Title	Acoustic Correlates of Devoiced Japanese Vowels : Velar Context
Author(s)	VARDEN, J. Kevin
Citation	明治学院大学英米文学・英語学論叢 = MEIJI GAKUIN UNIVERSITY THE JOURNAL OF ENGLISH & AMERICAN LITERATURE AND LINGUISTICS(125): 35-49
Issue Date	2010-03
URL	http://hdl.handle.net/10723/535
Rights	

Acoustic Correlates of Devoiced Japanese Vowels:
Velar Context

J. Kevin Varden

Abstract

This paper reports preliminary findings of research on the acoustic characteristics of voiceless obstruent/devoiced vowel sequences in Japanese. The starting point for the investigation is Beckman & Shoji (1984), who detailed the characteristics of /ʃ/ before /u/ and /i/, a difference used by listeners to identify the underlying vowel even in the face of total deletion. Overall results of the initial findings of the current study are consistent with those. The frication accompanying the voiceless velar stop is sufficiently colored by the underlying vowel to allow identification even when spectral information is not present. In addition, the more general fricativization of the voiceless velar stop is discussed, with spectral characteristics provided.

1. Introduction

1. 1. High Vowel Devoicing

High Vowel Devoicing (HVD) is the process whereby the high vowels /i/ and /u/ are produced without default sonorant voicing. HVD in Japanese is a hallmark of the standard dialect in particular and the Eastern Japan in general, although it has been described in western dialects as well (Imaizumi *et al.* 1999; Nagano-Madsen 2004). It is an active process applying to both well-established and new forms, and according to Imaizumi et al. (1999) is basically established in the language of children exposed to a devoicing dialect by the age of five. While discussed in detail elsewhere (Kondo 1997; Tsuchida 1997; Varden 1998), a brief description of Japanese HVD follows.

The early Generative Phonology analyses of Japanese HVD involved a context-sensitive feature change operating on high vowels between voiceless obstruents, or after a voiceless obstruent at the end of a clause (McCawley 1968). However, this descriptive analysis offers no explanatory value; in particular a change from [+voice] to [−voice] does not account for the wide range of observed voicing durations nor the large-scale frication often found at devoiced vowel sites. The analysis can be improved with Feature Geometry (Clements & Hume 1993), as in Figure 1.

Devoicing of the vowel is represented by delinking [+voice] from the vowel's Laryngeal node (Lar), with subsequent spreading of [+spread glottis] from the preceding aspirate stop or fricative[1]. The spread of [+continuant] ([+cont]) from the vowel to the stop consonant results in the allophonic affrication seen with both /t/ and /k/. Cases where vowel formants are pervasive throughout the consonantal frication, as well as the allophonic palatalization of /s/ before /i/ and labialization of /h/ before /u/, can be handled by spreading of the vowel features (not shown) from the Oral Cavity node (OC; Clements & Hume 1998) or its equivalent to the consonant's Supralaryngeal (SL) node; in cases where vowel information is lost, these vowel features can be

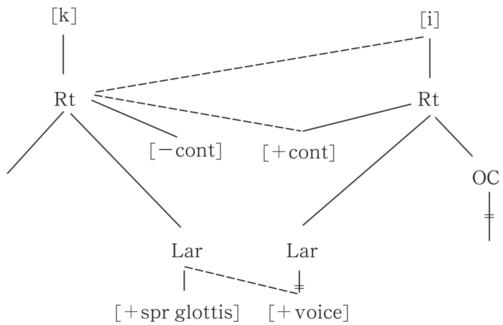


Figure 1 HVD as feature change in Feature Geometry; mora [k^xi]

delinked. Finally, for cases of full vowel deletion the vowel timing slot ([i] above) can be re-associated with the consonant's Root node (Rt), with or without spread of the vowel oral features to the consonant. This model adds explanatory value for both the source of the devoicing and the frication accompanying devoicing discussed below.

However, it does not pay due to the large variation in voicing and frication seen in the data. This variation can be handled by appealing to an Articulatory Phonology analysis (Browman & Goldstein 1986, 1989) where the individual glottal spreading gestures that produce aspiration or frication overlap with the voicing gestures for the vowels (Jun & Beckman 1993; Imaizumi et. al 1994). Figure 2 depicts, in order, fully voiced vowels, partially devoiced vowels, and fully devoiced vowels (convex curves represent glottal spreading gestures; concave curves represent voicing gestures).

As the glottal spreading gestures overlap the vowel voicing gestures, the duration of voicing decreases until total overlap produces a completely devoiced vowel. This overlap can also explain the devoicing of non-high and long vowels within a large corpus of spoken Japanese reported by Maekawa & Kikuchi (2005) and references therein. A similar overlap for the oral shaping gestures can explain the vowel formant structure riding on the frication often accompanying devoicing, as well as the variable amounts of frication produced at the vowel devoicing sites: As oral gesture overlap increases, the oral closure gestures of the obstruent can be shaped by the oral vowel gestures, resulting in a vowel-colored fricative.

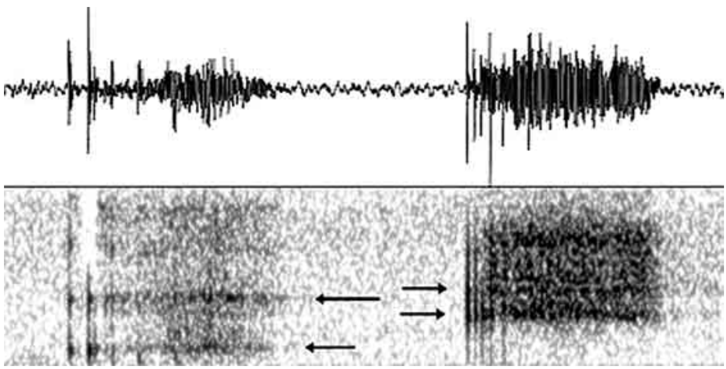
What this overlap analysis can not explain is vowel devoicing or affrication of stops at very slow SRs. It must therefore be added to a

Figure 2 HVD as overlap of glottal spreading gestures

phonological analysis of devoicing. HVD is, in short, a highly variable, goal-oriented set of processes, where a variety of articulatory strategies are used to produce the desired devoicing.

1. 2. Underlying vowel identity

Japanese vowels in devoicing environments can undergo several stages of reduction, from reduced voicing duration to total or partial devocalization to total deletion (Jun & Beckman 1993; Kondo 1997). The three components associated with vowel devoicing — loss of voicing, loss of vowel formants, and increase in frication — are not concomitant. Completely devoiced vowels can show significant formant structure on the frication at the C-V devoicing site, while other devoicing sites show no such structure. While Beckman & Shoji (1984) state that, for Japanese, formant structure is typically not observable in either spectra or spectrograms, this is highly segment- and speaker-dependent. Tsuchida (1994) found formant structure on [ʃi] and [ʃu] mora, with and without devoicing. Productions from the current data set as well clearly show formant structure (arrows) appropriate for the target [i] or [u] of the token *kuki* 'stem' by this participant. U is a devoiced [u] and I is a devoiced [i].



k^xU k^xI

Figure 3 Formant structure for kuki 'stem' with both vowels devoiced

The formant structure indicates appropriate vocalic gestures for the target vowel; the perceivable difference can easily be replicated by whispering the work *kuki*.

Although present in the above example, devoiced vowels can be reduced to very little or no formant structure, or are even deleted altogether, and yet the underlying vowel can be identified since in most dialects of Japanese as well as other languages high vowels trigger allophonic variation of the preceding obstruent: /ti/, /si/ and /hi/ surface as [tʃi], [ʃi] and [çi], respectively; /tu/ and /hu/ surface as [tsu] and [fu]. In all of these cases allophonic variation of a preceding obstruent reveals the underlying vowel. However, with /p/ and /k/ mandatory allophonic variation is not thought to occur; in addition allophonic neutralization occurs with /si/ and /syu/, which surface as [ʃi] and [ʃu]. What of these cases?

Beckman & Shoji (1984; hereafter B&S) addressed this last case by presenting listeners with excised samples of [ʃi] and [ʃu] from *shukan* vs. *shikan;* stimuli were presented to 14 participants in forced-choice scalar confidence rating trials identifying the initial mora as either [ʃi] or [ʃu]. Participants correctly identified 77% of the /si/ and 67% of the /syu/ mora that contained devoiced vowels; B&S demonstrated a robust velar coloring of the frication of [ʃu] due to coarticulation with the vowel even under conditions of complete vowel deletion, as posited by Ohso (1973) cited by B&S. Tsuchida (1994) achieved similar results, including specification of formant structure riding on the frication.

This still leaves the stops /p/ and /k/ to be accounted for. While the current data set does not include /p/, the same velar coloring of /k/ as B&S and Tsuchida (1994) found for /ʃ/ will be demonstrated. Rather than only velar coloring, though, for the four speakers considered here all repetitions of *kuki* had clear second formant structure riding on the frication due to coarticulation with the /u/.

1. 3. Velar stop fricativization & affrication

As discussed above, a voiceless velar stop is often pronounced as an affricate or a fricative; i.e. with a close oral closure. One non-invasive way to indirectly measure oral closure is to measure the resulting frication via a zero crossings trace. Zero crossings traces record the numbers of times the sound wave crosses the local zero y-axis; as oral closure increases so does frication noise, and hence so does the number of recorded zero crossings. While not being a direct measure of closure, the zero crossings trace gives us a means of quantifying the relative frication of neighboring segments of speech. An example is given below (top panel, waveform; middle panel, wideband spectrogram, bottom panel, zero crossings trace).

Note in particular the extreme frication associated with the voiced [i] across the second mora in the spectrogram and the accompanying high level of zero crossings — the level of zero crossings drops somewhat from the level for [k^x] during the lightly voiced vowel (35 ms of voicing as measured from a filtered waveform) but stays very high compared to the other vowels. Compare also the levels for the affricated /k/ in the first two mora to the aspiration of the /k/ in [ka]. By comparing zero cross-

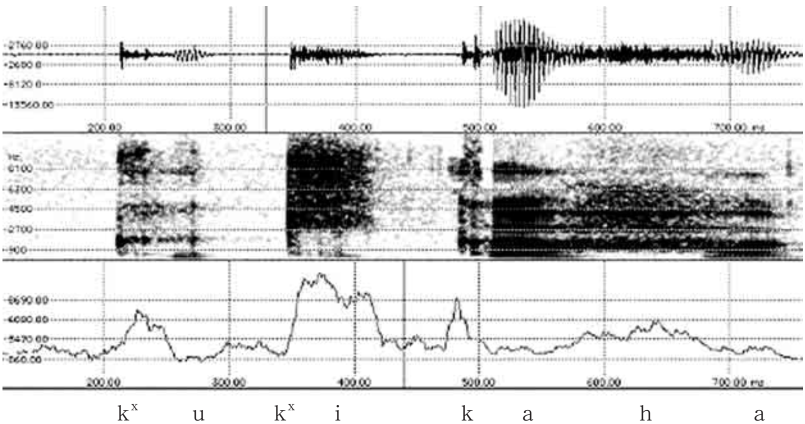


Figure 4 Waveform, wideband spectrogram and zero crossings trace of kuki 'stem'

ings of segments in this manner, the relative "noisiness" and therefore the relative degree of oral closure can be judged.

1. 4. Characteristics of frication

In the current data set, the velar stop /k/ was often fricativized or even spirantized, an alternation not unlike the spirantization of /g/ in many dialects of Japanese (Kawakami 1977). However the four participants under discussion only produced affricates. Although there is sufficient formant structure riding on the frication to recover the underlying [u], it is still worth asking into the characteristic of the frication itself. To this end statistics were generated for every mora of *kuki* that contained a devoiced vowel. In particular the first spectral moment (m 1; the spectral center of gravity) is often used to distinguish fricative place of articulation with, for instance, /s/ having a higher central frequency (Gordon 2002) than other fricatives. In addition, the second spectral moment (m 2; the standard of deviation) reflects the distribution of frication around m 1; for white noise sampled at 22.05 kHz (as these files are) the value of m 2 will be 3183 (Boersma & Weenink 2009). This measure can then be compared to the zero crossings measures as another indication of the sibilance of the consonant preceding a devoiced vowel.

2. Methods & Procedures

The experimental design of the study that generated the data set under consideration is discussed extensively in Varden (1998) and more succinctly in Varden (in press). The following gives a brief summary of the main points.

Data was collected from six female and four male Tokyo university students, ages 19–23 yrs. Stimuli consisted of ten sentences beginning with two-mora tokens containing two devoicable vowels, followed by either *to* 'and' or *ka* 'or'; e.g. *chichi to* 'crisis and'. Each speaker produced six repetitions of each sentence at slow, normal and fast SRs. The data set under consideration therefore consists of 3,600 potential devoicing sites inside tokens (10 tokens x 2 devoicable vowels x 6 repetitions x 3 SRs x 10 participants), as well as other devoicing sites within the non-token sentence material. The current study utilizes the data from four participants from various locations within the sentences, as well as all their productions of *kiki* 'crisis' and *kuki* 'stem'.

For the purposes of this study, zero crossing measures were taken for both C and V in 27 sample mora ([fu], [tʃi], [tsu], [shi], [su], [zu], [zo], [jo], [ki] and [ku]) containing both voiced and devoiced vowels. When the vowel was devoiced and heavily reduced, the single C-V envelope for the mora was divided roughly into thirds; the C measure was taken at the first third mark while the V measure was taken at the second third mark. Where both vowels were devoiced and there was no clear break between the first and second mora, the frication associated with the token was divided into halves, with each half being divided into a C and V region as above. Local transient peaks due to onset and offset closures were excluded.

The software Praat (Boersma & Weenink 2009) was used to take spectral slices from all devoiced vowel sites in productions of *kuki* using a FFT analysis with a range of 0–11 kHz and a 5 ms window. 20 ms of frication from the middle of each devoiced vowel site was sampled to minimize stop burst and coarticulatory effects, although it is impossible to fully avoid coarticulation due to the pervasive anticipatory coarticulation seen in the samples (B&S). 10 ms of the frication associated with aspiration was sampled from the /k/ of the following *ka* 'and' for comparison. From these spectra the first to fourth spectral moments (m 1 to m 4; mean, standard deviation, skewness, and kurtosis) were generated to investigate the differences in frication associated with the different underlying vowels. This paper discusses m 1 and m 2, the average weighted frequency of the frication and the distribution of frequencies around that center.

3. Results & Discussion

3. 1. Zero crossings & fricativization

As discussed above, a higher level of zero crossings is indicative of a close oral closure. The following box plot[2] from Varden (2009) gives the overall results of the zero crossings measures for data across all segments (see §1.3) and the four participants, grouped by SR. The left six boxes give values for mora containing devoiced vowels, while the right six boxes give values for mora containing voiced vowels. The left three boxes of each of these two groups represent measurements for consonants (C) and the right three boxes represent measures for vowels (V) by SR (f = fast, n = normal, s = slow).

Although appropriate statistical tests remain to be done, it is clear from the figure that devoiced vowels (second set of 3 boxes) displayed a great deal more frication than voiced vowels (fourth set of 3 boxes), with consonants preceding devoiced vowels showing on average a slightly higher degree of frication. It can also be seen that while 90% of the voiced vowels (the rightmost three boxes and points below) were produced with very low levels of zero crossings, some voiced vowels were produced with extreme frication (points above the boxes). This speaks to an independent control of oral closure, a co-articulation from a preceding noisy obstruent. The exact source of this extreme voiced vowel frication is left for further work.

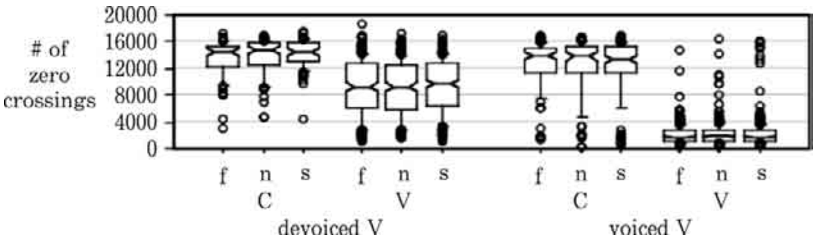


Figure 5 Zero crossing measurements for the sample, by participants and SR

Focusing now on the case of /k/ before /i/ and /u/, the following figures give composite data for all repetitions of *kiki* and *kuki*, by participant and vowel voicing condition. The four bars for each vowel voicing condition in each graph give the mean zero crossings for all mora of that condition; e.g. the first four bars of the figure for *kiki* 'crisis' give the mean zero crossings for, in order, all [k] (first bar) and [i] (second bar) in first mora containing a voiced [i], and all [k] (third bar) and [i] (fourth bar) in second mora containing a voiced [i]. Similarly the four bars in the devoiced condition give the means of all measures for [k] and [i] in first mora containing a devoiced [i] (first and second bars), and all [k] and [i] in second mora containing a devoiced vowel (third and fourth bars). Missing bars indicate no devoiced vowels produced; line segments at the tops of the bars indicate 95% confidence intervals.

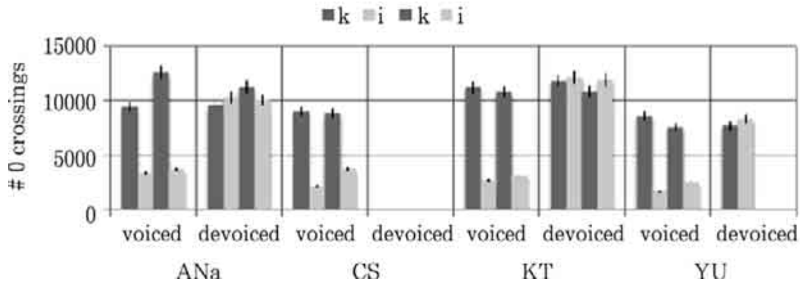


Figure 6 Zero crossings traces for *kiki* 'crisis' by participant and V voicing condition

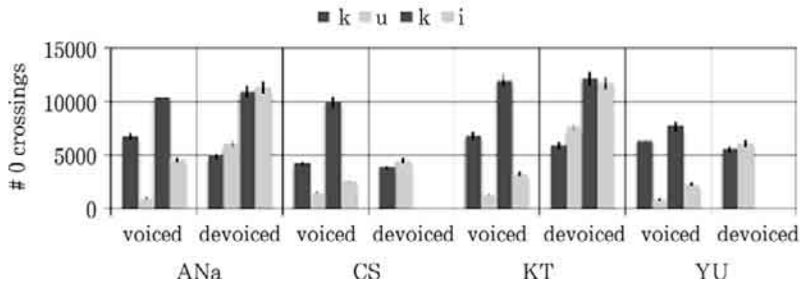


Figure 7 Zero crossings traces for *kuki* 'stem by participant and V voicing condition

Significant individual differences can be seen in the figures. For example, ANa produced slightly more frication in the first mora /k/ of *kiki* in the devoiced condition, but a higher degree of frication in the second mora /k/ in the voiced condition. The opposite situation occurs for *kuki*. Comparing across speakers shows even more inter-speaker variation. In particular, ANa and KT produced a great deal of frication in most contexts, independent of devoicing. For the devoiced vowels, YU produced less, while CS produced very little frication before what few vowels she did devoice. (For comparison, mean values for the [t] and [k] of the following syntactic particle typically ranged from 4100–5900 Hz for [to] and 5100–5850 Hz for [ka].) At least for two of the speakers, then, high levels of zero crossings are associated with [k] before [i] whether or not the vowel devoices, due to the close oral closure for [i]. This is reminiscent of the velar allophonic spirantization of /g/ that occurs in many dialects (Kawakami 1977). Mean zero crossings associated with [k] before [u] are also significant for these two speakers, although not so much as before [i], most likely due to the more open oral cavity for [u].

The close oral closure producing frication before a devoiced vowel is best viewed as an enhancement feature (Stevens et al. 1986); the increased frication facilitates vowel identity recovery by either carrying the underlying vowel's formant information or by carrying the allophonic frication characteristics of the underlying vowel. While this is not a mandatory change, e.g. does not occur in careful speech, it is interesting to speculate that allophonic variation of /k/ before high vowels might be in the process of joining the established variations of /h/, /t/, /d/, /s/ and /z/ for some speakers.

3.2. Velar coloring and spectral moments

Across the four participants the F 2 for devoiced [u] riding on frication manifests between 1600 to 1900 Hz, while the frication associated with devoiced or voiced [i] in the second mora typically had a lower bound of ≈2500 Hz, appropriate for the F 2 of /i/. Again, these formants were fairly well developed, more than just velar coloring due to coarticulation with /u/, and are indicative of pervasive anticipatory assimilation. The frication accompanying the aspiration of stops before voiced vowels also showed significant formant structure.

On the other hand, for cases where formant information is not preserved, pervasive frication of the velar stop suggests that vowel recovery might actually be a process of affricate identification: a palatalized [k^x] before /i/ vs. a velarized [k^x] before /u/. The standard procedure for identifying fricatives is examining the spectra and first spectral moment, the weighted center of gravity of the frication; for investigatory purposes spectral measures will be applied to the frication of the affricated stops.

The table below gives the average first spectral moment (m 1; the weighted average of the spectrum), the second spectral moment (m 2; the standard deviation), and counts (N) for both devoiced vowel conditions and the following particle *ka* 'or'. Note again that participants CS and YU did not devoice any second mora [i].

Table 1 Spectral moments for devoiced vowel sites, token *kuki*

	[kU]			[kI]			[ka]		
P	m 1(Hz)	m 2(Hz)	N	m 1(Hz)	m 2(Hz)	N	m 1(Hz)	m 2(Hz)	N
ANa	2024	*2110*	9	4869	*2255*	6	2163	*1810*	18
CS	1810	*1395*	6	—	—	—	2279	*1776*	18
KT	3060	*2513*	3	5966	*2131*	2	3022	*2439*	18
YU	2785	*2070*	14	—	—	—	2455	*2034*	18

As expected, the m 1 for vowels before /u/ had a much lower value than those before /i/, reflecting the respective velar and palatal coloring from the following vowels. We would also expect the values for [ka] to be higher, since the F 2 of [a] is between that of [u] and [i]. However, in retrospect the large variability seen in the spectral characteristics makes spectral moments less than useful for identifying the affricate; there appears to be simply too much variability in the relative strength of the various formants riding on the frication to achieve consistent differences. Closer examination of the frication across more contexts might lead to more generalizable results.

3. 3. Implications for HVD

Considering the widespread affrication of /k/ before both /u/ and /i/, it becomes tempting to say that for these speakers devoicing is actually occurring between voiceless affricates and fricatives. Certainly any analysis of HVD involving the spread glottis (Tsuchida 1997; Varden 1998) associated with both is on the right track. Speculatively, one might wonder whether the close oral closure producing the frication after allophonic [ç], [tʃ], [ts], [ʃ] and [k^x] preceded and hence contributed to the development of vowel devoicing by lowering the transglottal pressure differential, thereby making it more difficult to sustain vocal cord vibration. Alternatively, perhaps the allophonic variation of the voiceless stops developed in response to the need to maintain high frication in order to carry vowel formant information or frication coloring due to the underlying vowel, or both, i.e. as enhancement feature (in the sense of Stevens et al. 1986).

4. Further research

The following indications for further research are warranted. Since the affrication of /k/ before high vowels is so wide-spread in this data, it is worth pinning down the factors affecting this affrication for the entire current data set to see how they affect both the affrication process and devoicing. In addition, computation of spectral moments for all devoiced vowel sites for all tokens in the data set will be necessary to determine whether or not methods applicable to fricative identification can be applied to the affricated stops occurring at devoicing sites. Finally, perceptual studies in the manner of Beckman & Shoji (1984) to investigate devoiced vowel recovery in this context remain.

Notes

1 This analysis assumes underlying specification of [+spread glottis] for both aspirate stops as per Iverson & Salmons (1995) and fricatives (Vaux 1998).

2 Boxes contain the middle 80% of the data around the mean; means are given by medial lines. Notches in boxes represent the middle 50% distribution, while points outside the boxes are outliers.

References

Beckman, Mary, & A. Shoji (1984). Spectral and perceptual evidence for CV coarticulation in devoiced /si/ and /syu/ in Japanese. *Phonetica* 41: 61–71.

Boersma, Paul & Weenink, David (2009). Praat: doing phonetics by computer (Version 5.1.12) [Computer program]. Retrieved August 4, 2009, from http://www.praat.org/

Browman, Catherine P. & Louis M. Goldstein (1986). Towards an articulatory phonology. *Phonology Yearbook* 3: 219–252.

Browman, Catherine P. & Louis Goldstein (1989). Articulatory gestures as phonological units. *Phonology* 6: 201–231.

Clements, George & Elizabeth Hume (1993). The internal organization of speech sounds. In John Goldsmith, Ed. *A Handbook in Phonological Theory*. Cambridge: Blackwell: 245–306.

Gordon, Mathew, Paul Barthmaier & Kathy Sands (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association* 32: 141–174.

Hume, Elizabeth & David Odden (1996). Reconsidering [consonantal]. *Phonology* 13: 345–376.

Imaizumi, Satoshi et al. (1994). Vowel devoicing in Japanese dialogue between teachers and hearing-impaired or normal-hearing children: Listener adaptive characteristics of dialogue speech production. JASA 95 (5): 3012.

Imaizumi S., Fuwa K., Hosoi H. (1999). Development of adaptive phonetic gestures in children: evidence from vowel devoicing in two different dialects of Japanese. *JASA* 106. 2: 1033–44.

Iverson, Gregory K. & Joseph C. Salmons (1995). Aspiration and laryngeal representation in Germanic. *Phonology* 12: 369–396.

Jun, Sun-Ah & Mary E. Beckman (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 January 1–10, 1993, Los Angeles, CA.

Kawakami, Shin (1977) *Nihongo Onsei Gaisetsu* (*An Overview of Japanese Phonetics*). Tokyo: Oofuusha.

Kondo, Mariko (1997) *Mechanisms of vowel devoicing in Japanese*. Ph. D. dissertation, University of Edinburgh.

McCawley, James D. (1968). *The Phonological Component of a Grammar of Japanese.* Mouton de Gruyter: The Hague.

Maekawa, Kikuo & Hideaki Kikuchi (2005). Corpus-based analysis of vowel devoicing in spontaneous Japanese: an interim report. In Jeroen van de Weijer, Kensuke Nanjo & Tetsuo Nishihara, Eds., *Voicing in Japanese.* Berlin: Mouton de Gruyter. pp. 205–228.

Nagano-Madsen, Yasuko (2004). Phonetic characteristics of the Kochi dialect of Japanese: a comparison of the Yusuhara and Kochi-City dialects. *Africa & Asia* 4: 90–114.

Stevens, Kenneth N., Samuel Jay Keyser, et al. (1986). Toward a phonetic and phonological theory of redundant features. In J. S. Perkell & D. H. Klatt, Eds., *Invariance and Variability in Speech Processes.* Hillsdale, NJ: Lawrence Erlbaum. pp. 426–449.

Tsuchida, A. (1994). Fricative-vowel coarticulation in Japanese devoiced syllables: Acoustic and perceptual evidence. *Working Papers of the Cornell Phonetics Laboratory* 9: 183–222.

Tsuchida, Ayako (1997) *The Phonetics and Phonology of Japanese Vowel Devoicing.* Ph. D. dissertation, Cornell University.

Varden, J. Kevin (1998). *On High Vowel Devoicing in Standard Modern Japanese: Implications for Current Phonological Theory.* Ph.D. dissertation, Dept. of Linguistics, University of Washington. ＜http://www.meijigakuin.ac.jp/˜varden/devoicing/dissertation/＞

Varden, J. Kevin (2009). Acoustic correlates of devoiced vowels in Japanese. Poster presented at the 158th meeting of the Acoustic Society of America, San Antonio, TX, USA, October 26–30, 2009.

Varden, J. Kevin (in press). Overlap at the interface: High Vowel Devoicing in Japanese. *Karuchuru* 4.1.

Vaux, Bert (1998). "The laryngeal specifications of fricatives." *Linguistic Inquiry* 29(3): 497–511.