Working Papers of the Cornell Phonetics Laboratory
1994, vol. 9, pp. 183-222.

Fricative-vowel coarticulation in Japanese devoiced syllables: Acoustic and perceptual evidence

Ayako Tsuchida

The acoustic and perceptual characteristics of Japanese devoiced and voiced syllables, [ʃi] and [ʃɯ], were examined. Coarticulation effects of the following vowel were observed in the fricatives of both devoiced and voiced syllables, although the effects were weaker in devoiced syllables. Coarticulation effects of the following consonant were also observed in the fricatives in devoiced syllables, but not those in voiced syllables. Perception tests suggest that Japanese listeners use the fricative's center of gravity as a primary vowel cue, rather than the spectral peak in the fricative. Theoretical implications of the present results for different analyses of the Vowel Devoicing phenomenon are discussed.

1. Introduction

Japanese is often referred to as a language with devoiced vowels (Vance 1987). Vowel Devoicing[1] is usually observed when either of the two Japanese high vowels ([i] or [ɯ]) occurs between voiceless segments. For example, in the following words, the underlined vowels are devoiced: *tsutome* 'duty', *hikari* 'light'. It is said that typical devoiced vowels (such as in Cheyenne) show formant-like bands (Beckman and Shoji 1984). Han (1962b), in her study of acoustic properties of Japanese devoiced vowels, reports that in narrow-band spectrograms of a devoiced vowel, the harmonics were observed. They were less clear compared to a fully voiced vowel, yet still visually identifiable. When a high vowel follows a voiceless fricative, however, it is often observed that no characteristics of the vowel appear in spectrographic data (Kawakami 1977; Beckman and Shoji 1984). In spite of this absence of spectral characteristics, it has been shown that Japanese listeners can identify the underlying vowel following the fricative [ʃ] at better than chance level (Beckman and Shoji 1984). Beckman and Shoji hypothesize that anticipatory vowel coarticulation effects in the fricative provided a vowel cue. To justify their claim, it is necessary to examine the vowel coarticulation effects in the fricatives.

Previous work has shown that there is a clear anticipatory fricative-vowel coarticulation in English (Soli 1981). Spectral peaks that are affiliated with the formants of the following vowel were consistently observed at the end of the fricative. Specifically, the second formant peak in the fricative [ʃ] preceding [i] was 200-250 Hz higher than that preceding [u]. A perception experiment has shown that listeners were sensitive to the peak frequency

[1] I use the term 'devoicing' following the traditional descriptions of this process. The exact nature of this phenomenon, however, is yet to be uncovered as we will see below.

(Yeni-Komshian and Soli 1981). The spectral peak is argued to provide a strong vowel cue.

Unfortunately, there has been no systematic study on anticipatory fricative-vowel coarticulation in Japanese. Beckman and Shoji (1984) performed an acoustic study on Japanese [ʃi] and [ʃɯ] syllables in the devoicing environment, yet their data is limited. They examined the frequency of the fricative's lowest edge of the noise band in devoiced syllables and found that the fricatives followed by devoiced [i] had a significantly higher mean frequency values, compared to those followed by devoiced [ɯ]. Beckman and Shoji claim that the frequency values of the noise band provide a vowel cue. These frequency values for the fricatives followed by [i] and those followed by [ɯ], however, overlap greatly, thus it is not clear whether the values provided a strong vowel cue. As for the spectral peaks in the fricative, Beckman and Shoji examined only two tokens. Although their measurements show that the fricative preceding the vowel [i] had a higher peak frequency than the one preceding [ɯ], it is not clear if these peaks are indeed affiliated with the following vowel's formants. It is necessary to examine voiced syllables to see the relationship between the spectral peaks in the fricative and the vowel formants. A more extensive and systematic study on fricative-vowel coarticulation effects in Japanese is warranted, in order to understand how Japanese listeners identify a devoiced vowel, when it lacks spectral characteristics.

The present study investigates the anticipatory vowel coarticulation effects on the fricative in Japanese. Specifically, the occurrence of a spectral peak is examined in both voiced and devoiced syllables. The lower edge frequency of fricatives was also examined in both voiced and devoiced contexts. In addition, the coarticulation effects of the following segment are investigated. A subsequent perception study examines what Japanese listeners use as vowel cues.

The structure of this paper is as follows. In §2, previous studies on fricative-vowel coarticulation are reviewed in more detail. In §3 and §4, the acoustic experiments are discussed. In §5, previous studies on the perception of fricative-vowel coarticulation are reviewed. In §6, the perception experiment is presented. In §7, the theoretical implications of the present study's results are discussed. I conclude the paper in §8.

2. Previous studies on fricative-vowel coarticulation

2.1 English data

Soli (1981) investigated anticipatory vowel coarticulation effects on the sibilant fricatives [s, z, ʃ, ʒ]. The stimuli consisted of 12 bisyllabic FVFVF utterances produced by a male English speaker, where F was one of the fricatives and V was [i], [u] or [ɑ] (e.g. [sisis]). Multiple tokens were produced for each utterance. LPC spectra of the last 60 ms of the initial fricatives were computed every 10 ms using 25.6 ms Hamming window. In addition, LPC spectra of the last 60 ms of fricatives produced in isolation were computed. The analysis showed that fricatives preceding a vowel had a spectral peak affiliated with the second formant of the vowel. Spectral peaks between 1500 and 2000 Hz were observed and peak frequencies were 100-300 Hz higher for the fricatives preceding the high vowel [i]. Palatal fricatives also showed spectral peaks affiliated with the third formant of the following vowel. Soli (1981) reports that both the second and third formant had lower values for [ʃ(u)][2], compared to [ʃ(i)]. Among the three vowels, [ɑ] had most poorly defined formant peaks. For fricatives in isolation, second formant peaks were either poorly defined (alveolar fricatives) or invisible (palatal fricatives).

It is generally assumed that a narrow constriction in the oral cavity (as in fricative production) tends to cancel the effects of the natural resonances behind the constriction (Heinz and Stevens 1961). When a fricative is followed by a vowel, however, the constriction opens in anticipation of the vowel configuration. This allows formant peaks to appear in the last part of the fricative. Soli (1981) claims that the difference in fricative formants in various F(V) contexts can be accounted for by the principle of articulatory compatibility (Kozhevnikov and Chistovich 1965, cited in Soli 1981). The raised tongue position for the sibilants is more compatible with the tongue position for high vowels than for low vowels. The weaker coarticulation effect in the [ɑ] context is attributable to the incompatibility of tongue configurations.

2.2 Japanese data

Beckman and Shoji (1984) examined the acoustic characteristics of Japanese devoiced syllables, [ʃi] and [ʃɯ]. As stimuli, they used 22 words (11 minimal pairs) containing either [ʃi] or [ʃɯ] syllable in the devoicing environment (e.g., *ʃitʃoo* 'audition', *ʃɯtʃoo* 'assertion'). Each word, along with 12 filler words, was embedded in the frame sentence *kore-o* ____ *to yakusimasita* 'I translated this as _____'. These sentences were randomized and written in Japanese orthography three times in three different lists. Sentences containing the two members of a minimal pair did not occur next to each other. No test words occurred in the first or last four sentences on a page. Four speakers of Standard Japanese read the lists, producing 12 tokens (4 speakers X 3 lists) for each word. A wide-band spectrogram was made for each token. Beckman and Shoji examined if any

[2] Following Soli (1981), the stimuli are labeled together with the excised vowel.

voicing was present in the target syllable. When no voicing was observed, the lowest edge of the fricatives' noise band was measured. Unfortunately, the percent of the devoiced vowels is not reported. Beckman and Shoji note that some vowels were voiced. It is also not reported how the frequency of the lowest edge was determined.

Beckman and Shoji's measurements showed that [ʃ(ɯ)] had significantly lower mean edge frequency than [ʃ(i)] when there was no voicing in the target syllable (they do not give the exact values, but their graph shows mean frequency of approximately 2300 Hz for [ʃ(ɯ)] and 2750 Hz for [ʃ(i)]). They assume that this is due to a velar or labiovelar coloring for the vowel [ɯ]. Frequency values of the two syllable types overlapped greatly. Figure 1, taken from their article, shows the overlapping area.

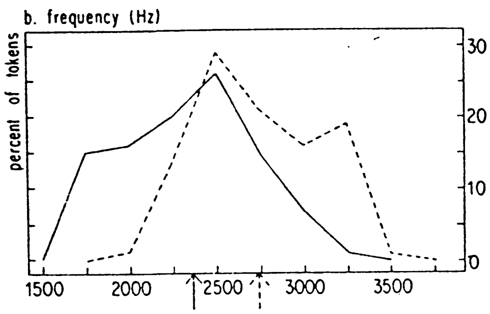

Figure 1. Relative frequency distribution of fricative's lower edge values: [ʃ(i)] (....) and [ʃ(ɯ)] (__). Arrows below abscissa indicate mean values (from Beckman and Shoji 1984: 66).

For their subsequent perception study (see §6 for more detail), Beckman and Shoji further examined one of the speaker's utterances of the words [ʃikaN] 'historical view' and [ʃɯkaN] 'subjectivity' (N is a placeless nasal). The words were produced by a male speaker. It was observed that [ʃ(i)] had higher formant peak frequencies than [ʃ(ɯ)][3]. However, it is unknown whether [ʃ(i)] always has a higher formant frequency than [ʃ(ɯ)], since the data is so limited. In addition, it is not clear if these peaks are affiliated with the formants of the following vowel, since Beckman and Shoji do not compare the values with formant frequency values of vowels. Even if we assume that the peaks in their study are indeed affiliated with the vowel's formants, there is another problem. It seems that they measured the third and fourth formant peaks for [ʃ(i)], while measuring the second and third formants peaks for [ʃ(ɯ)]. The exact values are not given, but Beckman and Shoji's figure (1984: 66) shows peaks around 2800 Hz and 3500 Hz for [ʃ(i)], and peaks around

[3] Beckman and Shoji (1984) refer to these peaks as K1 and K2.

1500 Hz and 2400 Hz for [ʃ(ɯ)]. The average F2 values[4] for Japanese [i] and [ɯ] are about 1950 Hz and 1400 Hz, respectively (Keating et al. 1984). The peaks for [ʃ(i)] in Beckman and Shoji's study are much higher than the second formant frequency. Thus, it is not clear whether a prominent second formant peak existed in [ʃ(i)] or not. The present study examines the acoustic characteristics of fricative-vowel coarticulation in a more systematic way.

3. Acoustic study

3.1 Methods

A. Stimuli

The acoustic properties of two types of syllables, those with a vowel in the devoicing environment and those with a vowel in the non-devoicing environment, were examined. The corpus consisted of 10 minimal pairs. Each pair differed only in the voicing of the consonant following the high vowel.

	devoicing environment	non-devoicing environment
[i]	[ʃito:] 'desperate struggle'	[ʃido:] 'private road'
	[ʃitʃo:] 'audition'	[ʃidʒo:] 'market'
	[ʃitai] 'corpse'	[ʃidai] 'private university'
	[ʃiho:] 'private law'	[ʃibo:] 'death'
	[ʃikjo:] 'market'	[ʃigjo:] 'commencement of work'
[ɯ]	[ʃɯto:] 'vaccination'	[ʃɯdo:] 'hand-operated'
	[ʃɯtʃo:] 'assertion'	[ʃɯdʒo:] 'emotion'
	[ʃɯtai] 'nucleus'	[ʃɯdai] 'theme'
	[ʃɯho:] 'technique'	[ʃɯbo:] 'leader'
	[ʃɯkjo:] merrymaking'	[ʃɯgjo:] 'training'

Table 1. Words used in the acoustic study.

All words contain three moras and have the same pitch accent pattern. These 20 words, along with 4 filler words were written in Japanese orthography in the frame sentence *kore-wa ___ to iu imi-desu* 'This means ___' in randomized order. No sentences containing two members of the minimal pair occurred next to each other. In addition, no test words occurred in the first and last two sentences on the list. Three female native speakers of

[4] These values were obtained by averaging F2 frequencies of high vowels produced by male speakers.

Standard Japanese read the list three times, producing nine tokens for each of the 20 stimuli (3 speakers X 3 repetitions). Speech was recorded in a sound-proof booth, using a cardioid microphone (Electrovoice, model RE 20) and a high-quality cassette recorder (Marantz, model PMD 222).

B. Analysis

The recordings were digitized on a SUN 3/160 computer at 12 kHz with a low-pass filter setting of 6 kHz and stored as files to be processed by the commercial software package WAVES+/ ESPS. Wide-band spectrograms and waveforms were made for each stimulus. The onset of the fricative was determined as the point immediately following the preceding vowel's periodicity. The offset of the fricative was defined as the point immediately prior to the initiation of the following vowel's first periodicity. When there was no vowel following, it was the point where frication in the waveform and spectrogram ended. The onset of the vowel was the same as the offset of the fricative. Vowel offset was defined as the offset of the second formant. When the second formant was not clear, offset of the first formant was used.

For each token, formant peaks for the last 40 ms of the fricative and the initial 30 ms of the vowel were computed every 10 ms, using 25.6 ms Hamming window. This was done by a program which selects the four most prominent peaks during the given time. In addition, an LPC spectrum (16 poles) for the final 25 ms of the fricative was derived. I looked for a significant prominence in the frequency range of approximately 2000-2500 Hz, which is slightly higher than the range found for male speakers in the previous study (Soli 1981), since the present study had female speakers only. When there was a prominent peak, its amplitude was measured. The lower edge of the noise band was measured from spectrograms at the end of the fricative. The lower edge was determined to be the point below which no energy was observed. In addition, the durations of the fricatives, vowels and the entire words were measured. Analyses of variance were performed for each set of data. Unless otherwise stated, vowel context (two levels) and voicing context (two levels) were the independent variables. When the analysis was limited to either voiced or devoiced syllables, paired two-tailed t-tests were performed.

3.2 Results

3.2.1 Vowel characteristics

A vowel was determined to be devoiced if there were no characteristics of any vowel observed in either spectrograms or waveforms. Table 2 shows the average devoicing rates for each syllable type.

	devoicing env.	non-devoicing env.
[ʃi]	100%	0%
[ʃɯ]	97%	0%

Table 2. Average devoicing rates[5].

When the vowel was followed by a voiced consonant, no devoicing occurred. Vowels followed by a voiceless consonant, on the other hand, were always devoiced with only one exception of [ɯ][6]. The devoicing rate was significantly higher for vowels in the devoicing environment [$\chi^2(1) = 158$, $p < .001$]. The vowel context did not have a main effect on the devoicing rate [$\chi^2(1) = .024$, ns]. Thus, we can conclude that vowel devoicing consistently occurred when a vowel occurred between voiceless consonants. These data are in agreement with results of other studies, which reported high devoicing rates for high vowels following a sibilant (Kondo 1993; Takeda and Kuwabara 1987).

3.2.2 Fricative characteristics

3.2.2.1 Formant peaks

An LPC spectrum for each token was computed for the last 25 ms of the fricative. A token was determined to have no peak, when there was no visible prominence in the second formant region (2000-2500 Hz). The following results were obtained:

[5] Vowels preceding /h/ were always voiced. It is known that intervocalic /h/ is frequently accompanied by vocal fold vibrations in Japanese (Yoshioka 1981). In fact, all instances of /h/ in the present study were voiced. We can assume that /h/ in test words behaved as a voiced consonant and devoicing did not occur. For this reason, test words with /h/ were thrown out.

[6] This token is not included in the following acoustic study, since a voiced vowel in the devoicing environment is observed to have different acoustic characteristics than a vowel in the non-devoicing environment (Kondo 1994).

	devoiced	voiced
[ʃ(i)]	47% (2490 Hz)	87% (2410 Hz)
[ʃ(ɯ)]	74% (2110 Hz)	98% (2190 Hz)

Table 3. Occurrence of a spectral peak and its mean frequency (averaged across speakers).

Table 3 shows that a spectral peak was sometimes observed even when the following vowel was devoiced. Interestingly, some voiced syllables lacked a peak. The overall peak occurrence rate, however, was significantly higher for voiced vowels (92%) than devoiced vowels (61%) [$\chi^2(1) = 34.14$, $p < .001$]. The vowel context had a main effect on the peak occurrence rate, too. The peak occurrence rate was significantly higher when the following vowel was [ɯ] (88%) than when it was [i] (69%) [$\chi^2(1) = 7.99$, $p < .005$].

As for the peak frequency values, voicing did not have a main effect [$F(1, 2) = .10$, $p > .75$]. The difference in the peak frequency between fricatives in devoiced syllables (2260 Hz) and those in voiced syllables (2290 Hz) was not significant. The vowel context, on the other hand, had a main effect on the peak frequency [$F(1, 2) = 81.61$, $p = .000$]. The frequency was higher for the fricatives followed by the vowel [i] than those followed by [ɯ], whether the vowel was devoiced or not. When averaged across devoiced and voiced syllables, the peak frequency was 2430 Hz for [ʃ(i)] and 2160 Hz for [ʃ(ɯ)]. The vowel by voicing interaction was not significant [$F (1, 3) = 3.05$, $p = .09$]. This indicates that when a devoiced syllable had a spectral peak, its frequency patterned in the same way as in voiced syllables.

For the tokens with a spectral peak, the amplitude of the peak was measured. The following results were obtained:

	devoiced	voiced
[ʃ(i)]	64.2 dB	69.9 dB
[ʃ(ɯ)]	70.9 dB	76.1 dB

Table 4. Amplitude of the spectral peak (averaged across speakers).

The peak amplitude was significantly smaller in devoiced syllables (67.7 dB) than that in voiced syllables (73.0 dB) [$F(1, 2) = 11.86$, $p = .001$]. This was the case for both [ʃ(i)] and [ʃ(ɯ)]. The vowel context also had a main effect on the amplitude of the spectral peak. The amplitude was significantly smaller when the fricative was followed by [i] (67.6 dB) compared to when it was followed by [ɯ] (73.9 dB) [F(1, 2) = 16.53, p = .000]. The vowel by voicing interaction was not significant [F(1, 3) = .027, p > .87].

The higher peak frequency for [ʃ(i)] suggests that these peaks are affiliated with the formants of the following vowel, as in the English data (Soli 1981). Let us now look at LPC spectra of the last part of the fricative juxtaposed with vowel formant frequencies. Figure 2 shows mean peak frequencies for voiced [ʃi] and [ʃɯ]:

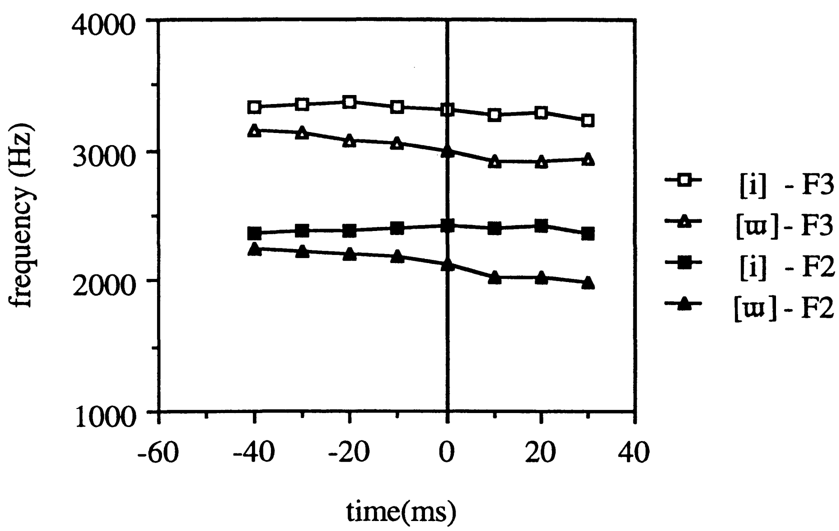

Figure 2. Mean fricative peak frequencies for voiced syllables. The vertical line marks the vowel onset.

The results are very similar to what Soli (1981) found for English syllables. We can see that the second formant peak of [ʃ(ɯ)] is decreasing towards the vowel onset, while that of [ʃ(i)] is slightly increasing. As for the third formant peak, [ʃ(i)] had stable higher peaks, while [ʃ(ɯ)] again showed a decrease in frequency. The second and third formant peaks for [ʃ(i)] were 200-300 Hz higher than those for [ʃ(ɯ)] in the last 10 ms of the fricative (frequency values for [ʃ(i)] and [ʃ(ɯ)] were 2400 Hz and 2180 Hz for F2, and 3330 Hz and 3040 Hz for F3, respectively). The differences were significant [t(14)=7.82, p=.0001 for F2; t(14)=10.03, p=.0001 for F3]. From this result, we can conclude that there is a clear fricative-vowel coarticulation effect in Japanese voiced syllables as was observed in English.

Similar coarticulation effects were observed in the devoiced syllables which had a peak in the fricative (that is, 61% of the devoiced syllables). Figure 3 shows mean fricative frequencies for those devoiced syllables. To see the relationship between the peak frequencies in the fricative and the vowel formant frequencies, the vowel formant values obtained from voiced syllables are juxtaposed:

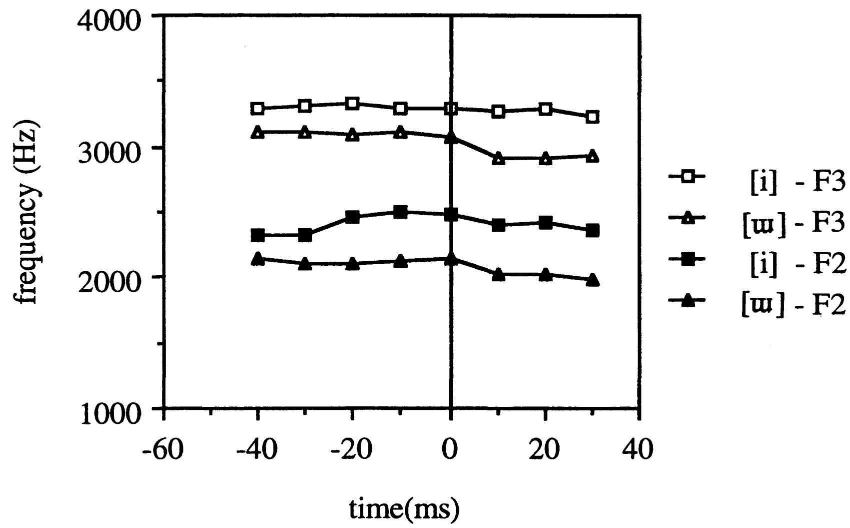

Figure 3. Mean fricative peak frequencies for devoiced syllables. The vowel formant values obtained from voiced syllables are juxtaposed. The vertical line marks the vowel onset.

Again, we see that both the second and third formant peaks had higher frequency values for [ʃ(i)] for the last 40 ms. The peaks were about 200-300 Hz higher for [ʃ(i)] in the last 10 ms of the fricative (frequency values for [ʃ(i)] and [ʃ(ɯ)] were 2520 Hz and 2110 Hz for F2, and 3280 Hz and 3100 Hz for F3, respectively). The difference between the formant peak values was highly significant [$t(7) = 6.19$, $p = .0004$ for F2; $t(5) = 3.20$, $p = .024$ for F3]. This suggests that the peaks observed in the fricative of devoiced syllables are also affiliated with the formants of the following vowel, although the amplitude of the peak was significantly smaller than voiced syllables, and not all fricatives in devoiced syllables had a peak.

3.2.2.2 Lower edge frequency

As was mentioned before, Beckman and Shoji (1984) observed that the lower edge frequency was lower for devoiced [ʃ(ɯ)] than devoiced [ʃ(i)], which may be explained as a coarticulation effect of the intended vowel.

In the present study, the lower edge frequency of the fricatives in both devoiced and voiced syllables were measured. The following results were obtained:

	devoiced	voiced
[ʃ(i)]	3080 Hz	2390 Hz
[ʃ(ɯ)]	2440 Hz	1830 Hz

Table 5. Lower edge frequency of the fricatives in devoiced and voiced syllables.

It was observed that when averaged across devoiced and voiced syllables, the frequency value of [ʃ(i)] was significantly higher than that of [ʃ(ɯ)] [$t(23) = 8.66$, $p = .0001$]. The lower edge frequency was 2730 Hz and 2160 Hz for [ʃ(i)] and [ʃ(ɯ)], respectively. Thus the lower edge value seems to be affected by the intended vowel as Beckman and Shoji observed. However, it was also observed that the frequency value of the fricatives in devoiced syllables was significantly higher than that in voiced syllables [$t(23) = 4.50$, $p = .0002$]. In fact, the lower edge value for devoiced [ʃ(ɯ)] was similar to that for voiced [ʃ(i)] (there was no significant difference between the cutoff frequencies of these two syllable types: $t(11) = .40$, $p > .698$). It is not clear how listeners can use the lower edge frequency values of devoiced [ʃ(ɯ)] as a vowel cue, when they do not differ from the values for voiced [ʃ(i)].

Let us now consider the distribution of lower edge frequency values. Beckman and Shoji (1984) observed that the frequency values of devoiced [ʃ(i)] and [ʃ(ɯ)] varied from 1500 Hz to 4000 Hz and overlapped largely, yet there was a significant difference between the mean values. In the present data, a similar distribution was found. The frequency values of devoiced [ʃ(i)] and [ʃ(ɯ)] ranged from 1500 Hz to 4800 Hz. Figure 4 shows the distribution of the lower edge values. Arrows below the abscissa indicate means.

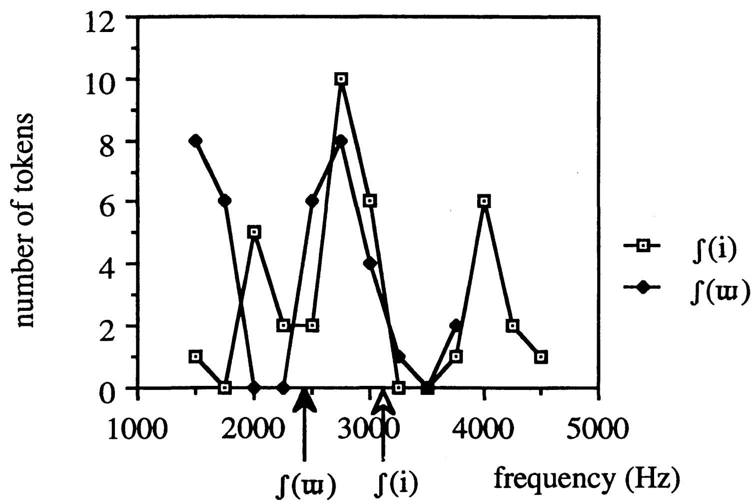

Figure 4. Distributions of lower edge values of fricatives in devoiced syllables.

The edge frequency was higher for [ʃ(i)] tokens. The average values were 3080 Hz for [ʃ(i)] and 2480 Hz for [ʃ(ɯ)]. They were significantly different [$t(11)= 4.51$, $p < .0009$]. Yet, as in Beckman and Shoji's study, a large area of overlap between the two syllable types was observed. Again, it is not clear how listeners can identify a devoiced vowel using the lower edge frequency values, when the values for [ʃ(i)] and [ʃ(ɯ)] overlap substantially. Fricatives in voiced syllables, on the other hand, did not have overlapping values. The lower edge frequency values of [ʃ(i)] ranged from 2040 Hz to 2930 Hz, while that of [ʃ(ɯ)] ranged from 1680 Hz to 2010 Hz.

3.2.3 Effects of the following consonant

3.2.3.1 Formant peaks

The waveforms and spectrograms of devoiced syllables showed no characteristics of a vowel following the fricative. The fricatives in devoiced syllables were immediately followed by a consonant. It is necessary to examine whether the following consonant had any effects on these fricatives.

Table 6 shows the peak occurrence rates and mean peak frequency in different consonant contexts[7]:

---

[7] Since vowels preceding /h/ were thrown out (see footnote 5), devoiced syllables have two consonant contexts, while voiced syllables have three.

	followed by [t]	followed by [k]
ʃ(i)	30% (2440 Hz)	100% (2510 Hz)
ʃ(ɯ)	65% (2110 Hz)	100% (2160 Hz)

Table 6. Occurrence of a spectral peak and its mean frequency (averaged across speakers) in devoiced syllables.

When the devoiced syllable was followed by [k], there was always a spectral peak. When the following consonant was [t], on the other hand, a spectral peak was observed significantly less often (47%) [$\chi^2(1) = 57.03$, $p < .001$]. Analysis of variance was performed with vowel context and consonant context (two levels) as independent variables. The mean peak frequency was slightly higher for syllables in the [k] context (2330 Hz) than those in the [t] context (2260 Hz), although the difference was not significant [$F(1, 2) = .71$, $p > .41$]. The amplitude of the peak also showed the effect of the following consonant:

	followed by [t]	followed by [k]
ʃ(i)	60.0 dB	74.0 dB
ʃ(ɯ)	67.7 dB	79.5 dB

Table 7. Amplitude of the peak in devoiced syllables followed by different consonants.

For both [ʃ(i)] and [ʃ(ɯ)], the amplitude was smaller when the following consonant was [t]. The consonant context had a main effect on the peak amplitude (when averaged across [ʃ(i)] and [ʃ(ɯ)], the amplitude was 64.1 dB for [ʃ(V)t] and 76.7 dB for [ʃ(V)k]): [$F(1, 2) = 29.59$, $p = .000$]. Thus, the present study shows that the peak in the fricative of devoiced syllables was affected by the context of the following consonant.

It is important to note that this kind of difference was not observed in voiced syllables. Table 8 shows the peak occurrence and mean peak frequency in voiced syllables:

	followed by [b]	followed by [d]	followed by [g]
ʃ(i)	78% (2330 Hz)	85% (2440 Hz)	100% (2400 Hz)
ʃ(ɯ)	89% (2230 Hz)	100% (2170 Hz)	100% (2200 Hz)

Table 8. Occurrence of a spectral peak and its mean frequency (averaged across speakers) in voiced syllables.

The consonant context did not have an effect on the peak occurrence rates [$\chi^2(2) = 2.26$, ns]. Analysis of variance was performed with vowel context and consonant context (three levels) as independent variables. The consonant context did not have a main effect on the peak frequency, either [$F(2,3) = .10$, $p > .91$]. The peak amplitude values were 73.8 dB, 72.8 dB and 72.8 dB for [ʃ(V)b], [ʃ(V)d] and [ʃ(V)g], respectively. Again, the consonant context had no main effect on the peak amplitude [$F(2, 3) = .16$, $p > .86$].

3.2.3.2 Effects of the following consonant on the lower edge frequency

It was shown that the peak occurrence rates and the peak amplitude were affected by the context of the following consonant. Interestingly, coarticulation effects of the following consonant were observed in the lower edge frequency values as well. Table 9 shows the lower edge frequency of the fricative in different contexts:

	followed by [t]	followed by [k]
ʃ(i)	3370 Hz	2190 Hz
ʃ(ɯ)	2760 Hz	1650 Hz

Table 9. Lower edge frequency of fricatives preceding a devoiced vowel averaged across speakers.

The mean lower edge frequency was 3070 Hz when the following consonant was [t], while it was 1920 Hz when the following consonant was [k]. For each vowel context, the lower edge frequency was higher when the following consonant was [t]. The following consonant had a main effect on the lower edge frequency values [$F(1,2) = 26.20$, $p = .000$] as well as the vowel context [$F(1, 2) = 6.64$, $p < .018$]. Since the values are affected by the following consonant as well as the vowel, it seems unlikely that lower edge frequency is used as a cue for devoiced vowels.

The fricatives followed by a voiced vowel, on the other hand, were not affected by the following consonant:

	followed by [b]	followed by [d]	followed by [g]
ʃ(i)	2230 Hz	2380 Hz	2430 Hz
ʃ(ɯ)	1830 Hz	1840 Hz	1820 Hz

Table 10. Lower edge frequency of fricatives preceding a voiced vowel averaged across speakers.

The lower edge was significantly higher for [ʃ(i)] than [ʃ(ɯ)] ([F(1, 2)=26.64, p=.000]), yet the difference among the syllables followed by different consonants was not significant [F(2, 3)=.31, p>.735]. These results suggest that when the high vowel is devoiced, the following consonant affects the initial fricative, but not when the vowel is voiced.

3.2.4 Duration

Let us now consider the duration of segments. Previous studies have shown that the duration of devoiced syllables is generally shorter than that of voiced syllables (Beckman 1982; Port et al. 1987). The duration of the fricative in a devoiced syllable is typically as short as the fricative duration in a voiced syllable. Similar results were obtained in the present study:

	devoiced syllable		voiced syllable	
	[ʃ] dur (ms)	V dur (ms)	[ʃ] dur (ms)	V dur (ms)
[ʃi]	116	0	114	47
[ʃɯ]	125	0	111	57

Table 11. Duration of fricatives and vowels in devoiced and voiced syllables.

The duration of the fricative was not longer for devoiced syllables compared to voiced syllables [F(1, 2) = 2.54, p > .12]. We can see that there was virtually no difference in fricative duration between devoiced and voiced syllables. The vowel context did not have a main effect on the fricative duration, either [F(1, 2) = .60, p > .44]. As for the entire syllable duration, voiced syllables were significantly longer than devoiced syllables [F(1, 2) = 24.00, p = .000]. Again, vowel context had no main effect on syllable duration [F(1, 2) = 1.27, p > .265].

It should be recalled that in the present study, unlike the English data (Soli 1981), some voiced syllables lacked a peak. One could assume that the occurrence of a peak may be related to vowel duration, since Japanese vowels are much shorter than English vowels.

Yet it does not seem to be the case. Sometimes a clear peak was observed even if a vowel was only 22 ms long, while the fricative in the same word in a different token had no peak when the vowel was 46 ms long. In addition, note that some fricatives in devoiced syllables had a peak. In these cases, vowel duration was 0 ms. Thus vowel duration does not seem to affect the occurrence of a spectral peak. Similarly, duration of the fricative seems irrelevant to the occurrence of the peak. Duration of the fricative in devoiced [ʃi] is almost the same as that in voiced [ʃi], but the peak occurrence rates were 47% and 87%, respectively. Thus, neither vowel duration nor fricative duration seems to affect the occurrence of spectral peaks.

3.3 Discussion

The present acoustic study has shown that there was a clear anticipatory coarticulation in syllables with voiced vowels. Spectral peaks, which are affiliated with formants of the following vowel, were observed as in the English data (Soli 1981). The difference in the second formant peak frequency was 220 Hz in the last 10 ms of the fricative. This was similar to the difference found in the English data (200-250 Hz).

In devoiced syllables, weaker coarticulation effects were observed in general. A greater number of tokens lacked a spectral peak when the vowel was devoiced. Even when a spectral peak was observed, it was significantly smaller in amplitude than that in the corresponding voiced syllables.

A greater number of tokens showed a formant peak when the fricative was followed by [ɯ] compared to [i]. This was the case for both voiced and devoiced syllables. We may assume that the tongue configuration for the [ʃ] constriction is more compatible with the tongue configuration for [ɯ] than for [i], since Japanese [ɯ] is centralized. The weaker coarticulation effect in the [i] context could be attributed to the less compatibility of tongue configurations, although it is necessary to look at articulatory movement in speech production in order to justify this assumption.

4. Experiment 2

In the present acoustic study, it was observed that both the spectral peak and the lower edge frequency were affected by different consonant contexts. When the following consonant was [t], the peak occurrence rate was lower and the peak amplitude was smaller, compared to [k]. The lower edge frequency was higher when the following consonant was [t]. It should be noted, however, that the tokens that were compared were not minimal pairs. It is possible that the segments following [k] or [t] had some effects on the initial fricative. Thus, it is necessary to see if the same effect is observed when the consonants are in the same environment.

4.1 Methods

A. Stimuli

The acoustic properties of two types of devoiced syllables, those followed by [k] and those followed by [t], were examined. The corpus consisted of the 6 minimal pairs, which differed only in the consonant following the high vowel:

	followed by [t]	followed by [k]
[i]	[ʃiteɴ] 'fulcrum' [ʃitoo] 'desperate struggle' [ʃitai] 'corpus'	[ʃikeɴ] 'personal opinion' [ʃikoo] 'enforce' [ʃikai] 'chairman"
[ɯ]	[ʃɯteɴ] 'main point' [ʃɯtoo] 'vaccination' [ʃɯtai] 'nucleus'	[ʃɯkeɴ] 'sovereignty' [ʃɯkoo] 'idea' [ʃɯkai] 'feast'

Table 12. Words used in Experiment 2.

All high vowels are in the devoicing environment. The pitch accent patterns are the same for all words. These 12 words, along with 6 filler words were written in Japanese orthography in the frame sentence *kore-wa* _____ *to iu imi-desu* 'This means _____' in randomized order. No sentences containing two members from the same group occurred next to each other. In addition, no test words occurred in the first and last three sentences on the list. Two female native speakers of standard Japanese read the list three times, producing six tokens for each of the 12 stimuli (2 speakers X 3 repetitions). The recording procedure was the same as in Experiment 1.

B. Analysis

The analysis procedure was the same as in Experiment 1. Methods of measurements were also the same as for Experiment 1. Analyses of variance were performed for each set of data. Vowel context (two levels) and consonant context (two levels) were independent variables.

4.2 Results

All of the vowels were devoiced in this experiment. Table 13 shows the occurrence of a spectral peak and the mean peak frequency:

	followed by [t]	followed by [k]
ʃ(i)	6% (2770 Hz)	67% (2420 Hz)
ʃ(ɯ)	22% (2100 Hz)	50% (2130 Hz)

Table 13. Occurrence of a spectral peak and the peak value in each syllable type.

Again, a spectral peak was observed more often in syllables followed by [k] (58%) than those followed by [t] (14%). The difference was significant [$\chi^2(1)= 22.18, p < .001$].

As in Experiment 1, the peak frequency was higher for the fricatives preceding [i] (2480 Hz) than those preceding [ɯ] (2120 Hz). The vowel context had a main effect on the peak frequency [$F(1, 2) = 31.58, p = .000$]. The peak frequency did not differ across consonantal contexts (2270 Hz for syllables followed by [t] and 2280 Hz for those followed by [k]) [$F(1, 2) = 3.49, p > .09$].

The peak amplitude was smaller for syllables followed by [t] (71.7 dB) compared those followed by [k] (86.6 dB). The difference was significant [$F(1, 2) = 22.23, p = .001$]. This result also agrees with the data obtained in Experiment 1.

Let us now see the lower edge frequency of these syllables. The following result was obtained:

	followed by [t]	followed by [k]
ʃ(i)	2730 Hz	1730 Hz
ʃ(ɯ)	2380 Hz	1150 Hz

Table 14. Lower edge frequency of fricatives followed by a devoiced vowel.

The result was similar to what was observed in Experiment 1. The lower edge frequency was higher when the devoiced vowel was followed by [t] (2550 Hz) compared to [k] (1440 Hz). Consonant contexts had a main effect on the lower edge frequency [$F(1, 2) = 149.14, p = .000$] as well as the vowel context [$F(1, 2) = 26.23, p = .000$]. Thus we can conclude that when the vowel is devoiced, the fricative shows coarticulation effects from the following consonant. It was mentioned before that in Beckman and Shoji's experiment, the lower edge frequency values varied from 1500 Hz to 4000 Hz. In their experiment, the high vowels were followed by [k, t, s, ʃ, tʃ, h, p]. It seems that the wide range of frequency values reflects the effect of different consonant contexts, although further investigation is necessary to determine the effects of consonants other than [t] and [k].

4.3 Discussion

In the present acoustic study, it was observed that devoiced syllables showed coarticulation effects from the following consonant, while the voiced syllables did not show such effects. The peak occurrence rate was significantly greater in the [k] context than the [t] context. Also, the peak amplitude was greater in the [k] context. Moreover, the lower edge frequency was significantly lower when the following consonant was [k] (the difference was 1110 Hz).

Interestingly, in their study of fricative-stop coarticulation, Repp and Mann (1981) found that the lower edge frequency of [ʃ] rises in [ʃt] cluster, while it decreases in [ʃk] cluster. At the offset of the fricative, they found a 600-800 Hz difference between the [ʃ] followed by [t] and that followed by [k]. Note that this was found in consonant clusters, which did not contain an intervening vowel. It should be further noted that Yoshioka (1981) found that the glottal movement and its adjustment during the production of the [ʃVC] sequence with a devoiced vowel resemble those for fricative-stop clusters in European languages. These data suggest that the devoiced high vowel following [ʃ] was transparent at the phonetic level. If the vowel is transparent, the fricative and the following consonant become adjacent to each other, as in fricative-consonant clusters in other languages.

5. Perception of fricative-vowel coarticulation

So far the acoustic evidence for fricative-vowel coarticulation in Japanese was discussed. The next question I addressed was if Japanese listeners are sensitive to these acoustic differences. A perception study was conducted to examine the perceptual evidence for anticipatory vowel coarticulation. In particular, it was investigated whether the presence of a spectral peak provides a vowel cue. Both voiced and devoiced syllables were examined. First, I review previous studies on the perception of fricative-vowel coarticulation.

5.1 English data

Yeni-Komshian and Soli (1981) conducted a perception study of fricative-vowel coarticulation, where the frication noise portion of English syllables was presented and listeners were asked to identify the excised vowel. The stimuli were made from the initial fricative in FVFVF sequences produced by a male and a female speaker, where F was one of the sibilant fricatives [s, z, ʃ, ʒ] and V was [i], [u] or [ɑ] (e.g. [sisis]). Each stimulus was edited so that all of them had the same amplitude and duration. Their results showed that listeners could accurately identify the excised high vowels, while identification rates for [ɑ] were near chance-level. Yeni-Komshian and Soli argue that the formant peaks in the fricative provide perceptual cues for the excised vowel. Specifically, the second formant peaks are claimed to be 'the best potential vowel cues' (Yeni-Komshian and Soli 1981: 973), since they vary consistently with the following vowel . The low identification rates for the vowel [ɑ] can be attributed to its weaker coarticulation effects on the fricative.

5.2 Japanese Data

Beckman and Shoji (1984) performed a perception experiment on Japanese devoiced vowels. The corpus for their study was 22 words containing [ʃi] and [ʃɯ] in the devoicing environment. Each word was embedded in the frame sentence *kore-o ____ to yakusimasita* 'I translated this as ______'. The utterance tokens were produced three times by four speakers, yielding 12 tokens for each of the 22 words. Beckman and Shoji made a stimulus tape by splicing all the utterance tokens in a random order. 14 listeners were asked to evaluate each token by circling the appropriate answer on a five-point scale, which ranged from 'definitely [the word containing /ɯ/]' through 'can't tell' to 'definitely [the word containing /i/]'. Chance level was 40%, since two out of five answers were considered to be correct. Some of the tokens had voiced vowels, yet as was mentioned before, it is not reported what percent of the vowels were voiced. Table 15 shows the percentage of correct responses:

	devoiced syllables	voiced syllables
[ʃi]	77%	86%
[ʃɯ]	67%	98%
mean	73%	93%

Table 15. Vowel identification rates (from Beckman and Shoji 1984: 67).

Beckman and Shoji's results show that listeners could identify the underlying vowel at a level significantly better than chance even for the devoiced tokens. This indicates that even when no voicing of a vowel was observed in the spectrogram, Japanese listeners could identify the intended vowel quite accurately. The overall identification rates, however, were higher for voiced syllables.

To determine the relationship between the recoverability of the vowel and the coarticulation effect in the fricative, Beckman and Shoji (1984) conducted another perception experiment using synthesized tokens. They modeled the male speaker's utterances of the words [ʃikaɴ] and [ʃɯkaɴ] which obtained the highest identification rate in the first experiment (ɴ is a placeless nasal). The high vowels were devoiced in both words. Beckman and Shoji made a continuum of eight stimuli, having the best [ʃikaɴ] and [ʃɯkaɴ] as the two endpoints. The six intermediate stimuli differed only in the values of two spectral peaks, with all other variables kept constant. The peak frequency values of the fricative in [ʃɯkaɴ] were much lower than that in [ʃikaɴ]. Beckman and Shoji made the test tape using the eight stimuli. It consisted of one practice block followed by eight test blocks, each containing four stimulus tokens. 14 native speakers listened to each stimulus token and chose either [ʃikaɴ] or [ʃɯkaɴ] by marking the appropriate column on the answer sheet. The results showed that the location of spectral peaks alone shifted the perception from 100% [i] identification to 100% [ɯ] identification. Beckman and Shoji (1984) claim that the listeners used the location of the spectral peaks in the fricative to recover the intended vowel.

Beckman and Shoji's first perception experiment used words embedded in the frame sentence as stimuli, thus it is not clear whether the listeners used anticipatory vowel coarticulation as a vowel cue. Spectral characteristics of the following segments may have provided vowel cues. The result of the second experiment, on the other hand, does show that the spectral characteristics in the fricative provided a vowel cue. However, it is not clear if the second formant peaks provided a robust vowel cue as in the English data. As was mentioned before, it seems that the peaks Beckman and Shoji used for [ʃ(i)] were affiliated with the third and fourth formants of the following vowel, while those for [ʃ(ɯ)] were affiliated with the second and third formants. [ʃ(i)] had peaks around 2800 Hz and 3500 Hz, and [ʃ(ɯ)] had peaks around 1500 Hz and 2400 Hz (as noted before, the average F2 values for Japanese [i] and [ɯ] are 1950 Hz and 1400 Hz, respectively (Keating et al. 1984)). Thus the token which received 100% [i] identification did not actually contain the second formant peak for [i]. In fact, one of the intermediate stimuli must have contained a spectral peak whose frequency was close to the second formant of [i], and yet received low

[i] identification rate. This suggests that Japanese listeners may not be as sensitive to the second formant peaks as English listeners are. It is necessary to investigate the perception of both [ʃ(i)] and [ʃ(ɯ)] with the second formant peak.

The present study examines the perception of fricatives with second formant peaks in voiced and devoiced syllables. In addition, the perception of fricatives without a spectral peak is investigated. Note that many devoiced syllables, as well as some voiced syllables, lacked spectral peaks in the present acoustic study (both in Experiment 1 and Experiment 2). It is important to examine if the listeners can identify the vowel when the fricative is peakless.

6. Perception Study

6.1 Methods

A. Stimuli

The frication noise portion excised from natural speech was used as stimuli so that listeners would not have access to any possible cues in the following segment. The experimental stimuli were obtained from one of the speakers' utterances that were used in the acoustic study. There were eight different syllable types: [ʃ] followed by voiced [i], voiced [ɯ], devoiced [i] and devoiced [ɯ], and each of them had two versions, one with a second formant peak and the other without a peak. Two tokens were chosen for each syllable type, except the peakless voiced [ɯ], which had only one token. There was only one instance of peakless voiced [ɯ][8]. Thus there were (2 X 8) - 1 =15 tokens. Table 16 shows the peak frequency values of tokens with a peak:

		token 1	token 2
[i]	voiced	2480	2360
	devoiced	2410	2610
[ɯ]	voiced	2170	2240
	devoiced	2090	2210

Table 16. Peak frequency values (in Hz) of tokens with a second formant peak.

[8] It should be noted that although this token did not have a spectral peak in the second formant region for the vowel [ɯ], it did have a peak at 2550 Hz, which is slightly higher than the second formant region of the vowel [i]. I included this token as a peakless voiced [ʃ(ɯ)] type, since all other instances of voiced [ʃ(ɯ)] had a clear peak in the second formant region for [ɯ].

The overall amplitude of each token was equalized so that all tokens had the same value. Amplitude range of the original stimuli was 42.9 dB to 49.2 dB. Average amplitude of fricatives followed by [i] was 44.6 dB, while that followed by [ɯ] was 46.3 dB. After the equalization, all tokens had 45.6 dB (this value was picked, since three tokens had this originally and it was close to the average). Duration of most tokens was 83-89 ms. For six extra long fricatives, the initial part was removed, so that the duration of fricatives did not vary too much. This did not affect the sound. The onset of these six tokens sounded natural.

Each digitized token was edited on the computer and the frication noise part was segmented. The segmented part corresponds to the part that was used for the duration measurement in the acoustic study. The test consisted of 10 repetitions of 16 tokens (the voiced [ɯ] without a peak was repeated twice so that each syllable type has the same number of tokens) in randomized order (thus 16 X 10 = 160 tokens). Segments were separated by 2 second interval. A practice trial was also prepared. In this trial, 10 different frication segments were presented in randomized order.

B. Procedure

Ten Japanese speakers of standard Tokyo dialect were tested. None of the subjects had a known history of either speech or hearing disorders. They were tested individually. The subjects were told that the utterance of [ʃi] and [ʃɯ] were made by one speaker and that only the frication noise portion of each syllable would be presented. Their task was to identify the excised vowel by pushing one of two buttons, which were labeled 'shi' and 'syu' in Japanese orthography. For half of the listeners, the left button was labeled 'shi' and the right button, 'syu'. For the other half, the left button was labeled 'syu' and the right button, 'shi'. The subjects pressed the buttons using their right index finger (they were all right-handed). The subjects were told to indicate as quickly and accurately as possible which vowel they thought they heard by pressing the corresponding button. Response accuracy was measured. The subjects listened to the tokens over headphones.

Each test began with a practice set, which had 10 tokens. All subjects had only one practice set. After that, the subjects were asked if they had any problems or questions about the experiment. Then the test was presented.

The analysis of variance was performed for the perception study. The independent variables were vowel context (two levels), voicing context (two levels) and peakedness (two levels). The dependent variable was the correct response rate (%).

6.2 Results

6.2.1 [i] vs. [ɯ]

The identification rate, when averaged across all listeners, showed no main effect of vowel context [F(1,3) = .074, p > .79]. [ʃ(i)] and [ʃ(ɯ)] were correctly identified 61% and 62% of the time, respectively.

6.2.2 Devoiced syllables vs. voiced syllables

The voicing of the following vowel had a main effect on the identification rates [F(1, 3) = 13.730, p = .000]. Surprisingly, the identification rate for devoiced syllables (69%) was higher than that for voiced syllables (54%). This was the case for both vowels. Table 17 shows the identification rates of devoiced and voiced syllables in different vowel contexts:

	devoiced	voiced
ʃ(i)	65%	59%
ʃ(ɯ)	74%	50%

Table 17. Identification rates for devoiced and voiced syllables in different vowel contexts.

Devoiced [ʃ(i)] tokens were correctly identified significantly more often than the voiced ones [t(19) = 2.29, p = .033]. Devoiced [ʃ(ɯ)] was also correctly identified significantly more often than the voiced ones [t(19) = 5.7, p = .0001]. The difference between the identification rates of devoiced and voiced tokens, however, was greater for the vowel [ɯ] (24% difference). The interaction of the voicing of the vowel and the vowel context was significant [F(1, 3) = 4.32, p = .041].

6.2.3 Peaked fricatives vs. peakless fricatives

The presence of a peak had a main effect on the overall identification rate [F(1, 3) = 16.053, p = .000]. Fricatives with a peak were correctly identified more often. The identification rate for fricatives with a peak was 70%, while that for peakless fricatives was 53%. Table 18 shows the identification rates for peaked and peakless tokens in different vowel contexts:

	peaked	peakless
ʃ(i)	61%	63%
ʃ(ɯ)	80%	44%

Table 18. Identification rates for peaked and peakless tokens in different vowel contexts.

The identification rates for the vowel [i] did not differ significantly whether the fricative had a peak or not [$t(19) = .87$, $p > .40$]. For the vowel [ɯ], on the other hand, peaked fricatives received significantly higher identification rates than peakless ones [$t(19) = 7.04$, $p = .0001$]. This resulted in a significant vowel by peak interaction [$F(1, 3) = 21.80$, $p = .000$].

6.2.4 Identification Rate for Each Syllable Type

Figure 5 shows the identification rates for each syllable type:

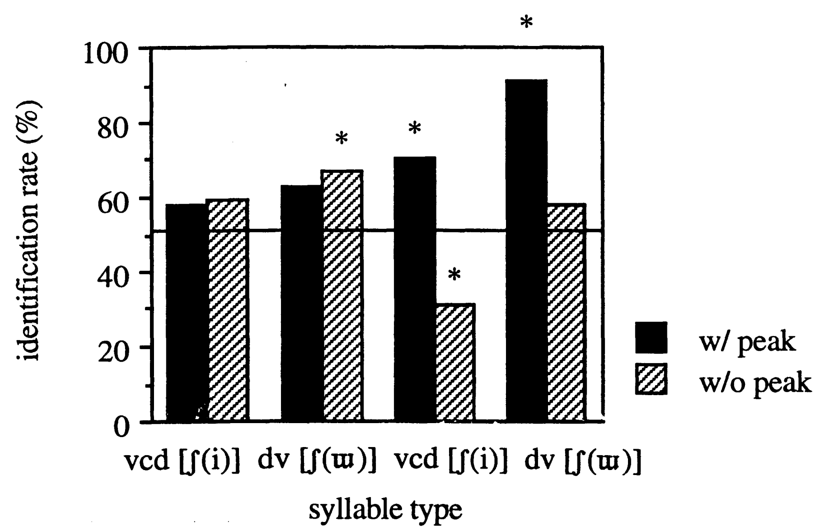

Figure 5. Identification rates for each syllable type (* indicates those which received significantly above or below chance level identification rates). The horizontal line represents chance level. 'Vcd' and 'dv' stand for voiced and devoiced, respectively.

As for fricatives with a peak, both voiced and devoiced [ʃ(i)] received near chance level identification. Devoiced [ʃ(i)] with a peak received a slightly higher identification rate than voiced [ʃ(i)] with a peak, yet the difference was not significant [$t(9) = 1.02$, $p > .33$]. Peaked [ʃ(ɯ)] tokens, on the other hand, were correctly identified at significantly better than chance level in both voiced and devoiced syllables. Moreover, devoiced [ʃ(ɯ)] with a peak received significantly higher identification rate than voiced [ʃ(ɯ)] with a peak [t(9) = 6.01, p = .0002].

Of the four types of peakless fricatives, only devoiced [ʃ(i)] was identified at significantly better than chance level. The identification rates for peakless voiced [ʃ(i)] and peakless devoiced [ʃ(ɯ)] were near chance level. Peakless voiced [ʃ(ɯ)] received a significantly below chance level identification rate. It should be noted, however, that this token had a clear peak at 2550 Hz (see footnote 8), which is slightly higher than a formant peak frequency for [i]. It can be assumed that the spectral peak was taken as a vowel cue for [i], which resulted in the low identification rate as [ɯ].

6.3 Discussion

The present perception study showed that fricatives in devoiced syllables were correctly identified significantly more often than those in voiced syllables. It was also observed that listeners identified peaked fricatives better than peakless fricatives.

Let us consider why fricatives in devoiced syllables received higher identification rates than those in voiced ones. Both peaked and peakless fricatives were correctly identified more often in devoiced syllables than in voiced syllables. As for the peaked tokens, one might assume that this was due to the peak frequency values. In the present experiment (see Table 16), the average peak frequency for [ʃ(i)] was higher (thus more [i]-like) for devoiced tokens (2510 Hz) than voiced tokens (2420 Hz). Also, the average peak frequency of devoiced [ʃ(ɯ)] tokens was 2150 Hz, which was lower (thus more [ɯ]-like) than the average peak frequency of the corresponding voiced tokens (2210 Hz). One might posit that more extreme frequency values of devoiced tokens resulted in higher identification rates. It should be recalled, however, that only [ʃ(ɯ)] received significantly higher identification rate for devoiced tokens than voiced tokens. The identification rates of voiced and devoiced [ʃ(i)] did not differ significantly. The peak frequency values alone do not account for this difference between [ʃ(i)] and [ʃ(ɯ)]. It is necessary to investigate other acoustic properties that caused this different identification rates.

For peakless fricatives, it is not clear why devoiced syllables received higher identification rates, either. One may assume that the lower edge frequency values were used as a vowel cue when fricatives lacked a peak. This indicates that lower edge frequency values provided stronger vowel cues for devoiced syllables than voiced syllables. However, as was shown in the present acoustic study, the lower edge frequency of devoiced syllables overlap greatly in the two vowel contexts, while that of voiced syllables showed distinct values for each vowel. Moreover, the edge frequency values were shown to be affected by the following consonant in devoiced syllables. Thus it is not plausible that the lower edge frequency values provide stronger vowel cues in devoiced syllables. Obviously it is necessary to investigate what acoustic properties were used as vowel cues, when spectral peaks were lacking.

Let us now consider the difference between the identification rates of peaked and peakless fricatives. The significantly higher identification rates for peaked fricatives seem to support Yeni-Komshian and Soli's (1981) claim, which assumes that a spectral peak provides a robust cue for the excised vowel. Peaks at lower frequencies seem to have been used as a cue for [ɯ]. It should be noted, however, that the presence of a peak was crucial for the identification of [ɯ], but not [i]. Identification rates of peaked and peakless [ʃ(i)] did not differ significantly, while peaked [ʃ(ɯ)] received significantly higher identification rate than peakless [ʃ(ɯ)]. In fact, the identification rates of peaked [ʃ(i)] were at chance level for both voiced and devoiced tokens. This indicates that spectral peaks do not provide a strong cue for [ʃ(i)]. This difference may be due to the lower peak occurrence rates in [ʃ(i)] syllables than in [ʃ(ɯ)] syllables. In the present acoustic study, it was observed that fricatives followed by [i] showed a spectral peak less often (69%). 88% of the fricatives followed by [ɯ], on the other hand, had a spectral peak. It may be the case that Japanese listeners generally use a spectral peak as a strong vowel cue for [ɯ], since it is constantly present, but not for [i], since it does not always appear.

6.4 Center of Gravity

So far we have considered a spectral peak as a strong vowel cue. Yet the fact that the peakless devoiced [ʃ(i)] received significantly better than chance level identification rate suggests that it is necessary to look at some other factors. Previous studies have shown that the center of gravity of vowels is related to vowel quality (Chistovich and Lublinskaya 1979; Chistovich et al. 1979). Center of gravity, or centroid 1st moment of spectral distribution, is a statistical measure which indicates the concentration of the energy distributions (Forrest et al. 1988). Frequency weighted by amplitude can be averaged as center of gravity. Chistovich et al. (1979), in their study of centers of gravity effects on vowel perception, report that synthesized vowels with a low center of gravity were almost always identified as the vowel [u]. Thus, the center of gravity of the fricatives was examined to see if there was any distinct pattern for each type of syllable. For each token, a sequence of spectra was computed. Each spectrum was computed with a 10-ms analysis window. Seven analysis intervals were obtained for each token. Figure 6 shows the centers of gravity of peaked fricatives in both devoiced and voiced syllables:

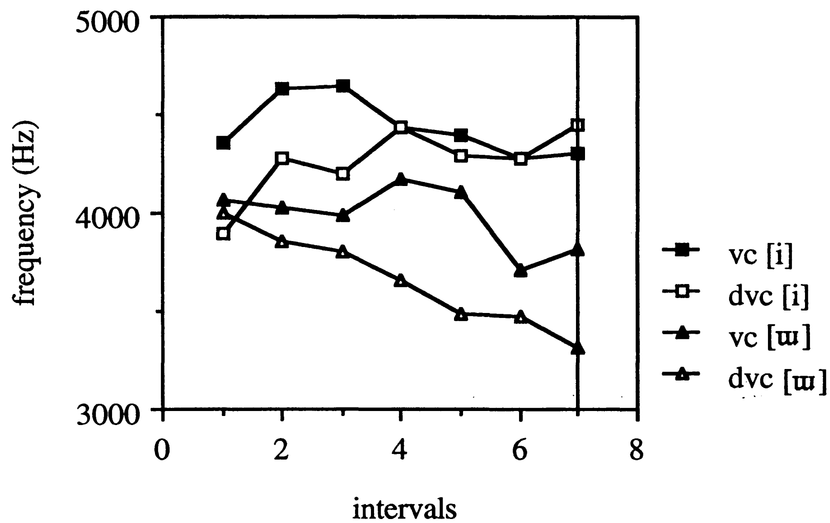

Figure 6. Centers of gravity of peaked fricatives in devoiced and voiced syllables. The vertical line indicates vowel onset.

Except at the beginning of the fricative, the mean frequencies of [ʃ(i)] were always higher than those of [ʃ(ɯ)] for both devoiced and voiced syllables. We can assume that this is a coarticulation effect from the following vowel. It should be noted that Figure 6 shows lower centers of gravity for devoiced [ʃ(ɯ)] than voiced [ʃ(ɯ)] for the entire period. Voiced and devoiced [ʃ(i)] tokens, on the other hand, had similar centers of gravity in the latter half of the fricative. Recall that in the present perception study, the identification rates of devoiced [ʃ(ɯ)] with a peak (91%) was significantly higher than voiced [ʃ(ɯ)] with a peak (70%). The difference between devoiced and voiced [ʃ(i)] with a peak, however, was not significant (63% for devoiced tokens and 59% for voiced tokens). If we assume that centers of gravity provide a strong vowel cue for Japanese listeners, the difference can be accounted for. We can assume that devoiced [ʃ(ɯ)] was accurately identified more often than voiced [ʃ(ɯ)], since lower centers of gravity provided a strong vowel cue. The identification rates of [ʃ(i)] did not differ, since their centers of gravity were similar.

Now let us consider the centers of gravity of both peaked and peakless [ʃ(i)] in devoiced and voiced syllables. Only devoiced peakless [ʃ(i)] token received a significantly better than chance-level identification rate. Figure 7 shows the centers of gravity of all types of [ʃ(i)]:

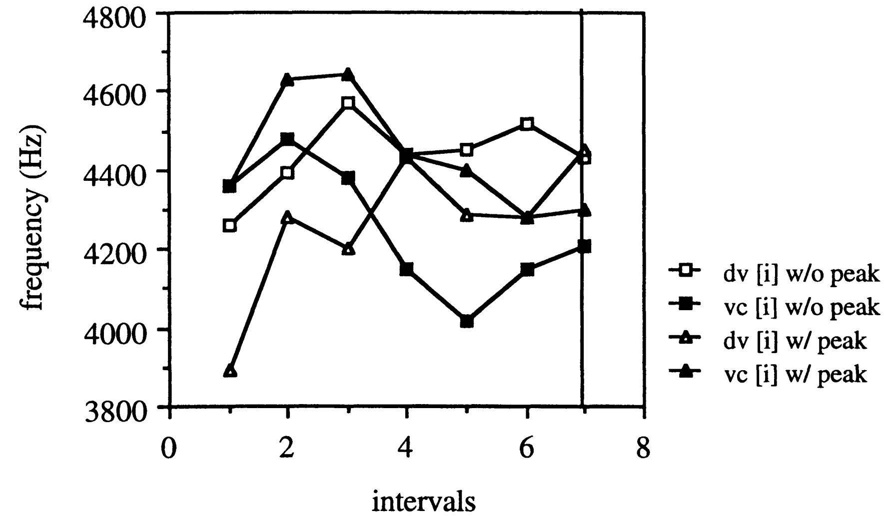

Figure 7. Centers of gravity of peaked and peakless fricatives in devoiced and voiced [ʃ(i)]. 'dv' and 'vc' stand for devoiced and voiced, respectively. The vertical line indicates vowel onset.

It can be observed that devoiced [ʃ(i)] without a peak maintained high centers of gravity for the entire period, while voiced [ʃ(i)] without a peak and devoiced [ʃ(i)] with a peak did not. Interestingly, voiced [ʃ(i)] with a peak had higher centers of gravity than devoiced [ʃ(i)] without a peak at the beginning of the fricative, and yet did not receive a high identification rate. This seems to suggest that the centers of gravity at the end of the fricative provides a stronger vowel cue than those at the beginning of the fricative.

Let us now consider the [ʃ(ɯ)] tokens. Figure 8 shows centers of gravity of both peaked and peakless [ʃ(ɯ)] in devoiced and voiced syllables:

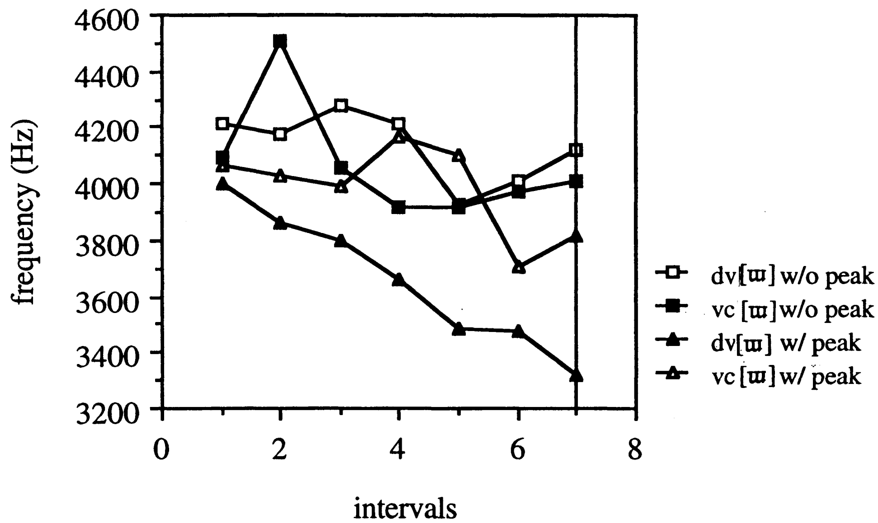

Figure 8. Centers of gravity of peaked and peakless fricatives in devoiced and voiced [ʃ(i)]. 'dv' and 'vc' stand for devoiced and voiced, respectively. The vertical line indicates vowel onset.

Figure 8 shows that at the end of the fricative, peaked tokens had lower centers of gravity than peakless ones. In the present study, peaked tokens received significantly above chance level identification rates, while peakless tokens received near or below chance level identification rates. It can be assumed that peakless tokens were identified less accurately, since their centers of gravity were higher.

Of the two types of peakless syllables, peakless devoiced [ʃ(ɯ)] received 58% correct response. Yet the identification rates of the two tokens differed greatly. The first [ʃ(ɯ)] token received only 44% correct rate (near chance level), while the second token received 71% (this was significantly above chance level). Figure 9 shows the centers of gravity of all peakless [ʃ(ɯ)] tokens (note that there was only one token for the peakless voiced [ʃ(ɯ)]):

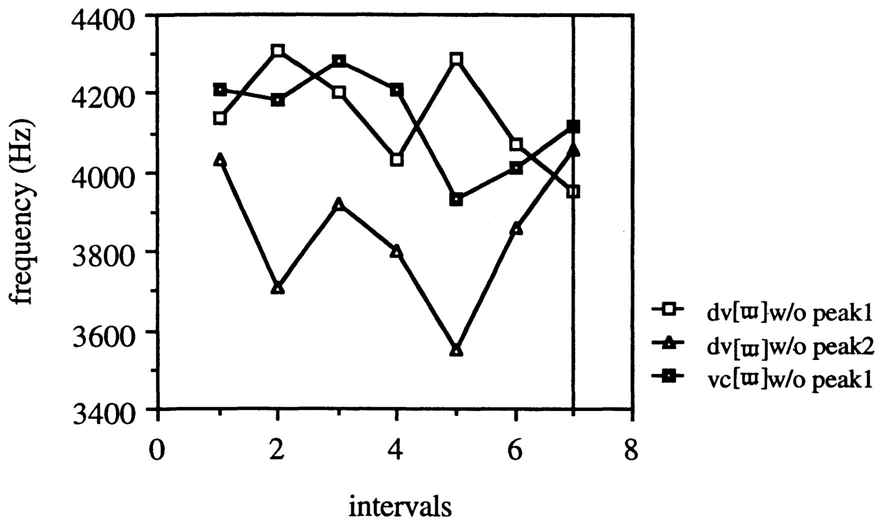

Figure 9. Centers of gravity of peakless devoiced and voiced [ʃ(ɯ)]. The vertical line indicates vowel onset. The numbers 1 and 2 indicate the first and second tokens.

It can be observed that the centers of gravity of the two devoiced [ʃ(ɯ)] tokens were quite different. The second token's frequency values were constantly low (thus more [ɯ]-like), while the first token's values were much higher (thus more [i]-like). The difference between the identification rates of the two tokens can be accounted for, if we assume that centers of gravity provided a vowel cue. It should be noted that voiced [ʃ(ɯ)] received a significantly below chance-level identification rate. Centers of gravity values of this token were higher (thus more [i]-like) than those of the second token of devoiced [ʃ(ɯ)]. In addition, it should be recalled that this token had a peak at a higher frequency. Both the peak location and center of gravity frequency values of this token seem to have been perceived as a cue for [i], yielding a lower identification rate.

To summarize, listeners seem to have used center of gravity as a vowel cue. When mean frequency was low, tokens were identified as [ɯ], which is in agreement with a previous study (Chistovich et al. 1979). The accurate identification of devoiced peakless [ʃ(i)] suggests that the centers of gravity at the end of the fricative provided a vowel cue.

One question arises: when there is a formant peak in the fricative, which do listeners use as a primary vowel cue, peak frequency or center of gravity? In general, when the center of gravity is high (low), the peak frequency is also high (low). For some tokens, however, the center of gravity was high, while the peak frequency is low (or vice versa).

Surprisingly, in such cases, it seems that center of gravity is used as a primary cue. When the two tokens of voiced [ʃ(i)]with a peak are compared, it is observed that one of them has a higher mean frequency than the other one for the entire period. The formant peak frequency, however, was higher for the token with lower center of gravity values (2480 Hz, while it was 2360 Hz for the other token). If listeners use the center of gravity as a primary vowel cue, the token with higher center of gravity should receive a higher identification rate. On the other hand, if listeners use the formant peak frequency as a primary cue, the token with a higher formant peak frequency should receive a higher identification rate. The results indicate that the center of gravity is used as a primary cue, since the identification rate was higher for the token with a higher center of gravity (65%, while the other token received 51%). In addition, for peaked fricatives, it was shown that centers of gravity values, rather than peak frequency values, account for the identification rate patterns of [ʃ(i)] and [ʃ(ɯ)] in different voicing contexts (see Figure 6). Thus the present study suggests that the peak frequency is not the strongest vowel cue in Japanese.

6.5 Comparison to Other Studies

This section compares the results of the present study with those of previous studies. Both in Yeni-Komshian and Soli (1981) and LaRiviere et al. (1975), it was reported that the identification rate for (voiced and peaked) [ʃ(i)] was near or below chance, while that for (voiced and peaked) [ʃ(u)] was significantly above chance level. In the present study, it was also observed that voiced [ʃ(i)] with a peak received a near chance identification rate, while the corresponding [ʃ(ɯ)] tokens received a significantly above chance level identification rate. Soli (1981) attributed the poor identification of [ʃ(i)] to the great prominence of the third formant peak, which was close to the second formant peak in frequency (2000 Hz for the second formant peak and 2500 Hz for the third formant peak). It was claimed that the third formant peak masked the second formant peak, obscuring the vowel cue. Yet this seems not to be the case for the present data. In the present acoustic study, the second and third peak frequencies were 2440 Hz and 3330 Hz, respectively. The difference in peak frequencies of the second and third formants was much greater. Thus it is not plausible that the third formant peak masked the second formant peak, obscuring the vowel identity. We may assume that Japanese listeners do not always utilize the high peak frequency as a cue for [i], since the peak in [ʃ(i)] is not always presence.

Let us now consider the devoiced tokens. In the present study, the identification rates for devoiced [ʃ(i)] and [ʃ(ɯ)] were 65% and 74%, respectively. Beckman and Shoji (1984) report that when a word with a devoiced syllable was presented in a frame sentence, the identification rates were 77% for [ʃ(i)] and 67% for [ʃ(ɯ)]. Although chance level for the present study was 50%, while in Beckman and Shoji's study it was 40%, the results of the two studies seem to be similar. This suggests that listeners identify the devoiced vowel mainly by the coarticulation effect observed in the fricative. It is not clear why the identification rate was higher for [ɯ] in the present study, while the vowel [i] received higher identification rate in Beckman and Shoji's study. It should be noted that average amplitude was equalized in the present experiment, which may have affected the identification rate.

7. Theoretical Implications of the Present Study

In the present study, both acoustic and perceptual evidence for fricative-vowel coarticulation effects in Japanese was observed. In this section, I examine the theoretical implications of the present study on different analyses of Vowel Devoicing. At the present time, several analyses of Japanese Vowel Devoicing are proposed. I am aware of four different types of analyses: high vowels are said to be (1) phonologically devoiced (Han 1962; Sawashima 1971; Shibatani 1990); (2) phonologically deleted (Ohso 1973; Beckman and Shoji 1984); (3) phonologically devoiced after stops but deleted after affricates and fricatives (Kawakami 1977) or (4) phonetically hidden (Beckman 1993). This section investigates how the following findings of the present acoustic study can be accounted for by each analysis: (1) anticipatory vowel coarticulation was observed in both voiced and devoiced syllables, but the effects were weaker in the devoiced syllables in that the peak occurrence rate was lower and the peak amplitude was smaller; and (2) coarticulation effects from the following consonant was observed in the devoiced syllables, but not in the voiced syllables.

7.1 Phonological Devoicing

Many researchers assume that the high vowels are phonologically devoiced in Japanese (Han 1962b; Sawashima 1971; Shibatani 1990). Devoiced vowels are considered as allophones of the voiced counterparts (McCawley 1968). Devoicing may be considered as an assimilation process, where the [-voice] feature of the neighboring consonants is assimilated.

The phonological devoicing analysis can account for the findings of the present study if we assume the following: a devoiced vowel has a weaker status than a fully voiced vowel in the phonology and it cannot serve as a syllable nucleus. As a result, a [ʃV.C] sequence becomes a fricative-devoiced vowel-stop cluster. A devoiced vowel does not affect the preceding fricative as much as a fully voiced vowel does, and its weaker status allows properties of the following consonant to appear in the fricative.

However, some problems remain. The devoicing analysis predicts that Japanese devoiced vowels should show some formant-like bands as typical devoiced vowels in other languages do. Yet, it is reported that when a devoiced vowel occurs after a fricative, often times no characteristics of vowels are found in spectrographic data (Kawakami 1977; Beckman and Shoji 1984). The present study also observed that devoiced high vowels after the fricative [ʃ] lacked any characteristics of vowels. The high vowels following a fricative seem to be deleted. The devoicing analysis alone does not explain why this happens. There must be other mechanisms that further reduce the vowel quality.

It should be noted that Vowel Devoicing shows gradient effects (Keating 1990; Cohn 1993). I use the term gradient to refer to changes over time during a segment, following Cohn (1993). Kondo (1993) reports that some high vowels were only partially voiced in her experiment. If all processes of Vowel Devoicing occurs in the phonology, no gradient effects should be observed. This means that a high vowel in Japanese should be either fully voiced or completely devoiced. We should not find a high vowel with partial voicing. Kondo's findings suggest that certain process of Vowel Devoicing applies in the phonetics. If we assume that the reduction of the vowel quality applies in the phonetics, gradient effects can be accounted for. Further investigation of the cases with partial devoicing is necessary to support the phonological devoicing analysis.

7.2 Phonological Deletion

Ohso (1973) assumes that the high vowels are deleted in the phonology by the following rule: [+high] V -> ø / [-vc] _ [-vc] or #. This rule deletes any high vowels that occur between two voiceless segments. Kawakami (1977, cited in Maekawa 1989), on the other hand, claims that high vowels are devoiced after stops, but deleted after affricates and fricatives. Sakuma (1929, cited in Maekawa) also mentions this distinction. The phonological deletion analyses, whether deletion applies to all vowels or only to vowels following a fricative, seem to have a problem in explaining the results of the present study. The analyses can account for the assumption that the high vowels following [ʃ] are transparent at the phonetic level. High vowels are deleted in the phonology, thus obviously they do not exist in the phonetics. [ʃVC] sequence yields fricative-stop clusters after the vowel is deleted. However, anticipatory vowel coarticulation effects (in the phonetics) cannot be accounted for, if high vowels are deleted in the phonology. If vowels are indeed deleted in the phonology, it is necessary to assume either that (1) spectral characteristics observed in fricatives actually reflect a phonological assimilation process; or that (2) some properties of vowels still remain after the deletion and affect the preceding consonant in the phonetics.

The first assumption indicates that spectral characteristics observed in fricatives should be the same whether the following vowel is deleted or not, since the assimilation process would have to precede vowel deletion. Yet the present study has shown that the effects of vowels on fricatives were weaker when the vowel is deleted. The amplitude of a spectral peak was smaller in devoiced syllables. In addition, if the effects in fricatives are due to a phonological assimilation process, the process should not affect only the last part of a fricative. As Cohn (1993) states, since 'phonological rules manipulate discrete, timeless segments', the entire fricative should show assimilation effects. Both Soli's study and the present study, however, show that the difference in peak frequency by the following vowel was greater in the last part of the fricatives. These data suggest that spectral characteristics observed in fricatives indeed reflect coarticulation effects.

The second assumption, which posits that some properties of vowels still remain after the deletion, seems to account for both vowel and consonant coarticulation effects. It can be posited that the following consonant can affect the initial fricative in devoiced syllables, since some properties of vowels remain in the structure. It can be further assumed that vowel coarticulation effects are stronger in voiced syllables, since vowels are physically present in the phonetics. The presence of vowels block the coarticulation effects from the following consonant. However, it is not clear what properties remain after deletion. In the present acoustic study, it was shown that many fricatives contain spectral peaks, which are affiliated with the formants of the following vowel. The formants are 'the results of the different ways in which the air in the vocal tract vibrates' (Ladefoged 1982: 175). This indicates that some information of vocal tract shape (tongue position) remains in the structure. At the present time, I am not aware of any other data that motivate the presence of vocal tract shape information in the phonology. Further investigation is necessary to determine whether there are any other pieces of evidence to support the phonological deletion analysis.

7.3 Phonetically Hidden

Recently, Jun and Beckman (1993) claimed that Vowel Devoicing is a purely phonetic process, contrary to the earlier claim in Beckman and Shoji (1984), where high vowels were assumed to be deleted in the phonology. Moreover, she argues that the high vowels are hidden, rather than deleted. The notion of 'phonetically hidden' segment is introduced in Browman and Goldstein (1990). Browman and Goldstein found that while in a production of the English token perfect memory, [t] seems to be deleted in casual speech, articulatory data actually show evidence of the alveolar gesture. Both the closure and the release of the alveolar gesture overlap the velar and bilabial gestures of the neighboring consonants, thus [t] is not heard. [t] is hidden. Similarly, Jun and Beckman (1993) claims that acoustic properties of devoiced Japanese high vowels are hidden, since the glottal opening gesture for the neighboring voiceless consonants overlaps the glottal closing gesture of the vowel. She states that this overlap is most commonly seen among high vowels, since they are the shortest in duration. She further claims that other vowels can also show devoicing when they are short.

It is not clear how the overlap analysis accounts for the transparency of a vowel in the phonetics. The significantly lower peak occurrence rates in fricatives in devoiced syllables seem problematic. If a vowel is hidden, a clear coarticulation effect of the vowel should be observed in the fricative, since the dorsal gesture must be the same as that in voiced vowels. The vowel is not heard since the glottal gesture is overlapped, yet that should not prevent formant peaks from appearing in the fricative. The present data can be explained only if we assume that glottal overlap affects the dorsal gesture. It is necessary to investigate both the glottal gesture and the dorsal gesture during the production of voiced and devoiced syllables in order to justify Beckman's claim.

There are some other problems with the overlap analysis. Kondo (1993) reports that Vowel Devoicing is observed in slow speech as well. This seems problematic, since vowel duration is longer in slow speech. Also, it should be noted that although non-high vowels sometimes devoice in Japanese, [e] is less likely to devoice than [a] and [o] (Sakurai 1966; Maekawa 1989). The overlap analysis does not explain why this is the case. Han (1962) reports that the longest vowel in Japanese is [a]. Why can't shorter [e] devoice more often? Finally, Beckman's analysis does not account for the fact that Vowel Devoicing does not occur in some dialects. We need to posit that overlap is a dialect-specific process.

At the present time, it is difficult to conclude which analysis is the correct one, since further investigation is necessary in various areas, as was discussed above. Through more extensive study, however, insight can be gained into the exact nature of Japanese Vowel Devoicing.

8. Conclusion

In the present study, acoustic and perceptual evidence of fricative-vowel coarticulation in Japanese were discussed. It was observed that there were clear anticipatory vowel coarticulation effects in Japanese voiced syllables, that are similar to those found in English (Soli 1984). Spectral peaks, which are affiliated with the second formant of the following vowel, were constantly observed in the last part of the fricative. Spectral peaks were also observed in some devoiced tokens and they seemed to reflect the second formant of the following vowel. Yet the amplitude of the peaks in devoiced tokens was significantly smaller than that of voiced tokens. In addition, many fricatives in devoiced syllables lacked spectral peaks. One of the striking findings of the present study is that fricatives in devoiced syllables showed coarticulation effects from the following consonant. Peak occurrence rates, peak amplitude and lower edge frequency varied significantly depending on the following consonant. Such variation was not found in fricatives in voiced syllables. This result, along with the findings in previous studies (Repp and Mann 1981; Yoshioka 1981), seems to suggest that devoiced vowels in the present study are transparent at the phonetic level.

In the perception experiment, it was examined if listeners could identify intended vowels when they were presented only the fricative noise portion. A previous study (Yeni-Komshian and Soli 1981) has shown that English-speaking listeners use a second formant peak as a strong vowel cue. The present data indicate that Japanese listeners also used a formant peak as a vowel cue. This seems to be the case whether the vowel was devoiced or not. Yet it was observed that the presence of a peak was essential for the identification of [ɯ], but not for [i]. It was argued that this may be because [ɯ] has a formant peak constantly, while many [i] tokens lack a peak. In addition, it was suggested that Japanese listeners used center of gravity as a vowel cue. Low center of gravity values were used as a cue for the vowel [ɯ]. It was argued that center of gravity is used as a primary cue, when a conflicting spectral peak was present in the fricative. These results suggest that what listeners use as a vowel cue can vary among different languages, even when similar coarticulation effects are found in the preceding segment. It can be assumed that since formant peaks in Japanese do not appear as constantly as in English, Japanese listeners use different tactics to identify vowels. Yet it is not clear if only Japanese listeners use the center of gravity as a vowel cue. It will be interesting to see whether English listeners are also sensitive to different center of gravity values.

9. References

Beckman, M. (1982) Segment Duration and the 'Mora' in Japanese. *Phonetica* 39, 113-135.

Beckman, M., and A. Shoji (1984) Spectral and Perceptual Evidence for CV Coarticulation in Devoiced /si/ and /syu/ in Japanese. *Phonetica* 41, 61-71.

Browman C.P., and L. Goldstein (1990) Tiers in Articulatory Phonology, with some implications for casual speech. In J. Kingston and M. Beckman (eds.) *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (pp. 341-376). Cambridge: Cambridge University Press.

Chistovich, L.A., and V.V. Lublinskaya (1979) The 'Center of Gravity' Effect in Vowel Spectra and Critical Distance Between the Formants: Psychoacoustic Study of the Perception of Vowel-like Stimuli. *Hearing Research* 1, 185-195.

Chistovich, L.A., R.L. Sheikin, and V.V. Lublinskaja (1979) "Centers of Gravity" and Spectral Peaks at the Determinants of Vowel Quality. In B. Lindblom and S. Ohman (eds.) *In Frontiers of Speech Communication Research* (pp. 143-157). Academic Press.

Cohn, A. (1993) Nasalisation in English: phonology or Phonetics. *Phonology* 10, 43-81.

Forrest, K., G. Weismer, P. Milenkovic, and R.N. Dougall (1988) Statistical Analysis of Word-initial Voiceless Obstruents: Preliminary Data. *JASA* 84, 115-123.

Han, M.S. (1962) The Feature of Duration in Japanese. *Onsei no Kenkyuu* 10, 65-80.

Han, M.S. (1962b) Unvoicing of Vowels in Japanese. *Onsei no Kenkyuu* 10, 81-100.

Heinz, J.M., and K.N. Stevens (1961) On the Properties of Voiceless Fricatives. *JASA* 33, 589-596.

Jun, S.-A., and M.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, Los Angeles, California.

Kawakami, S. (1977) *Nihongo Onsei Gaisetsu.* Tokyo: Oofuusha.

Keating, P.A. (1990) The Window Model of Coarticulation: Articulatory Evidence. In J. Kingston and M. Beckman (eds.) *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (pp. 451-470). Cambridge: Cambridge University Press.

Keating, P., M. Huffman, and E. Jackson (1984) Vowel allophones and the vowel formant phonetic space. *UCLA Working Papers in Phonetics* 59, 50-61.

Kondo, M. (1993) The Effect of Blocking Factors and Constraints on Consecutive Vowel Devoicing in Standard Japanese. Poster presented at Labphon 4, Oxford, 11-14 August, 1993.

Kondo, M. (1994) Mechanisms of Vowel Devoicing in Japanese. Paper presented at 1994 International Conference on Spoken Language Processing, 18-22 September, Yokohama.

Kozhevnikov, V.A., and L.A. Chistovich (1965) *Rech: Artikulyatsiya i Vospriyatiye* (Moscow-Leningrad, 1965). Trans. Speech: Articulation and Perception. Washington, D.C.: Joint Publication Research Service 30, 543.

Ladefoged, P. (1982) *A Course in Phonetics.* Harcourt Brace Jovanovich: New York.

LaRiviere, C.J., H. Winitz, and E. Herriman (1975) The Distribution of Perceptual Cues in English prevocalic fricatives. *J. Speech Hear. Res.* 18, 613-622.

Maekawa, K. (1989) Boin-no Musei-ka (Devoicing of Vowels). In Sugito, M. (ed.) *Nihongo-to Nihongo Kyooiku* 2 (pp. 135-153). Meiji Shoin.

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

Ohso, M. (1973) A Phonological Study of some English Loanwords in Japanese. in Studies in Phonology and Methodology. *Ohio State University Working Papers in Linguistics* 14, 1-26.

Port, R.F., J. Dalby, and M. O'Dell (1987) Evidence for mora timing in Japanese. *JASA* 81, 1574-1585.

Repp, B.H., and V.A. Mann (1981) Perceptual Assessment of Fricative-Stop Coarticulation. *JASA* 69, 1154-1163.

Sakuma, K. (1929) *Nihon Onseigaku (Japanese Phonetics).* [References are to the 1963 reprint. Tokyo: Kazama Shobou.]

Sakurai, S. (1966) Kyoutsuugo-no Hatsuon-de Chuui Subeki Kotogara (On Pronunciation of Standard Japanese). In Nihon Hoosoo Kyookai (ed.) *Nihongo Hatsuon Akusento Jiten* (pp. 31-43). Tokyo: Nohon Hoosoo.

Sawashima, M. (1971) Devoicing of Vowels. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics* 5, 7-13. Tokyo University.

Shibatani, M. (1990) *The Languages of Japan.* New York: Cambridge University Press.

Soli, S. (1981) Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation. *JASA* 70, 976-84.

Takeda, K., and H. Kuwabara (1987) Analysis and Prediction of devoicalizing phenomena (in Japanese). *Proceedings of Acoustic Society of Japan.* 105-106.

Vance, T.J. (1987) *An Introduction to Japanese Phonology.* Albany: State University of New York Press.

Yeni-Komshian, G., and S. Soli (1981) Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation. *JASA* 70, 966-75.

Yoshioka, H. (1981) Laryngeal Adjustments in the Production of the Fricative Consonants and Devoiced Vowels in Japanese. *Phonetica* 38, 236-251.