Lund University, Dept. of Linguistics
Working Papers 42 (1994), 117-127

Vowel Devoicing Rates in Japanese from a Sentence Corpus

Yasuko Nagano-Madsen

Introduction

The phenomenon of vowel devoicing has been described for a long time in Japanese linguistic and phonetic literature (Polivanov 1976 [1914], Han 1962, Sawashima 1971, Hirose 1971, Sugito 1988, Maekawa 1989 etc.). In recent years there has been a growing body of data with respect to the hierarchy and interaction among the factors responsible in inducing vowel devoicing in Japanese (Takeda & Kuwabara 1987, Kimura et al. 1988, Yoshida & Sagisaka 1990, Yoshida 1992, 1993, Kondo 1993, Kawai et al. 1993). Some of these studies have been carried out for technical use such as speech synthesis and recognition and are quantitatively more adequate than previous ones. Based on the results of these recent studies, in particular on Kawai et al. 1993, the present paper explores the extent to which the realization of vowel devoicing varies according to segmental contexts and accentuation using a sentence corpus. Although past studies have supplied much data on vowel devoicing factors in Japanese, potential variation due to segmental contexts and accentuation have not been explored from a sentence corpus.

Segmental contexts

Segmental factors are regarded as primary for vowel devoicing in Japanese because devoicing is restricted by segmental features to begin with. Other factors such as accent, position in a word, etc., which may affect the realization of devoicing can be considered as secondary. This treatment is explicit in the recent works by Kondo 1993 as well as Kawai et al. 1993.

A standard description of vowel devoicing is that the high vowels /i/ and /u/ tend to be 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 vibration in glottis [-voiced] for the adjacent consonants. However, it has been noted also that the feature [-voiced] alone is inadequate to cover all the instances of devoicing.

First, vowels may be devoiced even when the postvocalic consonant is phonologically voiced. Han 1962 gives examples when the high vowels are followed by a semi-vowel [y] as in *soo desuyo* 'that's right' where she adds that the prevocalic consonant in such a case is either [s] or [ʃ]. Takeda & Kuwabara's 1987 results show that the [ʃi] syllable were devoiced at considerably high rates (50%) before [g] and also before [n] and [m] though to much lesser extent. Yoshida & Sagisaka's analysis also includes [g], [m], [n], and [b] as postvocalic consonants and [y], [dz], [d], and [m] as prevocalic consonant though 96% of devoicing occured between voiceless consonants. Opposite to these cases is when high vowels do not become devoiced even between two phonologically voiceless consonants. Kawai et al.'s 1993 analysis indicates that such cases occur predominantly in the fricative-fricative environment but it is not uniform among all the fricatives.

Secondly, influence of other factors such as accent and position in word seems to vary depending on the segmental contexts. Yoshida & Sagisaka 1990 report that the influence of accent is stronger when the postvocalic consonant is [s] or [ʃ] where the devoicing rate was only 17%, while with postvocalic [k] the devoicing rate was 91%. Kawai et al. 1993 presents a thorough analysis of an accent dictionary concerning how the realization of vowel devoicing can vary depending on the segmental contexts and also depending on accent and position in word (cf. Table 1).

Their devoicing table indicates that segmental contexts can be divided into two major groups, one for which devoicing is realized on the basis of segmental contexts alone and the other for which factors of accent and position in word interact. The latter group also include several instances where the vowels aren't devoiced at all. The most relevant feature which divides different segments into these two groups looks like the manner of articulation of the postvocalic consonant. When a high vowel is followed by a stop consonant it always gets devoiced without further influence from factors of accent and/or position in word while the majority of postvocalic fricatives are affected by such factors. This observation is in good agreement with the results of some of the previous studies which have pointed out that the nature of the postvocalic consonant is the most decisive factor in determining vowel devoicing and that devoicing rate is higher

Table1. Devoicing table by Kawai et al. 1993, slightly modified.

	post- vocalic C									
syllable	p	t	k	tʃ	ts	s	ʃ	h	ç	f
pi	-	*	*	A	*	*	A	-	-	-
pu	*	*	*	*	*	*	*	X	-	*
tʃi	*	*	*	*	*	*	A	AΩ	Ω	X
tsu	*	*	*	*	*	*	*	X	X	X
ki	*	*	*	*	*	*	*	A	*	*
ku	*	*	*	*	*	*	*	A	*	*
su	*	*	*	*	*	AΩ	Ω	Ω	-	A
çi	*	*	*	*	*	*	A	AΩ	X	Ω
ʃi	*	*	*	*	*	AΩ	X	X	X	X
ʃu	*	*	*	*	*	Ω	X	AΩ	-	X
fu	-	*	*	*	*	*	*	Ω	X	-

X not devoiced
A devoiced unless it is accented
Ω devoiced unless it is word-initial
AΩ devoiced unless it is word-initial and accented
* devoiced without further conditions
- no data available

when a vowel is followed by a stop consonant than a fricative (Takeda & Kuwabara 1987, Yoshida & Sagisaka 1990).

Among the fricatives, [h] is strongest in preventing devoicing as all the vowels are either affected by additional factors or do not become devoiced at all when they are followed by this consonant. [ç] and [f], the other allophonic variants of /h/, are also quite strong followed by [ʃ]. [s] is closest to a stop consonant as it is least influenced by other factors among fricatives. The behaviour of [h] as strongly preventing devoicing may be due to the fact that [h] at intervocalic position is not totally voiceless but is articulated with some sort of vibration (Yoshioka et al. 1986). Similar behaviour of [ç] and [f] is understandable if we accept that these two sounds are not always articulated distinctively but often merge with [h]. Yoshida 1992, by using controlled material of bisyllabic words, also noted that devoicing rate is extremely low before [f], [ç], and [h] but in general before fricatives in comparison with stops and affricates.

Influence of the prevocalic consonant seems less obvious from Table 1. There is, however, an indication that the prevocalic consonant is also important as a segmental factor. If we compare the behaviour of vowels preceded by a stop consonant such as [p] and [k] with those preceded by a fricative, it can be seen that vowels preceded by a fricative are more likely to be influenced by the factors of accent and position in word. It is

interesting to note that the two affricates [tʃ] and [ts] behave more like a fricative than a stop when they precede a vowel while they behave basically the same as a stop consonant when they follow the vowel. This means that it is more appropriate to regard an affricate as a fricative when it is prevocalic and as a stop when it is postvocalic assuming that it is the immediately adjacent part of an affricate to the vowel which has the greatest influence on devoicing. Kimura et al. 1988 and Yoshida & Sagisaka 1990 both pointed out that devoicing becomes more frequent when the prevocalic consonant is a fricative. It would, however, be interesting to know how this segmental factor will interact with other factors.

The foregoing discussion has drawn a major division between stop consonants and fricatives with respect to the realization of vowel devoicing. It is interesting to note that Kawakami 1977 divides so-called devoiced vowels in Japanese into two types: when [ki, pi, ku, pu, ʃu, tʃu] occur before a voiceless consonant, they will have devoiced vowels. On the other hand syllables [ʃi, tʃi, çi, su, tsu, fu] before a voiceless consonant usually do not even possess a devoiced vowel. His division is apparently based on the nature of the prevocalic consonant and it is notable that the former group has, except for [(t)ʃu], stop consonants while the latter has fricatives. The division into stops and fricatives is also supported from a physiological study which reports that the mode of glottal opening attained during a devoiced vowel differs depending on whether the surrounding consonant is a fricative or a stop (Yoshioka et al. 1986).

Accent and tone

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 (see Yoshida 1993 for details).

Recent works using either database or controlled material, all agree that accent indeed has a strong influence in preventing devoicing (Takeda & Kuwabara 1987, Kimura et al. 1988, Yoshida 1993). Furthermore, Kimura et al. 1988 marked the hierarchy of accent and tone in influencing devoicing as accented < pre-accented < post-accented. It is not clear whether pre-accented vowel has H tone or L tone, but post-accented vowel should have a L tone. Yoshida & Sagisaka 1990 also notes 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[1]. However, it can be seen from Yoshida and Sagisaka's results that the difference in devoicing rates with respect to H or L tone is larger and consistent for the 2- and 3-mora words but not so remarkable for the 4- and 5-mora words across the three speakers. One possible hypothesis is that the influence of accent and tone appears stronger for a shorter word and utterance. Since all the previous works which have claimed the strong influence of accent have used words in isolation form, different results may be obtained from more natural connected speech material. Also, earlier works have not considered the segmental variation with reference to accent behaviour as reported in Kawai et al. 1993. In the present study, the effect of accent and tone are explored by bringing these two points into consideration.

Material

Data was obtained from the ATR Sentence Database. The database consists of 503 phonemically balanced short sentences which were read by 10 announcers of NHK (approximately 1 hour recording time for each speaker). Speech data are labelled on different layers and the information on devoicing was obtained from the phonemic layer and allophonic layer combined with accent/tone information obtained from another file which contains prosodic information. Devoicing is defined acoustically by the absence of voice bar.

Analysis and discussion

The total number of devoiced vowels among the 10 speakers is shown in Table 2.

Table 2. The total number of devoiced vowels in 503 sentences across 10 speakers. F=female, and M=male speakers.

speaker	FKN	MHT	FYM	FTK	MMY	MTK	MYI	FKS	MSH	MHO
tokens	607	637	650	678	712	716	728	739	744	876

[1]In the present paper, the terms pitch accent and tone H and L are used and are indicated as H* and H/L respectively. These terms correspond to the expressions accent kernel, accent H, and accent L in most of the previous papers on devoicing.

Segmental contexts

In order to have a closer look, data of a single speaker (MYI) was examined in detail. All the devoiced vowels were selected together with their consonantal environments and accentual environments. Table 3 shows the classification of the devoiced vowels according to the voicing feature. It can be seen that the large majority of the devoiced vowels occurred between voiceless consonants while only a minor amount of devoicing occurred between a voiceless consonant and a pause or a voiced consonant. The type of voiced consonant and the number of tokens which induced devoicing were as follows: [y] (9), [n] (9), [m] (8), [g] (7), [d] (5), [z] (3). In two instances, devoicing occured with a voiced prevocalic consonant and in both cases it was [z]. In one of them, it was also followed by a voiced consonant, i.e. devoicing occured between two voiced consonants in [hazuda]. As for the vowels, in addition to the expected /i/ and /u/, some tokens of /a/, /o/ and /e/ were devoiced in the present corpus but they were few in number.

Table 3. Analysis of the 728 devoiced vowels (speaker MYI) according to their segmental environments.

segmental contexts	percentage
voiceless consonant - voiceless consonant	87.5%
voliceless consonant - pause	7.0%
voiceless consonant - voiced consonant	5.8%

As a second step, the ten most frequent sequences which have a voiceless vowel were listed with their accent information (cf. Table 4). Then out of the 503 sentences, all the sequences which have the same segmental contexts were chosen and the devoicing rates were calculated. It is seen from the table that the majority of them have a stop as the postvocalic consonant and they are the ones which are expected to be devoiced without any influence from the accent and position in a word. Their actual devoicing rates are all very high, often being 100%, and indicate that there is a good match between an analysis of the accent dictionary (cf. Kawai et al. 1993) and the actual tokens obtained from a sentence corpus.

Accent and tone

Analysis of the 728 devoiced vowels from the speaker MYI is presented in Table 5 together with accent/tone information. It is seen that the largest proportion of devoiced vowels have a L tone followed by a H tone and finally by a pitch accent (H*). However, Table 4 shows that neither pitch

accent nor H tone has an overriding effect in preventing devoicing. Though there are some instances where the devoicing rates vary, no systematic influences due to accent and tone type can be observed.

Table 4. 10 most frequent sequences which included devoiced vowels and their devoicing rates in % according to accent and tone (speaker MYI). Figures in brackets show the number of tokens.

segmental sequence	overall devoicing rates (%)	devoicing rates (%) according to tone and accent		
		L	H	H*
ʃit	100 (87)	100	100	100
çit	100 (38)	100	100	-
tsuk	100 (38)	100	100	100
su#	100 (36)	100	100	100
ʃik	84 (33)	84	81	100
kus	96 (29)	100	95	100
fuk	100 (28)	100	100	100
kut	96 (26)	94	100	100
kit	95 (22)	100	85	100
kik	90(22)	100	66	100

Table 5. Analysis of the 728 devoiced vowels (speaker MYI) according to accent information.

tone and accent	L	H	H*
percentage	56.5%	35.4%	7.9%

Since the influence of accent on devoicing is reported to vary greatly depending on the segmental contexts (cf. Kawai et al. 1993), the devoicing rates of those sequences which have a fricative in postvocalic position were examined for the 10 speakers. All the potentially devoiceable sequences which fit into this category were selected and their actual devoicing rates were calculated according to accent/tone information (Table 6).

It is seen that the occurence of the [p]-syllables as well as [f] and [ç] as postvocalic consonants are rare. The devoicing rates for those sequences with fricative as postvocalic consonant are, in general, much lower than those sequences which have stop consonants in postvocalic position shown in Table 4. In particular, when [h] or [f] is in postvocalic position, devoicing rates are very low. The effect of [ç] is not very convincing because there is only one example of this segment as postvocalic consonant and because there was only one token for each speaker. But these results on the postvocalic fricatives agree with Yoshida's 1993 results which were

Table 6. Devoicing rates in % according to the postvocalic consonants (10 speakers pooled). The figures in brackets indicate the number of tokens.

syllable	post-vocalic consonant				
	s	ʃ	h	ç	f
pi	100 (10)	-	-	-	-
pu	-	0 (10)	-	-	-
tʃi	20 (30)	25 (20)	10 (120)	-	30 (10)
tsu	66 (90)	50 (110)	12 (110)	-	30 (10)
ki	61 (110)	61 (80)	0 (60)	-	-
ku	84 (330)	44 (220)	34 (230)	100 (10)	30 (10)
su	48 (70)	5 (20)	-	-	-
çi	99 (100)	-	-	-	-
ʃi	57 (200)	5 (20)	9 (160)	-	0 (10)
ʃu	-	15 (69)	0 (10)	-	-
fu	99 (100)	-	-	-	-

obtained from well-controlled word material. Most sequences show good correspondence to the analysis of an accent dictionary (Kawai et al. 1993) but there are some which deviate greatly from the prediction: they are [tʃis, puʃ, tsuʃ, kuʃ, kuf]. They are expected to be devoiced without any further conditions but they all showed much lower rates (less than 50%) of devoicing than expected.

In order to examine the influence of tone and accent, the devoicing rates were calculated according to accent information and totalled for [s, ʃ, h] and [ç/f]. The results show no systematic influence of tone and accent, i.e. influence is stronger in the order of H*>H>L, as indicated from the analysis of word database (Yoshida & Sagisaka 1990).

Conclusion

The vowel devoicing rates in a sentence corpus were explored with focus on segmental and accentual variations. The realization of vowel devoicing was found to vary greatly depending on the segmental contexts, in particular depending on whether the postvocalic consonant is a stop or a fricative. When [h] and [f], the allophones of /h/, are in postvocalic position, the devoicing rates were found to be very low. Devoicing rates were generally lower for [ʃ] than for [s] when they are postvocalic. These findings from a sentence corpus are in good agreement with the anlysis of an accent dictionary by Kawai et al. 1993. Although previous studies from words in isolation form have reported strong influence of accent in preventing devoicing, no overriding effect of pitch accent was found in the current material of sentence corpus. Likewise there was no systematic influence of

Table 7. Devoicing rates in % according to postvocalic consonant and accent/tone information (10 speakers pooled).

(a)

s	L	H	H*
pis	-	100	-
tʃis	0	30	-
tsus	20	72	-
kis	80	40	-
kus	75	87	100
sus	48	-	-
his	98	-	100
ʃis	23	86	45
fus	100	80	-
total	65	77	78

(b)

ʃ	L	H	H*
puʃ	0	-	-
tʃiʃ	0	50	-
tsuʃ	60	34	95
kiʃ	100	38	100
kuʃ	51	40	-
suʃ	5	-	-
ʃiʃ	5	-	-
ʃuʃ	22	0	-
fuʃ	100	-	77
total	28	35	93

(c)

h	L	H	H*
tʃih	5	12	40
tsuh	0	23	0
kih	0	0	0
kuh	22	37	-
suh	10	80	-
ʃih	12	8	
ʃuh	0	-	-
fuh	0	-	-
total	8	24	13

(d)

ç/f	L	H	H*
kuç	-	100	-
tʃif	30	-	-
tsuf	-	30	-
kuf	-	30	-
ʃif	0	-	-
total	15	53	-

tone and accent, i.e. H*>H>L, in influencing devoicing rates. It is indicative that the influence of accent and tone appear stronger on words in isolation form, and in particular on short words consisting of 2 or 3 moras (cf. Yoshida and Sagisaka 1990). However, the present study did not take into account some other factors which might have influenced devoicing, i.e. position in word and cases where the succession of devoiceable syllables are present. In order to confirm some of the findings from the present study, much further studies which control other factors are necessary.

Acknowledgements

This study was carried out as part of the project *Multilingual Prosodic Rules* which is sponsored by the Swedish Council for Research in the

Humanities and Social Sciences. I am grateful to ATR, in particular to Yoshinori Sagisaka and Nick Campbell, for giving me an opportunity to use the database. My sincere thanks are to Norio Higuchi for his kind guidance in organizing this work and to Ken Shimomura for programming.

References

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.

Kawai, Hisashi, Norio Higuchi, Tohru Simizu & 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.

Kawakami, Shin. 1977. *Nihongo onsei gaisetsu.* Tokyo: Ohusha.

Kimura, Osamu, Nobuyoshi Kaiki & 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.), *Nihongo no Onsei-Onin* 1, 135-153. Tokyo: Meiji Shoin.

Polivanov, Evgenij Dmitrievič. 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. 1988. 'Nihon no 8 toshi ni okeru boin no museika'. *Osaka Shoin Joshidaigaku Ronshu* 25.

Takeda, Kazuya & 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. 1993. 'Boin no museika ni eikyoo o ataeru shoyooso no kentoo'. 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.

Yoshioka, Hirohide, Anders Löfqvist, Hajime Hirose & René Collier. 1986. 'How voiceless sound sequences are organized in terms of glottal opening gestures'. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics* 20, 55-67. Tokyo University.