More Acoustic Traces of "Deleted" Vowels in Japanese [1]

ALICE FABER & TIMOTHY J. VANCE
Haskins Laboratory & Connecticut College

1. Introduction

1.1 Devoiced vowels in Japanese

A number of well-known languages exhibit vowel devoicing, including Korean, but the phenomenon has been studied most intensively in Japanese (Jun, Beckman, Niimi, & Tiede, 1998:1). The standard allophonic account for Japanese says that a short high vowel, /i/ or /ɯ/, is devoiced when it is flanked by voiceless consonants or, word-finally, when it is preceded by a voiceless consonant (Nihon Onsei Gakkai, 1976: 748; Alfonso, 1971: xxviii), but this is just a crude first approximation. For one thing, non-high vowels occasionally exhibit devoicing as well, although not nearly as often as high vowels (Han, 1962: 84–85; Vance, 1987: 48–49). Several additional factors may also affect the probability of devoicing in high vowels. For example, it has long been claimed that devoicing is inhibited in an accented syllable (Han, 1962: 81–82; Kuriyagawa & Sawashima, 1989; Imai, 1997: 48), although some studies have challenged this conventional wisdom (Haraguchi, 1984; Tsuchida, 1997; Kitahara, 1998), and it may be that there

[1] The research summarized in this report was supported by grant HD-01994 from the National Institutes of Health to Haskins Laboratories and by an endowed chair research allowance from Connecticut College. We would like to thank Louis Goldstein, Cathi Best, and Kathy Harris for their helpful discussion concerning this project; none of these colleagues bear any responsibility for flaws in our presentation or reasoning. We would also like to thank Mike D'Angelo and Philip Rubin for a helpful modification of the HADES software package.

1196

has been a diachronic weakening of the inhibiting power of accent in recent decades (Nihon Onsei Gakkai, 1976: 748; Kondo, 1997: 83–84). It has also been claimed that certain neighboring consonants favor devoicing more than others (Martin, 1952: 14; Han, 1962: 88–90), although the evidence for such differences is not very clear (Imai, 1997: 53). A third common claim is that devoicing in consecutive syllables is avoided (Martin, 1952: 14; Han, 1962: 91), but it can and does occur (Haraguchi, 1984; Imai, 1997: 50). When each of two or more consecutive syllables contains a short high vowel in the devoicing environment, the probability of devoicing in a particular syllable seems to depend on both accent and morphological structure (Vance, 1992; Kondo, 1997: 100–147).

There is no controversy about the relevance of tempo and style to the probability of devoicing in Japanese. Devoicing is less frequent in slower and more careful speech (Sakuma, 1929: 232-233; Kuriyagawa & Sawashima, 1989; Jun, Beckman, & Lee, 1998: 43). For example, professional teachers of hearing-impaired children devoice less in speech directed to such children than in speech directed to normal-hearing listeners (Imaizumi, Hayashi, & Deguchi, 1995).

There is also considerable variability in the phonetic realizations of devoiced vowels. Spectrograms sometimes show energy at the appropriate formant frequencies even in the absence of any voicing Han (1962: 83), and Jun, Beckman & Lee (1998: 50) sort tokens into three categories: voiced, partially devoiced, and completely devoiced. They define partial devoicing as "shorter than 30ms and show[ing] weak energy, at low frequencies only." Generally speaking, if the consonant preceding a devoiced vowel is a stop, an acoustic interval corresponding to the devoiced vowel is easily identifiable on waveforms and spectrograms, but if the preceding consonant is a fricative, such an interval is typically absent.

1.2 Evidence for the phonological presence of devoiced vowels

As far as phonological status is concerned, the consensus is that devoiced vowels, whether phonetically present or not, are allophones of the same phonemes as their voiced counterparts. This analysis certainly matches the intuitions of native speakers, and several lines of argument converge in supporting it. First, kana spelling makes no distinction between syllables with voiced and devoiced vowels; *tsuki* 'moon', which can be pronounced [tsʲkʲi], begins with the same letter as *tsugi* 'next', normally pronounced [tsɯgʲi]. Of course, orthography raises a chicken-and-egg problem: it may be that kana spelling simply reinforces intuitions that speakers would have anyway, but it may also be that spelling determines those intuitions to some extent (Vance, 1987: 54). Incidentally, in 1900 Ueda Kazutoshi proposed a novel romanization for Japanese that simply left out devoiced vowels, but it was swiftly rejected (Takebe, 1977: 305).

A second argument in favor of ascribing phonological status to devoiced vowels involves phonotactics. Bloch (1950) consistently treats devoiced vowels in certain environments as phonologically absent (Vance, 1987: 52–54), and as a result, his phonemic transcriptions are full of word-final obstruents and obstruent clusters. Not only are these transcriptions phonotactically unwieldy, they are highly counterintuitive (Han, 1962: 82).

A third argument for phonemic status for devoiced vowels rests on anticipatory coarticulation. Studies using English speakers have shown that articulatory activity for a segment may begin several phonological segments earlier (Bell-Berti & Harris, 1982; Keating, 1988; Boyce, Krakow, & Bell-Berti., 1991) and that this articulatory pattern is reflected in spectral characteristics of these preceding segments (Yeni-Komshian & Soli, 1981; Nittrouer, Studdert-Kennedy, & McGowan, 1989; Goodell & Studdert-Kennedy, 1993). For example, Japanese /ç/ has lower frequency spectral peaks and edges when followed by /ɯ/ than when followed by /i/, even when the vowel is devoiced and seems to have disappeared into the frication noise (Beckman & Shoji, 1984; Endo, 1998). Similarly, release bursts for /k/ have lower frequency spectral prominences when followed by /ɯ/ than when followed by /i/, again regardless of vowel voicing (Endo, 1998; Faber & Vance, submitted).

Presumably as a consequence both of phonotactics and of anticipatory coarticulation, native speakers of Japanese perceive devoiced vowels as vowels and can discriminate between devoiced /i/ and devoiced /ɯ/, although not quite as accurately as between voiced /i/ and voiced /ɯ/ (Beckman & Shoji, 1984; Endo, 1998). Phonetic transcriptions such as Alfonso's (1971: xxviii), in which, for example, /Qk/, /kik/ and /kuk/ are all transcribed identically as [k̚k], are simply inaccurate. The perceptible effects of anticipatory coarticulation, together with the variability of devoicing (which means that a high vowel in the devoicing environment can be pronounced as a voiced vowel), leave no real doubt about the phonological existence of devoiced vowels.

1.3 A gestural overlap account of Japanese devoiced vowels, and its implications

Jun & Beckman (1993) conclude, on the basis of both acoustic and physiological evidence, that the observed distribution of devoiced and apparently deleted short high vowels in Japanese results from the degree of overlap between the supralaryngeal gestures responsible for the quality of the vowels in question and the laryngeal abduction required for the flanking voiceless consonants. Given sufficient overlap, there will be no acoustic interval in which vocal fold vibration is evident, and it may not even be possible to identify an interval—even a voiceless interval—containing noise-excited formants. Nonetheless, the presence of these supralaryngeal gestures can be inferred from their acoustic effects on immediately preceding consonants: [ç] has lower-frequency spectral peaks when followed by /ɯ/ than when followed by /i/, just what would be expected on the basis of the differing supralaryngeal configurations required for the two vowels (Beckman & Shoji, 1984).

This is not the only potential spectral effect of these obscured supralaryngeal gestures on adjacent consonants. Studies of anticipatory coarticulation in English and in other languages reveal that articulatory movements for a given target may begin early enough to influence several preceding segments (e.g., Boyce, Krakow, & Bell-Berti, 1991). Thus, the configuration of the various articulators during any one phonological segment may be influenced by the gestural requirements of multiple segments in a word; further, the number of segments so affected is a function both of the gestural needs of those segments and of the speech rate (reflected in the number of segments occurring in a given temporal interval).

It would not be surprising, then, if initial /s/ in the English words *stay* and *stow* were to differ acoustically due to anticipation of the vowels /e/ and /o/, despite the intervening /t/. But, because of the intervening /t/, the /s/s in *stay* and *stow* might be less different from each other than the /s/s in *say* and *sew*. If these English words are compared with the superficially similar Japanese words *sute* and *suto*, the initial /s/ in the Japanese items might well be influenced to some extent by the word-final /e/ and /o/. However, this influence might be lessened by the presence of *two* intervening segments: /t/, as in the English items, and an additional /ɯ/. As a result of this difference in phonological structure, /s/ should be more similar in Japanese *sute* and *suto* than in English *stay* and *stow* (or in Japanese *se* and *so*); that is, /s/ should be less influenced by the final vowel in the longer Japanese items than in the longer English items.

The goal of the present paper is to investigate this hypothesis. We do this by comparing acoustic characteristics of fricatives in English and Japanese and by looking at the magnitude of the differential effects of vowels differing in frontness on the fricatives. The gestural overlap account of vowel devoicing in Japanese predicts that the fricatives in Japanese *sute* and *suto* should be less different than those in English *stay* and *stow*, because the former contain four phonological segments while the latter contain three. If acoustic differences between /s/ in *sute* and /s/ in *suto* are observed, that by itself would not count as evidence against our hypothesis; our hypothesis is that such differences would be consistently smaller in magnitude than differences between /s/ in *se* and /s/ in *so*.

2. Methods

2.1 Participants

Four speakers of Japanese and three of English were recruited through acquaintance networks. The Japanese speakers are all from the greater Tokyo area, while the English speakers are from various parts of the US; none speaks with a marked regional accent. All speakers were naive to the purpose of the recording, and the Japanese speakers thought that our purpose was to test their choice of accent patterns. Demographic details about the seven speakers are provided in Table 1.

Table1: Participants in the study

	Speaker	Gender	Age	From
English	E1	Female	16	Connecticut
	E2	Male	42	Missouri
	E3	Female	45	Rhode Island
Japanese	J1	Female	50	Tokyo
	J2	Male	42	Tokyo
	J3	Female	23	Tokyo
	J4	Female	39	Tokyo

2.2 Corpus

Japanese test items consisted of pairs of words differing in the vowel in the final syllable. In addition to monosyllables *se/so* and *shi/shu*, there were four pairs typically produced with devoiced vowels in the first syllable: *sute/suto*, *suki/suku*, *tsute/tsuto*, *shitee/shitoo*. The corpus also contained two pairs of words typically produced with fully voiced vowels in the first syllable: *tsugi/tsugu* and *sude/tsudo*. Three additional items, *fuku*, *to*, and *mitee*, were included to facilitate some additional comparisons. Five additional items (*bi*, *ne*, *kugi*, *kizu*, *puro*) served as foils. Subjects produced ten tokens of each item in a carrier sentence "*Sore kara* _____ *da yo*", in two separate randomizations. The contrasting pairs were intended to have unaccented first syllables, because of reports in the literature that accented vowels are less likely to be devoiced than unaccented vowels are. However, two speakers consistently produced *tsudo* with an accented initial syllable; one speaker was so unsure of what accent pattern to use that she consistently tried out both versions between repetitions of the full list. This speaker also produced several tokens of *sute* with initial accent (and a devoiced vowel).

English test items likewise consisted of pairs of words differing in vowel quality. In addition to the CVs *see/sue*, *say/sew*, and *she/shoe*, there were three pairs with initial clusters: *steep/stoop*, *stay/stow*, and *scale/skoal*. Two items, *tee* and *two*, were included to enable some further comparisons. An additional ten items (*bee*, *kick*, *day*, *cake*, *peak*, *doe*, *mane*, *light*, *boat*, *good*) served as foils. Subjects produced ten tokens of each item in the carrier "*It's a* ______, *Helene*", in two separate randomizations.

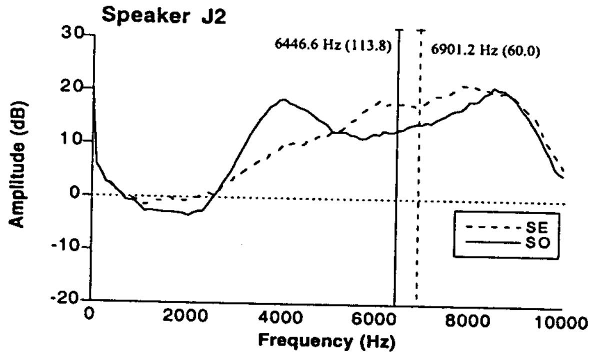
2.3 Data handling

All items were recorded on a Sony portable DAT recorder, and converted to Haskins PCM format (Whalen, Wiley, Rubin, & Cooper, 1990) at a 20kHz sampling rate for acoustic analysis. While some acoustic information about fricatives, especially /s/, might appear at frequencies higher than 10kHz, the maximum observable with a 20 kHz sampling rate, this was the maximum usable sampling rate for the HADES (Rubin, 1995) waveform editing and analysis software available to us.

Fricative spectra are acoustically complex, and typically involve one or more high-amplitude bands of aperiodic noise. In addition, there may be a narrower-band resonance that is contiguous with one of the formants of an adjacent vowel. As a result, it is extremely difficult to provide an adequate characterization of individual fricatives in terms of just one or two parameters. Choice of measurement parameter has in the past been based on a combination of measurement ease (e.g., location of spectral peaks or boundaries of the fricative noise), computational ease, and theoretical considerations. Our strategy in the present investigation is somewhat different (see also Evers, Reetz, & Lahiri, 1998). While we include some measures of the computationally-simple sort (location of peaks in the fricative noise, spectral center of gravity (=centroid: Nittrouer, Studdert-Kennedy, & McGowan, 1989; Nittrouer, 1995)), these measures were not always informative, for reasons we will explain. Therefore, we derived some additional measures from systematically comparing fricative spectra from tokens that we expected might be different.

As evident from Figure 1, which compares the spectra of 10 tokens each of speaker J2's *se* and *so* at the midpoint of each fricative, fricatives may differ in the frequency and amplitude of spectral peaks. These tokens do differ significantly in centroid, as marked by the two vertical lines, but, because the peaks are closer together for *se* than for *so*, the spectra are more different globally than they appear to be based simply on the centroid. Further, even on these averaged cross-sections, it would be difficult to mark the lower boundary of the fricative noise with any degree of reliability; especially for *se*. However, the increase in amplitude is much steeper for *so* than for *se*. As a result of this difference in slope, combined with the shift downward in frequency of the first spectral prominence, these spectra can be divided into three regions: a low frequency range containing very little energy, a mid frequency range in which there is apparently more energy present in *so* spectra than in *se* spectra, and a high frequency range in which

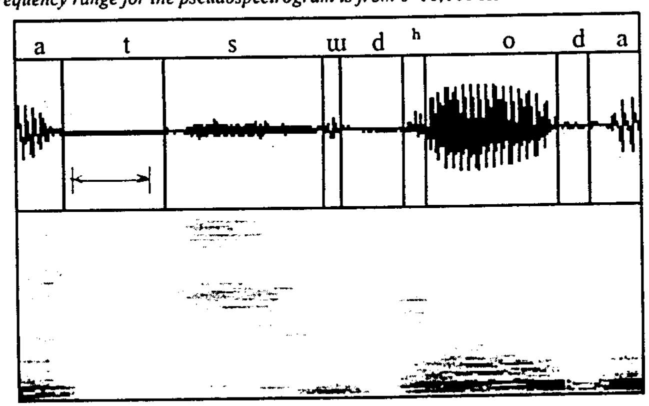
Figure 1: Averaged and smoothed spectral cross-sections for /s/ in 10 tokens of se *(dashed line) and 10 tokens of* so *(solid line), at fricative midpoint, produced by speaker J2.*



the reverse is true.

Because HADES can calculate centroids in batch mode, our first attack on analyzing the fricatives in the present corpus was to compare the centroids for the CV tokens in each speaker's corpus. If these were not different, we then proceeded to compare spectral cross-sections for these same tokens to see if there were spectral differences that would not be reflected in the centroids. If we could not identify a significant difference between spectra of CV tokens differing only in vowel (perhaps due to our limited sampling rate), we did not analyze that speaker's segmentally more complex tokens. However, if we found differences in the CV tokens, we then measured the more complex tokens, using the same measure, to determine whether vowel-conditioned differences, if any, were comparable in magnitude to those observed for the CV tokens.

Figure 2. Speaker J3: Sample waveform and pseudo-spectogram of a token of tsudo, *with phonetic transcription for orientation. Vertical lines represent the position of events in the token, from left to right: end of voicing for /a/, onset of frication for [s] in /ts/, end of frication (coinciding with onset of voicing for /ɯ/), beginning of closure for /d/, release of closure for /d/, onset of voicing for /o/, end of voicing for /o/, and release of closure for /d/ in the carrier phrase. The horizontal line in the closure interval for /ts/ represents 50 msec. The frequency range for the pseudospectrogram is from 0–10,000 Hz.*



Analysis proceeded as follows: Individual tokens were excised from larger files after it was verified that they had been read as intended, and that the target items were not preceded or followed by a pause. The first stage in analysis was to delineate segmental boundaries, based on screen displays of waveforms and pseudo-spectograms.

Figure 2 illustrates the maximal set of landmarks, represented by vertical lines in the waveform portion of the display; a phonetic transcription is provided for ease in identifying individual segments. Based on the locations of these boundaries, durations of individual segments were calculated. In addition, centroids were automatically calculated from DFT spectral cross-

sections[2] at five locations within each fricative: 15 msec after the onset of frication noise; 15 msec before the end of the frication noise; and, at three additional locations dividing the fricative into quarters. The edge measurements were offset 15 msec to avoid distortions based on formant excitation. If the CV tokens did not differ significantly in centroid frequency, complete spectral cross-sections at the same five frames were saved as text files, averaged, and plotted. On the basis of these plots, one of which appears in Figure 1, one or two spectral regions in which the two types appeared to contrast were identified. Using macros defined in Excel98 on the Macintosh, integrals[3] (representing the area underneath the spectral cross-section for these regions) were calculated for these frames.

In addition, formant frequencies were measured for two of the English speakers and one of the Japanese speakers. Because of the fronting of /u/ and, to a lesser extent, /o/ for many English speakers (Labov, 1991), it was important to verify that these two vowels did in fact differ from their front counterparts, especially at their onsets. If these vowels do not differ acoustically at their onsets, it would be unlikely that they would differentially affect the acoustics of preceding fricatives. For the Japanese speakers, the primary reason for looking at formant frequencies was to determine whether the formant frequencies for /ɯ/ in *tsugi/tsugu* and in *sude/tsudo* would be affected by the quality of the vowel in the second syllable.

3. Results

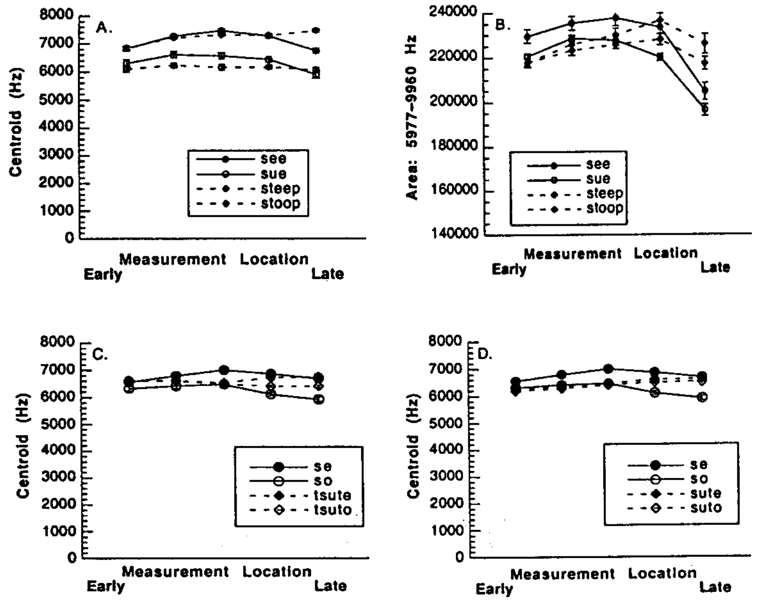
Some typical data are illustrated in Figure 3. In Figure 3A, each curve represents the mean centroids of the fricatives in ten tokens of one word produced by speaker E3, measured at the five locations indicated on the *x*-axis. The solid curves for *see* and *sue* are well separated, indicating that /s/ in *sue* indeed does have a lower frequency spectrum than that of /s/ in *see*. In addition, the dashed curves for *steep* and *stoop* are comparably separated. Figure 3B represents the integrals of the upper frequency range of the fricative spectra for the same words, produced by speaker E2. While these tokens are more variable, as evidenced by the size of the error bars associated with each point, the solid curves for *see* and *sue* are, nonetheless, distinct throughout the entire fricative. However, the dashed curves for *steep* and *stoop* are only distinct late in the fricative.

---

[2] DFT spectral cross-sections were calculated on windows of 1024 samples (50 msec), with an offset of 64 samples (3 msec). This relatively large temporal window was chosen to provide a high frequency resolution, as some of the hypothesized spectral differences between fricative types could have been of a relatively small magnitude. Centroids were calculated on the frequency range 1500–9999 Hz; the lower range was chosen to minimize the impact on the centroids of any voicing during the fricative as well as of any low frequency artifacts.

[3] Decibels represent relative rather than absolute amplitude, and the DFT routines in HADES sometimes return negative values, which confound the rather brute-force calculation of integrals used. To avoid these confounds, the DFT spectra were rescaled by adding 50 to each value, based on the minima observed in individual spectral slices.

Figure 3: A–Comparable difference in see/sue and steep/stoop (Speaker E3); B–Vocalic influence on /s/ in steep/stoop is comparable in magnitude to that in see/sue, but starts later in the fricative (Speaker E2); C–Influence of final vowel on [s] starts later in tsute/tsuto than in se/so and is of smaller magnitude (Speaker J4); D–There is no influence of the final vowel on /s/ in sute/suto, while there is for se/so (Speaker J4).



Figures 3C and 3D are comparable plots for tokens produced by speaker J4. The solid curves for the mean centroids of *se* and *so* show that the fricative spectra are influenced by the following vowel, at least in their second half. In contrast, the dashed curves for *tsute* and *tsuto* in Figure 3C show much less separation, and those for *sute* and *suto* in Figure 3D show none at all.

All such comparisons are summarized in Table 2. The number of plus signs provides a rough index of the magnitude of the differential effects of front and back vowels on fricative spectra, taking into account both the magnitude and the duration of differences like those illustrated in Figure 3. For the three English speakers, there was at least some acoustic difference between the fricatives in all word pairs for all speakers. Three of the six relevant comparisons showed approximately the same amount of difference in the longer pairs, and the other three comparisons showed somewhat less difference in the longer pairs than in the shorter pairs, but the fricatives in the longer pairs still differed based on vowel context.

Table 2: Results of study. A: English speakers; B: Japanese speakers. The number of +'s provides a rough indication of the extent to which the fricative in each pair reflects the influence of the last vowel in the word; '——' indicates no apparent difference.

A.	E1 area	E2 area	E3 centroid
say/so	+++	+	+++
stay/stow	++	+	+++
see/sue	+++	++	+++
steep/stoop	++	+	+++

B.	J1 area	J2 area	J3 centroid	J4 centroid
se/so	——	++	+++	++
sute/suto		+	——	——
tsute/tsuto		+	+	+
suki/suku		+	+	+
sude/tsudo		(+)	+	——
tsugi/tsugu		+	+	——
shi/shu	+++	+++	+++	+++
shitee/shitoo	——	(+)	——	——

For the Japanese speakers, in contrast, there was substantial acoustic difference between the fricatives in seven out of the eight CV pairs; Speaker J1's *se/so* showed no difference[4], either in the fricative centroids or in the calculated integrals in areas of the spectrum that appeared superficially different. For the remaining three Japanese speakers, none of the longer word pairs had fricatives that differed to the same extent as the fricatives in *se/so*, and, in four of the fifteen instances, there was no difference at all between the fricatives in the longer words. Likewise, for three of the speakers, there was no difference in the fricatives in *shitee* and *shitoo*, despite substantial difference between their fricatives in *shi* and *shu*; for J2, the fricatives in *shitee* and *shitoo* differed only slightly, and it is not clear that this particular difference would turn out to be statistically significant.

In the instances of longer Japanese words with initial fricatives not influenced by the vowel late in the word, the fricative shows a clear influence of the immediately following vowel, whether voiced or devoiced. That is, /ç/ in *shitee* and *shitoo* is, for all four speakers, more comparable to /ç/ in *shi* than to /ç/ in *shu*. As is evident from figures 3C and 3D, this is also the case for /s/ and /ts/. However, it is not the case that the supralaryngeal configuration for the later vowels has *no* influence earlier in the word. In fact, formant measurements taken during (voiced) /ɯ/ in J1's tokens of *sude/ tsudo* and *tsugi/tsugu* show that, even at the very beginning of the vowel, F2 is approximately 80 Hz higher in *sude* and *tsugi* than in *tsudo* and *tsugu*. At

[4] There are several possible accounts for this lack of difference. The first is that J1's fricatives did, in fact, differ in *se* and *so*, but that this difference was concentrated in areas of the spectrum above 10,000 Hz. The second is that because of dental abnormalities, her fricative spectra contained artifacts of sufficient magnitude to dilute any coarticulatory effects of the following vowel. Indeed, the spectral cross-sections for many of her tokens containing /s/ or /ts/ had an unusual relatively high amplitude peak at approximately 4,000 Hz.

the end of the vowel, immediately before the /g/ in *tsugi/tsugu* the difference is almost 500 Hz, while for the few tokens of *tsudo* with a long enough vowel for multiple measurements, F2 immediately before the /d/ is 175 Hz lower than immediately before the /d/ in *sude*. Thus, it is not that the final front and back vowels in these items do not exert a coarticulatory influence on preceding segments; rather, this influence does not extend into the initial consonant of the word, at least not measurably.

4. Discussion and Conclusion

The results of the present investigation are consistent with the gestural overlap view of Japanese devoiced vowels presented by Beckman and colleagues. Because the supralaryngeal gestures for short high vowels are *so* short—the ultra-short /ɯ/ in the token illustrated in Figure 2 is not atypical in our data, some of which are summarized in Table 3—it is relatively easy for these gestures to overlap totally with the flanking laryngeal abduction gestures required for adjacent voiceless obstruents. A schematic view of this overlap is presented in Figure 4 below.

Table 3: Mean duration (in msec) of first vowel in mitee, sude, tsudo, for four Japanese speakers (standard deviation in parentheses).

	J1	J2	J3	J4
mitee	83.3 (9.0)	73.2 (2.3)	51.3 (7.9)	71.2 (4.7)
sude	34.7 (9.1)	50.3 (17.8)	41.0 (12.3)	21.9 (5.9)
tsudo	26.7 (8.9)	53.2 (6.6)	34.2 (17.0)	34.9 (3.0)

The result of this overlap is not necessarily to extend the duration of the preceding consonant, but rather to shape its acoustic character (cf., Han, 1994); while the greater length of /ç/ in *shitee* relative to /m/ in *mitee* (see Table 4 on the following page) might give the impression that the duration of devoiced [i̥] is somehow incorporated into /ç/, some additional comparisons, also presented in Table 4, make clear that this is not the case.

Figure 4: Schematic of supra-laryngeal and laryngeal gestures for /sɯt/ in sute. Arrows to the right indicate the direction of greater constriction of the vocal tract at the relevant location. Because of the overlap of the two larygeal abduction gestures for /s/ and /t/, the glottis does not ever reach the degree of constriction necessary for voicing during /ɯ/.

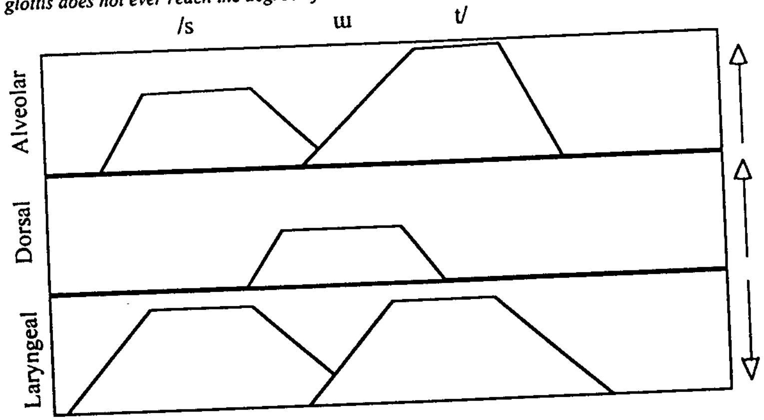


Table 4: Duration (in msec) of initial consonants in selected Japanese tokens (standard deviation in parentheses).

Speaker	*sute*	*so*	*sude*	*shitee*	*shi*	*mitee*
J1	138.39 (12.01)	106.28 (6.86)	139.78 (9.02)	125.78 (9.77)	128.55 (15.44)	77.90 (6.70)
J2	133.11 (15.95)	133.19 (9.58)	136.23 (13.94)	156.84 (18.60)	176.96 (30.02)	76.21 (16.62)
J3	83.30 (9.18)	88.59 (12.52)	113.94 (8.20)	98.93 (5.61)	123.18 (6.09)	76.30 (6.59)
J4	83.80 (5.20)	86.90 (9.63)	91.26 (16.34)	95.43 (10.93)	106.56 (10.25)	50.56 (4.75)

For speaker J1 only, [s] in *sute* is longer than [s] in *so*, while, for the other three speakers, there is absolutely no difference. Furthermore, [s] is virtually the same duration in *sute* and *sude* for J1 and J2, and substantially longer in *sude* than in *so* or *sute* for the other two speakers. Likewise, for all three of the four speakers, [ç] is shorter in *shitee* than in *shi*; for J1, there is no difference.

In any case, to the extent that Japanese morae are, in fact, of similar duration, it is not clear that onset consonants, which bear no metrical weight, ought to contribute to such isochrony or that surface acoustic duration is the appropriate measure thereof. Lehiste (1976) suggests that isochrony of phonological units, to the extent that it occurs, is a perceptual phenomenon. That is, units that are not physically comparable in duration are perceived by language users to be more comparable than they actually are. The duration of any given phonological segment, in Japanese as in other languages, depends on many factors, including at the very least phonological length, accent, phonological context, and intrinsic attributes of the segment. Listeners "parse out" these various factors, taking into account the contributions of each of these factors (Fowler, 1984). Thus, the perception of isochrony—of whatever level of phonological unit—should not be dependent solely on the surface durations of the phonological segments comprising that unit.

Our conclusions from the present study are simple: regardless of their surface duration, Japanese devoiced vowels maintain their supralaryngeal integrity, both in influencing the articulatory and acoustic characteristics of adjacent phonological units and in mediating longer-distance effects of one segment on another. The qualitative nature of this influence would be difficult to incorporate in a discrete, feature-changing account; rather, our findings complement those in other, superficially quite different, studies of anticipatory coarticulation (e.g., Fowler, 1980; Boyce, Krakow, & Bell-Berti, 1991) in supporting alternative models, such as that of Browman & Goldstein (1986, 1990; Surprenant & Goldstein, 1998), which explicitly specify the temporal extent of articulatory primitives. The superficial impression of vowel deletion results from the fact that the supralaryngeal gestures producing the vowels, whose presence is clearly deducible from their myriad acoustic consequences, may be fully masked by the supralaryngeal gestures associated with adjacent segments.

References

Alfonso, Anthony. 1971. *Japanese language patterns.* Tokyo: Sophia University.

Beckman, Mary & Atsuko Shoji. 1984. Spectral and perceptual evidence for CV coarticulation in devoiced /si/ and /syu/ in Japanese. *Phonetica* 41.61–67.

Bell-Berti, Fredericka & Katherine S. Harris. 1982. Temporal patterns of coarticulation. *JASA* 71.449–454.

Bloch, Bernard. 1950. Studies in colloquial Japanese IV: Phonemics. *Language* 26: 86–125

Boyce, Suzanne E., Rena A. Krakow, & Fredericka Bell-Berti. 1991. Phonological underspecification and speech motor organization. *Phonology* 8.219–236.

Browman, Catherine P. & Louis Goldstein. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3.219–252.

Browman, Catherine P. & Louis Goldstein. 1990. Tiers in articulatory phonology, with some implications for casual speech. In *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, ed. by John Kingston & Mary Beckman. Cambridge: Cambridge University Press.

Endo, Yasuko Amy. 1998. Phonetic perception of Japanese devoiced vowels by English speakers. Wellesley College honors thesis.

Evers, Vincent, Henning Reetz, & Aditi Lahiri. 1998. Crosslinguistic acoustic categorization of sibilants independent of phonological status. *Journal of Phonetics* 26.345–370.

Faber, Alice & Timothy J. Vance. Submitted. Perception of Japanese devoiced vowels by native speakers of English. Manuscript submitted for publication.

Fowler, Carol A. 1980. Coarticulation and theories of extrinsic timing. *Journal of Phonetics* 8.113–133.

Fowler, Carol A. 1984. Segmentation of coarticulated speech in perception. *Perception & Psychophysics* 36.359–368.

Goodell, Elizabeth W. & Michael Studdert-Kennedy. 1993. Acoustic evidence for the development of gestural coordination in the speech of 2-year-olds. *Journal of Speech and Hearing Research* 36.707–727.

Han, Mieko S. 1962. Unvoicing of vowels in Japanese. *Onsei no Kenkyû* 10.81–100.

Han, Mieko S. 1994. Acoustic manifestations of mora timing in Japanese. *Journal of the Acoustical Society of America* 96.73–82.

Haraguchi, Shôsuke. 1984. Some tonal and segmental effects of vowel height in Japanese. In *Language Sound Structure*, ed. by Mark Aronoff & Richard T. Oehrle. Cambridge: MIT Press.

Imai, Terumi. 1997. *Vowel Devoicing in Japanese.* Michigan State University master's thesis.

Imaizumi, Satoshi, Akiko Hayashi, & Toshisada Deguchi. 1995. Listener adaptive characteristics of vowel devoicing in Japanese dialogue. *Journal of the Acoustical Society of America* 98.768–778.

Jun, Sun-Ah & Mary E. Beckman. 1993. A gestural-overlap analysis of vowel devoicing in Japanese and Korean. Paper presented at the 67th Annual Meeting of the Linguistic Society of America, Los Angeles, CA.

Jun, Sun-Ah, Mary E. Beckman, Seiji Niimi, & Mark Tiede. 1998. Electromyographic evidence for a gestural-overlap analysis of vowel devoicing in Korean. *UCLA Working Papers in Phonetics* 96.1–42.

Jun, Sun-Ah, Mary E. Beckman, & Hyuck-Joon Lee. 1998. Fiberscopic evidence for the influence on vowel devoicing of the glottal configurations of Korean obstruents. *UCLA Working Papers in Phonetics* 96.43–68.

Keating, Patricia A. 1988. Underspecification in phonetics. *Phonology* 5.275–292.

Kitahara, Mafuyu. 1998. The interaction of pitch accent and vowel devoicing in Tokyo Japanese. In: *Japanese/Korean Linguistics 8*, ed. by David J. Silva. Stanford: CSLI.

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

Kuriyagawa, Fukuko & Masayuki Sawashima. 1989. Word accent, devoicing and duration of vowels in Japanese. *Annual Bulletin: Research Institute of Logopedics and Phoniatrics* 23.85–108.

Labov, William. 1991. The three dialects of English. In *New Ways of Analyzing Sound Change*, ed. by Penelope Eckert. New York: Academic Press, pp. 1–44.

Lehiste, Ilse. 1976. Isochrony revisited. *Journal of Phonetics* 5.253–263.

Martin, Samuel E. 1952. *Morphophonemics of Standard Colloquial Japanese.* Supplement to *Language* (Language Dissertation No. 47).

Nihon Onsei Gakkai, ed. 1976. Onseigaku daijiten. Tokyo: Sanshûsha.

Nittrouer, Susan. 1995. Children learn separate aspects of speech production at different rates: Evidence from spectral moments. *Journal of the Acoustical Society of America* 97.520–530.

Nittrouer, Susan, Michael Studdert-Kennedy, & Richard S. McGowan. 1989. The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *JSHR* 32.120–132.

Rubin, Philip E. 1995. HADES: A case study of the development of a signal analysis system. In *Applied Speech Technology*, ed. by A. Syrdal, R. Bennett, & B. Greenspan. Boca Raton, FL: CRL Press.

Sakuma, Kanae. 1929. *Nihon Onseigaku*. Tokyo: Kazama Shobô.

Surprenant, Aimée M. & Louis Goldstein. 1998. The perception of speech gestures. *Journal of the Acoustical Society of America* 104.518–529.

Takebe, Yoshiaki. 1977. Kokugo kokuji mondai no yurai. In *Iwanami Kôza Nihongo 3: Kokugo Kokuji Mondai*, ed. by Susumu Ôno & Takeshi Shibata. Tokyo: Iwanami.

Tsuchida, Ayako. 1997. Japanese vowel devoicing: Cases of consecutive devoicable syllables. Paper presented at the 8th Japanese/Korean Linguistics Conference, Ithaca, NY.

Vance, Timothy J. 1987. *An introduction to Japanese phonology.* Albany: SUNY Press.

Vance, Timothy J. 1992. Lexical phonology and Japanese vowel devoicing. In *The Joy of Grammar*, ed. by Diane Brentari, Gary N. Larson, & Lynn A. McLeod. Amsterdam: John Benjamins.

Whalen, D. H., Edward R. Wiley, Philip E. Rubin, & Franklin S. Cooper. 1990. The Haskins Pulse Code Modulation (PCM) system. *Behavioral Research Methods, Instruments, and Computers* 22.55–559.

Yeni-Komshian, Grace H. & Sigfrid D. Soli. 1981. Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation. *Journal of the Acoustical Society of America* 70.966–975.

04 03 02 01 00 5 4 3 2 1

Library of Congress Cataloging-in-Publication Data

Conference on Japanese/Korean Linguistics (1st : 1989 : University of Southern California)
Japanese/Korean Linguistics/edited by Hajime Hoji
p. cm.
Includes bibliographical references and index.
ISBN 1-57586-263-8
ISBN 1-57586-264-6 (pbk.)
1. Japanese language—Congresses. 2. Korean language—Congresses.
3. Japanese language—Grammar, Comparative–Korean–Congresses.
4. Korean language—Grammar, Comparative–Japanese–Congresses.
5. Linguistics–Congresses. I. Hoji, Hajime. II. Stanford Linguistics Association.
III. Center for the Study of Language and Information (U.S.) IV. Title.
PL503.C6 1989
495.6–dc20 90-2550
CIP

∞ The acid-free paper used in this book meets the minimum requirements of the American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

CSLI Publications reports new developments in the study of language, information, and computation. In addition to conference proceedings, our publications include lecture notes, monographs, working papers, revised dissertations, and Studies in Logic, Language and Information. Our aim is to make new results, ideas, and approaches available as quickly as possible. Please visit our website at
http://www.csli-publications.stanford.edu
for comments on this and other titles, as well as for changes and corrections by the author and publisher.

Japanese/Korean Linguistics

Volume 9

Edited by
Mineharu Nakayama &
Charles J. Quinn, Jr.

Published for the
Stanford Linguistics Association by
CSLI Publications
Center for the Study of Language and Information
Stanford, California