UCLA Working Papers in Phonetics, No.111 (2012), pp. 162-176

Japanese consecutive devoicing as a phonetic process: the relative contribution of conditioning factors and its speaker variability

Kuniko Nielsen
(nielsen@oakland.edu)

Abstract

In many dialects of Japanese, high vowels between voiceless consonants are often devoiced. This devoicing phenomenon is generally considered a phonological assimilation process. It is almost obligatory in the Tokyo dialect, except for some marked environments in which complete devoicing is often blocked. One such case is so called consecutive devoicing, where two or more consecutive vowels are in devoiceable environments. Although several accounts of consecutive devoicing have been proposed (e.g., Kondo, 2005; Yoshida, 2004; Tsuchida, 1997), the nature of its markedness, namely whether it is phonologically or phonetically driven, is still being debated. In order to provide a more comprehensive account of consecutive devoicing in the Tokyo dialect, the current study investigated the relative contribution of various factors on its occurrence as well as its across-speaker variability. The results revealed that the manner of consonants surrounding the second devoiceable vowel (C2-C3) had the strongest effect on the likelihood of consecutive devoicing among the factors tested, followed by the quality of vowel in the following mora (V3). Further, large and gradient speaker variability in the occurrence of consecutive devoicing was observed. These results indicate that consecutive devoicing is a phonetically driven process, rather than a phonological process as traditionally considered.

1. Introduction

In the standard and many other dialects of Japanese, short high vowels /i/ and /u/ tend to be devoiced or deleted when they occur between two voiceless consonants (Han, 1962; McCawley, 1968; Vance, 1987), e.g., /hikaku/>[çi̥kakɯ] 'comparison', /kiso/>[ki̥so] 'basis', /kokusai/>[koksai] 'international', /kikai/>[ki̥kai] 'machine'. This phenomenon of Japanese vowel devoicing was traditionally considered a phonological assimilation of the feature [-voice] (e.g., McCawley, 1968). More recently, alternative analyses were proposed by Tsuchida (2001) and Varden (1998) employing Optimality Theory (Prince and Smolensky, 1993) and Feature Geometry, respectively, in which Japanese voiceless vowels are specified as [+spread glottis] instead of [-voice]. Despite the difference in featural representations, Japanese vowel devoicing is considered to be phonologically driven.

The precise phonetic implementation of Japanese vowel devoicing is not homogeneous, however. It ranges from shortened (voiced) vowel, to devoiced vowel, to complete deletion (e.g, Beckman, 1982; Tsuchida, 1997; Kondo, 1997; Maekawa and Kikuchi, 2005). It has been reported that vowels in devoicing environments are normally deleted or shortened as opposed to 'devoiced' (Beckman,1982; Keating and Huffman, 1984). Based on these facts, Jun and Beckman (1993) proposed an alternative account of Japanese vowel devoicing as a gradient phonetic process involving overlap of glottal gestures. According to this view, Japanese (and also Korean) vowel devoicing is the result of extreme overlap and hiding of the vowel's (voiced) glottal gesture by the adjacent consonant's (voiceless) glottal gesture.

It is widely known that a number of both phonetic and phonological factors affect the rate of devoicing. It is more common in fast speech than slow speech, although it is not a fast speech phenomenon (Kuriyagawa and Sawashima, 1989; Maekawa and Kikuchi, 2005; Kondo, 2005). Consonantal context, especially manner of articulation**,** is also known to affect the devoicing rate (e.g., Maekawa and Kikuchi, 2005). Vowel quality has also been shown to affect the rate of devoicing (Maekawa and Kikuchi, 2005). Further, Yoshida (2004, 2006) reported that the rate of devoicing is higher when the following vowel is /a/ as opposed to /i/ or /u/. Other phonological factors such as syllable structure and position (Kondo, 2005), morphological/word boundary (Vance, 1987), moraic position (Kuriyagawa and Sawashima 1989), and location of pitch accent (Kitahara, 2001) have also been reported to influence the likelihood of devoicing. In addition, Japanese vowel devoicing is particularly prominent in the Tokyo dialect (= Standard Japanese), and thus there is a sociolinguistic register associated with this phenomenon (Imaizumi et al., 1999; Beckman, 1994).

1.1. Consecutive Devoicing

Japanese vowel devoicing in the aforementioned environments is almost obligatory, except for some marked environments in which complete devoicing is often blocked (e.g., Maekawa and Kikuchi, 2005). One of the most studied cases is so called consecutive devoicing, where two or more consecutive vowels are in devoiceable environments (Tsuchida, 2001; Kondo, 2005; Maekawa and Kikuchi, 2005), e.g., /kikikaesu/ > [ki̥ki̥kaesu] ‘ask again’; /zokuɕutsu/ > [zoku̥ɕu̥tsu] ‘one after another’. Consecutive devoicing has been claimed to be avoided potentially due to perceptual ease (Vance 1987, Maekawa 1989), although the precise nature of its markedness is still being debated. According to the analysis by Kondo (2005), consecutive devoicing is blocked by the Japanese syllable structure, and vowels in consecutive devoicing sites became voiceless only when resyllabification of the remaining consonants was possible. Her argument was based on the observations that high vowels in the single devoicing sites were almost always devoiced, while only some devoiceable vowels became voiceless in the consecutive devoicing sites, and that devoicing never occurred in three or more consecutive morae in her data. She further argued that none of the phonetic factors that are known to influence devoicing (e.g., manners of surrounding consonants, pitch accent) consistently block the devoicing in single devoicing sites, and thus phonetic analyses cannot explain the markedness of consecutive devoicing. In contrast, Yoshida (2004) and Tsuchida (1997) proposed that it is the phonetic environment of the vowels which determines the likelihood of consecutive devoicing. Yoshida (2004) reported that the manners of surrounding consonants had a clear effect on the rate of consecutive devoicing, while the effect of the position in the word was relatively weak. Tsuchida (1997), based on her observation that single site devoicing was complete and regular while consecutive devoicing is gradient and irregular, argued that consecutive devoicing is a phonetic process, whose likelihood is dependent on various phonetic factors.

1.2. Speaker Variability of Consecutive Devoicing

In the large corpus study of connected speech by Maekawa and Kikuchi (2005), consecutive devoicing was observed in 26.4% of all eligible cases. A Similar number is reported in Kuriyagawa and Sawashima (1989), where “Coincidental devoicing” (=consecutive) was observed in 25% of cases. What their data also showed is a clear individual difference among six speakers in both slow and fast speech: in fast speech, the rates of consecutive devoicing for two speakers were over 50% (41 words out of 80) for the HL accent type, while the rate of another speaker in the same setting was 2.5% (2 words out of 80). Mimatsu et al. (1998) also reported that the occurrence of consecutive devoicing differed to a high degree across their two participants (i.e., 47.8% and 74.8%). This wide cross-speaker variability suggests that the articulatory settings of individual speakers affect the rate of consecutive devoicing. However, in previous studies on Japanese devoicing (except for the corpus study by Maekawa and Kikuchi, 2005), the data are typically collected from a relatively small number of subjects and words, which makes it difficult to examine multiple phonetic/phonological factors and individual variability simultaneously.

1.3. The current study

Previous literature has revealed a number of phonetic and phonological factors which affect the likelihood of devoicing. However, the relative importance of these factors on consecutive devoicing is not well understood, and the nature of consecutive devoicing is still controversial. In addition, speaker variability in the rate of consecutive devoicing has also been reported, although the number of participants in previous studies is small and thus the degree and pattern of variability is not known. The purpose of the present study is to elucidate the nature of consecutive devoicing by addressing these issues. We conducted an experiment with a larger number of participants as well as lexical items compared to previous studies, and examined 1) the relative contribution of various phonetic and phonological factors on the likelihood of consecutive devoicing, and 2) the individual variability in the likelihood of consecutive devoicing.

Based on the existing literature, we can identify several different factors that may interact with consecutive devoicing. These factors are: 1) position of consecutive devoicing in a word (e.g., initial vs. medial), 2) manners of consonants surrounding the second devoiceable vowel (e.g., stop-fricative, stop-affricate, fricative-fricative), 3) vowel quality for the devoiceable vowels and the following vowel, and 4) accent pattern of the word. Based on the literature, we expect each of the factors to have the following influence. As for the position of consecutive devoicing, we would expect a higher likelihood of consecutive devoicing in word-medial position if consecutive devoicing requires resyllabification of a consonant cluster, as proposed in Kondo (2005). On the other hand, if consecutive devoicing is phonetically driven (Yoshida 2004 and Tsuchida 1997), a small effect of position is expected (as reported in Yoshida, 2006). As for the manners of consonants surrounding the second devoiceable vowel, we expect combinations such as fricative-stop, affricate-stop, and stop-stop to show higher occurrence of consecutive devoicing than fricative-fricative (Fujimoto,2005; Maekawa and Kikuchi, 2005). As for the vowel quality of the vowel following the consecutive devoicing site (=V3), we expect that low vowels in V3 position would show higher occurrence of consecutive devoicing (Yoshida, 2004). Accent-less words are also expected to show higher rate of devoicing when compared to words with pitch accent (Yoshida, 2004; Fujimoto, 2004). In addition to the factors above, the effect of number of mora (in a word) was also examined. It has been shown that mean duration of a mora decreases as the number of mora increases (Port et al.,1987), and that Japanese devoicing is more common when mora duration is short (Kuriyagawa and Sawashima, 1989). Thus we expect higher rate of consecutive devoicing among words with a greater number of mora.

2. Method

2.1. Participants

Twenty-four native monolingual speakers of Tokyo Japanese (age 18–50, mean 28.25, 12 females) with normal hearing served as participants for this experiment. None of the participants had lived outside of the greater Tokyo area (Tokyo, Chiba, Kanagawa, and Saitama), and none had daily exposure to foreign languages. They were recruited from the Tokyo University of Foreign Studies population, and were paid for their participation.

2.2. Stimuli Selection

The production list consisted of 110 Japanese words, including 1) 31 test words with consecutive devoicing environment, 2) 59 words with single devoicing environment, and 3) 20 Filler words containing no devoicing environment. Lexical frequency and familiarity were determined from the NTT database (Lexical properties of Japanese, Amano and Kondo, 2003). All words had high familiarity (> 5.0 on a 7-point scale). The test words were classified according to the aforementioned factors: (1) Position (Initial vs. Medial), 2) C2-C3 Manner (F-S, A-S, S-S, S-F, S-A, A-A, F-A, and F-F)[1], 3) V1 (/i/, /u/), 4) V2 (/i/, /u/), 5) V3 (/i/, /u/, /e/, /o/, and /a/), 6) Accent (Accented vs. Unaccented), 5) Mora (3-7). For a complete list of the test words in this experiment, see Appendix A.

[1] Although A-F (affricate-fricative) is also a possible combination, out stimuli did not include words in this category.

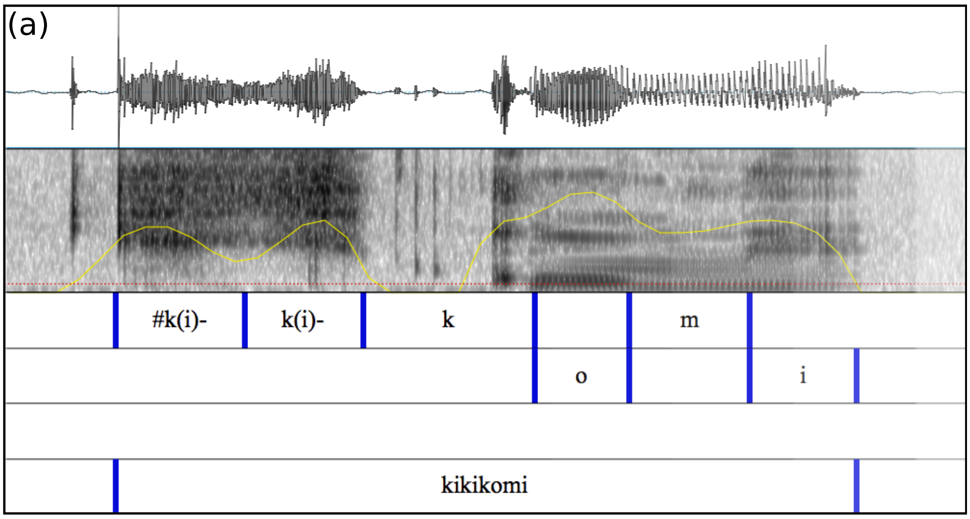
2.3. Procedure

The experiment was conducted in the Research Institute for Languages and Cultures of Asia and Africa (ILCAA) at the Tokyo University of Foreign Studies in Japan. After a brief explanation of the procedure by the author, each participant was seated in front of a computer in a sound booth. The experimental stimuli were presented using Matlab. The visual instruction (in Japanese) read as follows: "Please read aloud the words presented on the computer screen. Please pronounce them as naturally as possible." The words in the production list were visually presented one at a time, on a computer screen every 2.5 seconds, in random order. For each participant, the list was presented twice to elicit two tokens of each word. The participants' production was recorded through a Logitech A-0356A headset. Each participant's speech was digitally recorded into a computer at a sampling rate of 22100 Hz for later acoustic analysis.

2.4. Analysis

Speech signals were first segmented by the author and two research assistants using both waveforms and spectrograms in Praat (Boersma and Weenink, 2007). The rate of consecutive devoicing was then calculated for each test word and participant by obtaining the fraction of consecutive vowels that were completely devoiced/deleted. Figure 1 illustrates the waveform and spectrogram of the word /kikikomi/ 'getting information': As seen, the boundaries for consonants and vowels were marked (by the vertical lines with no break) in the tiers C and V, respectively, and each segment was labeled. Note that in some cases, it was difficult to determine a segment boundary, particularly in the case of complete vowel deletion between two identical segments as seen in Figure 1a. For these cases, the intensity reading in Praat (the thin line in the spectrogram in Figure 1) was used as a guide to determine the segmental boundary.

Statistical analysis of the data was based on linear mixed-effects modeling using the lmer function in the lme4 package (R Development Core Team, 2008). It is suitable for the current study because it allows us to include subject and items as crossed random effects, and also to combine random and fixed effects. Further, it is capable of coping with unbalanced data. Participant and Word were entered as random effects, while 1) Position, 2) C2-C3 Manner, 3) V1, 4) V2, 5) V3, 6) Accent, 5) Mora, and 6) Gender were entered as fixed effects. The dependent variable was the rate of consecutive devoicing. P-values were obtained from Monte Carlo Markov Chain estimate on the pvals.fnc( ) function.



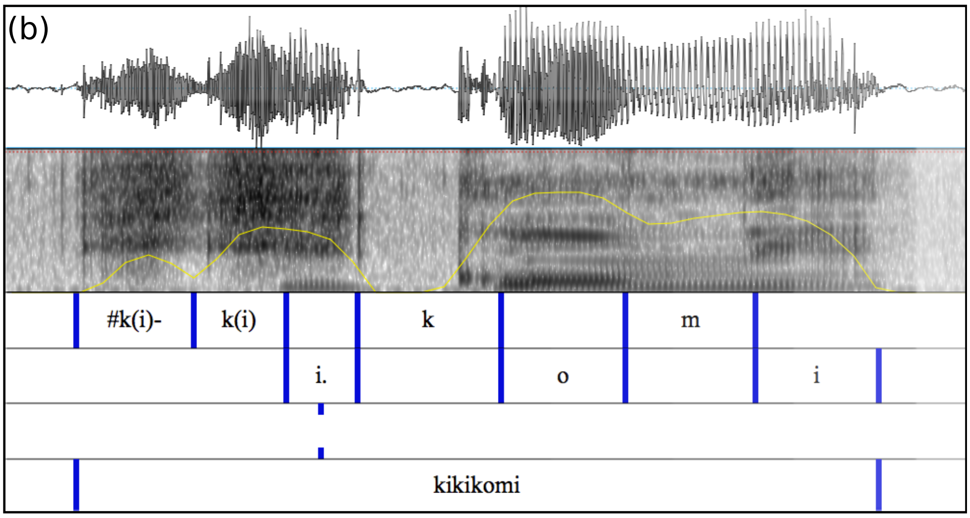


Figure 1: Examples of acoustic measurement for the word /kikikomi/ 'getting information'. 1(a) shows an example of consecutive devoicing where two consecutive vowels are deleted, while 1(b) shows a case where consecutive devoicing was blocked and only V1 (=[i]) was devoiced.

3. Results

The overall rate of consecutive devoicing was 22.7%, confirming previous studies that showed that consecutive devoicing is highly marked in Tokyo Japanese. At the same time, the data revealed that consecutive devoicing is not a fast-speech phenomenon, and does occur in the elicited laboratory-speech at a normal speech rate. Note that the devoicing rate is lower than the number reported in Maekawa and Kikuchi (2005) (26.4%), potentially due to the difference in speech style (isolated speech recorded in a laboratory vs. spontaneous and connected speech) as well as in size of the word set.

The effect of Position on the consecutive devoicing rate was significant [$t=-1.983$, $p<0.05$], showing that consecutive devoicing was more likely in word-initial position than in word-medial position. As can be seen in Figure 2, the rate of consecutive devoicing is higher for the word-initial position. At the same time, the difference between the two positions is rather small, suggesting that the position of devoicing site may not be a reliable predictor of consecutive devoicing. The implications of this result will be further discussed below. As expected, C2-C3 Manner (= the manner of articulation for C2 and C3) had a highly significant effect on the rate of consecutive devoicing [t=-5.479, p<0.001], and was shown to be the best predictor among the factors examined in the current study. As seen in Figure 3, the manner of articulation for the consonants surrounding V2 had a strong effect on the rate of consecutive devoicing, confirming the observation reported in Maekawa and Kikuchi (2005). As noted in previous studies, devoicing between two fricatives appears to be highly marked, while devoicing between fricative and stop as well as affricate and stop is highly probable (e.g., Maekawa, 1989; Kondo, 1997; Tsuchida, 1997; Yoshida, 2006).

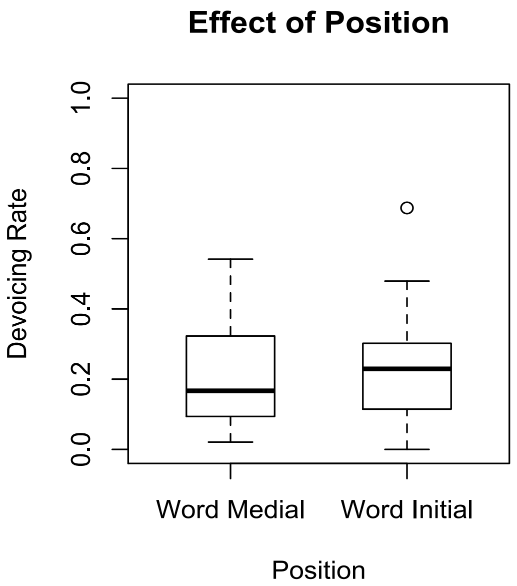


Figure 2: Effect of position on consecutive devoicing. The rate of consecutive devoicing was higher for word initial position, and the effect was significant.

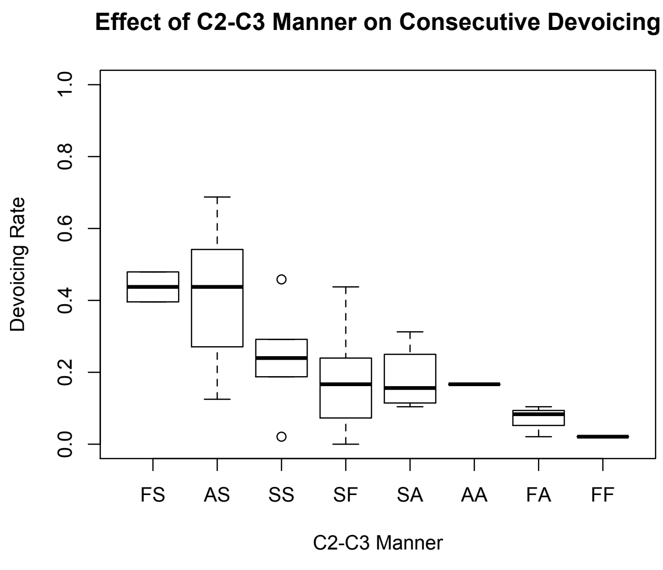


Figure 3: Effect of C2-C3 manner on consecutive devoicing. The manner of articulation for C2 and C3 had a significant effect on the rate of consecutive devoicing.

Neither V1 nor V2 had a significant effect on the rate of consecutive devoicing, although the effect of V2 showed a trend such that devoicing rate is higher when V2 is /i/ as opposed to /u/ [t=1.816, p=0.067]. The effect of V3 was significant [t=2.549, p<0.05], showing that likelihood of consecutive devoicing is influenced by the vowel quality of the following mora. Figure 4 shows the effect of V3 on consecutive devoicing: as seen, the rate of consecutive devoicing is higher when the following vowel (=V3) is not a close vowel. This result is consistent with Yoshida (2004, 2006), which showed that the rate of (single) devoicing is higher when the following vowel is /a/ as opposed to /i/ or /u/. The accentedness of the word also had a significant effect [t=-2.255, $p<0.05$], where unaccented words showed higher devoicing rate than words with pitch accent, again agreeing with the observation in Yoshida (2004) as well as Fujimoto (2004). Neither Mora nor Gender had a significant effect on the rate of consecutive devoicing [$t<1$, $p>0.1$]. Among fixed effects variables, only C2-C3 Manner and Accent had a significant interaction [t=2.046, $p<0.05$]: the effect of manner was stronger among unaccented words.

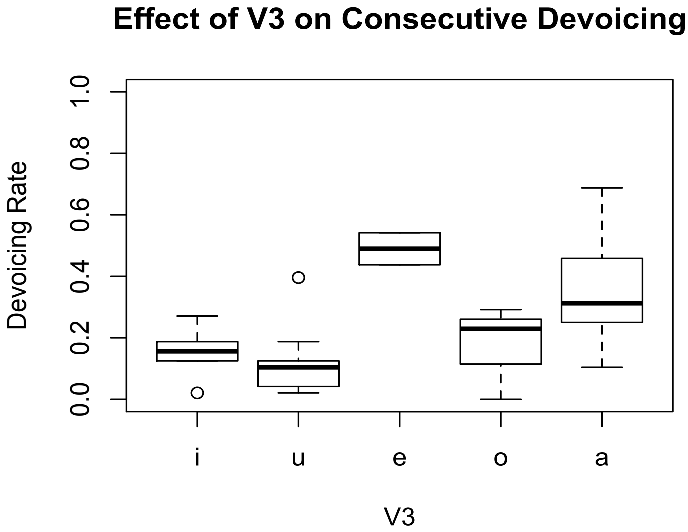


Figure 4: Effect of V3 on consecutive devoicing. The rate of consecutive devoicing was higher when the vowel following the consecutive devoicing site (=V3) was lower.

Our data also revealed that the two random effect variables, Participant and Word, had significant effects on the rate of consecutive devoicing. The model with Participant (as a random factor) had a better fit compared with a model without [χ^2 (1) = 165.38, $p< 0.0001$], and so did the model with Word [χ^2 (1) = 14.781, $p< 0.0001$]. The average rate of consecutive devoicing per participant (across all words) ranged from 2% to 63%, revealing that there is considerable cross-speaker variability in the rate of consecutive devoicing. Figure 5 shows the cross-speaker variability in the rate of consecutive devoicing for the twenty-four speakers, presented according to each speaker's average rate of consecutive devoicing across test words. As seen, the range is wide and its distribution is gradient. This shows that realization of consecutive devoicing is not categorical but gradient, indicating that consecutive devoicing is a phonetic process influenced by the articulatory setting of individual speakers. Figure 6 shows the word-level variability in consecutive devoicing, presented separately for C2-C3 Manner. As seen, the average rate of consecutive devoicing (across all participants) varies across words, ranging from 0% (*sukihoudai* 'self-indulgence') to 68% (*futsukayoi* 'hangover').

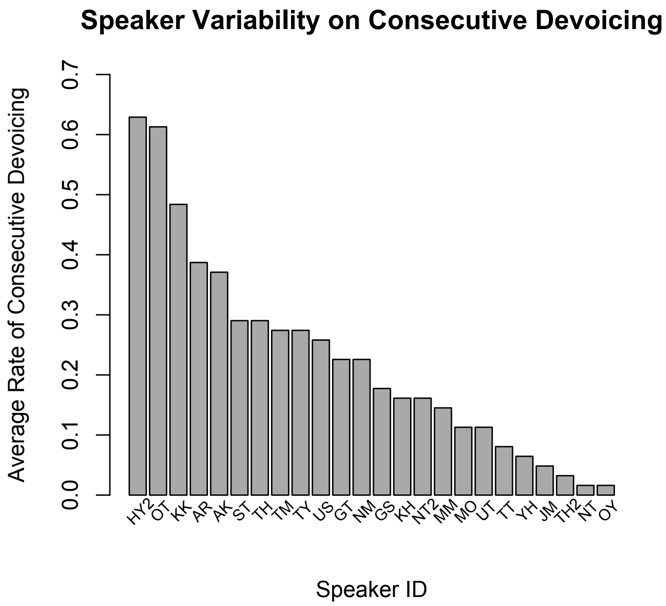


Figure 5: Speaker variability in the rate of consecutive devoicing. The average rate of consecutive devoicing per speaker, plotted from the highest to the lowest. The range was wide (63% to 2%) and its distribution was gradient.

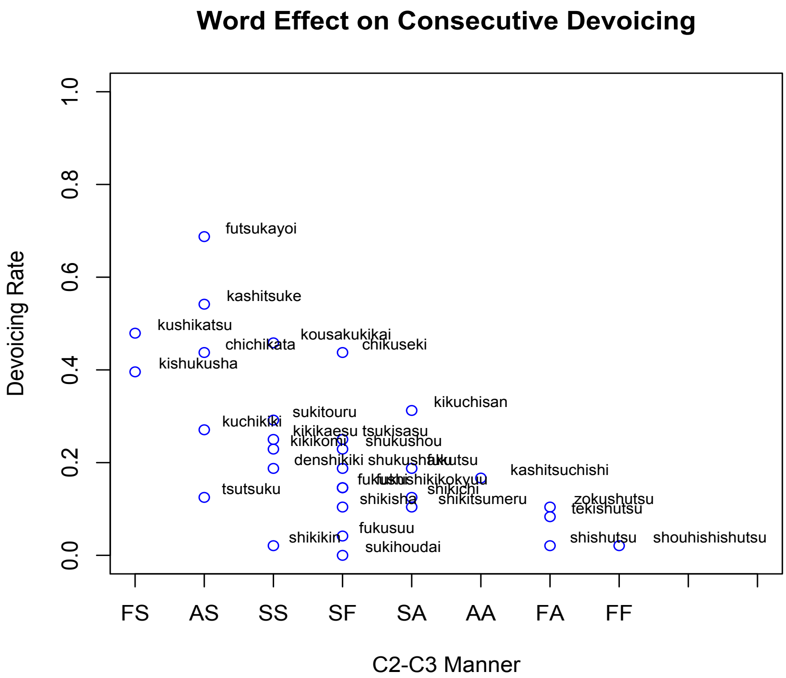


Figure 6: Word-level variability in consecutive devoicing, presented separately for C2-C3 Manner. The average rate of consecutive devoicing varied widely across test words, ranging from 68% to 0%.

Among the thirty-one test words with a possible consecutive devoicing environment, there was only one word which was never (consecutively) devoiced by any of the twenty-four speakers: *sukihoudai* 'self-indulgence'. Note that it is the only word which includes pre-/h/ devoicing (C3 is /h/ in this case), which has been shown to have a lower rate of devoicing (Fujimoto and Kiritani, 1998; Maekawa and Kikuchi, 2005). A closer inspection of the sound files revealed that the intervocalic /h/ in *sukihoudai* was produced as voiced by fifteen speakers, and as partially voiced by six speakers.

That is, of twenty-four speakers, only three produced /h/ as a purely voiceless fricative. Figure 7 shows an example of voiced /h/ in *sukihoudai*. As seen, the voicing bar continues throughout /h/ (as well as oscillation on the waveform), making the C3 voiced, subsequently creating an undevoiceable environment for the V2 (i.e., /i/ in [sukiɦo:dai]). The same pattern of /h/ voicing was observed in other words with a single devoicing site where a devoiceable high vowel is followed by /h/, such as *fuhai* 'corruption', *fuhou* 'news of a death', *shihai* 'domination', *and tehon* 'model'[2]. The implications of voiced /h/ on the patterns of Japanese devoicing will be further discussed below.

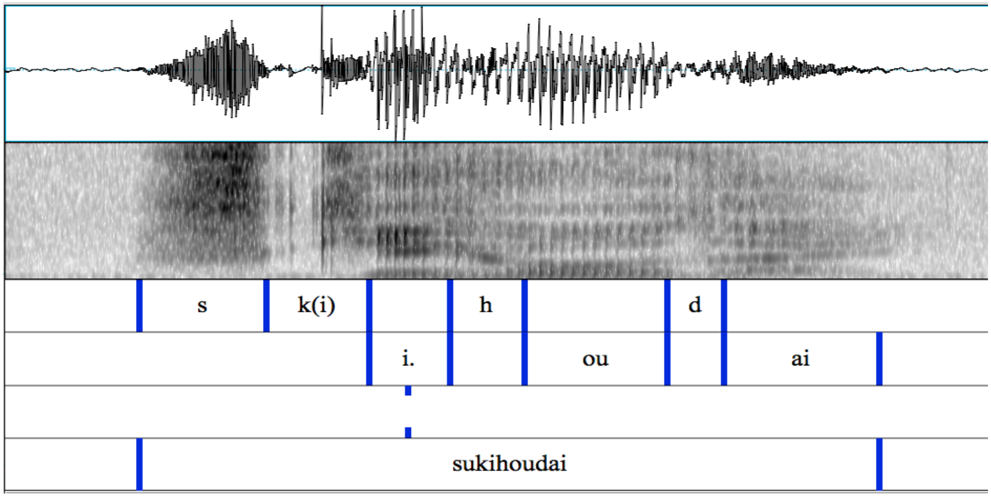


Figure 7: An example of voiced /h/ in *sukihoudai* 'self-indulgence'. The voicing bar on the spectrogram as well as the oscillation on the waveform continue throughout /h/.

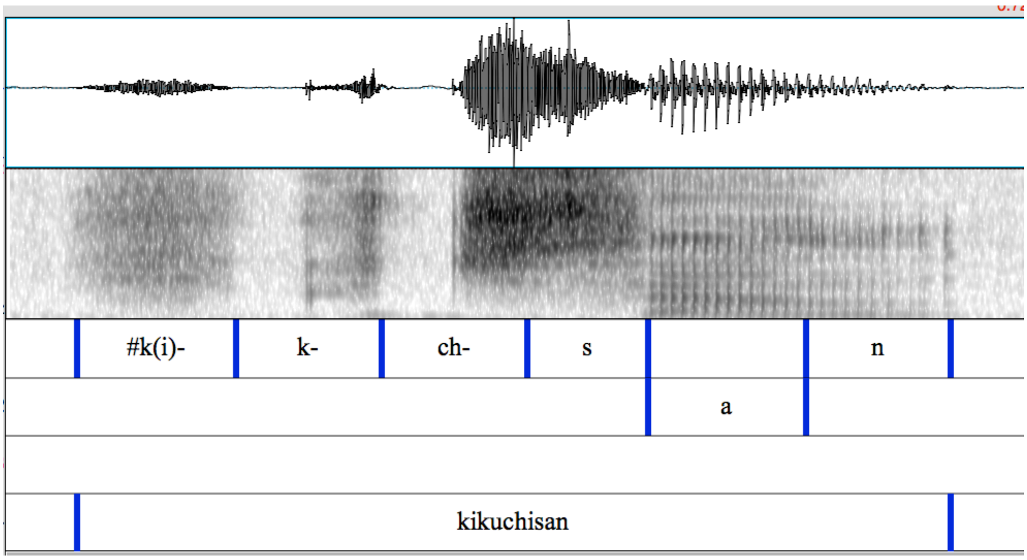


Figure 8: An example of tri-moraic consecutive devoicing in *kikuchisan* 'Mr./Ms. Kikuchi'. The first three vowels were deleted, resulting in CCCCVC.

Our data also revealed that some participants produced tri-moraic consecutive devoicing in word initial position (e.g., *kishukusha* 'dormitory'; *kichichisan* 'Mr./Ms. Kikuchi'; *kuchikiki*

[2] Note that this /h/ voicing appears to be limited to the allophone [h], but does not affect other allophones of the phoneme /h/ (i.e., [ç] and [ɸ]) in our data.

'middleman'). Figure 8 shows an example of *kichichisan*: as seen, the first three vowels were deleted, resulting in CCCCVC. This result shows that although tri-moraic consecutive devoicing is highly marked and less common than bimoraic consecutive devoicing, it does occur in elicited laboratory speech.

4. Discussion

4.1. Factors influencing consecutive devoicing

Our results showed that of all fixed-effect factors examined in the current study, manners of articulation for C2-C3 was the best predictor of the rate of consecutive devoicing, followed by V3, the accentedness of the word, and the position in a word (initial vs. medial). The data also revealed that V1, V2, number of mora, and gender of participants had no significant effect on the rate of consecutive devoicing. These results indicate that the likelihood of consecutive devoicing cannot be determined by a single factor, but is determined by a number of phonetic factors. The strongest factor, C2-C3 manner, is unarguably phonetic. The second strongest factor V3 is phonetically grounded as well: given that high vowels are inherently shorter than low or mid vowels in all languages, there is a greater probability that the glottal abductions of adjacent voiceless consonants will prevent full realization of the glottal adductions required for their voicing (= devoicing) (Lehiste, 1970). The accentedness of the vowel and position in a word are often considered phonological factors. However, realization of pitch accent involves various phonetic factors (or increased articulatory effort), such as longer vocal fold vibration, and increased pitch and intensity. In addition, phonetically speaking, word initial position is more optimal for vowel devoicing, because maximum glottal apertures for the voiceless obstruents are greater for word initial than word medial (Kuriyagawa and Sawashima, 1989; Fujimoto, 2004). To summarize, our results are in overall agreement with Yoshida (2004), which argued that in a consecutive environment, the phonetic environment determines the vowel to be devoiced.

4.2. Individual Variability

Our results also revealed wide and gradient cross-speaker variability in the rate of consecutive devoicing among the twenty-four participants. The average rate of consecutive devoicing across all words ranged from 2% to 63%, and speakers as a random effect factor (as opposed to a fixed effect factor) was shown to be a crucial element in our mixed-effects modeling. Of thirty-one test words, one speaker produced consecutive devoicing for twenty-five words, while two speakers produced consecutive devoicing for only one word (different word across the speakers). This finding is consistent with Kuriyagawa and Sawashima (1989), which reported clear individual differences in the rate of consecutive devoicing among six speakers. Previous studies have shown individual differences in laryngeal functions (e.g., Koegig et al. 2005; Hunson, 1997) indicating that speakers use different articulatory strategies in order to achieve the same phonological goals. The observed individual variability shows that the patterns of Japanese consecutive devoicing cannot be accounted for by a set of phonetic or phonological factors without factoring in individual differences, because consecutive devoicing appears to be in large part controlled by the articulatory setting of individual speakers. Further, the gradient pattern of individual variability found in our data provides further support for a phonetic interpretation of consecutive devoicing.

4.3. Consecutive Devoicing and Syllable Structure

As discussed earlier, Kondo (2005) argued that consecutive devoicing is blocked by the Japanese syllable structure, and that consecutive devoicing is possible only if the devoiced morae can be resyllabified. According to her analysis, a triple consonant cluster in onset position (e.g., shi̱ki̱chi) is illformed because it creates a trimoraic syallble which is unresyllabifiable. That is, simplex onset clusters (CCV) are bimoraic and are allowed word initially, but complex onset clusters (CCCV) are trimoraic and are not allowed in Japanese. This view makes an explicit prediction regarding the distribution of consecutive devoicing, namely, consecutive devoicing is favored in a word medial position compared to a word initial position, because a trimoraic syllable can be resyllabified into two syllables to avoid forming a superheavy syllable when they occur word medially (e.g., #CV.CCCV > #CVC.CCV), while they cannot be resyllabified if they occur word initially (#CCCV). Contrary to this prediction, our results showed a higher rate of consecutive devoicing in word-initial position. The effect was significant, however, it was not a strong predictor of consecutive devoicing in our data, and consecutive devoicing was observed in both word initial and medial position. Further, our data also showed a number of cases of consecutive devoicing where three consecutive vowels were deleted word initially, creating CCCCV strings, which would have been impossible according to Kondo's syllable-structure analysis, as the syllable would contain four morae.

Besides our finding that consecutive devoicing is not more probable in a word medial position (which is in agreement with Yoshida, 2006), there are some theoretical issues in Kondo's syllabification analysis. First, her finding on consonant lengthening (where consonants are longer in a devoiced mora) challenges her treatment of post-devoicing consonant sequence as a complex onset cluster. Syllable structure is known to influence the patterns of temporal stability, such as temporal alignment of onset clusters (Shaw et al., 2009). In complex onset languages such as English, the center (of the cluster) to anchor (e.g., vowel onset) interval is more stable across words than left or right edge to anchor interval (Browman and Goldstein, 1988), thus consonants become increasingly shorter as the onset cluster becomes more complex. On the other hand, in simplex onset languages such as Moroccan Arabic, right-edge-to-anchor is more stable, and thus consonant duration is independent of the complexity of onset (Shaw et al., 2009). If Japanese is a complex onset language (as implied in Kondo's analysis), we would expect post-devoicing consonants (which subsequently form clusters) to be shorter than consonants in non-devoicing environments. Given the reported consonant lengthening for devoiced mora, it is unlikely that the consonants flanking the deleted vowel(s) form a complex consonant cluster.

The second issue concerns Kondo's analysis of syllable weight. She argued that a triple consonant cluster in onset position (i.e., #CCCV) is illformed because it creates a trimoraic (and thus superheavy) syllable. However, except for geminates, onsets are not known to contribute to syllable weight, and an open syllable with triple consonant cluster is an unlikely form of superheavy syllable. Superheavy syallables are cross-linguistically marked, and often consist of a long vowel and coda (i.e., CVVC) (Gordon, 2006). Given that there is no evidence that onset consonants in devoiced morae contribute to syllable weight in Japanese, permitting simple cluster yet forbidding triple cluster due to the syllable weight seems to lack empirical justification.

Lastly, the Tokyo dialect of Japanese is known to be a "mora dialect", in which the basic phonological unit is the mora rather than the syllable (Kubozono, 1999; Poser, 1990; Otake et al., 1993). The aforementioned consonant lengthening indicates that mora timing is kept in devoiced morae (though it might be an undershoot), and thus it is unlikely that constraints motivated by Japanese syllable structure is responsible for the markedness of consecutive devoicing.

4.4. Pre /h/ Devoicing

In the feature classification by Halle and Stevens (1971), /h/ is assigned both [-voice] and [+spread glottis]. A high vowel which follows a voiceless obstruent and precedes /h/ is thus expected to go through devoicing in the existing analyses which consider devoicing as a phonological assimilation process of these features (e.g., McCawley, 1968; Tsuchida, 2001). However, previous studies have shown that vowel devoicing before /h/ is highly marked. For example, Maekawa and Kikuchi (2005) reported that the rate of pre /h/ devoicing was 33.88%, which was the lowest among all the voiceless obstruent contexts examined. In Fujimoto (2004), the rate of pre /h/ devoicing was 16% for Tokyo speakers (averaged over three consonants in C1 position, n=10), compared to nearly 100% for /k/, 99% for /t/, and 81% for /s/. Our data confirmed this pattern, showing that the overall devoicing rate in pre /h/ position was 12%, and in a word where /h/ is in C3 position of a consecutive devoicing environment (i.e., *sukihoudai*), the rate of consecutive devoicing was 0%.

As discussed in the results section, our acoustic analysis revealed that /h/ preceded by a devoiceable vowel (and followed by a mid or low vowel) was realized as voiced by the majority of participants, which would explain the reported low rates of vowel devoicing in pre /h/ position. This finding is also in agreement with previous articulatory studies which showed the /h/ voicing in intervocalic position. Sawashima (1969) reported that vocal fold vibration was observed throughout the duration of intervocalic /h/ in Japanese, while when /h/ follows a devoiced vowel, vibration was observed only partially. In Yoshioka (1979), intervocalic /h/ (in *'seehee*' and '*heehee*') was always accompanied by vocal fold vibration, although word initial /h/ was not. The similar patterns of voicing of /h/ has also been reported in English (e.g., Koenig, 2000; Koenig et al., 2008). The high probability of intervocalic /h/ voicing suggests that the phonetic environment of devoiceable vowel in pre /h/ position clearly differs from other devoicing environments, and that pre /h/ devoicing is a phonetically controlled phenomenon.

4.5. On the Mechanism of Consecutive Devoicing

As described in Introduction, Japanese vowel devoicing has been considered a phonological feature assimilation process on the basis of phonological constraints (e.g., McCawley, 1968; Vance, 1987; Kondo, 2005). Alternatively, Jun and Beckman (1993) argued that Japanese devoicing is a gradient phonetic process, and proposed a gestural overlap account of devoicing that is based on Browman and Goldstein (1989). More recently, Tsuchida (1997) and Fujimoto (2005) have proposed that Japanese devoicing is a combination of both phonological and phonetic processes. Tsuchida (1997), by examining the muscular activation for glottal abduction and glottal opening patterns for various voicing patterns of vowels, demonstrated that there are two distinct patterns of glottal gesture in devoicing, and proposed that there are two distinct mechanisms for Japanese devoicing: phonological and phonetic. The phonological devoicing is regular and complete, which includes devoicing of high vowels between voiceless consonants (except between voiceless fricatives and consecutive devoicing). The phonetic devoicing is a result of gestural overlap or undershoot, and thus it is irregular in occurrence and is gradient in the degree depending on phonetic factors. She noted that devoicing of non-high vowels, high vowels between voiceless fricatives, and consecutive devoicing fall in this category. Fujimoto (2005), by comparing the devoicing patterns of Tokyo speakers and Osaka speakers using photoelectric glottography (PGG), argued that devoicing in the Tokyo dialect is phonologically driven, where no voicing gesture is observed for the devoiced vowel (and there is one big glottal opening gesture for two consonants), and occurred regularly without speaker variation. On the other hand, devoicing in the Osaka dialect is phonetically driven, in which voicing gesture is observed for the devoiced vowel (and there are two glottal opening peaks for the two consonants surrounding the voicing gesture) even if the voicing is not realized, and the rate of devoicing varied among speakers from close to zero up to 100%.

Our results provide overall support for this dual-mechanism view of Japanese devoicing. As discussed earlier, the rate of consecutive devoicing varied greatly across phonetic environments as well as participants, confirming the claim that consecutive devoicing is a phonetic process. Our data also revealed that devoicing before /h/ was highly marked, likely due to the voicing of /h/. The realization of /h/ voicing was variable across participants, from 100% voicing to completely voiceless, indicating that pre /h/ devoicing is another case of phonetic devoicing, which is sensitive to its phonetic environment. At the same time, our data showed that most high vowels in canonical single devoicing environments are devoiced regularly by all participants with very little variability: 93% of the 110 sites of potential single vowel devoicing showed actual devoicing. This suggests that production of a voiceless vowel in such an environment is unlikely to occur at a later stage of speech production, but rather is assigned at the stage of speech planning without neural commands for voicing, and is thus phonological.

Future research is needed to elucidate exactly how these distinct processes derive, and how these devoiced vowels are represented in Japanese speakers' minds. If phonological devoicing is coded at an earlier stage of speech planning than phonetic devoicing, would the devoiced form be stored in lexical representations? Ogasawara and Warner (2009) showed that in word recognition tasks, listeners performed better when vowels were devoiced in the environment where vowel devoicing was expected, indicating that devoiced forms are indeed stored in lexicon. At the same time, Imaizumi et al. (1995) showed that professional teachers of hearing-impaired children reduced their vowel devoicing in order to improve their listeners' comprehension, indicating that speakers have access to a sub-lexical underlying representation (with no devoicing) in order to accommodate the needs of listeners. These findings indicate that devoiced vowels in Japanese are represented at multiple levels (i.e., lexical and sub-lexical), which is in agreement with Pierrehumbert (2002) and Nielsen (2011).

5. Conclusions

The current study investigated the roles played by various factors on the patterns of consecutive devoicing in Japanese and its individual variability, by examining data from twenty-four speakers. Mixed-effects modeling revealed that among the factors examined in the current study, the manner of consonants surrounding the second devoiceable vowel was the strongest predictor of the likelihood of consecutive devoicing, followed by the quality of vowel in the following mora, the accentedness of the word, and the position of consecutive devoicing in a word. Wide speaker variability was also observed, suggesting that the relative weight of these factors vary across speakers. Taken together, our results showed that consecutive devoicing in Japanese is a phonetic process, providing support for the claim that the phonetic environment of the vowel determines the likelihood of consecutive devoicing (Yoshida, 2004; Tsuchida, 1997). Further, our results provided additional support for the dual-mechanism account of Japanese devoicing (Tsuchida, 1997; Fujimoto, 2005), which claims that Japanese devoicing can be classified into two types of processes, phonological and phonetic.

6. Acknowledgments

This study was supported by an NSF Dissertation Improvement Grant (BCS-0547578, PI: Patricia Keating) and a UCLA Dissertation Year Fellowship. The author would like to thank Andries Coetzee for his constructive comments, and Yuko Abe and Peri Bhaskararao for their generous help with the data collection in Japan.

References

Amano, S., & Kondo, T., 2003. Nihongo-no Goi-Tokusei (Lexical properties of Japanese) Vol. 1-7, CD-ROM version, Sanseido, Tokyo.

Beckman, M. E., 1982. Segment duration and 'mora' in Japanese. *Phonetica, 39,* 113-135.

Beckman, M. E., 1994. When is a syllable not a syllable? *Ohio State Working Papers in Linguistics No. 44, Papers from the Linguistics Laboratory,* 50-69. Columbus, OH: The Ohio State University.

Beckman, M. E. & Shoji, A., 1984. Spectral and perceptual evidence for CV coarticulation in devoiced /si/ and /syu/ in Japanese, *Phonetica, 41,* 61-71.

Boersma, P. & Weenink, D., 2007. Praat: doing phonetics by computer (Version 4.5.25) [Computer program]. Retrieved May 7, 2007, from http://www.praat.org/

Browman, C. P., & Goldstein, L., 1989. Articulatory gestures as phonological units. *Phonology, 6,* 201-251.

Fujimoto, M., & Kiritani, S., 2003. Vowel duration and its effect on the frequency of vowel devoicing in Japanese: A comparison between Tokyo- and Osaka dialect speakers. *Proceedings of the 15th International Congress of Phonetic Sciences*, 3189-3192.

Fujimoto, M., 2004. Effects of consonant type and syllable position within a word on vowel devoicing in Japanese. *Proceedings of Speech Prosody 2006,* 625-628.

Fujimoto, M., 2005. Glottal opening pattern in devoiced tokens by an Osaka dialect speaker. *Journal of the Phonetic Society of Japan*, *9*, 50-59.

Gordon, M., 2006. Syllable weight: phonetics, phonology, typology. New York: Routledge.

Han, M. S., 1962. "Unvoicing of Vowels in Japanese," *Onsei no Kenkyuu 10*, 81–100.

Hanson, H. M., 1997. Glottal characteristics of female speakers: acoustic correlates. *Journal of the Acoustic Society of America, 101,* 466-81.

Imaizumi, S., Hayashi, A., & Deguchi, T., 1995. Listener adaptive characteristics of vowel devoicing in Japanese Dialogue. *J. Acoust. Soc. Amer. 98,* 768-778.

Imaizumi, S., Fuwa, K., & Hosoi, H., 1999. Development of adaptive phonetic gestures in children: Evidence from vowel devoicing in two different dialects of Japanese. *Journal of the Acoustical Society of America, 106,* 1033-1044.

Jun, S. & Beckman, M. E., 1993. A Gestural-overlap Analysis of Vowel Devoicing in Japanese and Korean. Paper presented at the 1993 Annual Meeting of the LSA, Los Angeles, 7-10 January, 1993.

Keating, P. A. & Huffman, M. K., 1984. Vowel Variation in Japanese. *Phonetica, 41*, 191-207.

Kitahara, M., 2001. *Category structure and function of pitch accent in Tokyo Japanese*. Doctoral dissertation, Indiana University.

Koenig, L. L., 2000. Laryngeal factors in voiceless consonant production in men, women and 5-year-olds. *Journal of Speech, Language Hearing Research. 43,* 1211–1228.

Koenig, L. L., Mencl, W. E., & Lucero, J. C., 2005. Multidimensional analysis of voicing offsets and onsets in female subjects. *Journal of Acoustical Society of America, 118,* 2535-2550.

Koenig, L. L., Lucero, J. C., & Mencl, W. E., 2008. Laryngeal and aerodynamic adjustments for voicing versus devoicing of /h/: A within-speaker study. *Journal of Voice, 22,* 709-720.

Kondo, M., 1997. *Mechanisms of Vowel Devoicing in Japanese*. Doctral dissertation. University of Edinburgh.

Kondo, M., 2005. Syllable structure and its acoustic effects on vowels in devoicing environments. In van de Weijer, J., Najo, K., Nishhara, T. (eds), *Voicing in Japanes,* Mouton de Gruyter, 229-245.

Kubozono, H., 1999. Mora and syllable. In N. Tsujimura (Ed.), *The handbook of Japanese linguistics*, Malden, MA: Blackwell Publishers, 31–61.

Kuriyagawa, F., & Sawashima, M., 1989. Word accent, devoicing and duration of vowels in Japanese. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, University of Tokyo, 23*, 85-108.

Maekawa, K. 1989. Boin no museika [Vowel Devoicing] in Japanese). In Sugito, M. (ed.) *Koza Nihongo to Nihongo Kyoiku, vol.2, Nihongo no Onsei, On'in*, 135-153. Tokyo: Meiji Shoin.

Maekawa, K. & Kikuchi, H., 2005. Corpus-based analysis of vowel devoicing in spontaneous Japanese: an interim report. In van de Weijer, J., Najo, K., Nishhara, T. (eds), *Voicing in Japanese*, Mouton de Gruyter, 205-228.

McCawley, J. D., 1968. *The Phonological Component of a Grammar of Japanese*. Mouton, The Hague.

Ogasawara, N., & Warner, N., 2009. Processing missing vowels: Allophonic processing in Japanese. *Language and Cognitive Processes*, *24*, 376–411.

Otake, T., Hatano, G., Cutler, A., and Mehler, J., 1993. Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language, 32,* 253-278.

Pierrehumbert, J. B., 2002. Word-Specific Phonetics. In C. Gussenhoven and N. Warner (eds.) *Laboratory Phonology VII*, Mouton de Gruyter, Berlin. 101-140.

Tsuchida, A., 1997. *Phonetics and Phonology of Japanese Vowel Devoicing.* Doctoral dissertation, Cornell University.

Tsuchida, A., 2001. Japanese vowel devoicing: cases of consecutive devoicing environments. *Journal of East Asian Linguistics* 10. 225–245.

Vance, T. J., 1987. *An introduction to Japanese phonology*. Albany, N.Y.: State University of New York Press.

Varden, J. K., 1998. *On High Vowel Devoicing in Standard Modern Japanese: Implications for Current Phonological Theory*. Doctoral dissertation, University of Washington.

Yoshida, N., 2004. *A phonetic study of Japanese vowel devoicing.* Doctoral dissertation, Kyoto University.

Yoshida, N., 2006. A phonetic study of Japanese vowel devoicing. *On'in Kenkyuu (Phonological Studies), 12,* 173-180.

Appendix A: List of test words

word	Japanese	Position	V1	V2	V3	Accent	Mora	C2-C3
chichikata	父方	Initial	i	i	a	0	4	AS
chikuseki	蓄積	Initial	i	u	e	0	4	SF
denshikiki	電子機器	Medial	i	i	i	1	5	SS
fukushi	福祉	Initial	u	u	i	1	3	SF
fukushikikokyuu	腹式呼吸	Initial	u	u	i	1	7	SF
fukusuu	複数	Initial	u	u	u	1	4	SF
fukutsu	不屈	Initial	u	u	u	0	3	SA
futsukayoi	二日酔い	Initial	u	u	a	0	5	AS
kashitsuchishi	過失致死	Medial	i	u	i	1	5	FA
kashitsuke	貸し付け	Medial	i	u	e	0	4	FA
kikikaesu	聞き返す	Initial	i	i	a	1	5	SS
kikikomi	聞き込み	Initial	i	i	o	0	4	SS
kikuchi-san	菊池さん	Initial	i	u	i	0	5	SA
kishukusha	寄宿舎	Initial	i	u	u	1	4	FS
kousakukikai	工作機械	Medial	u	i	a	1	7	SS
kuchikiki	口利き	Initial	u	i	i	0	4	AS
kushikatsu	串カツ	Initial	u	i	a	0	4	FS
shikichi	敷地	Medial	i	i	i	0	3	SA
shikikin	敷金	Initial	i	i	i	1	4	SS
shikisha	指揮者	Initial	i	i	a	1	3	SF
shikitsumeru	敷き詰める	Initial	i	i	u	1	5	SA
shishutsu	支出	Initial	i	u	u	0	3	FA
shouhishishutsu	消費支出	Medial	i	i	u	1	3	FF
shukushaku	縮尺	Initial	u	u	a	0	4	SF
shukushou	縮小	Initial	u	u	o	0	4	SF
sukihoudai	好き放題	Initial	u	i	o	1	6	SF
sukitouru	透き通る	Initial	u	i	o	1	5	SS
tekishutsu	摘出	Medial	i	u	u	0	4	SF
tsukisasu	突き刺す	Initial	u	i	a	1	4	SF
tsutsuku	つつく	Initial	u	u	u	1	3	AS
zokushutsu	続出	Medial	u	u	u	0	4	FA