AN ACOUSTIC STUDY OF DEVOICING OF THE GEMINATE OBSTRUENTS IN JAPANESE

Aki HIROSE & Michael ASHBY

University College London
a.h.hirose@gmail.com; m.ashby@phon.ucl.ac.uk

ABSTRACT

Historically, the phonological system of Japanese did not allow voiced geminate obstruents, and they can be found only in recent loanwords such as /baggu/ “bag” and /kiddo/ “kid”. However, the voicing of geminates in such loanwords is problematic, and seemingly voiceless pronunciations are often to be heard.

In a nonsense word study, three speakers read words exemplifying all potential voiced geminate obstruents, together with their voiceless and singleton counterparts, and measures were made of four possible voicing cues. The duration of closure voicing, and to a lesser extent F0 perturbation, suggest unvoicing of the geminates; but F1 transition resembles that for voiced sounds, while preceding vowels are actually longer before geminates than before singletons. Overall, it seems that laryngeal activity in geminates results from a pattern of deliberate control rather than the aerodynamic challenge of maintaining voicing during a long obstruent articulation.

Keywords: Japanese; loanword phonology; geminate; devoicing.

1. INTRODUCTION

This study was motivated by the observation that phonologically voiced geminate obstruents are often realised as voiceless in Japanese.

Historically, the phonological system of Japanese did not allow voiced geminate obstruents [3]. However, this constraint has been violated in recent loanwords, and there are now many examples of voiced geminates, giving rise to minimal pairs such as /baggu/ “bag” and /bakku/ “back”, /kiddo/ “kid” and /kitto/ “kit, ” so that voiced and voiceless geminate obstruents are in contrast in Japanese loanword phonology.

However, it is frequently observed that the voicing of geminates in those loanwords is completely lost and to native speakers' ears they sound identical to their voiceless counterparts. This phenomenon is especially marked among the loanwords which were introduced earlier in the history. In contemporary loanwords voiced geminates are sometimes retained, but often also replaced with voiceless ones.

This paper aims at confirming the phenomenon of apparent devoicing of voiced geminate obstruents by acoustic analysis of a variety of voicing cues.

2. EXPERIMENTAL PROCEDURE

2.1. Voicing cues

An experiment was designed to examine specific voicing cues. The following are the four main voicing cues examined in this experiment:

1) **Closure voicing**: voicing which continues into the consonant closure [2], [4].
2) **Fundamental Frequency (F0) of V1**: F0 of vowels tend to be higher when adjacent to a voiceless consonant than to a voiced one [8].
3) **V1 duration**: Vowels tend to be longer before a voiced consonant than a voiceless one [5], [8].
4) **First formant (F1) transition**: Voiced plosives are known to have a prominent F1 transition into the vowel following them [1].

2.2. Speakers and procedure

Three speakers (2 female, 1 male) were recruited. They were native speakers of standard (Tokyo) Japanese, aged between 20 and 30.

A script was prepared using ProRec[1] (Speech Prompt & Record System) 1.01. Simultaneous speech and laryngograph (Lx) recordings were made in an anechoic chamber. The Lx waveform was later used to measure the duration of voicing and to obtain the fundamental frequency. The sampling rate of the recorded tokens was 44.1 KHz, with 16 bit quantization.

The speakers were asked to pronounce a set of nonsense words which was displayed one by one on the screen in front of them. Each word was displayed in katakana, which is the Japanese orthography conventionally used for writing loanwords. The speakers were asked to pronounce each word with HLL pitch pattern, since this is the "default" pattern in loanword and nonsense word pronunciation [4] [7]. To make the utterances as natural as possible and to avoid them pronouncing the words hyper-correctly [4], they were asked to put each word in a sentence as follows:

Soko no [ ] choudai! *Pass me* [ ] *there!*

Each set was repeated three times, with 30 seconds break between each repetition. The order of the nonsense words was changed in each repetition.

2.3. Materials

The stimuli consist of 100 nonsense words, of which 70 words have C1V1C2C2V2 structure (*C=consonant; V=vowel;* e.g. /tabbu/) and 30 have C1V1C2V2 (e.g. /tabu/). V1 is always /a/, and C2 ranged over the possible voiced obstruents in Japanese and their voiceless counterparts, i.e. /b/, /d/, /ɡ/, [dʑ], /dz/ (voiced obstruents) and /p/, /t/, /k/, [tɕ], /ts/ (voiceless counterparts). V2 is /o/, /u/ or /i/, depending on the place of articulation of the preceding consonant (C2): /o/ was chosen for V2 when C2 is an alveolar (/t/, /d/); V2 = /i/ when C2 is an alveolo-palatal affricate ([tɕ], [dʑ]); V = /u/ when C2 is a bilabial (/p/, /b/,) velar (/k/, /ɡ/) or an alveolar affricate (/ts/, /dz/).

The choice of V2 was based on the rule of Vowel Epenthesis in Japanese loanword phonology: a. [o] is inserted after the alveolar stops: [t][d] (e.g. [pæd] 'pad' → [paddo]); b. [i] is inserted after palatoalveolar affricates: [tʃ][dʒ] (e.g. [pɪtʃ] 'pitch'→ [pittɕi]); and c. [ɯ] is inserted in all other contexts (e.g. [mæp] 'map' → [mappɯ]) (See [6] for more details on Japanese loanword phonology.)

Various consonants were chosen for C1 in the target group since initially we also wanted to see if the kind of consonant in C1 affects the devoicing of the voiced geminate consonants (C2 C2). It later emerged that there is probably no productive influence of this kind.

2.4. Measurements

The measurements of the recordings were performed using SFS [2] (Speech Filling System), and examples of measurement are shown in image file 1. Measurements were made in accordance with published procedures ([2], [4]).

3. RESULTS

The following table summarises the results from the experiment with the main findings in each acoustic cue.

Table 1: Summary of the main findings for each acoustic cue.

Acoustic cues	**Main findings**
Closure Voicing	Much SHORTER closure voicing for voiced geminates than voiced singletons
F0 of V1	F0 of V1 preceding a voiced geminate is lowered TO LESS EXTENT than when preceding a voiced singleton
V1 duration	LONGER V1 duration for voiced geminates than voiced singletons
F1 transition	BOTH voiced geminates and singletons show F1 transition

3.1. Closure voicing

One of the most important voicing cues is the voicing which continues into the consonant closure, which is identified as the presence of vocal fold vibration (Image file 1-a) during the hold phase (1-b) in our measurement. Table 1 shows (a) duration of the closure voicing; (b) duration of hold phase; and (c) the ratio of voicing duration with respect to duration of the hold phase.

Table 2: Duration of closure voicing (a) and of hold phase (b) and the ratio of (a) to (b) in each consonant type. The values are the average for the three subjects.

		a. Closure Voicing (ms)	b. Duration of hold phase (ms)	**c. Ratio (a/b) (%)**
Gem	VD	46.4	108.0	**47%**
	VL	6.9	114.2	**7%**
Sing	VD	38.0	44.0	**87%**
	VL	8.6	60.5	**16%**

(Gem=Geminates; Sing=Singletons; VD=Voiced; VL=Voiceless)

As can be seen from the table, there is a big difference in the closure voicing between voiced geminates and voiced singletons: while the average ratio of the closure duration with respect to the hold phase is as much as 87% for the voiced singletons, it is only 47% for the geminate counterpart, which is hardly more than half of the value for the singletons.

A spectrogram and Lx waveform of representative tokens of voiced singleton (2-a: /tado/) and geminate (2-b: /taddo/) are shown in Image file 2. The Lx waveform of the token /tado/ shows that the vocal fold vibration continues throughout the hold phase. On the other hand, the Lx waveform of /taddo/ shows that the vocal fold vibration continues only into the very beginning of the hold phase and then stops; in other words, the voicing is only partial.

The fact that voiced geminates have much shorter closure voicing than their singleton counterparts is strong evidence for the devoicing of the voiced geminates.

3.2. F0 of the preceding vowels

Fundamental frequency (F0) of a vowel tends to be higher when the vowel is adjacent to a voiceless consonant than to a voiced one, other factors affecting F0 being equal [8]. In order to find the influence on the F0 of the following consonants, the difference between F0 at the offset of the V1 (see Image file 1-d) and the peak F0 (= the highest F0 value in the vowel; see Image file 1-e) was measured, so that we could see how much the F0 drops when preceding different consonants. The values are shown in the table below, together with the average values of peak F0 and F0 at the offset of V1. Since the value of F0 can also be affected by the quality of the consonant (C1) preceding V1, the peak F0 here was defined as the highest F0 value after the mid-point of the vowel, so as to reduce influence from C1.

Table 3: F0 values of V1 – F0 difference (a), peak F0 (b) and F0 at the offset of V1 (c). The values are the average for the three subjects.

		a. F0 difference	b. Peak F0 (Hz)	c. Offset F0 (Hz)
Gem	VD	**17.7**	228.2	210.5
	VL	**10.8**	230.8	220.0
Sing	VD	**25.3**	230.8	205.5
	VL	**12.8**	233.0	220.2

When we look at the F0 difference, which shows how much the F0 drops before different consonants, we can see that the value is smaller for the voiced geminates than for voiced singletons, on average by 7.6 Hz. This indicates that the F0 before a voiced geminate does not get lowered as dramatically as when preceding a voiced singleton. This is a second type of evidence indicating devoicing of voiced geminate consonants. The difference is illustrated in Image file 3 which shows spectrograms and Fx contours from representative tokens.

3.3. Duration of the preceding vowel

Another possible acoustic cue for the voicing distinction is the length of the preceding vowel (V1 duration). Vowels tend to be longer before a voiced consonant than before a voiceless consonant ([5], [8]), and therefore, if their V1 duration was affected by the devoicing, the tokens for geminate consonants would show shorter V1 duration than their singleton counterparts.

However, when we compare between the voiced geminates and voiced singletons, we can notice that the V1 duration for the voiced geminate group is longer than that of for the voiced singleton group in all the individual results as well as the average of the three speakers. Thus, we must say that the examination of this particular voicing cue did not indicate devoicing of the phonologically voiced geminate consonants.

Table 4: Average duration of V1 in each consonant-type. The values are the average for the three subjects.

		Preceding vowel (ms)
Gem	VD	108.3
	VL	93.9
Sing	VD	90.1
	VL	78.4

3.4. F1 transition

The fourth voicing cue which was examined was the presence or absence of first formant (F1) transition, as voiced plosives are known to have an audible F1 transition -- F1 rising at the onset of the following vowel. Since the spectrograms of speakers 2 and 3 did not show F1 very clearly, the analyses done here were based on the data from speaker 1 only. Against our expectation, however, both the voiced singletons and geminates showed similar transitions in F1, while their voiceless counterparts showed no F1 transition.

This means that, in terms of F1 transition there is no indication of a difference in voicing between the singletons and geminates. However, since the analyses were based on the data from only one speaker, more data is needed to confirm the result.

Image file 4 (singletons) and 5 (geminates) show spectrograms of representative tokens for each consonant type with attention called to the F1 transition.

3.5. Devoicing of high vowels

One interesting finding that is worth exploring further is in connection with the devoicing of high vowels. High vowels [i, ɯ] in Japanese are often devoiced when surrounded by voiceless consonants. In case of the speakers from our experiment, on average as much as 82% of all the voiceless tokens (among which voiceless geminates = 85%; voiceless singletons = 79%) had a following high vowel devoiced. On the other hand, none of the high vowels which follow voiced singletons were devoiced, and only 10% of all the tokens for voiced geminates had their following high vowels devoiced. What is more, for speaker 2 who had the smallest degree of closure voicing for voiced geminates, while ca. 96 % of all the tokens for phonologically voiced geminates were completely devoiced in terms of closure voicing, only 6% of the tokens had their following high vowels devoiced. This appears to suggest that the devoicing of high vowels in Japanese may be conditioned by phonological rather than purely phonetic considerations.

4. DISCUSSION

This paper addresses experimental data on voicing cues in single and geminate obstruents to test the hypothesis that voiced geminate obstruents in Japanese are likely to be devoiced.

Two out of the four acoustic cues examined did not support the hypothesis. But closure voicing, which is arguably the most important voicing cue, did show very clear evidence for devoicing. F0 behaviour in the preceding vowels also points to a devoicing gesture. Overall, results show that voiced geminate obstruents in Japanese are likely to be devoiced at least partially.

The question arises: what is causing the devoicing of the geminates? Is it merely a physical effect that results from the difficulty or impossibility of maintaining the vocal hold vibration for the whole duration? In fact, some researchers such as Kawahara [4] argue that the devoicing of the geminate occurs because "[…] maintaining glottal vibration during a long obstruent closure is aerodynamically challenging."

However, the data from our experiment suggests that the devoicing is more like something which is deliberately done by the speakers, rather than merely the result of the physical impossibility of maintaining vocal fold vibration for the longer duration of consonantal closure.

If it was merely because of the physical impossibility of keeping vocal fold vibration throughout the longer hold phase, we would expect that the voicing would be maintained at least as long as in singletons, since we know that the voicing can continue at least as long as singletons. However, when we look at the data for the closure voicing in (*3.1*), we realise that the voicing in the geminates often ceases much earlier than that in the singletons. This can be seen when comparing the figures in Image file 2a and b: the oscillation of the Lx waveform for the voiced geminate /taddo/ in 2-b stops around 20ms after the onset of C2, while it continues for the whole duration of the consonantal closure which lasts for around 42.2 ms in /tado/ in 2-a. This suggests that the devoicing in geminates is something the speakers do deliberately.

Early cessation of voicing could be explained as a distinctive pattern of larynx control, characteristic of geminates, and one of the properties, in addition to length, which serve to distinguish between geminate and singleton.

5. REFERENCES

[1] Chen, M. and Alwan. A. "On the Perception of Voicing for Plosives in Noise." EUROSPEECH 2001, Aalborg, Denmark, Vol. 1, pp. 175-178.

[2] Espoto, A. & Di Benedetto, M.G. (1999) Acoustical and Perceptual Study of Gemination in Italian Stops. Journal of Acoustical Society of America. Vol. 106 (4) Acoustical Society of America.

[3] Ito, J. & Mester, A. (1999) "The Phonological Lexicon." *The Handbook of Japanese Linguistics:* pp.62-100. Ed. Tsujimura, N. Blackwell Publishers.

[4] Kawahara, S. (2005) Voicing and Geminacy in Japanese: An Acoustic and Perceptual Study. UNOP 31, 87-120.

[5] Kingston, J. and Diehl, R.L. (1994) Phonetic Knowledge. *Language,* Vol. 70, No. 3, pp. 419-454.

[6] Kubozono, Haruo (2002) "Prosodic Structure of Loanwords in Japanese: Syllable Structure, Accent and Morphology." *Journal of Phonetic Society of Japan,* Vol.6 No.1: pp.79-97. Nihon Onsei Gakkai.

[7] Kubozono, H. (1999) "Mora and Syllable." *The Handbook of Japanese Linguistics:* pp.31-61. Ed. Tsujimura, N. Blackwell Publishers.

[8] Maddieson, I. (1997) *Phonetic Universals*. The Handbook of Phonetic Sciences: pp. 619-639. Ed. Hardcastle, W.J. & Laver, J. Blackwell Publishers.

---

[1] Available on-line from http://www.phon.ucl.ac.uk/resource/prorec/

[2] Available on-line from http://www.phon.ucl.ac.uk/resource/sfs/