Inter-Speaker Variations in the Production and Distribution of Devoiced Vowels in Japanese

Yasuko Nagano-Madsen
Japanese Section, Department of Oriental Languages, University of Gothenburg
Västra Hamngatan 3, 41117 Göteborg, Sweden.
Tel: 031-7734844, Fax: 046-104210,
e-mail :Yasuko.Nagano_Madsen@ling.lu.se

1.Introduction

1.1 *Vowel devoicing in Japanese*

The phenomenon of vowel devoicing has been one of the central issues in phonetic and phonological descriptions of Standard Japanese (Polivanov 1976 [1914], Han 1962, Sawashima 1971, Hirose 1971, Sugito 1988, Maekawa 1989 etc.). At the same time, Japanese has long been nearly an exclusive source of information regarding this phenomenon in linguistic literature. In recent years, however, there has been a growing body of data on vowel devoicing from languages other than Japanese including Montreal French (Cedergren and Simoneau 1985), Korean (Kim 1993, Jun and Beckman 1994), and Turkish (Jennedy 1994), providing an opportunity for cross-linguistic comparison of vowel devoicing. These studies indicate that the phonetic conditions in which vowels are devoiced across several languages are strikingly similar, devoiced vowels are usually high vowels in a voiceless environment.

As far as Japanese is concerned, however, vowel devoicing is primarily a matter of phonology because a devoiced vowel is not a mere by-product of fast or careless speech but is present even in careful slow speech. Sawashima (1971), observing active effort of widening of the glottis for the production of devoiced vowels in Japanese, states that it is not a mere skipping of the phonatory adjustments for the vowel. The fact that devoiced vowels are indicated in pronunciation dictionaries also indicates the status of devoiced vowels within the Japanese phonology.

However, the traditional representation of vowel devoicing as a categorical change from [+voice] to [-voice], has recently been challenged. Sun-Ah and Beckman (1994) argue that devoicing processes such as those found in Japanese and Korean can be better accounted for as a general phonetic process of undershoot, where a vowel's glottal gesture overlaps with a preceding voiceless consonant. They claim that traditional phonological representation cannot explain the typically gradient character of devoicing seen in their data. They consider devoicing as another example of the kind of gradient phonetic effect resulting from more or less subtle adjustments to the magnitude and timing of otherwise invariant

gestural specifications, as in Browman and Goldstein's account of casual speech processes in English (Browman and Goldstein 1990).

One problem in representing vowel devoicing is that it seems to involve much greater variations in its exact phonetic manifestation. One can compare it, for example, with the allophonic variants of /t/ which take the forms [tʃ] and [ts] before /i/ and /u/ respectively, which are much strictly controlled. In addition, realization of vowel devoicing is known to differ both across various dialects of Japanese as well as across speakers within the same dialect (Sugito 1969, 1988, Yoshida and Sagisaka 1990, Yoshida 1992, Nagano-Madsen 1994). Again, no such variations are known for the allophonic variants of /t/.

1.2 *Vowel devoicing factors*

Segmental factors are primary for vowel devoicing in Japanese while other factors such as accent, position in a word etc., which may affect the realization of devoicing, can be considered as secondary. A standard description of vowel devoicing in Japanese is that the high vowels /i/ and /u/ are devoiced between the two voiceless consonants, or between a voiceless consonant and a pause. The description indicates that there are two major requirements for the vowels to be devoiced in Japanese. They are (1) tongue height [+high] of the vowel in question and (2) the lack of periodic glottal vibration [-voiced] in the adjacent consonants. Other factors which have been claimed to influence vowel devoicing include the manner of articulation of prevocalic and postvocalic consonants, accentuation, the presence or absence of another devoiceable vowel in an adjacent syllable, position in the word, utterance size and speech tempo among others.

The influence of accent in the realization of vowel devoicing in Japanese has been known for a long time (cf. Sakurai 1985 for a detailed description). By and large, it has been described that a potentially devoiceable vowel tends to resist devoicing once it is accented. Influence of accent, however, can appear in a different way by shifting accent to either the preceding or the following vowel. Many works using word in isolation form agree that accent indeed has a strong influence in preventing devoicing (cf. Nagano-Madsen 1994 for details).

Only recently have quantitative studies been undertaken into interaction between these factors. Takeda and Kuwabara (1987) followed by Yoshida and Sagisaka (1990), both using ATR word database, have attempted to establish a hierarchy among these factors while Kawai et al. (1993) presented an analysis of the NHK Accent Dictionary. Based on Kawai et al.'s analysis, Nagano-Madsen (1994) suggested that the factor of accent interacts with the manner of the postvocalic consonant. Kondo (1993), in her controlled material, found that the presence or absence of another potentially devoiceable vowel in an adjacent syllable has a critical influence in the realization of vowel devoicing, since devoicing of two successive vowels was not favoured. Some of these findings have opened a new area of close examination in the study of vowel devoicing in Japanese.

1.3 *Inter-speaker differences*

Although devoiced vowels are indicated in dictionaries like the NHK Accent Dictionary, considerable differences have been reported between the professional radio and T.V. announcers. Yoshida and Sagisaka (1990) reported that the total number of devoiced vowels read by three NHK announcers was 538, 601, and 734 respectively in 5449 words. In Nagano-Madsen's analysis of ATR sentence database consisting of 503 sentences, the total number of devoiced vowels varied from 607 to 876 between ten NHK announcers.

Though inter-speaker differences was not the focus of their study, Yoshida and Sagisaka (1990) state that some factors contribute more to the inter-speaker differences. Speakers were found to differ in the distribution of devoiced vowels according to (1) the nature of the prevocalic consonant and (2) whether or not the vowel is word-final. No effects were found from (1) the nature of postvocalic consonant and (2) accent. However, these findings were found not to be applicable to their sentence material in the same study. To date, it is still largely unclear how speakers differ in the production and distribution of devoiced vowels in Japanese.

1.3 The scope of present study

Based on the foregoing discussion, the present study examines the inter-speaker variation found in the production and distribution of devoiced vowels in Tokyo Japanese. As a first step, three female speakers, non-professional but teachers of Japanese, age between 48-59, were analysed using controlled corpus consisting of nonsense words embeded in a career sentence. The controlled factors include (1) prevocalic and postvocalic consonant, (2) accent, and (3) presence or absence of another devoiceable vowel in the same word. Prior to the distribution analysis, the phonetic realization of potentially devoiceable vowels were classified into four acoustic prototypes based on the spectrographic patterns.The distribution of acoustic prototypes was then analysed for each speaker.

2. Material and analysis

The corpus consisted of nonsense words having three CV syllables, where the segments in the first CVC were varied systematically. All possible combinations of devoiced vowels were included (cf. Table 2). The nonsense words such as *pipaka, pitaka, pikaka,* were embedded in a carrier sentence "Korewa ___ desu " (this is ___). When the consonant of the second syllable starts with [tʃ, ts, ʃ, ç, f], the following vowel becomes also potentially devoiceable, giving two possibilities for devoiced vowels in succession. Therefore, for those having two potentially devoiceable vowels, another set of words with [ga] was prepared as well (as [pitsuga] in addition to [pitsuka]). Speakers were asked to read these words with two accentual patterns, one having H*L (such as the place name Akita) and the

other LH (such as the place name Mitaka). Speakers exercised few sentences before the recordings.

The entire corpus consisted of 330 words, containing 440 tokens of high vowels in a voiceless environment. The material written in Japanese Katakana was read by three speakers of Tokyo Japanese, all female, aged 48-59 years. All speech material was digitized at 20kHz sampling rates using the Soundscope programme installed on a Macintosh. Acoustic analysis was carried out spectrographically.

3. Results

3.1 Acoustic manifestation

Based on the acoustic analysis of the high vowels produced by the three speakers, four types of acoustic realizations were identified in order to characterize the variations in production. It should be noted that not all of them are clear-cut divisions. It was easy to separate C-type from A-,B-, and D-types by the presence or absence of regular voicing. Difference between A-, B-, and D-types are gradient, varying most consistently in intensity but also in duration and timing in overlapping with the preceding frication phase. A-type has a regular voicing, intensity, and duration while B-type and D-type are reduced successively in that order. A typical sample for each type is shown in Figure 1.

A. With regular voicing and vowel formant, strong intensity.
B. Regular voicing is detectable but with greatly reduced intensity, usually also reduced in duration.
C. Frication phase only.
D. A further reduced state of type-B, sometimes close to a complete deletion.

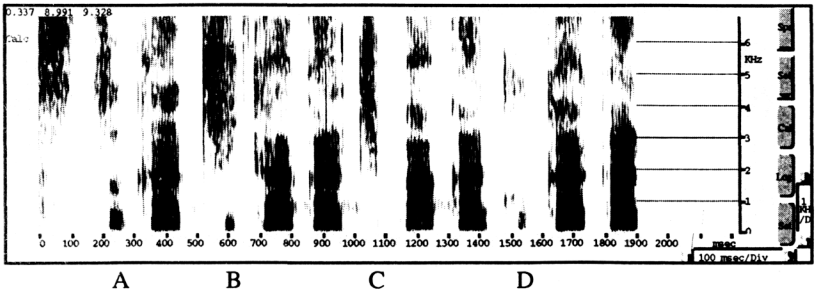

Figure 1. Sample tokens of type-A, B, -C, and -D in the extracted words [sutsuka], [kikaka], [kipaka], and [kukaka] respectively (all with a LHH accentual pattern, speaker S.B.).

4.2 Distribution

Analysis of the 440 high vowels between voiceless consonants is presented in Table 2 for each speaker. In the following discussion, the types B, C, and D are regarded as devoiced vowels while type A is a fully voiced vowel. The devoicing rates of the three speakers are in good agreement with the figures reported previously for Tokyo speakers (Sugito 1988, Yoshida 1992). The exact acoustic manifestation, however, seems to differ across speakers. Although all the three speakers preferred the type-C (with frication phase only), speaker S.B. used type-B (Regular voicing is detectable but with greatly reduced intensity, usually also reduced in duration.) more frequently than other speakers and particularly for words with LH accent pattern. It can be seen that speakers differ not only in the number of devoiced vowels but also in the acoustic realization type. Type-D (a further reduced state of type-B, close to a complete deletion.) was least common, being not even present for speaker K.K. It is interesting to note that most of the-D types occurred with the [ku] syllables (nine out of twelve) and the rest for [pi] or [pu].

speakers realization type	K.K.	N.T.	S.B.
A	40.0	48.9	43.9
B	9.6	2.7	24.3
C	50.3	46.8	31.4
D	0	1.6	0.3

Table 2. The acoustic realizations of the 440 high vowels between voiceless consonants for the three speakers (%).

In the present material, the vowel in question appeared in the initial syllable of CVCVCV, which was embedded in a carrier sentence. When the vowel was devoiced after a stop consonant, its presence was usually detectable in the much longer duration of aspiration (or frication noise), and the difference in quality (/i/ or /u/) was sometimes shown as different spectral patterns. Syllables such as /pi, pu, ki, ku/ usually had this pattern with or without weak voicing towards the end of the frication phase. When the syllable was unaccented, it tended to have reduced duration and intensity. Differences in quality between [ʃi] and [ʃu] with devoiced vowels, appeared to be most difficult to capture both spectrographically and auditorily.

	post- vocalic C									
syll.	p	t	k	tʃ	ts	s	ʃ	h	ç	f
pi	B	B	B	A	A	B	A	A	A	A
pu	B	B	C	B	B	B	B	B	B	B
tʃi	C	C	C	B	B	C	A	A	A	B
tsu	C	C	C	A	A	B	A	B	A	A
ki	C	C	B	B	B	C	B	B	B	B
ku	B	B	D	B	B	D	B	D	B	B
su	C	C	C	B	B	B	A	A	A	A
ʃi	C	C	C	B	B	B	B	A	A	A
ʃu	B	C	B	A	A	B	A	B	A	A
çi	A	C	C	A	A	B	A	B	A	A
fu	C	B	B	B	A	C	B	B	A	B

Table 2. The distribution of acoustic prototypes for S.B. with a LH accent pattern.

Distribution of these different types of acoustic realizations were tabulated for all of the speakers. Table 2 shows an example from a speaker S.D. for the H*L accent pattern. Results of the realization of vowel devoicing was examined with reference to the postvocalic consonants, with or without accent, and with presence or absence of another potentially devoiceable vowel in the following syllable.

4.3 Influence of postvocalic consonant and accent

Figure 2 shows the distribution of acoustic manifestation of the high vowels for the three speakers with respect to the nature of postvocalic consonant and accent information. The most striking regularity is that all the speakers drew a division between [s] and [ʃ] for both accented and unaccented versions. For the speakers N.T: and K.K., the fricative [s] and the affricates [tʃ, ts] behave the same as stop consonants [p, t, k], while S.B. had two further divisions, one for the affricates [tʃ, ts] and one for the fricative [s]. In addition, K.K.'s results differed from the other two speakers in treating [h] in the same way as stops and affricates.

For all three speakers, devoiced vowels before the stop consonants [p, t, k] were predominantly C-types in both accented and unaccented positions. N.T. and K.K. showed basically the same behavior for [tʃ] and [ts], while S.B. differed. Limited influence of accent was obvious for S.B. and K.K., for whom vowels in unaccented position were always reduced in the magnitude of acoustic manifestation for mainly fricatives as the postvocalic consonant.. This was not true for N.Y., however, who showed no such systematic reduction due to accentuation.

When devoiced vowels were found before a fricative, the phonetic realization never included the C-type except for the tokens with [h] for K.K. The acoustic manifestation of the so-called devoiced vowels can be divided into at least two major types, which appear to be conditioned by the postvocalic consonants.

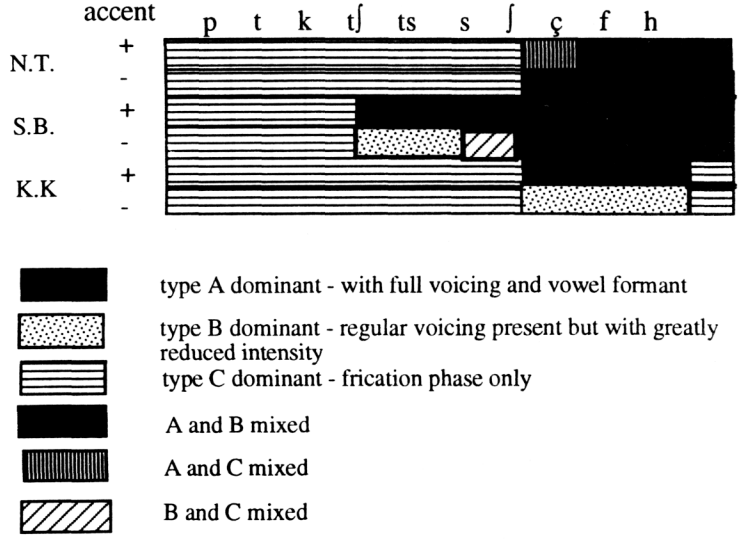

Figure 2. Distribution of the acoustic vowel type with reference to the postvocalic consonant and accent information.

4.4 In words with consecutive devoicing possibilities

Kondo (1993) indicated that the presence or absence of another potentially devoiceable vowel in the same word plays an important role, pointing out that the devoicing of two successive syllables are avoided. In the present study, influence of this factor was examined by comparing the vowels in two groups. In the first group, only one vowel occurred in a voiceless environment while in the second group, two successive vowels occurred in a voiceless environment.

Figure 3 shows the acoustic manifestations of high vowels in two consecutive devoicing environments with reference to postvocalic consonant and accent information. It can be seen that the devoicing of both vowels is not common, it was found only in K.K.'s tokens involving affricates as postvocalic consonant with LH accent type. There seems to be, however, no single and straightforward rule in chosing which of the two vowels to be devoiced or not devoiced at all. A systematic influence of accent was found only in N.T.'s tokens involving [tʃ, ts, ʃ]

as postvocalic consonants, where accented, be it H* or H, were chosen not to be devoiced. For S.B. and K.K., there is a slight tendency to devoice the second vowel regardless of accent type. For all the speakers, there is a strong tendency not to devoice the vowel when followed either by [ç] or [f].

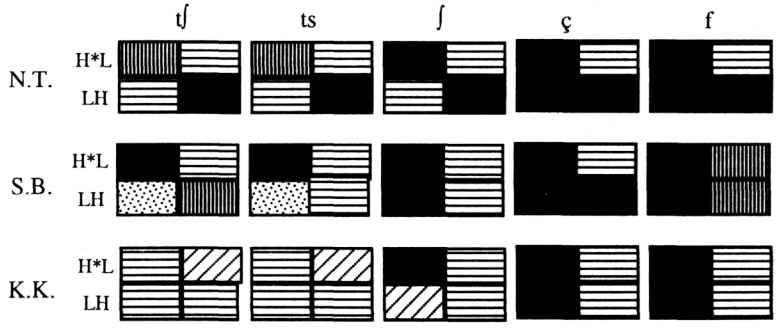

Figure 3. Acoustic manifestation of the high vowels in two consecutive voiceless environments with reference to the postvocalic consonant and accent type.

Here again difference between stop and fricative seems to play a role in determining the realization of devoiceable vowels. As far as the acoustic signal is concerned, when a devoiced vowel is integrated in the frication phase, there will be a continuum of two frication phases if the following syllable starts with a fricative while if a stop or affricate follows, there will be an occlusion after the frication phase. A succession of two frication phases without voicing seems to be avoided.

5. Summary

The results of the present study showed the variations found in the production and distribution of the devoiced vowels in Japanese by three speakers. Despite the fact that no single speaker used exactly the same type of manifestation and distribution, certain regularities were impressive.

The influence of prevocalic and postvocalic consonants has been discussed much in the literature and most studies agree that the manner of the postvocalic consonant is responsible for the variations found in the realization of vowel devoicing. It means that the feature [-voice] alone is not adequate for the postvocalic consonants. Explicitly or implicitly, it has been indicated that the devoicing rate is higher when a vowel is followed by a stop than a fricative (Takeda and Kuwabara. 1987, Yoshida and Sagisaka. 1990, Yoshida 1992, Kawai et al. 1993).

By and large, the results of the present study agreed with this line of observation. Furthermore, it has been shown that the degree of reduction differs

depending on the segmental and accentual context for some speakers. The most consistent regularity across the three speakers was associated with the division between [s] and [ʃ], for which a drastic change in manifestation was observed. For the postvocalic [s] as well as stop consonants, the acoustic manifestation was predominantly C-type while for [ʃ] and other fricatives, it was B or A.

The results of present study showed no overriding effect of accent in preventing devoicing since the majority of the vowels were devoiced even in an accented syllable. Effect of accent appeared differently for the three speakers. For S.B. and K.K., it interacted mainly with the postvocalic consonant while for N.T., it influenced devoicing when there were two devoiceable vowels. Yoshida & Sagisaka (1990) noted that the rate of devoicing declines in the order of L>H>H*, i.e. it becomes less if the vowel has a H tone and even less if it has a pitch accent. In the present study, two of the speakers (N.T. and S.B) showed a slight tendency for such a hierarchy while the speaker K.K.'s manifestation did not show any.

A greater variation was found as to which of the two potentially devoiceable vowels should be devoiced. While N.T. respected accent as an influential factor, S.B. and K.K. respected the position of the vowel in question more. Furthermore, there was an interaction with the postvocalic consonant. As in Kondo (1993), devoicing of both high vowels seems to be avoided.

Since the present study has examined the devoiced vowels in word-initial position, the obtained results may not apply for the devoiced vowels in other positions. Much further study is needed in this respect.

Acknowledgements

This work is partially financed by the Swedish Council for Research in the Humanities and Social Sciences (HSFR).

References

Browman, Catherine and Louis Goldstein. 1990 Tiers in articulatory phonology, with some implications for casual speech. In Kingston and Beckman, eds., *Papers in Laboratory Phonology I: Between the Grammar and the physics of Speech*, 341-376.

Han, Mieko. 1962 Unvoicing of vowels in Japanese. *Study of Sounds* 10, 81-100.

Hirose, Hajime. 1971 The activity of the adductor laryngeal muscles in respect to vowel devoicing in Japanese. *Phonetica* 23, 156-170.

Jannedy, Stefanie. 1994 Prosodic and segmental influences on high vowel devoicing in Turkish. In *Proceedings of the Fifth Australian Internaitonal Conference on Speech Science and Technology*, 674-679. Perth.

Kawai, Hisashi, Norio Higuchi, Tohru Simizu and Seiichi Yamamoto 1993. A study of a devoicing rule for speech synthesis of Japanese (in Japanese). *Proceedings of Acoustic Society of Japan 1993 Oct.,* 243-244.

Kim, Hyun-Gi, Seiji Niimi and Hajime Hirose. 1993 Devoicing of vowel in Korean. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics* 27, 151-155. Tokyo University.

Kimura, Osamu, Nobuyoshi Kaiki and Atsugo Kito. 1988 A study on devoicing rule in speech synthesis by rule (in Japanese). *Proceedings of Acoustic Society of Japan 1988 Mar.,* 137-138.

Kondo, Mariko. 1993 The effect of blocking factors and constraints on consecutive vowel devoicing in Standard Japanese. Poster presented at Labphon 4, Oxford, 11-14 August 1993.

Maekawa, Kikuo. 1989 Boin no museika. In Miyoko Sugito (ed.), *Nihon-go no Onsei-Onin* 1, 135-153. Tokyo: Meiji Shoin.

Nagano-Madsen, Y. 1994 Vowel devoicing rates in Japanese from a sentence corpus. *Working Papers* 42, 117-127, Department of Linguistics and Phonetics, Lund University.

Polivanov, E.D. 1976 [1914] Nihongo Ryukyugo onsei hikaku gaikan. In S. Murayama (ed.). *Nihongo Kenkyuu*, 126-147. Tokyo:Kobundo.

Sakurai, Shigeharu. 1985 Kyootsuu-go no hatsuon de chuui subeki kotogara. Appendix to NHK (ed). *NHK Accent Dictionary*. Tokyo.

Sawashima, Masayuki. 1971 Devoicing of vowels. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics* 5, 7-13. Tokyo University.

Sugito, Miyoko. 1969 A study of voiceless vowels in Tokyo and Osaka dialects (in Japanese). In *Study of Sounds XIV*, 249-264.

Sugito, Miyoko. 1988 Nihon no 8 toshi ni okeru boin no museika. *Osaka Shoin Joshidaigaku Ronshu*, no.25.

Sun-Ah, Jun. and Beckman Mary.E. 1994 Distribution of devoiced high vowels in Korean. In the *Proceedings of 1994 International Conference on Spoken Language Processing*, 479-482. Yokohama.

Takeda, Kazuya and Hisao Kuwabara. 1987 Analysis and prediction of devocalizing phenomena (in Japanese). *Proceedings of Acoustic Society of Japan 1987 Oct.,* 105-106.

Yoshida, Natsuya. 1992 Boin-museika no washa-kan ni kansuru onseigaku-teki kenkyuu. Unpublished manuscript. Department of Linguistics, Kyoto University.

Yoshida, Natsuya & Yoshinori Sagisaka. 1990. Factor analysis of vowel devoicing in Japanese (in Japanese). *ATR Technical Report* TR-1-0159. ATR (Interpreting Telephony Research Laboratories), Kyoto.