Listener–Adaptive Adjustments in Speech Production Evidence from Vowel Devoicing

S. Imaizumi* and A. Hayashi**
***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

Listener-oriented adaptation of speaking style appears to affect various stages in speech production processes. Our previous analyses (Imaizumi et al.. 1993a, b) of dialogues between professional teachers and normal-hearing (NH) or hearing-impaired (HI) children found that the teachers tended to use simpler and shorter sentences for the HI children than for the NH ones. They also reduced their speaking rates by inserting longer pauses at phonological phrase boundaries and producing longer syllable durations. The teachers also reduced their vowel devoicing, probably to improve listener comprehension (Imaizumi et al., 1994b, c, 1995).

The purpose of the present paper is to elucidate acoustical and perceptual effects of listener dependent adjustmenting speaking style by analyzing vowel devoicing in dialogue between teachers and NH or HI children. Vowel devoicing has been analyzed from various points of view through physiological (Hirose et al., 1970; Hirose, 1971; Sawashima, 1969, 1971; Yoshioka, 1981; Yoshioka et al., 1982; Sugito and Hirose, 1988), acoustical (Beckman, 1081; Han, 1962a, b; Kuriyagawa and Sawashima, 1989), phonological (Shibata, 1955) and phonetic (Jun and Beckman, 1993; Kondo, 1994) analyses. Its situation-dependentness, however, has not been clarified enough. The main focus of the present study is to model vowel devoicing as a result of goal oriented adjustments of speaking style within some phonological, phonetic, and acoustical constraints.

Method

Recording of Dialogues

Dialogues were recorded during a simple picture-searching game through which a teacher attempted to assess the speech communication ability of a HI or NH child.

Two different panels were prepared, A or B, 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 all the names one by one.

The question was fixed as "Donokoga /CVCVCV/ desuka?" (Who is /CVCVCV/ ?), where /CVCVCV/ represents the name of a picture. If the teacher mistakenly used a different form of question, the sample was not used.

Recording of Read Speech

To clarify the differences between dialogue and read speech, a read list was also recorded and analyzed. Six teachers read the target sentences "Donokoga /CVCVCV/ desuka?" five times at fast (RDF), normal (RDN) and slow (RDS) tempos. The abbreviation RD is used to represent the read speech tokens.

Analyzed Samples

The names of the pictures consisted of target moras used to analyze the structure of the dialogues. Each name consisted of three moras. that is, /$C_1V_1C_2V_2C_3V_3$/, where one mora is formed by one consonant C and one vowel V.

Six types of moras were analyzed, i.e., AffIni, FriIni, $StoMed_1$, AffMed, FriMed, and $StoMed_2$, which represent the manner of articulation of the component consonant (fricatives. affricates. or stops) and mora position (initial or medial). Accent was placed on the initial mora.

AffIni and FriIni were the initial moras followed by /ki/, while $StoMed_1$ was /ki/ following the narrow vowels /i/ or /u/. Both the initial and medial moras could be devoiced for AffIni, FriIni, and $StoMed_1$. All the moras in AffMed, FriMed, and $StoMed_2$ were preceded by an open vowel /a/. $StoMed_1$ and $StoMed_2$ were treated separately because devoicing can be different depending on whether the preceding mora can be frequently devoiced ($StoMed_1$) or not ($StoMed_2$).

Subjects

Six professional teachers and seven HI or NH children participated in the test. All were speakers of the Tokyo dialect of Japanese.

Measurements

All the target moras, totalling 2740, were examined using an acoustic analysis system[1]. For each target mora sample, M_m, the length of the unvoiced segment, U_m, and that of the voiced segment, V_m, and their sum, $L_m = U_m + V_m$, were measured. M_m was classified as "Voiced" unless $V_m = 0$. For each sample, the classification variable *Voice* was either "Devoiced" or "Voiced."

Modeling

Four classification variables (*Mode*, *Mora*, *Teacher* and *Voice*) were defined as follows: *Mode* with five levels (HI, NH, RDF, RDN, RDS), *Mora* with six levels (AffIni, FriIni, $StoMed_1$, AffMed, FriMed, $StoMed_2$), *Teacher* with six levels (TE1 – TE6), and *Voice* with two levels (Devoiced, Voiced). Each mora sample was characterized by the continuous variables U_m, V_m, and L_m, and by the classification variables of *Mode*, *Mora*, *Teacher*, and *Voice*.

A four-dimensional contingency table, F_{ijkl}, was constructed first. F_{ijkl} represents the frequency of samples classified at the cell, C_{ijkl}, of the i-th *Mode*, j-th *Mora*, k-th *Teacher*, and l-th *Voice*. The devoicing rate, $Pr(C_{ijkl})$, was calculated as the ratio of the number of devoiced moras to the total number of moras for each cell, i.e., $Pr(C_{ijkl}) = F_{ijkl} / (F_{ijkl} + F_{ijk2})$.

Two statistical models were constructed, i.e., a generalized linear model (GLM) describing the relationship between the mora length and classification variables, and a logistic regression model predicting the devoicing rate using the mora length and classification variables (SAS, 1989).

Perceptual Analyses

The perceptual characteristics of the tokens were analyzed using the semantic differential method. As previously reported, 24 pairs of adjectives were used as 9-point dipole rating scales. The listening subjects were 8 normal hearing students. The tokens used were 30 samples of /Donokoga hikita desuka?/ spoken by the 6 teachers in the five modes. Obtained rating scores were 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 and Discussion

Mora Length Adjustments

The ANOVA obtained by the GLM procedure showed that L_m was significantly affected by the four classification variables (*Mode*, *Mora*, *Teacher* and *Voice*).

Figure 1 shows bar plots for the total mora length, L_m, with respect to *Mode* vs. *Mora*. As shown in Fig. 1, the teachers significantly lengthened the moras in speech directed to the HI children vs. the NH children and read speech. FriIni had the longest mora length which was significantly longer than the other mora groups. There was no significant difference in L_m between StoMed1 vs. StoMed2, AffIni vs. AffMed.

Devoicing Rate Variations

Figure 2 shows the predicted logistic regression curves for the devoicing rate of the HI, NH, and RD tokens. As L_m increases, the predicted devoicing rate clearly decreases more for HI

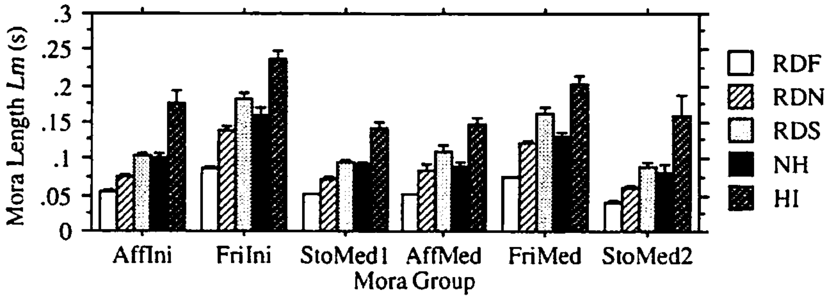

Figure 1. Interaction bar plot between Mode vs. Mora for mora length L_m.

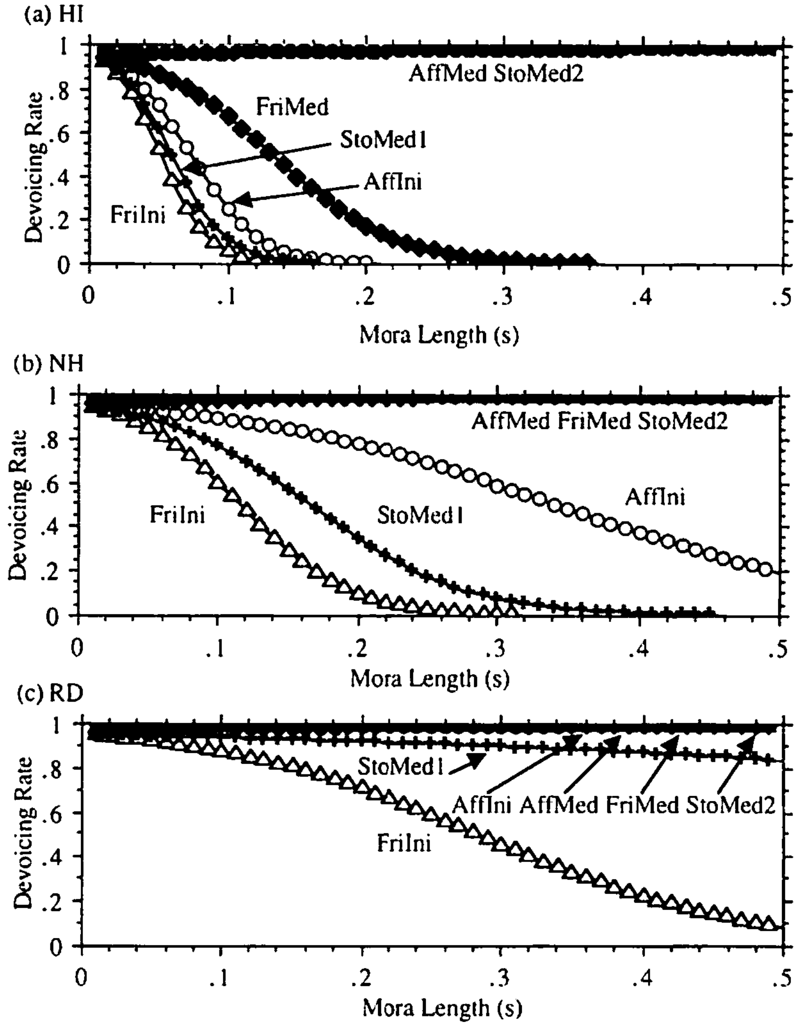

Figure 2. The predicted devoicing rate by CATMOD logistic regression analysis for the modes of (a) HI, (b) NH, and (c) RD. The mode RD includes RDF, RDN and RDS.

than for NH, and even more than for RD, which confirms our previous report (Imaizumi et al., 1995).

Some common tendencies, however, were observed regardless of mode. The accented initial mora groups, FriIni and AffIni, tended to have a lower devoicing rate than the medial mora groups, FriMed, AffMed and StoMed2. FriIni, the accented initial moras followed by /ki/, had the lowest devoicing rate, while AffMed and StoMed2, the medial unaccented moras preceded by an open vowel /a/, had the highest devoicing rate.

Furthermore, the devoicing rate was significantly different between $StoMed_1$ and $StoMed_2$. Both had an unaccented medial /ki/. $StoMed_1$ was preceded by a high vowel which was frequently devoiced, while $StoMed_2$ was preceded by an open vowel /a/ which was seldom devoiced. This significant difference in the devoicing rate cannot be accounted for by mora length variation because there was no significant difference in L_m. Thus, this result suggests that the devoicing rate depends on the devoicing probability of the preceding vowel.

These common tendencies observed regardless of mode may be explained by a phonological rule proposed by Kondo (1994). She showed that vowels tended not to be devoiced consecutively over two moras to avoid creating series of consonant clusters at the surface level, which is not a favored structure in Japanese.

From our point of view, these common tendencies further suggest that the teachers reduced the devoicing rate more in the HI vs. NH samples, and even more in the HI vs. RD samples, within phonological constraints. The mora groups, AffMed and StoMed2, which were highly devoicable from the phonological point of view were kept devoiced even when the teachers tried to talk carefully to the HI children, while the others were highly voiced. The teachers did reduce the devoicing rate so as to enhance the contrasts between the highly devoicable mora groups and the others within the phonological constraint of Japanese. Recalling our previous report (Imaizumi et al., 1995), listener-oriented adaptation of devoicing occurred within the phonological and phonetic (Jun and Beckman, 1993; Kondo, 1994) constraints of Japanese.

Perceptual Characteristics

The perceptual profiles of tokens could represented by four factors F1, F2, F3 and F4. F1 represents the contrast between discomfort ("Rough, Uneasy, Busy,") and pleasant ("Easy, Kind, Friendly, Restful, Polite"), corresponding to the perceptual difference between the RD and the other modes (NH and HI). F3 represents the contrast between "Slow, Stiff, Unnatural, Intelligible, Strong" and "Busy, Lifeless, Tense, Rough, Dull," corresponding to the differences between HI and the other modes (RD and NH). F2 and F4 could be interpreted as representing differences between the teachers. These results suggest that listener-oriented adaptation in speaking style produced significant perceptual effects.

Conclusion

The teachers did reduce the devoicing rate more in the HI than NH and RD samples in such a manner that contrasts between the highly devoicable mora groups and the others were enhanced within the phonological and phonetic constraints of Japanese. Listener-oriented adaptation in speaking style created significant acoustical and perceptual effects.

Acknowledgements

Sincere gratitude is extended to all the teachers and students at Ohji Elementary School. This research was supported by a Grant-in-Aid for Scientific Research on Priority Areas of "Spoken Dialogue," Ministry of Education, Science and Culture, Japan.

References

Beckman, M., and Shoji, A. (1984). "Spectral and perceptual evidence for CV coarticulation in devoiced /si/ and /syu/ in Japanese," *Phonetica* 41, 61–71.

Han, M. S. (1962a). *Japanese Phonology - An Analysis Based Upon Sound Spectrograms* (Kenkyusha, Tokyo, Japan).

Han, M. S. (1962b). "Unvoicing of vowels in Japanese," *Study Sounds* 10, 81–100.
Hirose, H., Simada, Z., and Fujimura, O. (1970). "An electromyographic study of the activity of the laryngeal muscles during speech utterances," *Annual Bulletin* (Research Institute of Logopedics and Phoniatrics, University of Tokyo, Japan) 4, 9–25.
Hirose, H. (1971). "The activity of the adductor laryngeal muscles in respect to vowel devoicing in Japanese," *Phonetica* 23, 156–170.
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.
Imaizumi, S., Hayashi, A., and Deguchi, T. (1995), "Listener adaptive characteristics of vowel devoicing in Japanese dialogue," *J. Acoust. Soc. Am.*, in press.
Jun, S-A. and Beckman, M. (1993). "A gestural-overlap analysis of vowel devoicing in Japanese and Korean," *Annual Meeting of the Linguistic Society of America*, (Los Angeles, USA, 1993).
Kondo, M. (1994). "Mechanisms of vowel devoicing in Japanese," *Proc. of ICSLP 94* (Yokohama, 1994), 1, 61–64.
Kuriyagawa, F., and Sawashima, M. (1989). "Word accent, devoicing and duration of vowels in Japanese," *Annual Bulletin* (Research Institute of Logopedics and Phoniatrics, University of Tokyo, Japan) 23, 85–108.
Lindblom, B. E. F. (1989). "Phonetic invariance and the adaptive nature of speech," in *Working models of human perception*, edited by B. A. G. Elsendoom and H. Bouma (Academic Press, London), 139–173.
Nooteboom, A. G., and Eefting, W. (1992). "To what extent is speech production controlled by speech perception? Some questions and some experimental evidence," in *Speech Perception, Production and Linguistic Structure*, edited by Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka (IOS Press, Tokyo, 1992), 439–449.
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.
SAS Institute Inc. (1989). *SAS/STAT User's Guide*, Ver. 6, Fourth Edition (Cary, NC, USA).
Sawashima, M. (1969). "Devoiced syllables in Japanese - A preliminary study by photoelectric glottography -," *Annual Bulletin* (Research Institute of Logopedics and Phoniatrics, University of Tokyo, Japan) 2, 11–20.

Sawashima, M. (1971). "Devoicing of vowels," *Annual Bulletin* (Research Institute of Logopedics and Phoniatrics, University of Tokyo, Japan) 5, 7–14.
Shibata, T. (1955). "Museika," in *Kokugogaku jiten* ("Devoicing," in Japanese Dictionary), (Kokugogaku Gakkai, Tokyo, Japan), p. 899.
Sugito, M. and Hirose, H.: "Production and perception of accented devoiced vowels in Japanese." *Annual Bulletin* (Research Institute of Logopedics and Phoniatrics, University of Tokyo, Japan) 22, 21-40.
Yoshioka, H. (1981). "Laryngeal adjustments in the production of the fricative consonants and devoiced vowels in Japanese," *Phonetica* 38, 236-251.
Yoshioka, H., Löfqvist, A., and Hirose, H. (1982). "Laryngeal adjustments in Japanese voiceless sound production," J. *Phonetics* 10, 1-10.