS

3.1. Developmental characteristics

As shown in Figure 2 for the emotion task, the correctness score significantly increased with age for "sarcasm" and "banter" phrases which have inconsistent linguistic and emotional valences, although the score of junior high school students did not reach the level of adult subjects. Male subjects showed significantly lower correctness score than female subjects at age of Middle and High. The sex difference, however, was not significant for the adult subjects.

3.2. Brain imaging results

The male subjects showed significant activation in more cortical areas than the female subjects for the both tasks. For the male subjects, significant activation ($P_{PWE\text{-}corr}$<0.05) was found in the bilateral middle temporal gyri including the superior temporal sulci, the bilateral superior frontal gyri, the left posterior cerebellar lobe and the left inferior frontal gyrus. On the other hand, for the female subjects, the left posterior cerebellar lobe was the only area, the activity of which was significant at $P_{PWE\text{-}corr}$<0.05 corrected for multiple comparison.

The activated cortical areas were similar between the male and female subjects for the language task when compared to the control task. There was no cortical area activated significantly when comparison was made between the male and female subjects for the language task. When compared to the female subjects, the male subjects showed significantly stronger activation in only one cortical area in the right frontomedian cortex ($P_{PWE\text{-}corr}$<0.05).

3.3. Acoustic characteristics of speech materials

Acoustic characteristics varied depending on Manner and Meaning. As illustrated in Figure 3, ANOVA on F0 range revealed that the interaction effect of Manner and Meaning

(F(1,160)=4.97, p=0.027), the main effects of Manner (F(1,160)=63.21, p<0.0001) and Meaning (F(1,160)=9.98, p<0.005) were significant. The post-hoc test revealed that the warmhearted utterances had 104Hz wider F0 range than the coldhearted utterances (p<0.0001), and the linguistically negative phrases had 41Hz wider F0 range than the linguistically positive phrases (p=0.02). The warmhearted utterances of negative phrases had 70Hz wider F0 range than the warmhearted utterances of positive phrases (p=0.004), while coldhearted utterances had no significant differences between linguistically positive and negative phrases (p=0.27).

The interaction effect of Manner and Meaning was also observed for the Maximum F0 and the utterance length.

4. Discussion

The correctness score, as shown in Figure 2, significantly increased with age for sarcasm and banter phrases which have inconsistent linguistic and emotional valences, and the score of junior high school students did not reach at the level of adult subjects. These results suggest that the ability to understand speaker's hidden but true intensions by separating and integrating linguistic and emotional valences develop slower than the rate estimated by the conventional theory of mind tests based on a false belief. This also suggests that the emotion task developed here is useful to measure the speech communication ability, or the ability to understand speakers' mind for verbal acts.

The present results suggest that neural resources responsible for speakers' mind reading are distributed over STS, inferior frontal regions, medial frontal regions and the posterior cerebellum. Although several previous studies have suggested [7, 8], neither the right-hemisphere predominance in the emotion task nor the left-hemisphere predominance in the language task was observed in this study. One possible interpretation of the present results is that the linguistic as well as emotional processes may automatically be induced as cooperative and interactive processes, and the differences may exist only at a high processing stage to integrate the linguistic and emotional processes to generate a coherent interpretation of the speaker's mind.

The above-mentioned slow maturation of the ability to understand speakers' mind may indicates that it is a neural system to develop during the school age, which accomplish higher information processing than emotional prosody processing and than linguistic semantic processing. It may be an integration process for various conflicting information to generate a coherent interpretation of the mental states of interlocutors including speakers and their owns as listeners. One possible candidate of such neural system is the MPFC [11], which may play a general role of inferring interlocutors' mental states, and such a function of MPFC may support the social skill for smooth speech communication.

As shown in Figure 3, significant interaction effects of language and emotion were observed on the acoustic characteristics of utterances, such as F0 range, and also on the perceptual behavior evaluated by response time and judgment correctness. These results obtained in our experiment suggest that emotion modulates linguistic processes not only in speech production but also in speech perception, and such modulations may differ between the sexes particularly in perceptual processes [9-11]. These results support, in part, previous studies suggesting that emotional prosody modulates perceptual word processing and that the time-course of this modulation differs for males and females, that is, women make an earlier use of emotional prosody during word processing as compared to men [11-15].

5. Conclusion

The present results suggest that neural resources responsible for speakers' mind reading are distributed over STS, inferior frontal regions, medial frontal regions and the posterior cerebellum. The neural processes for language and emotional prosody significantly interact to mentalize verbal acts. The neural mechanism to interpret speaker's real intensions from spoken phrases with inconsistent linguistic and emotional valences develops rather slowly over the junior high school age, although females develop faster than males.

6. References

[1] Baron-Cohen, S.; Leslie, A. M.; & Frith, U., 1985. Does the autistic child have a "theory of mind"? *Cognition, 21 (1)*, 37-46

[2] Leslie, K. R; Johnson-Frey, S. H.; Grafton, S. T., 2004. Functional imaging of face and hand imitation: towards a motor theory of empathy. *Neuroimage, 21(2)*, 601-607.

[3] Siegal, M.; Carrington, J.; Radel, M., 1996. Theory of mind and pragmatic understanding following right hemisphere damage. *Brain Lang* 53(1), 40-50.

[4] Ferstl, E.C.; von Cramon, D.Y., 2002. What does the frontomedian cortex contribute to language processing: coherence or theory of mind? *Neuroimage* 17(3), 1599-612.

[5] Frith, C. D.; Frith U., 1999. Interacting Minds − A Biological Basis. Science 286, 1692-1695.

[6] Statistical Parametric Mapping: SPM2, 2002. Developed by members & collaborators of the Wellcome Department of Imaging Neuroscience http: //www.fil.ion.ucl.ac.uk/ spm/ spm2.html.

[7] Imaizumi, S., Mori, K., et al., 1997. Vocal identification of speaker and emotion activates different brain regions. *NeuroReport* 8(5), 2809-2812

[8] Imaizumi, S.; Mori, K., Kiritani; S., Hosoi, H.; Tonoike, M., 1998. Task-dependent laterality for cue decoding during spoken language processing. *NeuroReport* 9 (5), 899-903.

[9] Imaizumi, S.; Homma, M.; Ozawa, Y.; Maruishi, M.; and Muranaka, H., 2004. Gender differences in the functional organization of the brain for emotional prosody processing, Speech Prosody 2004, 605-608.

[10] Imaizumi, S.; Homma, M.; Ozawa, Y.; Maruishi, M.; Muranaka, H., 2004. Gender differences in emotional prosody processing -An fMRI study- *Psychologia* 47(2), 113-124.

[11] Homma, M.; Imaizumi, S.; Y., Maruishi, M.; and Muranaka, H., 2006. The neural mechanisms for understanding self and speaker's mind from emotional speech: an event-related fMRI study, Speech Prosody 2006.

[12] Frost, J.A.; Binder, J.R.; Springer, J.A.; Hammeke, T.A.; Koelsch, S.; Maess, B.; Grossmann, T.; Friederici, A.D., 2003. Electric brain responses reveal gender differences in music processing. *Neuroreport* 14(5), 709-713.

[13] Kotz, S.A.; Meyer, M.; Alter, K.; Besson, M.; von Cramon, D.Y.; Friederici, A.D., 2003. On the

lateralization of emotional prosody: an event-related functional MR investigation. *Brain Lang* 86(3), 366-376.

[14] Morris, J.S.; Scotto, S.K.; Dolan, R.J., 1999. Saying it with feeling: neural responses to emotional vocalizations. *Neuropsychologia* 37, 1155-1163.

[15] Schirmer, A.; Kotz, S.A.; Friederici, A.D., 2002. Sex differentiates the role of emotional prosody during word processing. *Brain Res Cogn Brain Res* 14(2), 228-233.

[16] Schirmer, A.; Kotz, S.A., 2003. ERP evidence for a sex-specific stroop effect in emotional speech. *J. Cognitive Neuroscience* 15(8), 1135-1148.

Acknowledgments

We would like to express our sincere gratitude to all the subjects participated in this study. Without their contribution, this study could not be carried out.

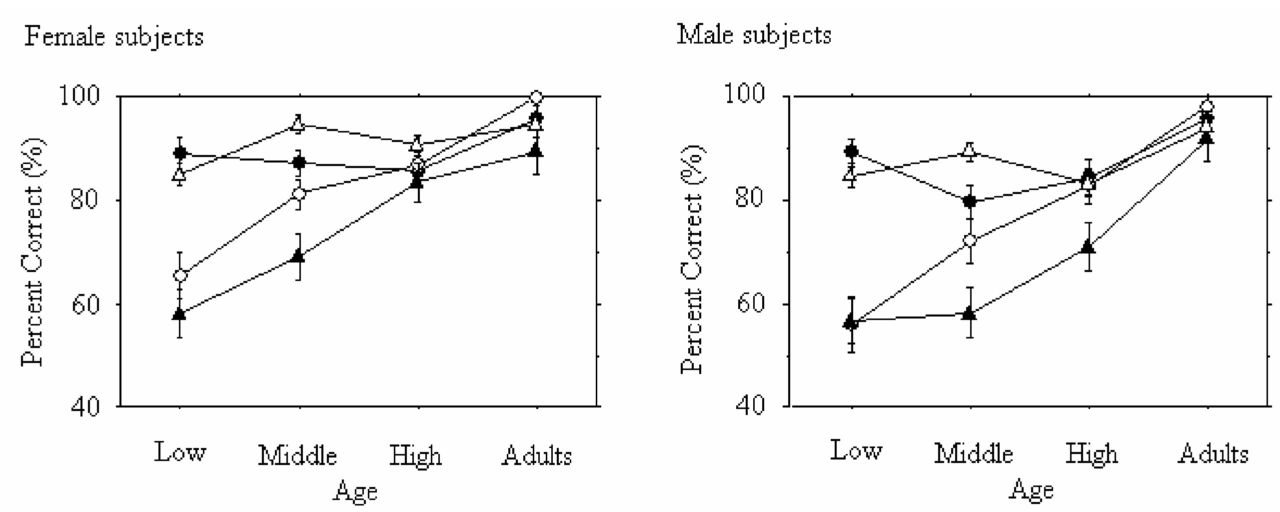

Figure 2: Development of the mind-reading ability. ●:praise(language +, emotion +), ○:sarcasm (language +, emotion -), ▲banter (language -, emotion +),□ blame (language -, emotion -).

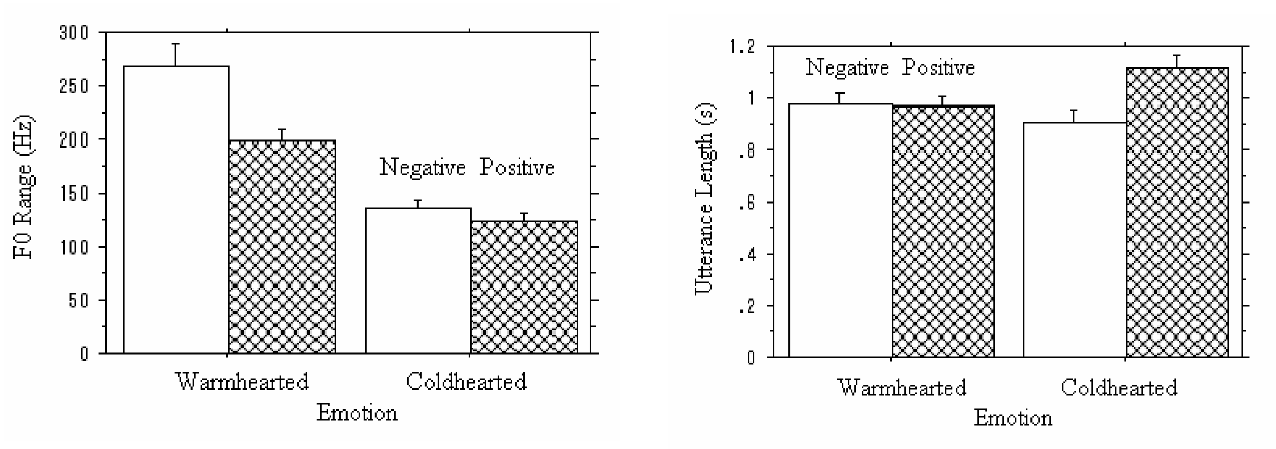

Figure 3: (a, left) *F0 range with error bars*. (b, right) *Utterance length with error bars*. Spoken phrases directed to adults. Empty bar: Negative linguistic meanings, Meshed bar: Positive linguistic meanings.