Listener–adaptive Characteristics in Dialogue: Effects of Temporal Adjustments on Emotional Aspects of Speech

S. Imaizumi*, A. Hayashi** and T. Deguchi**
*Institute of Logopedics and Phoniatrics, University of Tokyo
7-3-1, Hongo, Bunkyo-ku, Tokyo, 113 Japan
**Faculty of Education, Tokyo Gakugei University
4-1-1, Nukuikitamachi, Koganei-shi, Tokyo, 184 Japan

Introduction

Interlocutors seem to adapt their speaking style to various aspects of situations, particularly to their dialogue partners. Our previous analyses of dialogue between professional teachers and normal-hearing (NH) or hearing-impaired (HI) children found that teachers tended to use simpler and shorter sentences for HI children than toward NH ones, which subsequently induces more responses (Imaizumi *et al.*, 1993a, 1993b). They also inserted longer pauses at phonological phrase boundaries, stretched syllable duration, and reduced their speaking rate.

Comparing "clearly-read speech" intended for the hearing-impaired to conversational speech, Picheny *et al.* (1985, 1986) concluded the following: (1) The speaking rate decreases and the intelligibility substantially increases; (2) vowels in the clearly-read speech are reduced to a lesser extent in comparison with conversational speech; (3) stop bursts, as well as all word-final consonants, are released more frequently in clearly-read speech than in conversational speech; and (4) the root mean square (RMS) intensities for obstruent sounds, particularly consonants, are greater in clearly-read speech.

These reports suggest that interlocutors adapt their speaking style to their listeners' hearing/speech capacities in order to help them understand dialogue.

This paper reports the effects of listener–adaptive adjustments of the temporal structure of dialogue on the perceived emotional aspects of speech. The emotional aspects of dialogue seem to be significantly affected by temporal structureand may be very dependent on speaker-listener interactions, although such aspects of dialogue have not been intensively studied so far.

Method

Dialogue was recorded during a simple picture-searching game through which a teacher attempted to assess the speech communication ability of a HI or NH child. The game was played as follows.

Two different panels were prepared, A or B, with each displaying 11 pictures (illustrations) of boys/girls labeled with their names. The A panel was set in front of the child and a copy of it in front of the teacher. The teacher instructed the child to point to a picture as fast as possible after a name was called out. The teacher randomly called out the all names one by one.

Before the game, each teacher was instructed to explain the game using his/her own words and sentences. The only fixed question was "Donokoga /CVCVCV/ desuka?" (Which child is /CVCVCV/ ?), where /CVCVCV/ represents the name in a picture. If the teacher mistakenly used a different form of question, the sample was not used.

Digital audio tape and video recordings of 14 teacher-child pairs were carried out in a sound-proof testing room. The teacher initially explained the task to each child and began the game using Panel A (Game A), and then after a short pause used Panel B (Game B). One microphone was placed about 20 cm away from the teacher's mouth and another about 20 cm away from the child's mouth.

To clarify the differences between dialogue and read speech, a read passage was also recorded and analyzed. Seven teachers read the target sentences "Donokoga /CVCVCV/ desuka?" five times as fast and as clearly as possible. Recording was carried out using the same method as for the dialogue. The abbreviation RD is used to represent the read speech tokens.

The material used has been described in other place (Imaizumi *et al.*, 1993c, 1994b, 1994c).

Seven professional teachers, seven hearing-impaired children (HI) and seven normal-hearing children (NH) participated in the test. They were speakers of the standard Tokyo dialect of Japanese.

Using an object-oriented acoustic analysis system (Imaizumi *et al.*, 1993c, 1994a), the temporal structure of the dialogue, turn taking, and the temporal characteristics of the speech waveform were analyzed.

The speech directed to the normal-hearing (NH) and the hearing-impaired (HI) children will be referred to as the NH and the HI tokens, respectively, in the remainder of this paper.

In order to clarify the perceptual effects of temporal adjustments in dialogue, the perceptual characteristics of the RD, NH and HI tokens were analyzed using the semantic differential method. As shown in Fig. 1, 24 pairs of adjectives were used as 9-point dipole rating scales. The listening subjects were 8 normal hearing students. The tokens used were 21 samples of "Donokoga hikita desuka? " spoken by the 7 teachers in the three modes RD, NH and HI. A total of 168 x 24 rating scores was analyzed by a principal factor analysis, and then a regression analysis was carried out to extract any significant correlations with the temporal structure of the speech.

Results

Perceptual differences

In order to test the significance of the effects of mode (RD, NH and HI) and teacher (A, B,..., F), an analysis of variance (ANOVA) was carried out for the rating scores on the 24 dipole scales. For most of the scales used, the main effects of mode, teacher and their interaction were significant at the 1% level. Fisher's PLSD test revealed that the differences between RD and NH or HI (read tokens versus dialogue speech) were significant for all the rating scales, but that the differences between NH and HI were significant only for some of the rating scales.

A principal factor analysis was used to extract a few principal factors which could account for the differences in a compact way. Four extracted factors had the accounting–for-rates of 72.4, 5.2, 3.7, and 2.4%, and these factors accounted for 83.7% of the total variance in the rating scores.

Figures 1 (a) and (b) show the factor loading of the four extracted factors after Varimax orthogonal rotation. Factor 1 (F1) represents the emotional contrast between discomfort ("Awful, Rough, Uneasy, Busy, etc.") and pleasant ("Easy, Kind, Friendly, Restful, Polite, etc."). Factor 2 represents the contrast between "Strong" and "Dull, Lifeless." Factor 3 represents the contrast between "Slow, Stiff, Unnatural, Intelligible, Strong" and "Busy, Lifeless, Tense, Rough, Dull" Factor 4 represents the contrast between "Lifeless, Deep" and "Stiff, Kind, Unnatural."

The analysis of variance (ANOVA) for the rotated factors revealed the following.

(1) For Factor 1, mode (p<0.0001) and teacher (p=0.0029) and their interaction (p<0.0001) had significant effects. Factor 1 mainly represents the differences in the emotion scores between the read tokens (RD) and the dialogue tokens (HI and NH). It was found that the read tokens (RD) except those of Teacher E tended to be perceived as "Awful, Rough, Uneasy and Busy." On the other hand, the dialogue tokens (HI and NH) were perceived as "Easy, Kind, Friendly, Restful, Polite." The HI tokens had values intermediate between the RD and NH tokens, although there were some differences depending on the teachers.

(2) For Factor 2, only teacher had a significant main effect (p=0.0048). This was also true for Factor 4, for which only teacher had a small but a significant main effect (p=0.0149). These two factors represent the speaker-dependent differences in the emotional rating scores.

(3) For Factor 3, mode (p<0.0001) and teacher (p<0.0027) had significant effects, but their interaction (p=0.1562) had no significant effect. Fisher's PLSD test revealed that the differences between HI and the other modes (RD or NH) were significant (p<0.0001), but that the difference between RD and NH was not significant. This factor represents the differences between the HI tokens and the other modes (RD and NH). The contrast can be seen in "Slow, Stiff, Unnatural, Intelligible, Strong" and "Busy, Lifeless, Tense, Rough, Dull."

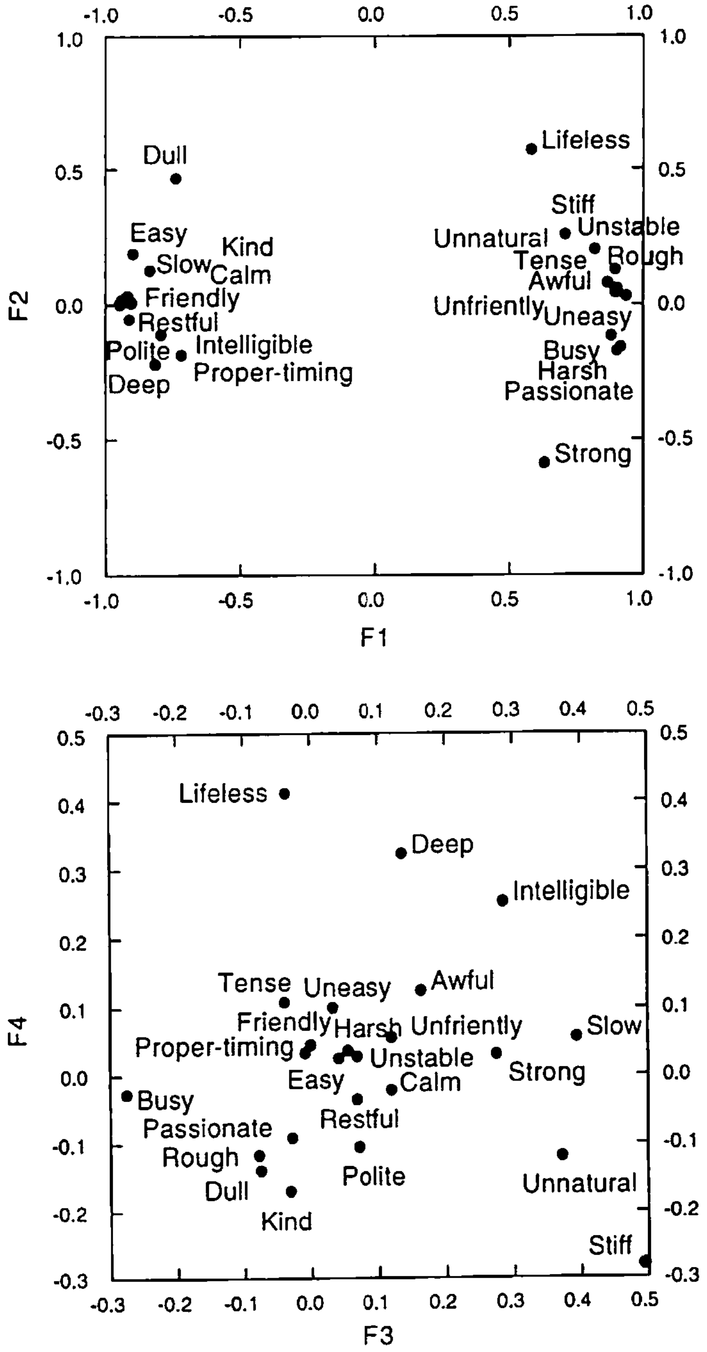

Fig. 1 (a) above and (b) below. The factor loading for the four extracted factors after the Varimax orthogonal rotation.

Effects of Temporal Structure

In order to determine the effects of the temporal structure of the tokens on the emotion ratings, various parts of the tokens, whole sentences, words, moras, pauses, voiced/unvoiced/silent segments were measured, and regression analyses were carried out to obtaine the coefficients of determination (the squared correlation coefficients). The results can be summarized as follows.

(1) The total length of the tokens accounted for 55% of the total variance represented by Factor 1. The total length of the tokens was very important in explaining the differences between RD and the other modes (NH and HI) which were represented by Factor 1. The total length of the tokens accounted for 42% of the variance in Factor 3 which represents the differences between HI and the other modes.

(2) The length of the target names accounted for 53% of the total variance in Factor 3. This measure was very important in explaining the differences between HI and the other modes (RD and NH). This measure also accounted for 32% of the variance in Factor 1 which represents the differences between RD and the other modes.

(3) The length of the final part of the target sentences, /desuka/ in /donokoga CVCVCV desuka?/, played a very important role in accounting for Factor 1 which represents the emotional contrast between the RD tokens and the other modes (NH and HI). As mentioned above, Factor 1 (F1) represents the emotional contrast between discomfort ("Awful, Rough, Uneasy, Busy, etc.") and pleasant ("Easy, Kind, Friendly, Restful, Polite, etc.").

Discussion

Our previous study (Imaizumi *et al.*, 1993c, 1994b) revealed the following results. 1) Teachers significantly reduced their devoicing rate in HI tokens comparing to RD and NH tokens. 2) Teachers significantly lengthened the voiced and unvoiced segment lengths in HI tokens compared to RD and NH tokens. 3) Mora length accounted for 53% of the total variance in the devoicing rate. 4) When the effect of mora length on the devoicing rate was excluded, the HI tokens had a 10% lower residual devoicing rate than the NH, which was statistically significant.

Based on these results, we previously concluded that teachers reduce their devoicing by not only lengthening the intervals for successive voicing and devoicing gestures but also by resizing component gestures to some extent, when talking to the hearing-impaired (Imaizumi *et al.*, 1994b).

The present study revealed that the emotional profiles of tokens could represented by four factors F1, F2, F3 and F4. F1 represents the emotional contrast between discomfort ("Awful, Rough, Uneasy, Busy, etc.") and pleasant ("Easy, Kind, Friendly, Restful, Polite, etc."), and it also represented the emotional difference between the RD and the other modes (NH and HI). F3 represents the emotional contrast between "Slow, Stiff, Unnatural, Intelligible, Strong" and "Busy, Lifeless, Tense, Rough, Dull," and F3 also represents the differences between HI and the other modes (RD and NH). F2 and F4 could be interpreted as representing speaker-dependent differences among the teachers.

The present study also revealed that listener-adaptive temporal adjustments of dialogue significantly affect the emotional profiles of speech as perceived by listeners. For instance, the total length of the tokens accounted for 55% of the total variance in F1, which represents the emotional difference between RD and the other modes (NH and HI). It also accounted for 42% of the variance in F3 which represents the emotional differences between HI and the other modes. The length of the target names accounted for 53% of the total variance in F3 which represents the emotional differences between HI and the other modes (RD and NH). The length of the final part of target sentences, /desuka/ in /donokoga CVCVCV desuka?/, plays a very important role in accounting for F1 which represents the emotional contrast between the RD tokens and the other modes (NH and HI).

Summarizing these results, we may conclude that listener-adaptive temporal adjustments of dialogue affects not only the segmental characteristics of speech, such as devoicing and vowel undershoot, but also prosody-related characteristics of speech, such as the emotional profiles reported on here.

Our previous study (Imaizumi *et al.*, 1993 a, 1993b) revealed that teachers tend to use simpler and shorter sentences, inducing more responses, toward the hearing impaired than toward the normal-hearing. These results indicate that a listener-oriented adaptation of speaking style seems to affect various stages, including grammatical encoding, prosody forming and segmental structuring processes.

Conclusions

The effects of listener-adaptive temporal adjustments in dialogue on the perceived emotional characteristics of speech were investigated by acoustically and perceptually analyzing the speech of teachers directed to hearing-impaired (HI) or normal-hearing (NH) children. Read tokens (RD) were also analyzed for comparison. Acoustical analyses showed the following. 1) The emotional profiles of dialogue could be represented by four factors F1, F2, F3 and F4. 2) F1 represented the emotional differences between the read tokens (RD) and the dialogue speech (NH and HI), which were mainly a contrast between discomfort versus pleasant. 3) F3 represents an difference between HI and the other modes (RD and NH), which could be interpreted as the emotional contrast between "Slow, Stiff, Unnatural, Intelligible, Strong" and "Busy, Lifeless, Tense, Rough, Dull". 4) F2 and F4 could be interpreted as representing speaker-dependent differences among the teachers. 5) The temporal structures of the tokens significantly accounted for the above-mentioned emotional profiles.

These results suggest that the listener-adaptive temporal adjustment of dialogue affects not only the segmental characteristics of speech, such as devoicing, but also prosody-related characteristics of speech such as emotional profiles.

Acknowledgments

We are sincerely grateful to all of the teachers and children at Ohji Elementary School, Tokyo. This research was partially supported by a Grant-in-Aid for Scientific Research for Priority Areas on "Spoken Dialogue," Ministry of Education, Science and Culture, Japan.

References

Imaizumi, S, Hayashi, A. and Deguchi, T. (1993a). "Planning in speech production: Listener adaptive characteristics," Jpn J. Logopedics and Phoniatrics 34, 394-401 (in Japanese).

Imaizumi, S, Hayashi, A. and Deguchi, T. (1993b). "Listener adaptive characteristics in dialogue speech," in Proceedings of the International Symposium on Spoken Dialogue, edited by K. Shirai, T. Kobayashi, and Y. Harada, (Waseda University Printing, Tokyo, Japan), 279-282.

Imaizumi, S, Hamaguchi, S. and Deguchi, T. (1993c). "Vowel devoicing in teachers' speech directed to the hearing-impaired / normal-hearing children. -Do teachers avoid devoicing to help hearing-impaired understand dialogue?-." Comminication Disorder Research, 22, 7-20 (in Japanese).

Imaizumi, S., Hartono, A., Niimi, S., Hirose, H., Saida, H., and Shimura, Y. (1994a). "Evaluation of vocal controllability by an object oriented acoustic analysis system," J. Acoust. Soc. Jpn (E) 15, 113-116.

Imaizumi, S, Hamaguchi, S. and Deguchi, T. (1994b). "Vowel devoicing in Japanese dialogue between teachers and hearing-impaired or normal-hearing children: Listener adaptive characteristics of dialogue speech poduction," J. Acoust. Soc. Am., 95(5), 3012.

Imaizumi, S., Hayashi, A., and Deguchi, T. (1994c), "Listener adaptive characteristics in dialogue speech -Effects of temporal adjustment on emotional aspects of speech-," in *Proceedings of ICSLP 94* (Yokohama, Japan, Sept. 1994), Acoust. Soc. Jpn., 4, 1967–1970.

Picheny, M. A, Durlach, N. I., and Braida, L. (1985). "Speaking clearly for the hard of hearing: Intelligibility differences between clear and conversational speech," J. Speech and Hearing Res. 28, 96-103.

Picheny, M. A., Durlach, N. I., and Braida. L. (1986). "Speaking clearly for the hard of hearing: Acoustic characteristics of clear and conversational speech," J. Speech and Hearing Res. 29, 434-446.

Picheny, M. A., Durlach, N. I., and Braida, L. (1989). "Speaking clearly for the hard of hearing: An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech," J. Speech and Hearing Res. 32, 600-603.