Vowel reduction and voicing judgment of following stops – comparison between Japanese and English speakers

Makiko Aoyagi and **Masahiko Komatsu**
Sophia University, Tokyo Japan
E-mail: m-aoyagi@sophia.ac.jp, koma2@splab.ee.sophia.ac.jp

ABSTRACT

Vowel devoicing in Japanese has gradient characteristics in production and perception. English also shows gradient devoicing in the vowel weakening process. A perceptual experiment on stop consonant voicing in (C)V_1CV_2 was conducted on Japanese and English speakers to test and compare the effects of the preceding vowel V_1, gradually reduced from fully voiced, long with closure voicing to completely devoiced. Five types of /ki/ as CV_1 were spliced with a *da-ta* VOT continuum as CV_2. Subjects' judgment of the spliced *kita/kida* variations over a VOT range was compared with that of *ta-da* VOT variations. For both groups, strong effects of closure voicing and vowel devoicing were observed, and voicing judgment shifted toward voiceless with the gradual reduction of the preceding vowel, indicating the parallelism of phonetic, gradient vowel reduction and its perception.

1. INTRODUCTION

The gradient characteristics of vowel devoicing in Japanese have been reported by means of acoustic and production data, [1,2,3]. These studies generally take the stance that gestural overlap accounts for the gradient variations of devoicing phenomena.

In terms of perception, Aoyagi [4] demonstrated that the effect of devoiced vowels is also gradient in Japanese. In the experiment, five types of CV_1 (*ki-*) were prepared, differing in closure voicing, vowel duration, relative amplitude and voicing. These five CVs represented a series of vowel reduction (no quality change in Japanese) from a full vowel with closure voicing to a completely devoiced vowel. The five CVs were then spliced with a *da/ta* VOT continuum as CV_2 with two closure durations. Subjects' judgment between the spliced *kita* and *kida* along a VOT range was compared with that of *ta* and *da* in isolation. The results revealed that there were strong effects of the preceding vowel (V_1) with an accompanying closure duration effect — a full vowel with closure voicing induced voiced judgment and a devoiced vowel voiceless judgment in the following stop regardless of the VOT values. More importantly, however, the voicing judgment was not affected by the "presence or absence" of voicing of the preceding vowel (V_1) but in fact by the "degree" of voicing. As the vowel was gradually reduced, the voicing judgment shifted towards voiceless also gradually; i.e. the more reduced the vowel, the more voiceless judgment caused. This suggested that a completely devoiced vowel did not hold exclusive status, often represented phonologically with [–voice], prompting the following consonant to be perceived voiceless, but rather that a devoiced vowel was the end point of a continuum of vowel reduction in perception as well.

1.1 Devoicing in Japanese and English

English also demonstrates vowel devoicing in the vowel weakening process, as in Ch<u>i</u>cago and c<u>o</u>mmunication [5,6,7,8]. This phenomenon is usually referred to as (schwa) deletion, but acoustic and production data show weakening variations and oppose the phonological treatment of deletion [5,9]. In addition to this phonetic variation, English shares its devoicing environment with Japanese as seen in *sh<u>i</u>kaku,* "square" and *k<u>o</u>maru,* "be troubled"; and devoicing in Japanese and English could both be treated on the same basis of gestural overlap [5]. On the other hand, devoicing differs between the two languages in its frequency and dependency on speech rate and style; e.g. typical Japanese devoicing such as in *k<u>i</u>ta*, "north" (a high vowel surrounded by voiceless consonants in the unaccented initial syllable and a non-consecutive devoicing environment) is independent of speech rate [10], whereas English devoicing is affected by speech rate and style [7].

1.2 Perception effect of preceding sounds in English

Among others, closure duration and properties of the preceding sound, such as voicing, the termination manner and the amplitude decline, affect the voicing judgment of the following consonant in English as well [11,12]. The experiment in [11] used a "whispered" vowel preceding a stop across a word boundary. Thus the perceptual effect of a devoiced vowel has not been observed in English.

1.3 Objective of Study

Given the phonetic nature of devoicing shared by the two languages and the differences noted above, it is of research interest to test how English speakers respond to a devoiced vowel and other variations of full and reduced vowels. As a first attempt, we tested English speakers on the same stimuli previously used for Japanese.

2. EXPERIMENT

2.1 Stimuli

The stimuli were the same as those in [4] and described as follows (see [4] for further details).

The words *ta*, *da*, *kita* and *kida* (in the LH pitch pattern; with and without devoicing for *kita*) were recorded many times by a native Japanese speaker and digitized (sampling at 22 kHz, 16-bit quantization) for later editing. For

ISBN 1-876346-48-5 © 2003 UAB

naturalness, the variations below were derived from samples of *da* and *kida* that had the same F0 in their [a].

2.1.1 Da-ta continuum

The *da-ta* continuum was created using the method by [11]: by removing the voicing lead of *da*, and replacing each of the initial glottal cycles of *da* with an aperiodic segment of *ta* of equal duration. An amplitude ramp was given at the voicing onset to avoid unnaturalness. Seven stimuli were made with the VOT values of 0, 7, 13, 18, 23, 28 and 33 ms.

2.1.2 Five variations of the preceding syllable ki-

(1) "FullV[+CLV]": The first syllable of *kida* was extracted up to the point where closure voicing comes virtually to zero just before the burst of the following stop. This syllable had initial frication (69 ms), a full vowel duration to oral closure (82 ms) and long closure voicing (58 ms) with 2 ms to reach the burst.

(2) "FullV[−CLV]": Next the closure voicing was removed to leave 10 ms after oral closure. An amplitude ramp was given to avoid sudden cut-off.

(3) "ShortenedV": The vowel of FullV[−CLV] was then shortened to 51 ms by removing some glottal cycles from the stable section.

(4) "3-cycleV": The vowel was further shortened to 22 ms with three recognizable glottal cycles left. The shorter duration and lower amplitude of 3-cycleV are typical of reduced, partially devoiced vowels [5].

(5) "DevoicedV": Finally, a syllable with a devoiced vowel of equal duration to the total duration of 3-cycleV (91 ms) was excerpted from the *kita* ([ki̥ta]) samples. This naturally pronounced devoiced vowel was used instead of an extension of the frication part of 3-cycleV, which lacked vowel coloring and sounded unnatural.

The two extremes (1) and (5) had vowels typical before a voiced and voiceless stop respectively (the "extreme vowels"), and the others between them, (2), (3) and (4), (the "intermediate vowels") had vowels reduced in steps.

2.1.3 Splicing and closure duration

The five preceding syllables *ki-* and each of the seven *da-ta* VOT variations were spliced with two closure durations (CD) of 60 and 100 ms, making 70 stimuli (5×7×2=70). These values of closure duration were the average of voiced and voiceless stops of the recorded samples.

2.2 Subjects and procedure

28 Japanese speakers from around Tokyo ("Jpn," 9 males and 19 females, ages 20-43, average 28) and 10 English speakers who have studied Japanese or lived in Japan for or less than 4 months ("Eng," 8 males and 2 females, ages 16-45, average 25) participated in the experiment.

The subject heard five times each of the 70 types of stimuli (5×70=350) randomly arranged on a PC and click-chose the word heard from "probably *kita*," "*kita*," "*kida*" or "probably *kida*." Then the subject heard five times each of the seven *da-ta* VOT variations (5×7=35) and likewise judged between *da* and *ta*. After practice the subject proceeded at his/her pace.

"*kita*" and "Probably *kita*" were combined to count as voiceless judgment, and "*kida*" and "probably *kida*" as voiced judgment. Each of the 77 stimuli (70 *kida/kita* and seven *da-ta*) was judged 140 times by Jpn and 50 times by Eng for their voicing judgment.

3. RESULTS

3.1 Voicing judgment by Jpn

Figures 1, 2 and 3 show the rates of voiced responses by Jpn for the *da-ta* VOT variations in isolation and the spliced *kita/kida* variations.

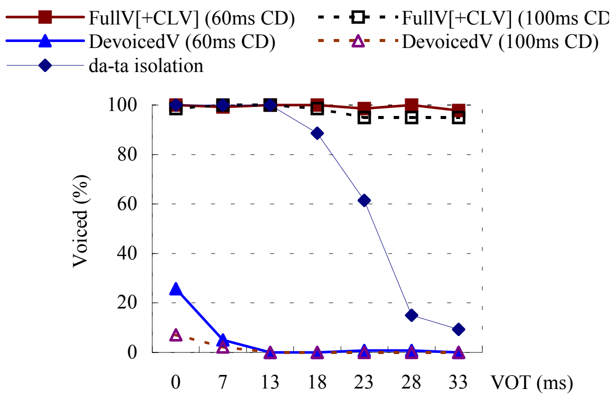

Figure 1: Jpn Voiced responses – ta-da in isolation and with extreme vowels

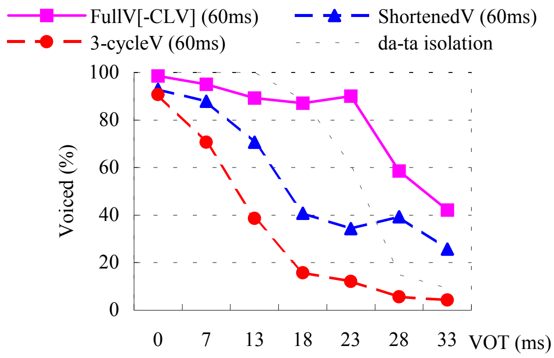

Figure 2: Jpn Voiced responses – intermediate vowels (60ms CD)

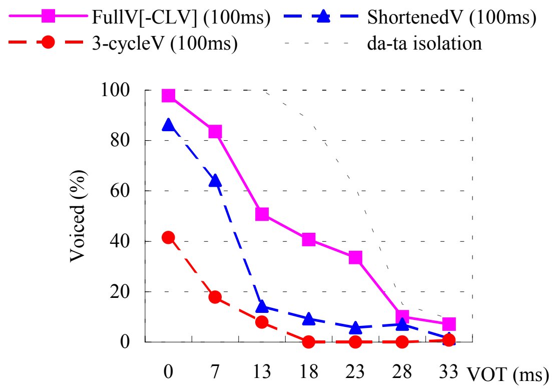

Figure 3: Jpn Voiced responses – intermediate vowels (100ms CD)

In Figure 1, the *da-ta* VOT variations in isolation show a steep decline with the approximated categorical boundary at 24.2 ms. In contrast to this baseline, the results for the extreme vowels, FullV[+CLV] and DevoicedV, stay near 100% and 0% mostly over the VOT range, showing strong effects of the preceding vowels. Some differences between the two closure durations seen near the margins are not significant, except the DevoicedV tokens at 0 ms VOT present the effect of closure duration, with the shorter closure of 60 ms for a higher voiced response rate than its longer counterpart of 100ms, $t(27) = 4.84$, $p < .01$.

Figure 2 depicts cases with the intermediate vowels FullV[−CLV], ShortenedV and 3-cycleV with 60 ms CD. Figure 3 shows the same with 100 ms CD. They all respond to VOT and show a rightward decline; longer the VOT, the more decreased voiced response rates. In addition, the voiced response rates shift downward as the vowel's duration shortens and the relative amplitude decreases from FullV[−CLV] to ShortenedV to 3-cycleV. Furthermore, The three vowel cases with the longer CD (100 ms) in Figure 3 show lower voiced response rates than their respective counterparts (60 ms) in Figure 2. An ANOVA on the subjects' voiced responses (averaged over VOT) revealed significant main effects of the preceding vowel condition, $F(2, 54) = 193.79$, $p < .001$, and of the closure duration, $F(1, 27) = 226.82$, $p < .001$, with interaction between the factors, $F(2, 54) = 7.02$, $p < .01$.

3.2 Voicing judgment by Eng

Figures 4, 5 and 6 show the rates of voiced responses by Eng for the *da-ta* VOT variations in isolation and the spliced *kita/kida* variations.

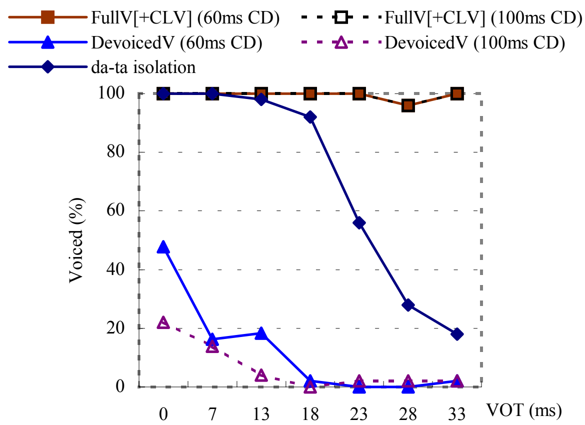

Figure 4: Eng Voiced responses – da-ta in isolation and with extreme vowels

Figure 4 shows the baseline of *da-ta* variations in isolation with a steep decline and the estimated categorical boundary at 23.5 ms, quite similar to the one for Jpn in Figure 1. It is also seen that FullV[+CLV] caused virtually all responses to be voiced as for Jpn. In contrast, DevoicedV induced voiceless responses to a great extent. However, voiced responses increased at the shorter VOT values 13, 7 and 0 ms, and the increases are more apparent than the Jpn cases. At the shorter VOT values, except 7ms, also observed is the closure duration effect with the shorter CD (60 ms) causing more voiced responses than the longer CD (100ms), $t(9) = 2.714$, $p < .05$ for 13ms and $t(9) = 2.689$, $p < .05$ for 0 ms.

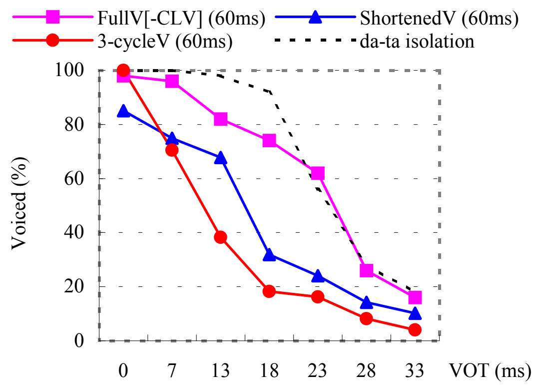

Figure 5: Eng Voiced responses – intermediate vowels (60ms CD)

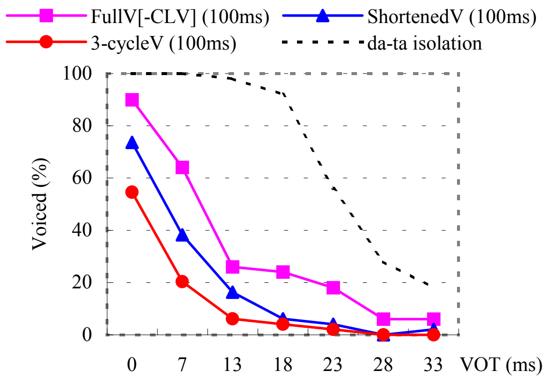

Figure 6: Eng Voiced responses – intermediate vowels (100ms CD)

Figure 5 shows cases with the intermediate vowels FullV[−CLV], ShortenedV and 3-cycleV with 60 ms CD. Figure 6 shows the same with 100 ms CD. The results present the same tendencies as the Jpn results: (i) the intermediate vowels (Figures 5 and 6) show greater response to VOT than the extreme vowels (Figure 4); (ii) as the vowel reduces, the voicing judgment shifts toward voiceless as seen in both Figures 5 and 6; and (iii) the longer closure (100 ms, Figure 6) shows more voiceless responses than the shorter closure (60 ms, Figure 5). An ANOVA on the subjects' voiced responses (averaged over VOT) revealed significant main effects of the preceding vowel condition, $F(2, 18) = 44.527$, $p < .001$, and of closure duration, $F(1, 9) = 86.247$, $p < .001$, with no interaction.

4. DISCUSSION

The lines for the *da-ta* VOT variations in isolation by Jpn and Eng are generally similar in shape and categorical boundary. The lines do not reach 0% as the continuum retains voicing properties of *da* and the replaced aperiodic segments from *ta* may not have voiceless properties strong enough to override those of voiced, especially for Eng.

For both Jpn and Eng the effects of the extreme vowels FullV[+CLV] and DevoicedV are observed. This is analogous to "anticipation of coarticulation" discussed in [11]. At the shorter VOT values 13, 7 and 0 ms, Eng is less influenced by DevoicedV than Jpn. The effects of DevoicedV and VOT interact, and Eng is more responsive to VOT and Jpn to DevoicedV. This could be a natural

ISBN 1-876346-48-5 © 2003 UAB

reflection of Eng requiring longer VOT for voicelessness. However, examining the VOT baselines of Jpn and Eng both achieving 100% at those values, the difference could also be attributed to the difference in susceptibility to this contextual effect, i.e. the magnitude of the perceptual effect of a devoiced vowel. It is not clear whether the difference comes from the stimuli being Japanese or the difference in devoicing phenomena between the two languages. Despite the difference, the general, and fairly strong effect of the devoiced vowel is present for Eng as well. Additionally, post-experiment interviews revealed that most of the Eng subjects do not devoice [i] in *kita* in either their unguided utterance or repetition led by a model utterance with devoicing. Still the devoiced vowel had a clear perceptual effect on Eng as well, causing a substantial shift of voicing judgment from the baseline toward voiceless.

Jpn and Eng also share the general pattern of gradual lowering of voiced judgment in accordance with the gradual reduction of the preceding vowel. It is not only the completely devoiced vowel that causes voiceless perception to the following stop, but also reduced vowels have the same effect though to accordingly reduced extents. The shift from FullV[+CLV] to FullV[−CLV] can be due to loss of closure voicing, the faster amplitude decline at the closure, or their combination. The further shifts to ShortenedV and to 3-cycleV can be explained by other factors associated with vowel reduction such as shorter duration and lower relative amplitude (no quality change in Japanese). The production pattern of a shorter, fast-closing vowel before a voiceless stop and a longer, lingering closure voicing before a voiced stop [12,13, 14] is reflected in perception.

This effect of the preceding vowel interacts with another effect of closure duration both in Jpn and Eng. “Shorter closure duration for voiced and longer closure duration for voiceless” is also reflective of observations in production of intervocalic stops, and consistent with other perceptual studies [12,15,16,17]. Closure duration often shows complementary relations to preceding vowel duration. Such proportional analyses including durational proportions of closure duration and closure voicing, should follow.

5. CONCLUSIONS

In light of phonetic nature of the devoicing phenomena, this study tested the effects of preceding vowels, gradually reduced from fully voiced, long with closure voicing to completely devoiced, on the voicing judgment of the following stop on Japanese and English speakers. Devoicing in Japanese, especially typical kinds, is less rate-dependent than in English; e.g. full devoicing of [i] in *kita* is almost exceptionless even in slow, read speech. As English devoicing is also phonetic, and possibly more phonetic than Japanese, a phonetic, gradient response to the gradually reduced vowels by English speakers was all the more plausible and expected. In fact, that was the case. It was not the “presence/absence” of voicing but rather the “degree” of voicing that caused increased voiceless judgment to the following stop. Voiceless judgment increased as the vowel reduction proceeded, indicating parallel relations between gradient vowel reduction including devoicing and its perception both in Japanese and English, which present similar and dissimilar properties of devoicing. Also the study suggests that coarticulatory characteristics are reflected in perception.

* I extend my gratitude to Tsutomu Sugawara and Mike Milward for their advice on this study. Part of the study was funded by Grant-in-Aid for Scientific Research of Ministry of Edu., Cul., Sport, Sci. and Tech. of Japan (KAKENHI: 14710381).

REFERENCES

[1] Kondo, M., *Mechanisms of vowel devoicing in Japanese*, Ph.D. diss., University of Edinburgh, 1997.

[2] Yoshioka, H., ”Laryngeal adjustments in the production of the fricative consonants and devoiced vowels in Japanese, *Phonetica*, 38: 236-251, 1981.

[3] Fujimoto, M., Murano, E., Niimi, S. and Kiritani, S., “Correspondence between the glottal geture overlap pattern and vowel devoicing in Japnaese,” *Proc. of ICSLP 98*, 1997.

[4] Aoyagi, M., “Contextual effect on voicing judgment of following stop consonant in Japanese,” *Proc. of ICSLP 02*: 1129-1132, 2002.

[5] Beckman, M., “When is syllable not a syllable?” in Otake, T., Cutler, A. (eds.), *Phonological Structure and Language Processing, Cross-linguistic studies*, 1997.

[6] Keating, P. “Phonetic and phonological representation of stop consonant voicing,” *Lang.*, 60: 286-319, 1984

[7] Pilchi, H. *Manual of English Phonetics,* Mouton de Gruyter, 1994.

[8] Ladefoged, P. and Maddieson, I, *The sounds of the World's Lanauages*, Blackwell, 1996.

[9] Davidson, L., “Tomato/Tmato, Potato/Ptato: Constraints on Schwa Deletion in English” poster presented at LabPhon 8, Yale U., June 27-30, 2002.

[10] Varden, J. K., *On high vowel devoicing in standard modern Japanese: implications for current phonological theory*, Ph.D. diss., Univ. of Washington, 1998.

[11] Repp, B. and Lin, H. B., “Effects of preceding context on voice-onset-time category boundary,” *J. Exp. Psych: Hum. Percep. & Perform.*, 17: 289-302, 1991.

[12] Parker, F., “The coarticulation of vowel and stop consonants,” *Journal of Phonetics*, 2: 211-221, 1974.

[13] Homma, Y., “Durational relationship between Japanese stops and vowels,” *J. of Phonet*, 9: 273-281, 1981.

[14] Sugito, M. and Kanda, Y., “Nihongowasha to cyugokugowasha niyoru nihongo no musei oyobi yusei haretsuon no onkyoteki tokucho” (Jpn. stop consonants by Jpn and Chi.), *J. of Osaka Shoin Wom. U.*, 24, 1987.

[15] Aoyagi, M., “Vowel devoicing and voicing judgment of the following stop by Japanese and English speakers”, *Sophia Linguistica*, 44/45: 91-108, 1999.

[16] Lisker, L., “Closure duration and the intervocalic voiced-voiceless distinction in English,” *Language*, 33-1: 42-49, 1957.

[17] Ando, H. & Kashino, M, “Nihongo heisashiin no chikaku niokeru muonkukancho no yakuwari” (Effects of blank duration on percep. of Jpn stop consonants,” *Proc. Acoust. Soc. of Japan*: 361-362, Oct. 1989.