3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

MECHANISMS OF VOWEL DEVOICING IN JAPANESE

Mariko Kondo

Department of Linguistics, University of Edinburgh
1F, Adam Ferguson Building, 40 George Square
Edinburgh, EH8 9LL, Scotland, U.K.

ABSTRACT

The intensity measurements of vowels in devoicing environments suggest that high vowel devoicing in Japanese is a natural vowel reduction process resulting from glottal gestural overlap. Earlier work showed that high vowel devoicing appears almost compulsory at the normal speaking rate so long as there is no devoiceable vowel in an adjacent mora. High vowels can sometimes remain voiced in devoicing environments, but their intensity is smaller than that of ordinary vowels. However, in the consecutive devoicing environments, voiced vowels in devoicing environments are not weakened. Two different mechanisms seem to control vowel devoicing depending on the environments.

1. INTRODUCTION

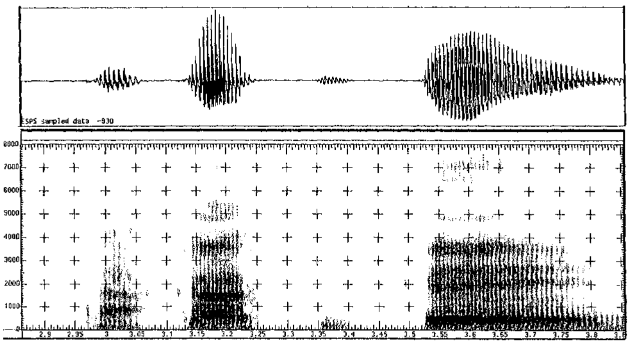
In Standard Japanese, the high vowels /i/and /u/ become voiceless either between voiceless obstruents or between a voiceless obstruent and a pause. An earlier study (Kondo, 1993) [2] showed that high vowel devoicing was almost exceptionless so long as there was no devoiceable vowel in adjacent morae. In words with consecutive devoicing possibilities, devoicing was conditioned by accent location, the type of preceding consonants and the presence of internal word boundary. Consecutive devoicing over three morae was uncommon at normal speaking rate.

Phenomena called vowel devoicing or similar to vowel devoicing have been reported in many languages other than Japanese (Cheyenne, Comanche, English, French, German, Greek, Hebrew, Korean, Polish, Swabian, Shoshoni, etc.). Many of them are in similar environments to those in Japanese and occur optionally in fast speech as a result of vowel reduction. However, in Japanese it has been said that vowel devoicing is not an optional fast speech rule; it occurs at slow tempo as well as normal tempo.

Vowel devoicing is not a clear cut distinction between voiced and voiceless. There are quite a few marginal cases. Vowels in devoicing environments are phonetically realised as fully voiced vowels, partially voiced vowels and completely voiceless vowels. Partially voiced vowels are characterised as having shorter duration; on the speech pressure waveform, periodic waves are repeated between two and seven times, whereas for fully voiced vowels periodic waves are repeated more than ten times. Partially voiced vowels are also generally weaker than fully voiced vowels; their intensities are lower than those of fully voiced vowels.

However, it is not very easy to draw a distinction between fully voiced and partially voiced vowels. Therefore partially voiced vowels were treated as voiced vowels in the previous study. In fact, vowels which were categorised as fully voiced in devoicing environments were very often weaker than vowels in non-devoicing environments, regardless of their durations. For example, /u/ of /takakuke:e:/in figure 1 is devoiceable surrounded by voiceless consonants but was voiced in this particular utterance. However, its intensity was lower than that of other vowels in the same word. This implies that high vowels tend to be weakened between voiceless consonants, and they are finally reduced to voiceless.

Figure 1. An example of intensity weakening
/takakuke:e:/ 'diversification'



t a k a k ɯ k e:

2. HYPOTHESES

High vowels in devoicing environments are usually devoiced if there is no devoiceable vowel in an adjacent mora. Devoiceable vowels can sometimes be voiced, but they vary phonetically from fully voiced to almost voiceless. If it is natural for high vowels to be devoiced, when they remain voiced, they may be phonetically different from vowels in non-devoicing environments. Hypotheses of the experiment are:

(1) Voiced vowels in devoicing environments tend to be weaker than ordinary vowels.
(2) There must be a difference in the degree of weakening depending on the single or consecutive environments since they showed different devoicing rates.
(3) As devoicing is part of the vowel reduction process similar to the phenomenon in other languages, the vowels must be weakened more at a faster tempo.

3. EXPERIMENTAL METHODS

An experiment was carried out to measure the intensities of vowels in different devoicing and non-devoicing environments. The intensity of each vowel was measured. Three Standard Japanese speakers (one male and two female) read the six test words listed below: (a) three with single devoicing site and (b) three with consecutive devoicing possibilities. Devoiceable vowels are underlined.

(a)	/taisjok<u>u</u>teate/	'retiring allowance'
	/kamot<u>u</u>seNpak<u>u</u>/	'cargo boats'
	/takasak<u>i</u>simiN/	'theTakasaki citizen'
(b)	/h<u>u</u>k<u>u</u>sjok<u>u</u>keNsa/	'dress inspection'
	/sjok<u>u</u>h<u>i</u>setujak<u>u</u>/	'a cut in food expenses'
	/haisj<u>u</u>t<u>u</u>kidjuN/	'exhaust limit'

It should be noted that /N/ is a moraic nasal and /sj/ sequence is realised as [ɕ].

The test words were embedded in a carrier sentence " kore wa _____ to yomimasu" ('This is pronounced _____'). Each speaker pronounced each test word three times at random at three different tempi: fast (1.7 seconds), normal (2.5 seconds) and slow (3.3 seconds). The speakers adjusted their speaking tempi according to tones sounding in their headphones with a 3 seconds pause between sentences. The subjects were asked to time the first mora /ko/ of /kore/ at the first beep, and the last mora /su/ of /yomimasu/ at the second beep.

Different tempi were set up in order (1) to examine intensity of vowels in devoicing sites at different tempi and (2) to trigger voicing of some vowels in devoicing sites. As an earlier study showed, high vowels were almost always devoiced at normal speaking rate if there was no devoiceable vowel in an adjacent mora. Devoiceable vowels may undergo devoicing process at a normal speaking rate but may remain voiced at a slower rate. Secondly, comfortable speaking rates vary from speaker to speaker. So the set normal speaking rate may be faster than a comfortable rate for some people but may be felt to be slower than normal for others.

4. RESULTS AND DISCUSSION

4.1 Voicing rates

As predicted, not all vowels in devoicing environments were voiced, but some devoiceable vowel remained voiced as showen in Table 1.

Table 1. The number of voiced vowels in single devoicing environments and in consecutive devoicing environments (maximum number = 18)

subject	environment	fast	normal	slow
A	single	4	0	1
	consecutive	6	7	10
B	single	3	3	12
	consecutive	4	14	12
C	single	4	6	8
	consecutive	11	12	13

Subjects B and C showed high voicing rates even for words with one devoicing site, which was contradictory to the results of Kondo (1993).[2] This was probably because normal speaking rates of subjects B and C were faster than the set 'normal' tempo. They voiced less at the fast tempo and more at the slow tempo.

4.2 Difference of vowel intensity

The intensities of vowels in devoicing environments and in non-devoicing environments were compared using a T-test. The average intensities of vowels in non-devoicing environments excluding word-initial and word-final morae were calculated. When a subject inserted a pause between a test word and the carrier sentence, the vowels in the first and last morae of the test words were very often weakened. Therefore, the data from the vowels in word-initial and final morae were excluded. The average intensity was compared with the intensity of voiced devoiceable vowels in the same word.

As shown in Table 2, the intensity of voiced devoiceable vowels was significantly lower than that of ordinary vowels at all tempi except for C. Kondo (1993,[2] 1994)[3] showed that devoicing rates and the duration of devoiced morae were largely dependent on the environment in which they occur; i.e. with single devoicing or consecutive devoicing environments. Therefore, the intensities of vowels were compared by their environments.

Table 2. T-test of intensity difference between vowels in devoicing environments and in non-devoicing environments (one-tailed)

subject	tempo	mean difference	SD	T (df)	P-value
A	fast	8.93 dB	5.33	5.30 (9)	P<0.001
	normal	3.23 dB	1.93	4.44 (6)	P<0.005
	slow	6.97 dB	5.97	3.87(10)	P<0.005
B	fast	4.05 dB	5.01	2.14 (6)	P<0.05
	normal	2.82 dB	2.24	5.18(16)	P<0.001
	slow	1.98 dB	2.36	4.09(23)	P<0.001
C	fast	2.00 dB	4.49	1.78(15)	P<0.05
	normal	1.47 dB	3.64	1.71(15)	P<0.1
	slow	1.31 dB	3.03	1.94(20)	P<0.1

Table 3. T-test of intensity difference between vowels in single devoicing sites and non-devoicing environments (one-tailed)

subject	tempo	mean difference	SD	T (df)	P-value
A	fast	10.27 dB	3.36	6.12 (3)	P<0.005
	normal				
	slow	(16.99 dB)			
B	fast	6.45 dB	2.31	4.83 (2)	P<0.025
	normal	5.01 dB	2.13	4.09 (2)	P<0.05
	slow	3.35 dB	2.16	5.37(11)	P<0.001
C	fast	6.58 dB	4.52	3.25 (4)	P<0.025
	normal	3.74 dB	3.96	2.32 (5)	P<0.05
	slow	2.61 dB	1.04	7.12 (7)	P<0.001

Table 4. T-test of intensity difference between vowels in consecutive devoicing environments and in non-devoicing environments (one-tailed)

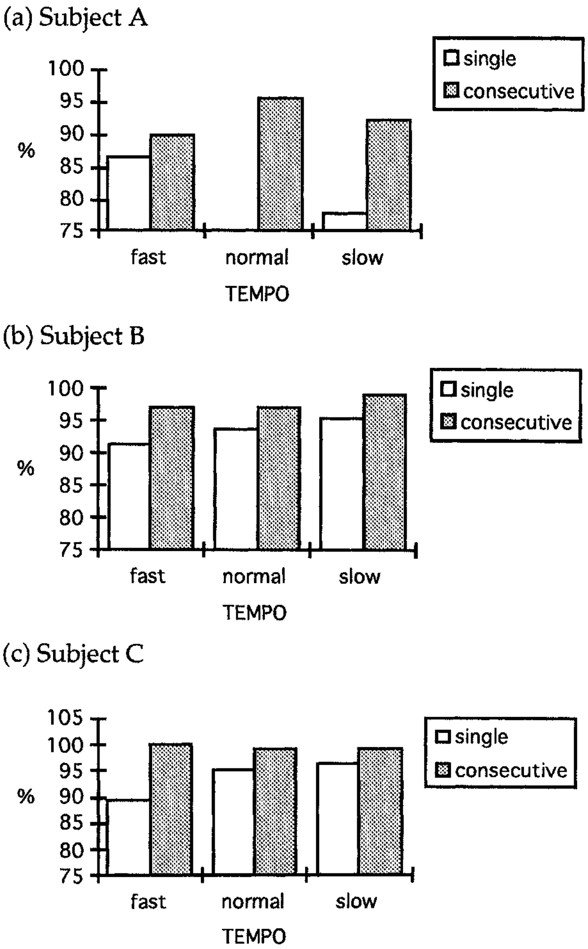
subject	tempo	mean difference	SD	T (df)	P-value
A	fast	8.04 dB	6.49	3.04 (5)	P<0.025
	normal	3.23 dB	1.93	4.44 (6)	P<0.005
	slow	5.97 dB	5.23	3.61 (9)	P<0.005
B	fast	2.26 dB	6.05	0.74 (3)	P>0.1
	normal	2.35 dB	2.04	4.31(13)	P<0.001
	slow	0.60 dB	1.69	0.49(11)	P>0.1
C	fast	-0.08 dB	2.62	-0.1(10)	P>0.1
	normal	0.67 dB	3.10	0.74(11)	P>0.1
	slow	0.50 dB	3.65	0.50(12)	P>0.1

When there was only one devoicing site in a word and the devoiceable vowels were voiced, the intensities of the devoiceable vowels were significantly lower than those of non-devoiceable vowels at all tempi (Table3). In consecutive devoicing environments, however, when devoiceable vowels were voiced, their intensities were not necessarily lowered (Table4). In some cases, the intensities were higher than those of non-devoiceable vowels. Different mechanisms seem to control vowel devoicing.

4.3 The effect of speaking rate

Voiced vowels in devoicing environments were generally weaker than vowels in non-devoicing environments in single devoicing condition (Figures 2b and 2c). They were weakened most at the fast tempo and least weakened at the slow tempo for subjects B and C. Subject A did not voice devoiceable vowels in single devoicing condition at the normal rate at all and voiced only once at the slow rate (Figure 2a). However, voiced devoiceable vowels were not weakened very much in the consecutive devoicing environments. Their intensities of some of them were in fact higher than those of other vowels.

Figure 2. The intensity ratios of vowels in devoicing environments comparing to the average intensity of vowels in non-devoicing environments



4.4 Vowel devoicing process

Jun and Beckman (1993) [1] studied vowel devoicing in Korean with relation to glottal gestural overlap and suggested that the same process applied to Japanese vowel devoicing. In Korean, too, high vowels are devoiced between voiceless stops. Korean has three types of voiceless stops: namely aspirated stops, fortis stops and lenis stops. These stops show different glottal openings and vowel devoicing rates were largely dependent on the type of preceding stops. Their results showed that devoicing was more prevalent when vowels were preceding by aspirated stops which showed a large glottal-opening gesture and less when preceded by fortis stops for which the glottal opening was much smaller and the glottis was closed well before the oral release. For vowels preceded by lenis stops, devoicing occurred more when the preceded lenis stops were in phrase initial position but less when lenis stops were in phrase medial position. Lenis stops showed different glottal movements depending on their position in a phrase: (a) when they occur phrase initially, the glottis shows a fairly large opening to start but is nearly closed when a following vowel starts, and (b) when they occur phrase medially, there is no apparent opening of the glottis. Therefore, Jun and Beckman concluded that vowel devoicing in Korean was the result of glottal gestural overlap of neighbouring vowel and consonants.

Jun and Beckman applied the same theory for Japanese vowel devoicing based on Yoshioka's study of the posterior cricoarytenoid (abductor) and interarytenoid (adductor) in Japanese (Yoshioka, 1981) [4]. His electromiographic data showed the blending gesture of the abductor and adductor when a vowel was devoiced between voiceless consonants. When a vowel was voiced between voiceless consonants, there were two separate movements of the abductor for the voiceless vowels, and the adductor showed a movement for the vowel. However, when a vowel was devoiced, the abductor showed only one big movement between the voiceless consonants and the movement of the adductor was suppressed between the voiceless consonants. Therefore, they suggested that vowel devoicing in Japanese, as in Korean, is the result of glottal gestural-overlap.

The results of my intensity measurements support the gestural-overlap account of vowel devoicing. Vowels in devoicing environments could be voiced, but voiced devoiceable vowels were weaker than ordinary vowels. This implies that it is natural for high vowels to undergo devoicing, rather than remaining voiced between voiceless consonants as a result of glottal gestural-overlap over the consonants and vowel. This is also supported by the facts that generally voiced devoiceable vowels were weakened more when the speaking rate was faster in single devoicing environments.

However, vowel devoicing seems to involve other factors as well as glottal gesture since not all devoiceable vowels became voiceless in consecutive devoicing environments. Moreover, voiced devoiceable vowels in consecutive devoicing environments were not weakened. If vowels were devoiced consecutively over two morae, it would create series of consonant clusters on the surface level, which is not a favoured structure in Japanese. The vowels can be blended into adjacent voiceless consonants but some constraints block some vowels to undergo devoicing process so that the preferred syllable structure is retained on the surface level.

5. CONCLUSION

Vowel devoicing seems to be primarily the result of laryngeal assimilation of high vowels and their adjacent voiceless consonants. The vowels could be voiced but they are weakened. That suggests that vowel devoicing is a final state of vowel weakening processes. However, the data of consecutive devoicing cases suggest that devoicing appears to be influenced by Japanese syllable structure.

REFERENCES

[1] Jun, S.-A. and Beckman, M. (1993) "A gestural-overlap analysis of vowel devoicing in Japanese and Korean", Paper presented at the 1993 annual Meeting of the Linguistic Society of America, Los Angeles, California, 7-10 January 1993

[2] Kondo, M. (1993) "The effect of blocking factors and constraints on the consecutive vowel devoicing in Standard Japanese", Poster presented at the 4th conference of Laboratory Phonology, Oxford, England, 11-14 August, 1993

[3] Kondo, M. (1994) "The durational compensation of segments within a mora in Japanese", Paper presented at the Spring meeting of the Linguistic Association of Great Britain, Salford, England, 5-7 April, 1994

[4] Yoshioka, H. (1981) "Laryngeal adjustments in the production of the fricative consonants and devoiced vowels in Japanese", *Phonetica* 38, 236-251