Acoustics'08 Paris
June 29-July 4, 2008
www.acoustics08-paris.org

Neural mechanism for understanding speakers' intentions: developmental analyses of children with and without communication disorders

Satoshi Imaizumi

The Prefectural Univ. of Hiroshima, 1-1 Gakuen-machi, 723-0053 Mihara, Japan
imaizumi@pu-hiroshima.ac.jp

Ability to understand speakers' intentions through linguistic contents and affective prosody is examined for children between 5 and 12 years-old with and without communication disorders. Four types of spoken short phrases, expressing admiration, sarcasm, blame and humor or joke, were presented. For each stimulus, the subjects were asked to choose between two cards, one with a written word "praise" and another with "no praise" accompanied by corresponding drawings. For children without any communication disorders, the percentage of the correct judgment of speaker intentions was high and stable for admiration and blame phrases which have congruent linguistic and affective valences. It was significantly low for 6-year-old children and increased with age for the sarcastic or joking phrases which have incongruent linguistic and affective valences. The percent correct was significantly lower for autistic children than normally developing children particularly for the phrases with incongruent valences. Although a significant difference was found between children with severe- and mild-hearing impairment, no significant difference was found in the percent correct between the congruent and incongruent phrases for them. Based on these results together with brain activation analyses using fMRI, a model of neural mechanism for understanding speakers' intentions is discussed.

1 Introduction

Phrases with positive linguistic meanings may convey negative meanings when uttered with coldhearted emotion. For instance, "It's wonderful" uttered with evident sarcastic prosody 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 a warmhearted prosody with a friendly feeling. Listeners have to correctly interpret emotional or affective prosody and integrate it with literal meaning of the speech act to understand hidden but true intentions of the speaker.

The ability to understand speaker's intention seems to play an important role to establish smooth relationships with others in speech communication. Little has been known, however, about how linguistic and emotional/affective information is integrated in the brain to understand speaker's intent, and how this brain mechanism develops and matures in what stage of life.

Affective or emotional prosody of speech is well known to contribute to convey speaker's intention. It is also well known that infants and very young children are sensitive to prosodic variations in speech [1]. Newborns rely on prosodic cues to recognize their mother's voice [2]. They can discriminate between the rising and falling intonation contours of Japanese words [3].

Although the high sensitivity of infants to speech prosody has been widely acknowledged, it is also well known that judgments requiring an understanding of the speaker's intention from linguistic stimuli remain difficult for children under 6 [4-9].

In order to resolve such controversy, we also analyzed development of the mentalizing ability testing the young children with normal development and those with communication disorders. We also analyzed brain activities based on functional MRI measurements during when subjects judged linguistic and emotional meanings of the phrases, and discussed possible neural mechanisms for mentalizing speaking acts.

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 like you," and 40 phrases with negative linguistic meanings, such as "I hate you," were uttered warmheartedly or coldheartedly by four speakers 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, "admiration/praise" as the linguistically and emotionally positive phrases, "sarcasm" as the linguistically positive but emotionally negative phrases, "banter/joke" as the linguistically negative but 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) as a between-subject measure.

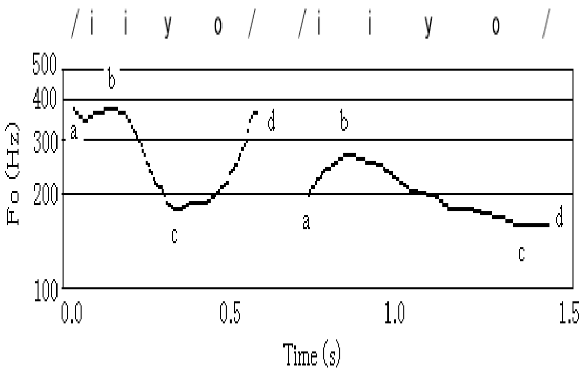

Fig. 1 F_0 contour of two utterances /iiyo/ (meaning "okay") with warmhearted emotion (left) and with coldhearted emotion (right). A: initial F_0, b: maximum F_0, c: minimum F_0, d: final F_0, F_0 range= b-c, length=time difference between d and a.

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 and correctly 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 affective Manner (Warmhearted vs. Coldhearted), literal 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.

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) [10].

2.4 Developmental analyses

In order to develop a screening test for early detection of communication problems in children, the intention-reading ability from speech was measured for 446 listening subjects using spoken phrases directed to children based on the two simplified tasks, I and II. Task I was to judge if the speaker praises you or not. Task II was to judge if the speaker blames you or not.

We also tested 20 with autism (ASD), 11 with attention deficit/hyperactivity disorder (ADHD), 24 with hearing impaired (HI), and 24 normally developing children (NDC) using a computerized screening system. The average age was 9.9, 10.5 and 9.5 without any significant differences between the groups.

3 Results

3.1 Developmental characteristics

As shown in Figure 2, the correctness score obtained from the 446 normally developing children significantly increased with age for "sarcasm" and "banter" phrases which have incongruent linguistic and emotional valences.
The correctness score of the autistic children (ASD) was significantly lower than the normally developing children (NDC) and ADHD groups for the phrases with incongruent valences. Although a significant difference was found between children with severe- and mild-hearing impairment, no significant difference was found in the percent correct between the congruent and incongruent phrases for them.

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 (PPWE-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 PPWE-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 (PPWE-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.

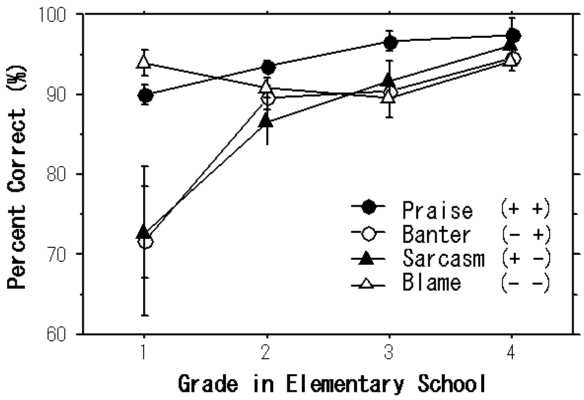

Fig. 2 Development of the intention-reading ability.

4 Discussion

The intention-reading ability measured as the correctness score, as shown in Figure 2, is high even in 6-years-old children for the congruent phrases which have congruent linguistic and emotional valences. For the incongruent phrases, such as sarcasm and banter, which have incongruent linguistic and emotional valences, it is significantly low for the 6-years-old children compared to the older children, and it increases with age. From the age-matched comparison, it is revealed that the intention-reading ability is significantly low in the autistic children for the incongruent phrases.

This result suggests that it is easy for children to understand speaker's intention from congruent phrases which are uttered with prosody congruent to the literal meanings, and is not for incongruent phrases. If a spoken phrase has prosody which is incongruent with its literal meaning, listeners must integrate incongruent properties to generate a coherent interpretation of speaker's intention. This is not easy and takes years for children to mature, particularly for phrases depend on the mental state of speaker. Intentions expressed in congruent phrases such as praise or blame refer basic emotions, while those expressed in congruent phrases such as sarcasm or banter are more complex and might be higher than the basic emotions, and are harder to understand for young children. Children have to understand speakers have their own intentions, which can be expressed through affective prosody of spoken phrases, which may differ from their literal meanings. To understand speakers' intention properly, children must acquire "mentalizing" ability, by which they understand the mental state of speakers which are not necessarily simple and are not necessarily expressed in a simple straightforward manner. The autistic children seem to have difficulty in "mentalizing" such complex incongruent expressions.

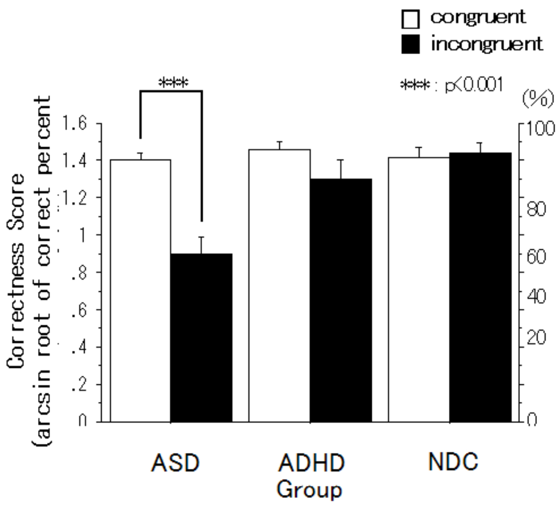

Fig. 3 The correctness score of ASD, ADHD, and NDC groups for congruent and incongruent phrases.

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" [11-14] or "mentalizing"[15]. Most conventional theory of mind tests based on a false belief estimate that "mentalizing" ability starts to mature by early stage of life as age 5. The present results suggest that children around this age start to mature to understand speaker's intentions even from incongruent expressions such as sarcasm and banter. Although the high sensitivity of infants to speech prosody is a necessary condition to acquire speech communication ability, it is not enough to understand speakers' intentions which are not necessarily expressed in a simple straightforward manner.

The present results suggest that neural resources responsible for speakers' intention reading are distributed over STS, inferior frontal regions, medial frontal regions and the posterior cerebellum. Although several previous studies have suggested [16-19], 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 early school age, which

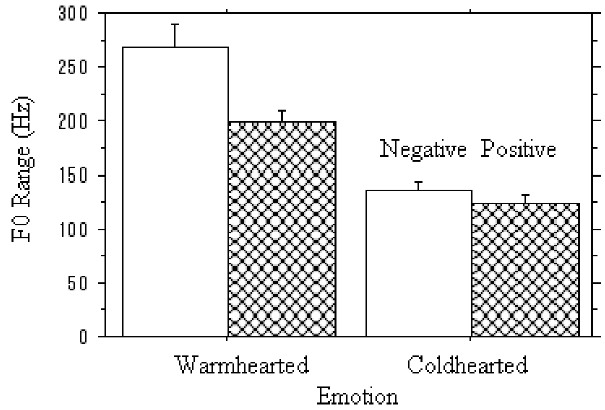

Fig. 4 F_0 range with error bars.
Spoken phrases directed to adults. Empty bar: Negative linguistic meanings, Meshed bar: Positive linguistic meanings.

accomplishes 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 [20], 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 [7, 8, 21-23]. 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 [7, 8, 21-23].

5 Conclusion

The ability to understand speakers' intentions from speech was significantly low for 6-year-old children than elder children, and increased with age for the sarcastic or joking phrases which have incongruent linguistic and affective valences. The percent correct was significantly lower for autistic children than normally developing children particularly for the phrases with incongruent valences. The neural resources responsible for understanding speakers' intention are distributed over STS, inferior frontal regions, medial frontal regions and the posterior cerebellum, which have been reported as neural resources contributing for theory of mind or mentalizing the minds of others. The neural processes for language and emotional prosody significantly interact to mentalize verbal acts. The neural mechanism for interpreting speaker's intensions from speech continues to mature even during the early school age.

Acknowledgments

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

References

[1] A. Fernald, "Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages," *Child Development* 64, 657-674 (1993).

[2] A. J. DeCasper, W. P. Fifer, "On human bonding: Newborns prefer their mother's voice," *Science* 208, 1174-1176 (1980).

[3] T. Nazzi, C. Floccia, J. Bertoncini, "Discrimination of pitch contour by neonates," *Infant Behavior and Development* 21, 779-784 (1998).

[4] B. P. Ackerman, "Contextual integration and utterance interpretation: The ability of children and adults to interpret sarcastic utterances," *Child Development* 53, 1075-1083 (1982).

[5] B. P. Ackerman, "Form and function in children's understanding of ironic utterances," *Journal of Experimental Child Psychology* 35, 487-508 (1983).

[6] B. P. Ackerman, "Children's sensitivity to comprehension failure in interpreting a nonliteral use of an utterance," *Child Development* 57, 485-497 (1986).

[7] S. Imaizumi, M. Homma, Y. Ozawa, M. Maruishi, H. Muranaka, "Gender differences in the functional organization of the brain for emotional prosody processing," *Speech Prosody* 2004, 605-608 (2004).

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

[9] V. Laval, A. Bert-Erboul, "French-speaking children's understanding of sarcasm: the role of intonation and context," *J. Speech Lang. Hear. Res.* 48(3), 610-20 (2005).

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

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

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

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

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

[15] C. D. Frith, U. Frith, "Interacting minds – A biological basis," *Science* 286, 1692-1695 (1999).

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

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

[18] S. A. Kotz, M. Meyer, K. Alter, K.M. Besson, D. Y. von Cramon, A. D. Friederici, "On the lateralization of emotional prosody: an event-related functional MR investigation," *Brain Lang.* 86(3), 366-376 (2003).

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

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

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

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

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