

CHAPTER 2

LOSS OF VOWEL VOICING IN JAPANESE

2.1 INTRODUCTION TO CHAPTER 2

This chapter presents an overview of the various characterizations of loss of high vowel voicing in Japanese.

The organization of this chapter is as follows. §2.2 presents an overview of previous work done on vowel devoicing in Japanese. It begins with the traditional phonological feature-changing formulation of the rule and notes problems with that formulation. It continues with the alternative gestural overlap account of devoicing presented in Jun & Beckman (1993) that is based on Browman & Goldstein (1986, 1990, 1992). The analysis of Tsuchida (1994, 1997) is then reviewed, which contends that the apparent discrepancies between the phonological and gestural accounts of devoicing are best resolved by positing both processes at work in Japanese—a gestural overlap account that causes loss of voicing of non-high vowels and high vowels between fricatives, and a phonological process involving the spread of the feature [+spread glottis] from the preceding voiceless obstruent that causes devoicing of high vowels in other devoicing environments. Evidence from the current data set is also presented supporting and refining this analysis.

§2.3 continues with an aspect of Japanese vowel devoicing that apparently has not yet received proper attention in the literature, the fricativization accompanying devoiced vowels that is prevalent in the current data set and routinely observable in media and public speech. Vowels in the devoicing environment are often fricativized following coronal and velar obstruents, consistent with the report of the ‘obstruentized’ vowels of Uyghur reported on in Hahn (1991) and other fricativized vowels in Standard Chinese (Ladefoged & Maddieson 1990, 1996). In the current data set this fricativization occurs both with voiced and devoiced vowels, implying independent manipulation of the oral closure of the voicing gestures.

§2.4 attempts to clarify the terminology associated with the devoicing and fricativization processes, and §2.5 gives geometric representations of the various changes that vowels in Japanese can undergo. Finally, §2.6 summarizes this chapter.

2.2 HIGH VOWEL DEVOICING

The devoicing of high vowels is one of the hallmarks of Japanese in the Kanto region surrounding Tokyo. Its description has a long history in the literature (e.g. Sakuma 1929; Bloch 1950; Martin 1952). According to Maekawa (1989), citing Miyajima (1961), the earliest reference to devoicing is apparently a Latin grammar of Japanese by D. Collado in 1632. This section will discuss the application of the rule of High Vowel Devoicing (HVD), beginning with factors said to affect its application.

2.2.1 FACTORS AFFECTING THE APPLICATION OF HVD

There are many factors affecting the application of HVD, including vowel height, consonantal environment, position in the word, phrase and utterance, the quality of the vowel in the following mora, the number of consecutive devoicable vowels, pitch accent location, formality of speech, rate of speech, dialectal and ideolectal variation, generational variation, and grammatical structure of the word (Kondo 1997; see §4.5 of that work for an excellent overview). This section will review the factors investigated in this study and the literature that has reported them.

2.2.1.1 SPEECH RATE

HVD was characterized as a fast speech rule, as opposed to a casual speech rule, by Hasegawa (1979). That is to say, it was classified as a rule dependent on Speech Rate (SR) and not variations in formality. Previous research has provided evidence for this SR effect (Kuriyagawa & Sawashima 1989, Maekawa 1990, Kondo 1993, Jun & Beckman 1993). In fact, it has been shown that non-high vowels are also devoiced by some speakers at faster speech rates (Maekawa 1990; Kondo 1993).

However, as noted by both Kondo (1997) and Tsuchida (1997), in actuality it is very difficult to categorize HVD as a fast speech rule—Sakuma (1929) maintains that if a word contains only one devoicable vowel, it must be devoiced; Bloch (1950) and Kawakami (1977) both maintain that high vowels in the devoicing environment are voiced only in unnaturally careful speech. In addition, Vance (1987:55) suggests and Beckman (1994:57) states outright that high vowels are devoiced even in the most carefully read speech. Kondo (1994) notes that one speaker in that study did

not voice any of the 9 vowels in non-consecutive devoicing sites at the determined normal SR, and only 1 out of 9 vowels at the slow SR. Han (1994: 76) reports that devoicing of [i] and [u] were quite variable within that group of participants as well, with some participants devoicing many vowels at all SRs. The participants in Varden & Sato (1996) and Kondo (1997) also showed loss of high vowel voicing at slow SRs.

While a SR effect was still observed in Varden & Sato (1996) and the data in the current study, these observations and studies suggest that the ‘fast speech’ requirement of HVD is being lost for at least younger speakers of the Tokyo dialect. It was this loss of SR that initially prompted the study being reported on in this work. As will be seen in subsequent chapters, many vowels were devoiced at even the slowest SR utilized in this study.

2.2.1.2 PITCH ACCENT PLACEMENT

Perhaps the factor most discussed in the literature is the effect of lexical pitch accent placement on the application of HVD. Traditionally it has been held that accented vowels tend not to be devoiced (McCawley 1977; Haraguchi 1977); accented vowels are either not voiced, or the accent is shifted to another vowel. However, it has been noted that this accent shift appears to be disappearing, especially for younger speakers (Sugito & Hirose 1988; Kondo 1997; Tsuchida 1997: 23-31; Kitahara 1997, 1998; Varden 1997). In particular, in work unfortunately not well known in the west, Sugito (1966, 1969, 1971) notes that the loss of voicing on an accented vowel results in the lost pitch being determined by the level of pitch on the following vowel.¹ Additionally, perhaps the most authoritative accent dictionary of Standard Japanese, NHK (1985), lists both voiced and devoiced variants for many accented vowels. Further, Tsuchida (1997: 26) claims that the only interaction still seen between accent and loss of voicing for younger speakers is in borrowed vocabulary. Finally, Kitahara (1997, 1998) provides evidence from phrasal pitch contours that when accented high vowels were devoiced, the pitch accent location had indeed been lost and had not been shifted to another mora.

¹Thanks to Tsutomu Sato for bringing Professor Sugito’s earlier papers to my attention.

It will be noted that in the present study it also appears that accent and loss of voicing did still interact to some degree—of the 10 participants examined in this study, all but 2 showed at least some variability in pitch accent placement accompanied by variable devoicing of vowels (i.e. voicing the 1st vowel of a token and devoicing the 2nd in one repetition, and then devoicing the 1st vowel and voicing the 2nd in another repetition of the same token). Further research is necessary to determine how much of an influence the experimental setup had on this variability.

Overall the devoicing of accented vowels is in general supported by the current data set, and will be discussed more fully in Ch. 3. The devoicing of accented vowels, together with the alternative strategies available for avoiding devoicing, indicate that there will be no significant effect of pitch accent placement on vowel devoicing. Effects of pitch accent placement on the devoicing of vowels in a later chapter will therefore not be checked.

2.2.1.3 SOCIOLOGICAL FACTORS

It would appear that sociological factors are now exerting the greater influence on the application of the rule than either SR or accent placement. Beckman (1994: 66) notes both the spreading of devoicing in the ‘prestige’ dialect of Tokyo, and that purposeful voicing of sentence-final vowels (i.e. suppression of devoicing) in the Osaka region seems to be used to exclude speakers from the Tokyo region. Imaizumi et al. (1995) notes a decrease in the number of vowels devoiced when teachers are interacting with hearing-impaired children then when they are interacting with normal-hearing children. Perhaps surprisingly, Yuen & Hubbard (1997) found that sociological factors (gender, speaking style, etc.) played a greater role in the number of vowels devoiced by the participants in that study than either segmental environment or pitch accent placement. And it is the impression of the author that sentence-finally the [u] of the copula *desu* and *-masu* (a non-past verb ending) are voiced only in situations when an impression of formality or respect is desired (e.g. during newscasts, or when talking to those of higher social rank or strangers). It would appear that sentence-final devoicing of these vowels, at least for many speakers, should now be considered ubiquitous, with its suppression to indicate formality or respect being the marked case.

2.2.1.4 MORaic POSITION OF THE VOWEL

Another factor affecting the application of HVD that has been noted in the literature is that vowels in the 1st mora of two-mora words tend to devoice more than the vowels in the 2nd mora (Kuriyagawa & Sawashima 1989). This finding was supported by the current data set, although the effect was not consistent for all tokens or at all 3 SRs utilized in this study. §6.4.1 will discuss this further.

2.2.1.5 CONSONANTAL ENVIRONMENT

The consonantal environment that the devoicable vowel is found in can also have an effect on its application. Han (1962b) suggests that a preceding fricative has a greater devoicing influence than an affricate, which has a greater influence than a stop. According to Takeda & Kuwabara (1987), high vowels are more likely to devoice after fricatives and affricates than after plosives. Simada et al. (1991) found that some speakers tend to devoice vowels after plosives, while others tend to delete them. Consonantal effects observed in this study will be noted in following discussion.

Other factors affecting the application of HVD such as the number of devoicable high vowels in a row and focus accent will not be discussed in this work. However, see Kondo (1994, 1997) for extensive discussion of devoicing in consecutive devoicing environments; see also Vance (1987) Ch. 6 for a more complete overview of factors affecting devoicing.

The various formulations of HVD that have been proposed in the literature will now be reviewed.

2.2.2 HVD AS A FEATURE-CHANGING RULE

HVD has been traditionally characterized as a phonological (i.e. feature-changing) rule (e.g. McCawley 1968: 127) which could be formulated as follows.

$$V \begin{matrix} [+high] \end{matrix} \rightarrow [-voice] / \begin{matrix} C \\ [-voice] \end{matrix} \text{ ————— } \left\{ \begin{matrix} C \\ [-voice] \\ \# \end{matrix} \right\}$$

Figure 2.1 A traditional formulation of the rule of HVD.

That is to say, high vowels devoice between voiceless obstruents, or after a voiceless obstruent and before a pause.

Devoicing before a pause was included in this rule primarily to account for the devoicing of the sentence-final [u] of *desu* (the copula) and *-masu* (a formal verb ending) that was discussed above. This sentence-final devoicing will not be discussed in this thesis due to its almost universal application at all SRs and levels of formality, except, as mentioned before, when the speaker wishes to present a formal impression by consciously suppressing devoicing.²

It has long been recognized that HVD can occur at slower SRs and in more formal speech (e.g. Sakuma 1929; Bloch 1950), indicating that it is a phonological process: if a change in pronunciation occurs at a SR slow enough for there to have been sufficient time for any existing articulatory gestures to be realized,³ then the change must have happened before the articulatory instructions were issued—there must have been a phonological change. Examples occurring at the slowest SRs will be presented below that show exactly this situation.

Looking at the rule above, and as discussed in other work (e.g. Jun & Beckman 1993), the rule as formulated here predicts that only two types of vowels will be produced: 1) vowels specified for [+voice] and hence produced with full voicing; and 2) vowels specified for [-voice] and hence produced without any voicing (i.e.

²Reflective of the wide-spread application of sentence-final devoicing of [u] is the fact that 2 of the 3 Taiwanese students learning Japanese reported on in Varden & Sato (1996) devoiced all sentence-final [u] in that study; the remaining participant retained voicing on only a portion of them.

³The term ‘sufficient time’ refers to the minimal time required for the articulatory trajectories of a given articulation to be realized; see Löfqvist (1997 §4.3) for discussion.

whispered vowels). The rule as formulated above implies that no other aspect of the vowel production will change other than the normal adjustments in temporal duration due to speech rate (SR) or prosodic changes.

However, the characterization of devoiced vowels as simply lacking vocal cord vibration suggested by Figure 2.1 above is not accurate in at least two respects.

First, as noted in Jun & Beckman (1993), the rule as stated above does not account for the fact that the amount of voicing evidenced during vowel production is gradient. The traditional formulation of the rule predicts either fully voiced vowels or fully devoiced vowels. However, the voicing of a vowel does not begin at full intensity or end with a sudden decrease to zero; the intensity of the glottal vibrations increases over a period time, typically maintains some plateau, and then decreases over a period of time. This gradiency increases at faster SRs for high vowels between voiceless obstruents (Beckman 1994). Globally, the duration of voicing that can be found in multiple repetitions of a given vowel ranges from the full duration of the vowel to the slightest amount of voicing. As little as one cyclic glottal pulse at a frequency appropriate for the voicing of a speaker has been observed in the data in this study and elsewhere (Beckman & Shoji 1988; Kondo 1993). The range of voicing durations that can be found in the production of vowels at a given SR would seem to be too great to be due to adjustment of segment, mora, and word durations because of SR and stylistic variation (for discussion of temporal adjustments, see Kawasaki 1983; Port et al. 1987; Campbell 1992; Han 1994; Pirello et al. 1997; Farnetani 1997).

Second, the rule as stated above does not reflect the large-scale articulatory changes of the devoiced vowels routinely observable in public speech and found in this study. The devoiced vowels found in this study are not whispered vowels; their mode of production is quite distinct from vowels produced during normal whispering. Vowels produced during whispering exhibit no appreciable closure of the oral tract. The air rushing through the spread glottis provides the energy for the resonance of the oral tract that is associated with the production of the vowel (Catford 1977: 96; Ladefoged 1996: 107-108), with the clarity of the whispering often being strengthened apparently by the glottal tension utilized for 'breathy voice'. In contrast, the closure associated with the production of vowels undergoing

HVD is that of a fricative, a fricative closure with enough oral tract vowel configuration maintained to provide the perception of a vowel: frication caused by the air rushing out of this newly-formed closure provides the energy for the resonance of the oral cavity that results in formant structure formation.

In addition to the articulatory changes taking place in the oral cavity, Tsuchida (1994, 1997) notes that the glottal aperture of these vowels matches that of a fricative, not that of a vowel. There is a significantly larger glottal opening associated with both true fricatives and devoiced vowels attributed to the assignment of the feature [+spread glottis] to both; in the case of fricatives, underlyingly; in the case of devoiced vowels, by phonological rule. Indeed, it seems that Tsuchida (1997) comes quite close to calling the devoiced vowels seen in that study fricatives.

For the two reasons given above, the characterization of HVD as a spread of the feature [voice] is therefore not consistent with the observed data. The fricativization mentioned above will be discussed further in §2.3. An answer to the first of these criticisms, that a feature-changing rule cannot account for the observed gradiency of voicing, is found in Jun & Beckman (1993).

2.2.3 AN ALTERNATIVE ACCOUNT: GESTURAL OVERLAP

As noted in Jun & Beckman (1993), the Gestural Score framework proposed in Browman & Goldstein (1990, 1992) provides an explanation for the gradiency of voicing duration observed both locally in one vowel production and globally across repetitions. In the Gestural Score account, all vowels are phonologically specified as [+voice]; devoicing of vowels is due to the instructions for voicing being overlapped by the preceding or following instructions to produce a voiceless consonant. This overlapping of articulatory instructions is shown in Figure 2.2 below for the word *kiki* ‘crisis’. (The spreading symbols represent instructions to the glottis to inhibit or stop voicing, and the closing symbols represent instructions to the glottis to initiate voicing. They are not meant to represent actual vocal cord movement or degree of glottal spread or closure.)

contexts at the same SR have been observed. Examples of this range of voicing duration values will be provided from the current data set as well.

Imaizumi et al. (1995: 775~776), in their discussion of the gestural overlap account of devoicing presented in Jun & Beckman (1993), delimit three ways that overlapping of the vowel voicing gesture can be achieved: 1) the devoicing gestures of the surrounding voiceless obstruents can be temporally shifted toward the vowel due to an increase in SR⁴; 2) the surrounding devoicing gestures can be lengthened⁵; or 3) a combination of the two. The 3 strategies are represented graphically below (after Imaizumi et al. 1995: 775, Figure 4).

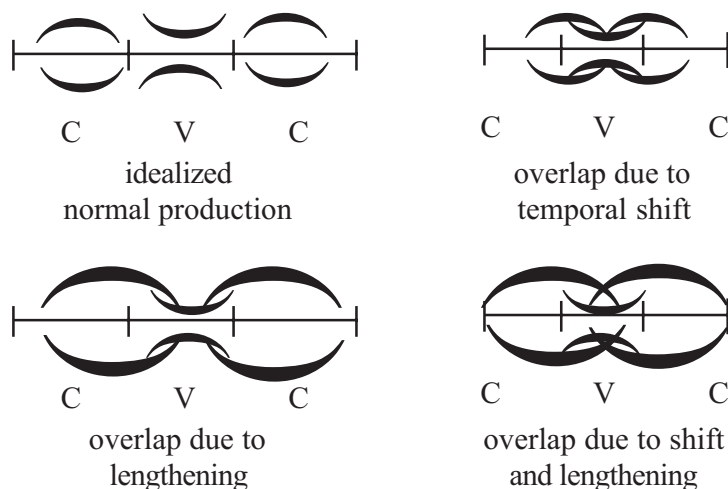


Figure 2.3 Three strategies for achieving overlap of vowel voicing gestures by surrounding devoicing gestures

While Imaizumi et al. (1995) does not provide evidence for which strategy was employed by the participants in their study, evidence will be presented below that both overlapping and strengthening was utilized by the participants in this study.

⁴Imaizumi et al. 1995 attributes the temporal shift of the devoicing gesture to an increase in SR. This implies that the devoicing gesture is more robust than the voicing gesture of the vowel since it is the voicing gesture that is overlapped and not vice versa. This point will be accepted here as well.

⁵Imaizumi et al. (1995) use the term ‘strengthened’, but the net effect of strengthening a devoicing gesture is an increase in duration of the gesture.

For much of the devoicing observed in this study a temporal shift in the location of the frication associated with the preceding obstruent can be seen; it appears that a temporal shift in the centering of the frication—and hence the glottal spreading responsible for the frication—was responsible for the devoicing of the vowel. Further, it will be seen that only the preceding obstruent’s glottal spreading gesture appeared to be shifted; the frication associated with the following obstruent remained fixed with respect to the succeeding vowel within token productions of similar duration. Therefore the devoicing of the vowel will be attributed to the glottal spread of the preceding voiceless obstruent, with the glottal spread associated with the following voiceless obstruent not being actively involved in the devoicing.

At least one participant in this study appears to have utilized the second strategy mentioned above; in her productions spectrographic evidence points to a lengthening of the glottal spreading gesture. Indeed, in some productions by some participants the frication associated with the token-initial obstruent is seen to continue through to the 2nd mora, with no clear medial stop closure or release being made. These cases will be discussed further below, with representative productions being given.

However, characterizing devoicing as a phonetic overlap of glottal spreading gestures also fails to account for the most basic of observations regarding HVD: vowels often devoice even when the SR must be slow enough to allow full achievement of articulatory targets.

The fundamental problem with extending the gestural overlap account of devoicing depicted in Figure 2.2 above to all instances of HVD is that it predicts that high vowels will devoice only when the SR is fast enough for the devoicing gestures to completely overlap the voicing gestures. As noted in Tsuchida (1997), this characterization of HVD is not tenable. HVD occurs even at slow SRs where there must be more than sufficient time for the voicing gesture to be realized. Devoicing at slow SRs has been noted in the literature (Kondo 1993, 1994, 1997; Varden & Sato 1996) and can be seen in the current data set as well. The fact that a vowel devoices even at slow SRs indicates that the voicing gesture must not have been present at the phonetic level—that the instructions to voice the vowel were changed before the vowel was pronounced.

There would seem to be one possible analysis involving gestural overlap that would allow devoicing to occur at even slow SRs. This would be an analysis where the glottal spreading gestures of the surrounding voiceless obstruents could be freely realigned within the temporal domain of the segments production. If the glottal spread associated with a voiceless obstruent were allowed to be centered at any point of time in the production of the obstruent, it could also be centered toward the end of a obstruent's production. This would result in the glottal spread continuing into the following vowel. An overlap of the vowel's voicing gesture could occur, and the vowel could be (at least partially) devoiced. In another repetition of the obstruent and vowel, the obstruent's glottal spreading gesture could be centered more toward the beginning of the obstruent's production. This could result in the glottal spreading gesture being completed when the production of the following vowel began. No overlap of the vowel's voicing gesture would occur; the vowel would not be devoiced.

However, based on other studies (Kim 1970; Kingston 1990; Kingston & Diehl 1994) and the data that will be presented below, attributing devoicing to the free alignment of the glottal spreading gesture does not seem tenable. The data in this study shows that the alignment of the glottal spreading gesture associated with a voiceless fricative or affricate, as indicated by the location of the frication associated with that obstruent, is not found in a range of temporal locations as would be expected if alignment of the glottal spreading gesture were free. The location of the frication associated with a consonant in the current data set is fixed. In addition, no reports of free alignment of the frication associated with a segment's glottal spreading gesture appear to have been made in any of the literature involving glottal spreading. When a voiceless fricative or affricate precedes a voiced vowel, the frication is centered on the fricative component of the consonant. When a voiceless fricative or affricate precedes a voiceless vowel, the frication is centered midway between the consonant and the vowel site. This is consistent with both the data and analysis of Kingston (1990), Iverson & Salmons (1995), and Tsuchida (1997). The temporal shift of the glottal spreading gesture is therefore thought to be due to a phonological process that will be discussed below, not free alignment of the spreading gesture.

There is one other point which argues against free temporal alignment of the glottal spreading gestures. If there were great enough freedom in aligning the spreading gesture with the rest of the segmental articulation so that shifting the spreading gesture's point of alignment caused the devoicing at slower SRs, the strong prediction is made that all vowels—high, mid and low—would be subject to this type of devoicing at slower SRs. There seems to be no phonetic reason to allow the temporal alignment of the glottal spreading gesture to be freer for high vowels than for low vowels.⁶ The fact that non-high vowels do not devoice at slower SRs as high vowels do argues against free temporal realignment of the devoicing gesture.

Analyzing devoicing as a phonological process does provide a ready answer as to why generally only high vowels devoice, however—historical reassessment of a phonetic phenomenon as a phonological one. The close oral closure accompanying high vowels makes it much more likely for air flow to drop below the minimum requirement for voicing to be sustained (Locke 1979); in Japanese and several other languages, this apparently has led to the grammaticalization of fricativization/devoicing process for high vowels.

Disallowing a phonetic account of HVD at slow SRs leaves only a phonological account; namely, that some sort of feature adjustment is responsible for the devoicing observed at slow SRs. However, the fact that non-high vowels devoice more at faster SRs (Maekawa 1990) and that gradiency of voicing is seen both within individual productions and across many productions indicates that a phonetic overlap of the voicing gesture is also present.

The apparent discrepancy in the data is reconciled in Tsuchida (1997), who proposed that both phonetic and phonological devoicing are present in Japanese.

⁶Kate Davis (p.c.) has pointed out to me that the inherent lower pitch of mid and low vowels is manifest by varying degrees of vocal fold tension, and this varying tension may allow greater freedom in alignment. I am unaware of any studies indicating this has been observed or argued against.

2.2.4 HVD AS BOTH PHONETIC AND PHONOLOGICAL

Tsuchida (1997) provides particularly strong experimental evidence for the phonological devoicing of high vowels between plosives. In that study, electromyographical (EMG) measurements of muscle activity and fiberoptic observation of the accompanying glottal spreading were made during the production of both voiced and devoiced vowels. Tsuchida found that when high vowels devoiced between plosives, there was only one associated large spike of activity of the laryngeal muscle involved in glottal spread, the posterior cricoarytenoid (PCA) muscle, and only one large accompanying glottal spreading gesture. These findings supported Sawashima (1971), which also found one PCA activation accompanied by one large glottal spread when high vowels devoiced between plosives.

In contrast, the participant in the Tsuchida (1997) study generally did not devoice high vowels between fricatives. When high vowels were preceded and followed by voiceless fricatives, two clear activations of the PCA and two clear glottal spreading gestures were observed, one for each consonant preceding and following the vowel. (The same two activations of the PCA and two accompanying glottal spreading gestures were also found for non-high vowels in all devoicing environments.)

Tsuchida (1997) therefore argued for both phonological and phonetic devoicing in Japanese: in the case of high vowels between voiceless plosives, a phonological spread of the feature [+spread glottis] from the preceding voiceless obstruent; in the case of high vowels between voiceless fricatives and non-high vowels in all devoicing environments, a phonetic overlap that causes loss of voicing and devoicing at faster SRs. Indeed, this proposal was anticipated as far back as Bloch (1950: 136), who stated that while “[voiceless [i] and [u] are] paralleled, especially in slow or careful speech, by an otherwise identical synonymous phrase containing [voiced [i] or [u]] instead”, voiceless [ɑ] and [o] occur “in rapid speech only”. And as discussed in Kondo (1997), the results of Maekawa’s (1990) studies into devoicing of /u/ vs. devoicing of /a/ and intervocalic voicing of /h/ also implied two processes at work: one influenced by SR and the other independent of it.

As for the mechanism responsible for devoicing, Beckman (1994: 58-59) states that devoicing can be attributed to closeness of the oral constriction. When there is a close

oral constriction maintained at the release of the preceding consonant, as there is for fricatives and affricates, the air pressure that has built up in the oral tract due to the closure cannot be vented as quickly when the segment is released as if the oral tract were relatively open. This maintenance of the air pressure in the oral tract above the larynx makes it more difficult for voicing to be initiated, since there must be a minimal pressure difference across the larynx for voicing to be able to take place (Lieberman & Blumstein 1988: 101-103). When the vowel following a fricative or affricate is a high vowel, itself made with greater closure than mid or low vowels, the effect of closure blocking the venting of air pressure will compound, and the likelihood that the realization of the voicing gesture will be completely blocked increases. This leads to the prediction that high vowels will devoice more readily after fricatives and affricates, a prediction born out by Takeda & Kuwabara (1987) and Kondo (1993) (as cited by Beckman 1994). Indeed, Kondo (1994, 1997) has found good evidence of vowels in environments that are adjacent to other vowel devoicing environments (i.e. consecutive devoicing environments) undergoing reduction in intensity even when they are not devoiced. This suggests that devoicing or deletion of vowels in these consecutive devoicing contexts can be seen as an end result of an overall vowel reduction process (see Kondo 1997 Ch. 8 for discussion).

Taken together with the glottal spreading overlap analysis of Jun & Beckman (1993), the prediction is that adjustment of two parameters can be responsible for devoicing of vowels: 1) adjustments to the glottal spreading associated with an adjacent voiceless obstruent; and 2) adjustments to the oral closure of the vowel due to an adjacent fricative or affricate, whether they are phonemically or allophonically derived. Further work is needed to completely delimit the two processes.

Another analyses involving both phonological and phonetic derivation of the same acoustic end in a given language is the proposal made for English by Hayes (1992) in a comment on Nolan (1992). The same argument is made for English /s/ to /š/ assimilation in Holst & Nolan (1995).

Nolan (1992), using palatolographic measurement of productions of the English string *late calls*, noted that many speakers produce a complex coronal/velar

articulation for the medial [t k] cluster. However, the coronal articulation of the complex segment is weakened to varying degree, while the velar articulation is not; the velar component is ‘robust’. Hayes (1992) accounted for this by positing both a binary phonological process and a gradient phonetic one: the phonological process spreads the velar place of articulation forward to the coronal to produce the complex segment, while a phonetic process weakens the coronal portion of the complex segment to varying degree—the gradiency of the weakening is the result of the process being due to overlap of gestures.

Holst & Nolan (1995) provided further evidence for both processes in the production of /s/ and /š / clusters both within and across a syntactic boundary. Their data indicated that within syntactic boundaries, /s/ fully assimilated to a following /š /, indicating phonological assimilation due to feature spreading, while across syntactic boundaries the assimilation was gradient, indicating phonetic overlap of gestures.⁷

In regard to Japanese, Jun & Beckman (1993) and Varden & Sato (1996) have reported gradient amounts of voicing associated with high vowels that are not between voiceless fricatives. This gradiency of voicing durations at a given SR is observed in the current data set as well.

Returning now to the phonological devoicing posited by Tsuchida (1997), recall that the temporal shift of the glottal spreading gesture of a consonant preceding a devoiced vowel is due not to overlap or free alignment of glottal spreading gestures, but instead is due to a phonological spread of [+spread glottis] from the obstruent to the vowel. This double-linking of the feature then leads to its temporal realignment to the midpoint between the obstruent and the vowel. This realignment of the frication associated with fricatives and affricates can be seen in the current data set. A representative pair of productions is given in the figure below.

⁷ But see Browman (1995) for reinterpretation of their data as only phonetic overlap.

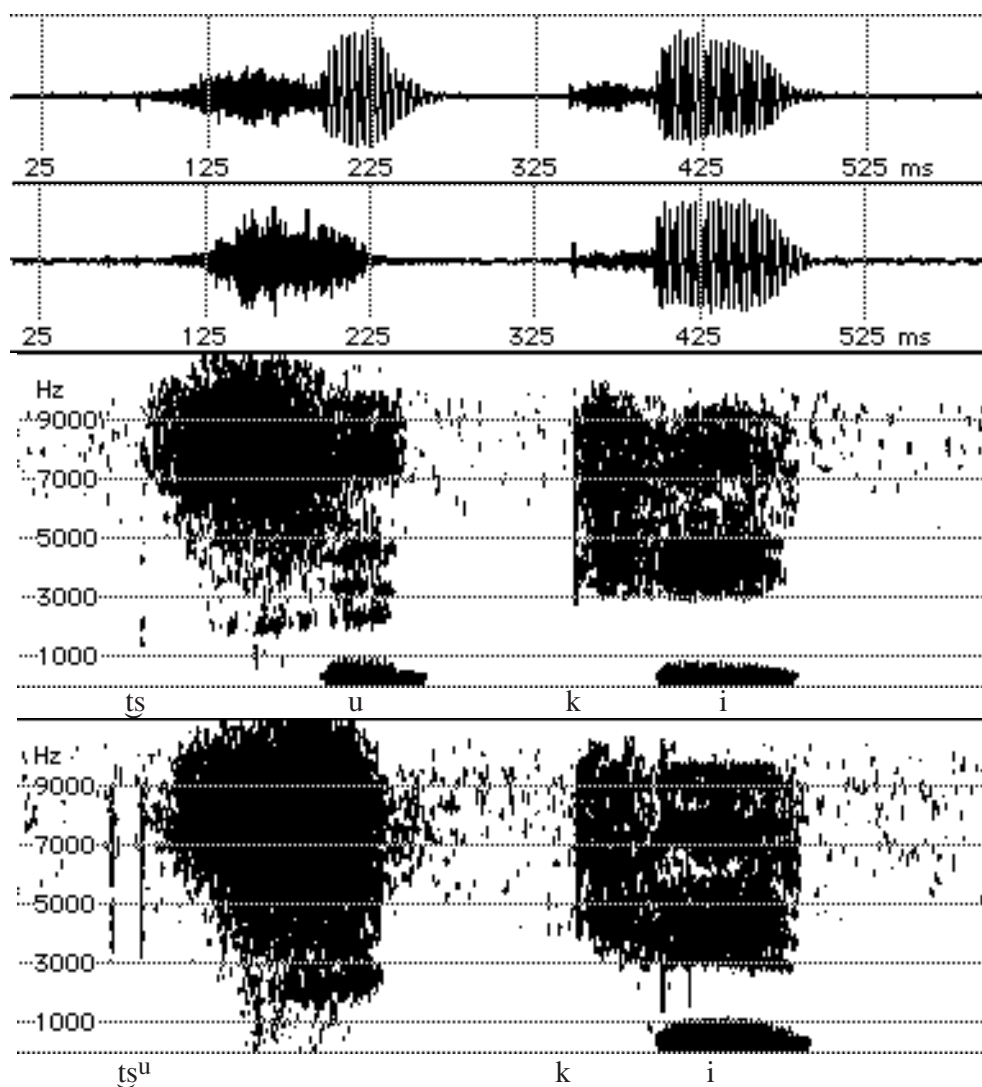


Figure 2.4 Waveforms and spectrograms for voiced and devoiced 1st vowel of *tsuki* 'moon', participant HK.

The 1st vowel of the second production is devoiced, even though the 556 ms token duration is toward the long end of all productions and indicates a fairly slow SR (this participant's productions averaged 554 ms for the slow repetitions, 383 ms for the normal repetitions, and 309 ms for the fast). In contrast, the first production of the same token with a token duration of 558 ms shows 78 ms of voicing (as measured from the filtered waveform) associated with the 1st vowel.

Notice that the various components of the two signals are temporally fairly closely aligned: the weak release of the initial affricate [ts] as evidenced by the vertical striation to the left of the spectrograms⁸, the onset of the closure of the medial stop [k], its release, the amount of aspiration of the second stop (i.e. the VOT), the onset of the second vowel's voicing, and the onset of the closure of the following clitic's initial stop [t] are all fairly well temporally aligned in the two productions.

The developed formant structure on the frication of the first vowel in the first production and the weak formant structure on the first vowel in the second production shows that there was a tongue-shaping gesture at the vowel site of both productions. This indicates that the tongue-shaping gesture is independent of voicing of the vowel. The equivalent duration of the 1st mora in both productions also speaks to the retention of the vowel's timing slot, eliminating an analysis where the vowel is deleted after a coloring of the frication of the obstruent for this and similar productions.

The VOT associated with the 2nd mora vowel is approximately the same in both productions, indicating that the glottal spreading gesture associated with the release of the medial stop [k] is of the same duration and has the same temporal alignment in both. As noted earlier, this is taken to indicate that the preceding consonant (i.e. the 1st mora consonant) is responsible for devoicing, and that the glottal spreading gesture associated with a following consonant is not involved in devoicing of a vowel that precedes it.

Turning now to the frication associated with the 1st mora fricative, it can be seen that the duration of the frication is approximately the same in both productions—approximately 120 ms and 110 ms, respectively. This indicates that the glottal spreading gestures were also of about the same duration, arguing against an analysis involving a lengthening of the glottal spreading gesture in this case, whether or not it is strengthened.

⁸For the 2nd repetition, the 2nd of the token-initial closure releases was used in taking the token duration measurement.

Instead, note that the peak of the frication of the affricate in the devoiced vowel production is temporally shifted when compared to the location of the peak frication in the voiced vowel production. This is consistent with what Kingston (1990: 427) noted in discussion of s-voiceless stop clusters in languages like English (e.g. the initial cluster in stop). It was observed that glottal spreading in these clusters reached its maximal width at a point that "...is close to the boundary between the two oral articulations, a temporal compromise between the early peak [glottal spread] of the fricative and the late peak [glottal spread] of the stop." This is in contrast with the location of the maximal spread for voiceless fricatives and plosives when they are not found in clusters—in the case of voiceless fricatives, the maximal spread is usually aligned with the middle of the segment; for voiceless aspirate stops, it is aligned with the release of the closure (Browman & Goldstein 1986; Kingston 1990, Goldstein 1990; Tsuchida 1997).

This midpoint location of the shared spreading gesture also led Iverson & Salmons (1995) to posit the sharing of the feature [+spread glottis] in voiceless fricative/stop clusters in English, similar to the spread of the same feature posited in Tsuchida (1997). Iverson & Salmons (1995) posit an underlyingly marked specification of [+spread glottis] for voiceless stops in English, rather than the redundant specification of [-voice] that is usually assumed for this class of segments.⁹ Since both voiceless fricatives and voiceless stops are then specified for this feature, the OCP (McCarthy 1987) forces the dual specification of this feature to coalesce in this cluster-initial environment. The temporal alignment of the shared feature is thought to be triggered by the feature sharing and determined by the two segments' midpoint.

As mentioned in §1.4.3, in this analysis the degree of observed aspiration in English is attributed to the independent prosodic strengthening of segments caused by English stress assignment (Nespor & Vogel 1986; Halle & Vergnaud 1987; Kingston & Diehl 1994). Japanese aspiration generally patterns that of English (Homma 1980;

⁹Aspirate voiceless stops are then the default case in Iverson's proposal, with aspiration being enhanced according to prosodic position. This is quite reasonable since it is the VOT, largely concomitant with the strength of aspiration, that is used to distinguish voiced and unvoiced stops in English (Lisker & Abramson 1964).

1981) and can be plainly seen in the current data set; the differences between aspiration in English and Japanese can be attributed to different prosodic strengthening strategies in the two languages.

Independently using observation of glottal spreading and activation of the muscles causing it, Tsuchida (1997) posited devoicing of high vowels between voiceless fricatives as a spread of the feature [+spread glottis] from the voiceless fricative to the vowel. The sharing of the feature would then trigger a realignment of the spreading gesture to midway between the two segments. This realignment was evidenced in that study in both the plots of muscular activity and in the plots of the resulting glottal spread, and was held to be responsible for the devoicing of high vowels in this environment.

In light of the discussion of Iverson & Salmons (1995) above, a slight revision of the laryngeal system given in Tsuchida (1997: 53, Table 2.7) is required; here voiceless stops have been merged with voiceless fricatives and affricates, and /h/.

Table 2.1 Modified laryngeal classification system for Japanese.

	voiceless obstruents, [h]	voiced obstruents	voiced vowels	devoiced vowels	[ʔ]
[voice]		√	√		
[s.g.]	√			√	
[c.g.]					√

Let it be noted that an attempt to account for the observed glottal spreading in many studies (e.g. Tsuchida 1997) has not yet been attempted.

Returning to the spectrograms in Figure 2.4 above, it can also be seen that the frication associated with the mora containing the voiced vowel is centered between the release of the consonant and the onset of the vowel; i.e. the frication is centered on the latter half of the initial affricate, the [s] portion of the affricate. This same alignment is seen in all [tsu] morae containing a voiced vowel in this study. In contrast, the peak of the frication associated with the devoiced vowels can be seen to be aligned within the vowel site, consistent with the phonological coalescence of the

initial consonant's and vowel's specification for [+spread glottis] posited by both Tsuchida (1997) and Iverson & Salmons (1995).

In addition to the devoicing at slower SRs, alternate voicing and devoicing of vowels at the same SRs can be seen in the following two repetitions of *chiki* 'friend(s)' by participant ANa. In the first production the first vowel has been devoiced, whereas in the second, the second vowel has been devoiced. (The 1st panel gives the waveform; the 2nd, the pitch trace; and the 3rd, the spectrogram.)

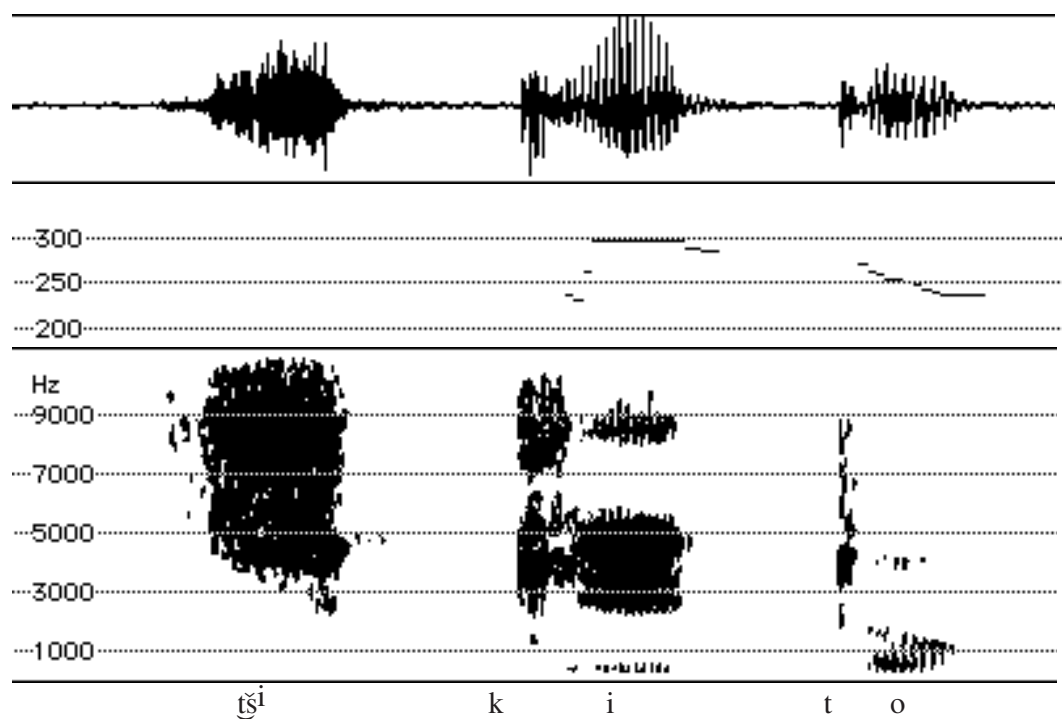


Figure 2.5 Alternate devoicing of vowels at same SR, 1st vowel devoiced (participant ANa).

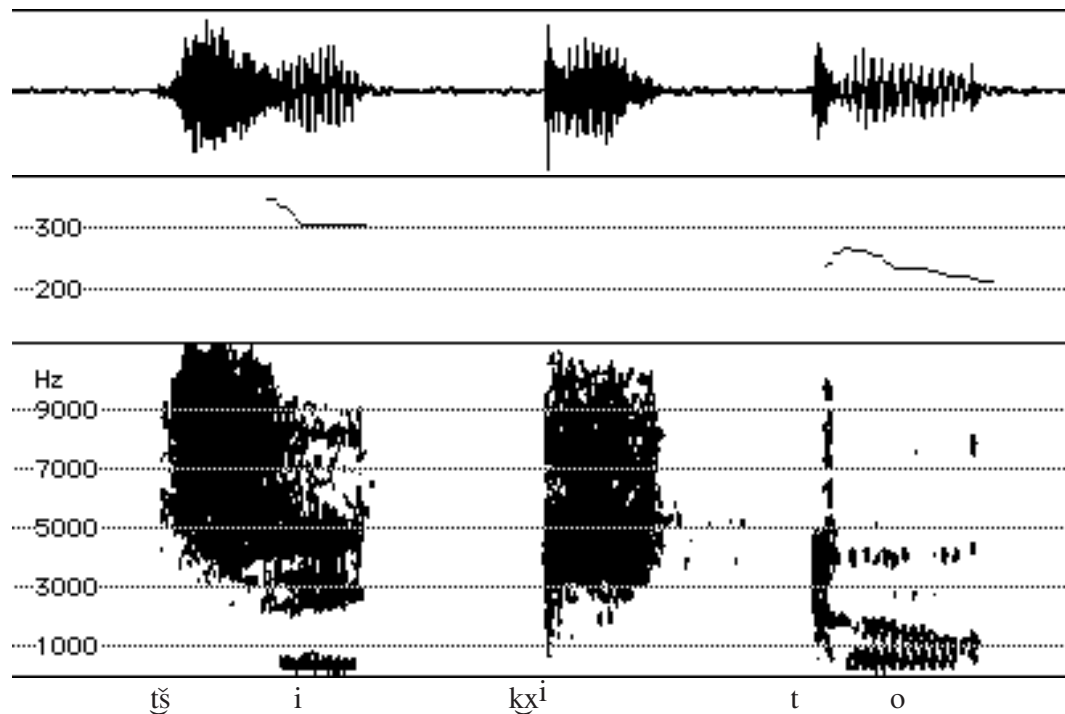


Figure 2.6 Alternate devoicing of vowels at same SR, 2nd vowel devoiced (participant ANa).

Note again the close alignment of the formant structure of the first vowel and the stop closure and releases in both productions. Again, the large difference in the two productions, aside from f_0 activity, is in the duration and centering of the frication on the first vowel, and the duration of frication on the second vowel. The approximately equal durations of both morae argue against an overlapping gesture analysis based on SR; there seems to be no plausible explanation why voicing of the vowel in each mora would have time to be initiated and sustained in one production but not the other. A phonological analysis again provides a ready answer, however—in the first production, the participant simply chose to accent the second vowel and devoice the first; in the second production, she chose to accent the first vowel and devoice the second.¹⁰

¹⁰It appears that, for this participant, pitch accent placement for the token *chiki* is not completely fixed, possibly due to areal influences from adjacent dialects and/or the infrequent use of this word—the commonly-used word for ‘friend(s)’ is *tomodachi*. In 2 of the 18 repetitions she accented the 1st mora; the 1st mora

The characteristics of the frication in both devoiced vowels are also quite consistent with a devoicing analysis utilizing the spread of the feature [+spread glottis]. In the first production, again the duration and temporal location of the frication are consistent with the sharing of the feature spread from the voiceless fricative. The VOT observed in the second production, as with the *tsuki* example above, provides the same analysis for the devoicing of the second mora as well. The longer frication observed is entirely consistent with the spread of the feature [+spread glottis] from the aspirate stop to the vowel, again resulting in a shared specification for the feature that is temporally aligned midway between the two segments; i.e. substantially after the release of the stop.

However, it must be noted that temporal shift of the glottal spreading gesture is not observed for all productions for all participants. In particular, participant TO showed much less frication associated with her productions, and it appears that the period of frication lengthened rather than was shifted.

The following figures contrast two productions of the token *tsuki* ‘moon’ by participant TO.

vowel was voiced and the 2nd mora vowel was devoiced. In 15 of the 18 repetitions she accented the 2nd mora; this vowel was voiced, and the 1st vowel was devoiced. In 1 repetition both vowel were voiced. There is no difference in the meaning of the productions due to different placement of the accent other than the participant did not devoice the accented vowels.

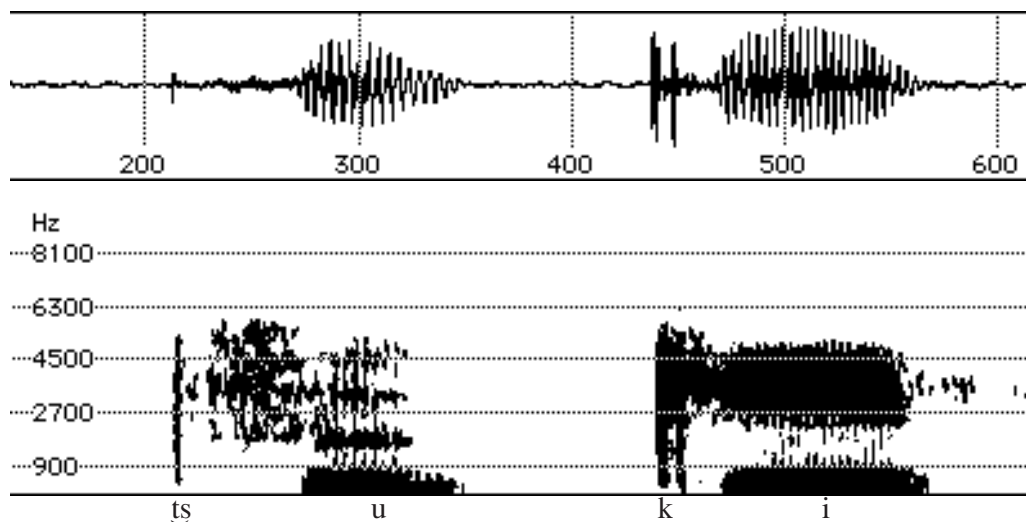


Figure 2.7 Waveform and spectrogram for voiced 1st vowel of *tsuki* ‘moon’, participant TO.

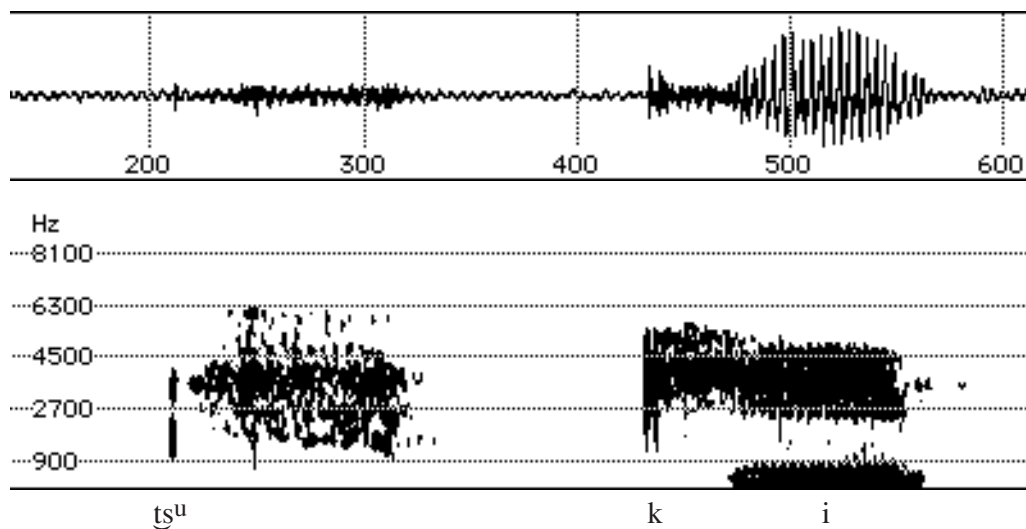


Figure 2.8 Waveform and spectrogram for devoiced 1st vowel of *tsuki* ‘moon’, participant TO. Again, the first vowel of the second production is devoiced, even though the 499 ms token duration is toward the long end of all productions and indicates a fairly low SR (this participant’s productions averaged 485 ms for the slow repetitions, 434 ms for the normal, and 319 ms for the fast). In contrast, the first production of the same token with a token duration of 501 ms shows 77 ms of voicing (as measured from the filtered waveform) associated with the first vowel.

However, notice that the frication associated with the initial obstruent of the 1st production is not shifted in the 2nd production. Instead, it is lengthened so that it continues from shortly after the release of the affricate on through the vowel site. This indicates the glottal spreading gesture associated with the obstruent is also lengthened in duration. It is thought that the cause of this lengthening is phonological in nature; this will be pursued in the next section.

It would appear from this and other similar productions from this participant that the individual strategies used to align the glottal spreading gesture show variation both across speakers and, as will be seen below in the next section, for different obstruents. For some speakers, the glottal spreading gesture associated with the preceding obstruent is temporally shifted, realigning at the midpoint between the obstruent and the vowel. For others, the glottal spreading gesture of associated with the preceding obstruent is lengthened, lasting the duration of the obstruent's and the vowel's production. And as noted above in the discussion of Imaizumi et al. (1995) in §2.2.3 of this work, some productions by some participants displayed frication lasting throughout the entire token, with no clear medial stop being produced. Representative productions of this will also be given in the next section.

This chapter will therefore adopt Tsuchida (1997), that vowel devoicing is caused by the spreading of the feature [glottal spread] from a preceding obstruent to a following vowel and subsequent realignment of the spreading gesture to the midpoint between the obstruent and the vowel. However, it has been noted that realignment of the gesture does not always occur; some speakers evidently utilize a lengthening strategy that achieves the same end.

Further, it was noted that phonetic overlap is active during the production of high vowels in all environments. While the overlap can not be responsible for devoicing at slower SRs, it is responsible for the gradient amounts of voicing durations observed, and may be responsible for devoicing at faster SRs. This would be consistent with the increased number of devoiced vowels seen at the higher SRs, as noted in the previous chapter.

In addition to these production arguments, statistical evidence from the current data set will be presented in Ch. 6 that shows quite clearly the data being separated into two distinct groups. The immediate conclusion is that the two groups of data were generated by two separate processes; i.e. phonetic overlap of voicing instructions and a phonological deletion of the voicing instructions.

2.2.5 SUMMARY OF ARGUMENTS FOR BOTH PROCESSES

The following four considerations were discussed above:

- 1) devoicing of high vowels between plosives occurs even at slow SRs where there is ample time to initiate any existing articulatory instructions;
- 2) the frication accompanying devoiced vowels is either temporally shifted to the mid-point of the obstruent and vowel site, or lengthened so that it continues through the vowel site;
- 3) there is only one glottal muscle activation/spreading gesture when high vowels between plosives are devoiced, but two glottal muscle activations/spreadings when non-high vowels and high vowels between fricatives are produced (Tsuchida 1997); and
- 4) observed voicing values are gradient for any given SR (i.e. token duration) for both high vowels (current data) and non-high vowels (Maekawa 1990).

Based on these considerations, it is concluded that both processes affecting voicing are required to account for the data observed in this and other studies. In summary, this work follows Tsuchida (1997) in claiming two loss of voicing processes at work in Japanese:

- 1) a phonological process is responsible for the devoicing of high vowels between voiceless plosives (i.e. a spread of [+spread glottis] from a preceding stop or fricative) which may occur at all SRs; and
- 2) a phonetic variation of the overlapping of voicing instructions by the glottal spread associated with the preceding consonant which increases with SR is responsible for a loss of voicing duration in all vowels; it is this phonetic mechanism that causes devoicing of non-high vowels and high vowels between fricatives at fast enough SRs.

Including a phonetic reduction of voicing durations in the analysis of the data for high vowels as well as non-high vowels leads to the prediction that there will be high vowels that will lose all their voicing due to overlapping of glottal gestures at high SRs. Indeed, this may account for the statistically significant increase in the number of devoiced vowels seen in the previous chapter in the discussion of the effect of SR on whether or not a vowel is devoiced.

2.3 HVD AND FRICATIVIZATION

One characteristic of Japanese high vowels that to the best of my knowledge has not yet received proper attention in the literature was mentioned above in the context of realignment of the glottal spreading gesture: the fricativization of vowels that usually accompanies devoicing, and is widely observable in the media and public.

Indeed, fricativization of vowels is not limited to Japanese.

2.3.1 OTHER ATTESTATIONS OF FRICATIVIZED VOWELS

While rare in the world's languages, fricativized vowels have been reported in the literature. Ladefoged and Maddieson (1990, 1996) briefly discuss the fricativized allophones of /i/ that occur after the apical post-alveolar fricatives and palatal affricates in Standard Chinese. Perhaps the best phonetic description of fricativized vowels comes from Hahn (1991), a grammar of the Turkic language of Uyghur.

In the Northwestern dialects of Uyghur, high vowels are devoiced and then fricativized in initial [ʔVC] syllables. In the Southwestern dialects, the environment

has broadened to apply between any voiceless consonants. This can be seen in the following data from Hahn (1991) (after Kaisse 1992).

Northern dialects: between ? and C

?it → ?iʃt → ?ʃt ‘dog’
 ?uka → ?uø^wka → ?ø^wka ‘younger brother’
 ?ütläš → ?üø^wtläš → ?ø^wtläš ‘Roast!’

Southwestern dialects: between C and C

pit → piʃt → pʃt ‘louse’
 kitap → kiʃtap → kʃtap ‘book’

Figure 2.9 High Vowel Devoicing and Obstruentization in Uyghur.

Notice that after accompanying the vowel devoicing, there is what appears to be from the close transcription an epenthetic fricative following the vowel. The devoiced vowel can then be completely reduced, leaving only the syllabic fricative that in the case of reduced round vowels retains the vowel rounding. In the words of Hahn (1991: 34):

“Unlike in Japanese (e.g., *shitsurei* ‘impoliteness’), vowel devoicing in Uyghur is a matter of simultaneous progressive and regressive assimilation. To the untrained ear, the resulting “whispered” vowels seem either very short and indistinct or altogether absent.”

The simultaneous progressive and regressive nature of the assimilation can be seen in the derivations above, where the derivation of the fricative combines the continuancy and place of articulation (coronal for /i/; labial for /u/) of the preceding vowel and the consonantality of the following consonant.

In order to argue against Kaisse’s (1992) analysis of this fricativization as a spread of the feature [+consonantal], Hume & Odden (1996) suggest that the transcriptions given by Hahn (1991) are due to a ‘perceptual phonologization’ of the vowel-colored frication accompanying the devoiced vowel.¹¹ Laboratory analysis of this language is

¹¹It will be noted that Hume & Odden appear to have misinterpreted Hahn (1991) in at least one respect. They assert that:

needed to determine whether or not the presence of the fricatives remains at slower SRs; if they are present at slower SRs, as the fricatives in Japanese are, they must also be phonetically or phonologically derived. It is suspected that they are, since Hahn does not mention a SR connection in discussion of voiceless vowels.

Despite the differences in the devoiced vowels of Japanese and Uyghur noted by Hahn, there are also some striking similarities. Lacking knowledge of previous discussion in the literature, some time will be spent on the mode of production of Japanese devoiced vowels.

2.3.2 MODE OF PRODUCTION OF JAPANESE FRICATIVIZED VOWELS

Fricativized vowels are quite readily heard on any newscast or in any public setting. They can be easily produced by simply starting with an exaggerated articulation of [i] and slowly releasing the affricate [t͡ɕ], or starting with an exaggerated articulation of [u] and slowly releasing the affricate [t͡ɕ]. In most cases even when Japanese high vowels completely devoice there is enough resonance associated with the spirantized release of the vowel to distinguish clearly between an [u] and an [i] (cf. the release of the affricates [t͡ɕ] and [t͡ɕ] when you do not begin with an exaggerated vowel articulation; see also discussion of the preservation of vowel information after vowel deletion via allophonic variation of the preceding consonant in Modern Greek in Kaisse 1985 and 1990: 141).

Devoiced productions are then often fricative closures with accompanying oral cavity configurations that allows the formation of formant frequencies. The fricative closure produces high frequency turbulent noise, as any fricative does; this noise can be seen in both wide-band spectrograms and in zero crossing traces¹² of the vowel

“However, Hahn, also lists the noun /isim/ ‘the name’ as an example of devoicing without a fricative, despite the fact that this word is of the same phonetic class as words such as /it/ ‘dog’ which illustrate the fricative.” (p. 368)

However, what is not noted is that the example /isim/ occurs together with the very short (3 lines of text), unrelated §4.3.2 Devoicing. There is no reference to vowel fricativization one way or the other in this section.

¹²Zero crossing traces show the number of times that the sound wave cross the $y = 0$ axis. For low-frequency, highly periodic sounds such as vowels, the number of

reduction sites. The following figure shows a production of *kiki* ‘anxiety’ where the first vowel is devoiced and the second vowel is voiced (top panel: waveform; middle panel: wide-band spectrogram; bottom panel: zero crossings trace).

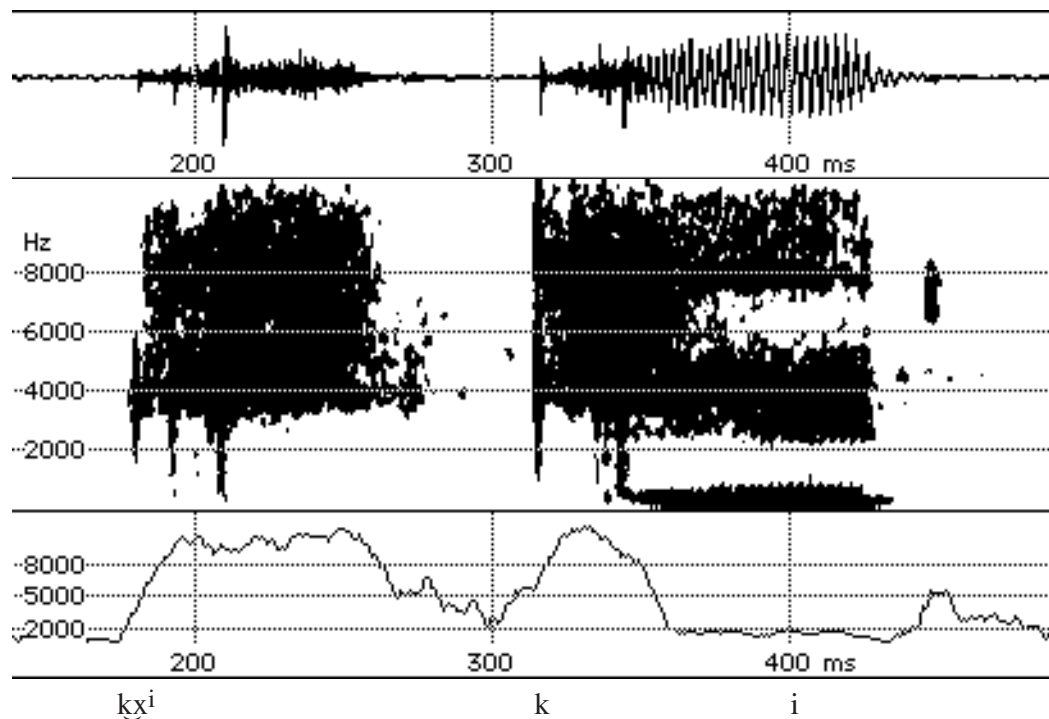


Figure 2.10 Waveform, zero crossings trace, and spectrogram for production of *kiki* ‘anxiety’ by participant ANa.

In the spectrogram (bottom panel), the release of the [k] at the beginning of the 2nd mora shows a fairly noisy velar release, with the frication accompanying the release resulting in approximately 35 ms of VOT. This is a typical velar plosive production before a voiced vowel for this participant. In contrast, the frication associated with the release of the 1st mora’s initial plosive continues throughout the vowel site. This indicates the production of an affricate closure—compare the production of the affricate [tʃ] in the first mora of the example in Figure 2.7. Although not particularly

times the wave crosses the $y = 0$ axis will be relatively low (e.g. a pure A above middle C will cross the axis 440 times per second). In contrast, the high-frequency turbulence of fricatives will cross the $x = 0$ axis many thousands of times per second.

clear in the spectrogram, the presence of the formant structure on the frication is quite audible; its supposition results in the vowel-colored velar affricate [kx¹].

The difference between the two velar releases can also be seen in the zero crossings traces (middle panel) associated with the two morae in the production above. Notice how the number of zero crossings remains quite high throughout the first mora, while in the second the number of zero crossings drops to a substantially lower level (in this production, approximately 100 crossings a second as calculated with a 10 ms analysis window) as soon as the vowel voicing is initiated. This is typical of productions involving both devoiced vowels and voiced vowels; that is, the presence of frication usually accompanies devoicing (i.e. frication increases as vowel voicing durations decrease).

This corelationship can be seen by comparing the second mora of the following productions, where the frication accompanying the release of the 2nd-mora plosive decreases much more quickly as the vowel voicing increases. In Figures 2.11 to 2.13, the top panel is the waveform for the token; the 2nd panel shows voicing and lower harmonic activity via a band-pass filtered (100 Hz and 1000 Hz cutoffs) token waveform; the 3rd panel gives the wide-band spectrogram; and the final panel gives the zero crossings trace.

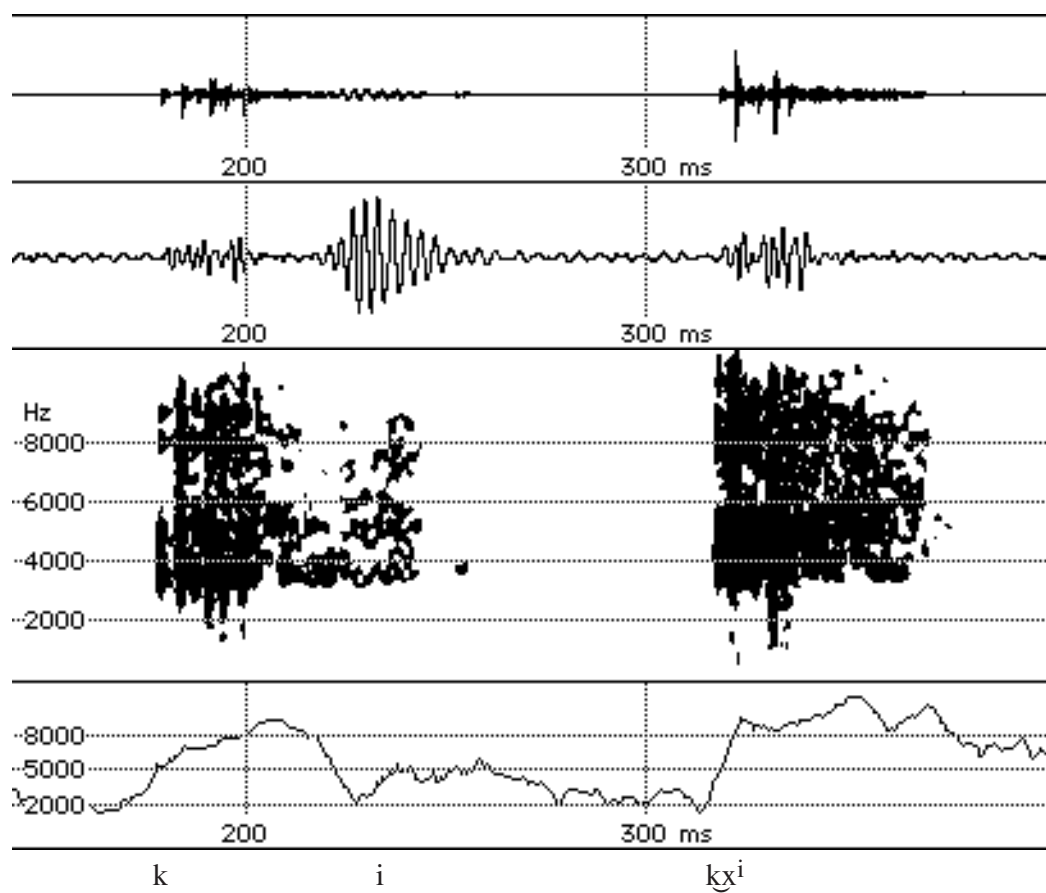


Figure 2.11 Production of *kiki* ‘anxiety’ by participant ANa showing heavy frication accompanying devoiced 2nd vowel.

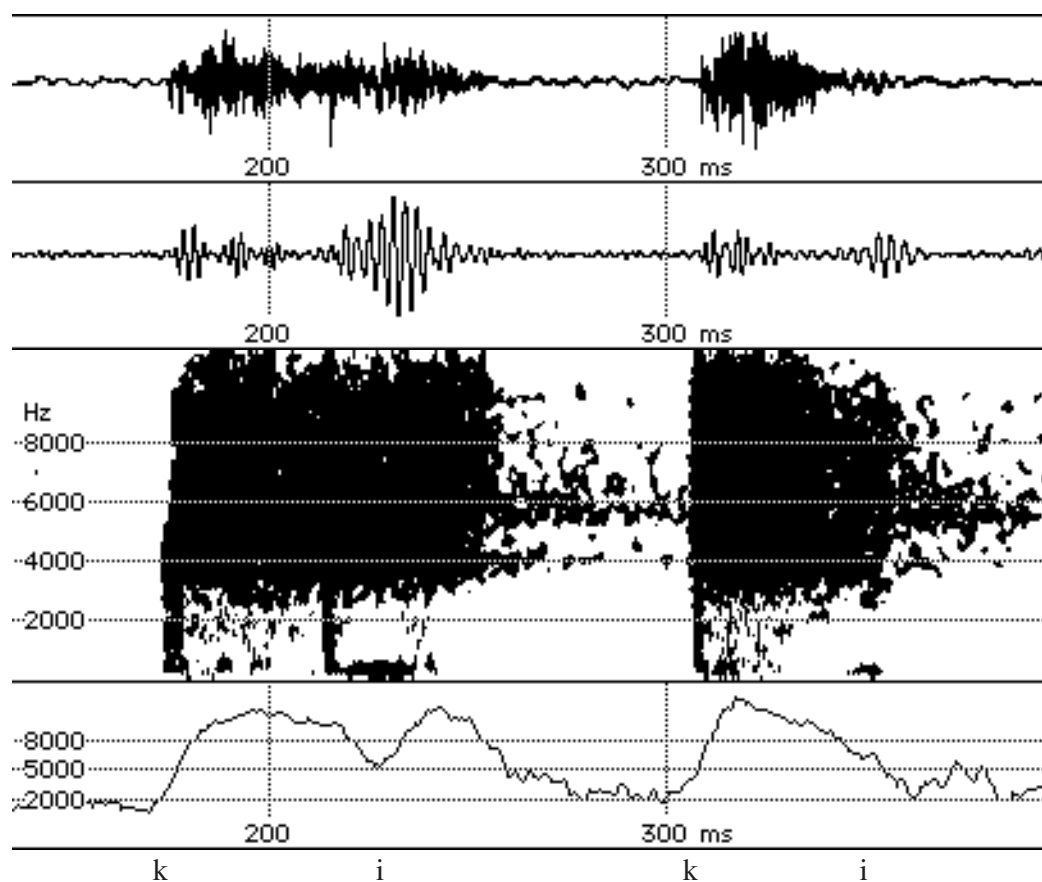


Figure 2.12 Production of *kiki* ‘anxiety’ by participant ANa showing decreasing frication accompanying lightly voiced 2nd vowel.

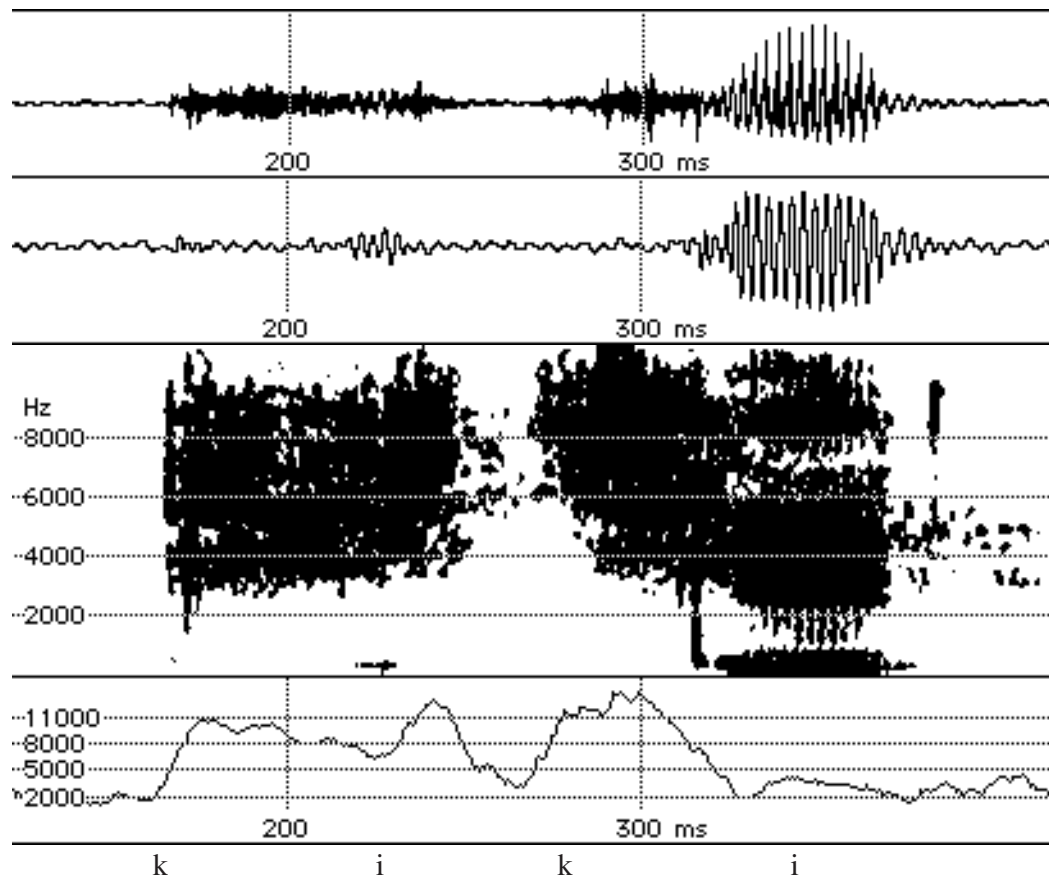


Figure 2.13 Production of *kiki* ‘anxiety’ by participant ANa showing frication accompanying heavily voiced 2nd vowel.

The relationship between devoicing and frication can be seen in the 2nd mora of all three productions. In the 1st production, the number of zero crossings is quite high throughout the vowel site; in the 2nd, it decreases throughout the light vowel voicing (faintly visible in the spectrogram; 27 ms as measured from the waveform); in the 3rd, the number of zero crossings drops to a much lower level during the stronger voicing.

It can also be seen from the small valley in the zero crossings trace above the voicing of the 1st mora vowel in each production that the strength of the medial stop’s closure also plays a role in how spirantized the vowel is. In Figure 2.12 above, the medial stop closure is not as abrupt as it often is in productions in this study; there is an increase in both the frication before the closure and the number of zero

crossings (visible in the zero crossings trace). Finally, in Figure 2.13 it can be seen that the frication associated with the medial closure is quite strong, resulting in a high peak in the zero crossings trace.

This co-occurrence of frication and reduced voicing duration is considered to be evidence of oral closure during the production of the token. The combination of the velar closure and the close oral production of the high vowel facilitate the production of the affricate release. It is thought that this oral closure due to the production of the affricate release increases supra-glottal pressure to the point that voicing is inhibited. Note that this is quite different than the interpretation in the cases above that were attributed to the lengthening of the glottal spreading gestures as per Imaizumi et al. (1995). In this case the trigger is evidently not an adjustment of a glottal spreading gesture, but instead an adjustment of the oral closure—the phonological production of an affricate velar release. The resulting phonetic increase in airflow obstruction results in the decreased voicing.

However, it can also be seen from other productions that higher levels of frication, indicated both by high-frequency activity in the spectrogram and high numbers of zero crossings, can occur with strongly voiced vowels. This can be seen in the following two productions. (Figure 2.14 is a production at the normal SR, Figure 2.15 at the fast.)

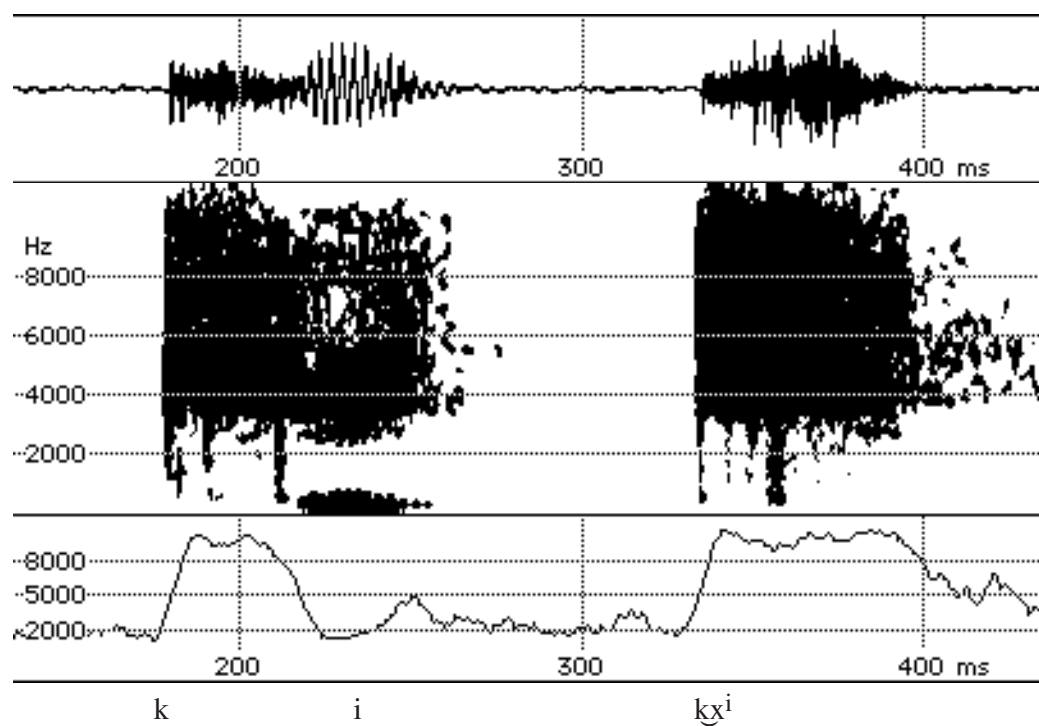


Figure 2.14 Production of *kiki* ‘anxiety’ by participant ANa showing lower frication on voiced 1st vowel.

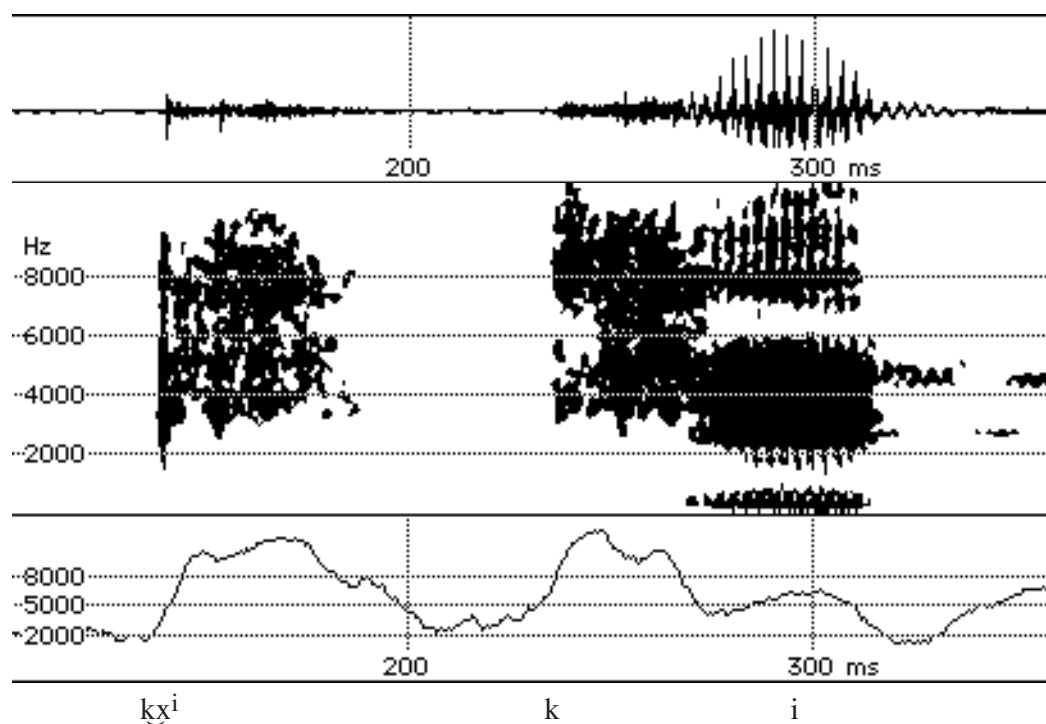


Figure 2.15 Production of *kiki* ‘anxiety’ by participant ANa showing heavy frication on heavily voiced 2nd vowel.

The spectrogram in Figure 2.14 shows that the frication associated with the 1st mora is quite heavy throughout, but the zero crossings trace shows that the number of zero crossings (indicative of frication) falls substantially after the beginning of the vowel voicing to a level of about 2000, and then increases again at the proximity of the oral closure for the medial stop. In contrast, in the second production, the number of zero crossings remains at a higher level through the production of the vowel even though the vowel is much stronger voiced than in the first production. This would seem to indicate that while oral closure and duration of vowel voicing are closely linked, there can be independent oral closure without significant devoicing. The limits as to how much frication can coexist with voicing remain to be seen once the full data set has been subjected to this investigation.

On a final note regarding fricativization, note the almost total loss of the medial stop closure in Figure 2.13. Indeed, in some productions the medial closure was entirely lost, so that in effect one long affricate containing two concentrations of formant activity was observed. A representative production is presented below.

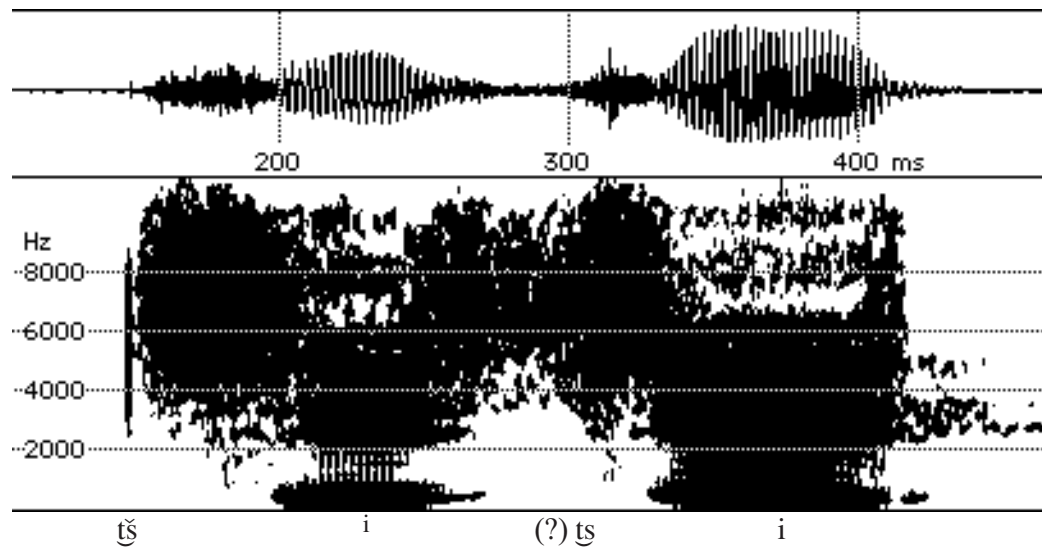


Figure 2.16 Production of *chichi* ‘father’ by participant CS showing no full closure of the medial stop.

The many repetitions showing evidence of frication continuing throughout the closure site for this and other participants speaks against this being simply a speech error. Indeed, even with the lack of closure the token is still readily identifiable. Rather, coordinated yet independent control of the oral closure is indicated.

To what degree manipulation of the oral closure affects voicing by reducing the air flow and hence the vocal fold vibration, remains for future study involving analysis of the current data set and interpretation of EMG and glottal aperture inspection.

The next section will attempt to clarify the terminology surrounding HVD based on the above observations and those of other works.

2.4 CLARIFICATION OF TERMINOLOGY

As discussed above, the traditional formulation of the rule of HVD does not accurately reflect the articulatory changes that often occur with both voiced and devoiced vowels. Evidence that other researchers have observed this phenomenon is found in the wide range of terminology used to describe ‘devoicing’ of Japanese

vowels discussed in Beckman (1994: 57): ‘deletion’ or ‘deleted’ (Bloch 1950; Ohso 1973; Beckman & Shoji 1984); ‘two grades of devoicing’ (Maekawa 1990); ‘voiced’, ‘completely devoiced’, and ‘partially devoiced’ (Jun & Beckman 1993); and ‘fully voiceless’ (Kondo 1993). Beckman (1994) adds the term ‘devocalization’. And as discussed above, Hahn (1991) uses the term ‘obstruentization’ in discussing a very similar process in the Southwestern dialects of Uyghur where high vowels between voiceless consonants surface as voiceless fricatives, while Ladefoged & Maddieson (1996) describe the ‘fricativized’ vowels of Asia.

This work assumes that the wide-ranging terminology that has been employed in the literature reflects observation of the variations in articulatory reduction. As noted in Kondo (1997), Jun & Beckman (1993), and Beckman (1994), vowel devoicing in Japanese is akin to vowel reductions in other languages. In particular, Kondo (1997) showed variable reduced intensity of vowels in devoicing environments. It seems reasonable to view the many labels given Japanese high vowels as indicative of the various observed stages of reduction.

This section attempts to account for the observed changes with the appropriate geometrical representation. It follows Ladefoged & Maddieson (1996) in using the term ‘fricativized’ when discussing the vowel-colored fricatives that are found in the current data set, since this is the phonetic reality as observed from the productions in the previous sections. It reserves the more usual term ‘whispered’ to indicate production lacking only specification for [-voice] in order to avoid the ambiguous term ‘devoiced’, since production of a truly devoiced (i.e. whispered) vowel is an articulatorily different process than the productions observed here. This differs somewhat from the definitions used in other works (e.g. Yuen & Hubbard 1997) that collapse devoicing and whispering.

From a production viewpoint it appears that high vowels in Japanese reduce in three independent yet interconnected ways. They can be devoiced, either totally due to a shift of the glottal spreading gesture or partially due to lengthening of the frication associated with increased closure of the oral constriction. They can also be fricativized as a result of the increased oral closure, whether or not they are fully devoiced. Finally, the formant structure associated with the tongue-shaping gesture

of the vowel can be reduced in degrees, disappearing entirely for many tokens involving velar stops followed by devoiced vowels. As noted before, however, the full deletion of the formant structure is quite rare in the current data set.

More commonly, the simultaneous application of an increased oral closure and overlapping of the voicing results in a fricativized, devoiced vowel—a fricative that still carries the vowel information in the form of formant structures superimposed on the frication of the segment occupying the vowel site. However, since voiced vowels carrying high amounts of frication were also observed in the current data, it is assumed that the two processes operate independently of each other.

Therefore the following definitions of the various possible vowels in Japanese will be held to be accurate:

- 1) **fully voiced vowels** that are produced with voicing for the full duration of formant activity; the energy for the vocal resonance is provided by the vocal cord vibration;
- 2) **partially voiced vowels** that are produced with voicing for only a portion of the formant activity; frication is often present due to the increase in oral constriction with the energy for the formant structure being provided by both the vocal cord vibration and frication of the increased closure.
- 3) **whispered vowels** that are fully formed vowels lacking only vocal cord vibration which do not display the heavy frication associated with an increased oral closure; the energy for the formant structure development is provided by glottal friction (primarily the second formant frequency—see Ladefoged 1996: 107-108 and Ní Chasaide & Gobl 1997 for discussion);
- 4) **fricativized vowels** that are produced with increased frication due to increased oral closure; these vowels may or may not display vowel voicing¹³, and the increased oral closure may or may not result in the loss of the medial stop;

¹³The different set of formant frequencies produced with these vowels due to the reduced oral cavity will be noted, but the exact values found in the current data set

5) **devoiced vowels** that are produced with no voicing but formant structure is present on the frication associated with the glottal spreading gesture of the preceding obstruent or the frication due to an increased oral constriction; the frication provides the energy for the development of the formant structure; and

6) **deleted vowels** where there is no trace of formant structure or voicing remaining, the duration of the mora containing the vowel site may or may not be substantially reduced.

While the specification for both fricativized and devoiced vowels seems redundant since devoiced vowels by the definition above are a subclass of fricativized vowels (i.e. the vowel information is carried by the frication of the preceding obstruent), this dual specification is necessary to encompass the fact that voiced vowels can carry frication even though the main source for formant structure energy is the vibration of the vocal cords.

The following section will attempt to describe these changes with appropriate geometrical representations of the changes the vowels undergo.

2.5 GEOMETRIC REPRESENTATIONS OF VOWEL TYPES

In order to approach appropriate representations for all of the various vowel states delineated above, sample representations are given here for the first vowel in the word *chichi* [tʃitʃi] ‘father’. For clarity, only the first vowel of each word will be dealt with in the following discussion. It will be understood that the vowel is eligible to undergo HVD; i.e. is followed by a voiceless obstruent.

2.5.1 VOICED VOWELS

The first two types of vowels listed above are those vowels that display voicing, that are fully or partially voiced. Any high vowel that retains voicing—i.e. a high

will be left for further research. See Ladefoged (1996 ch. 7) for discussion of the effects of cavity shape on formant frequencies, and Kondo (1997) §8.4.5 for analysis of F1 and F2 for that data set.

vowel that does not undergo HVD because it is not in a reducing environment or because the speaker simply chooses not to voice it (e.g. Figure 2.4 above)—must be underlyingly specified for voicing. The gradency in voicing durations discussed above are then due to phonetic processes at work on the following underlying representation.

The underlying representation of a voiced vowel is presented below.

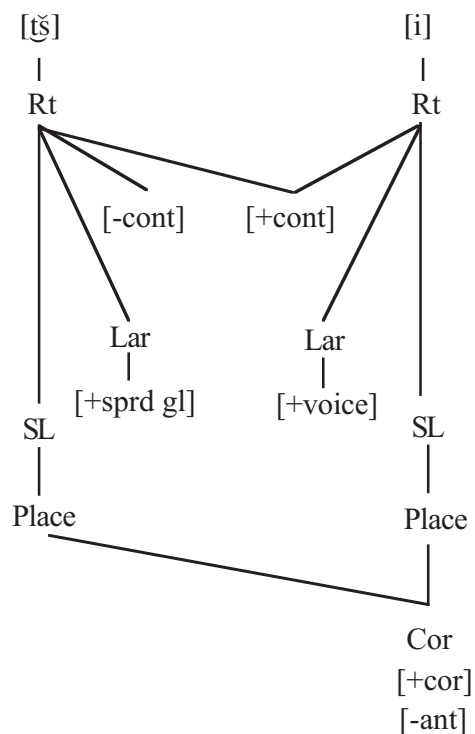


Figure 2.17 Geometric representation of a voiced vowel.

This representation is the same one given for the allophonic variation in Figure 1.6 after the derivation of the allophone [t͡ʃ], and can be observed in a number of the spectrograms presented in previous sections (e.g. Figure 2.4). Note that as discussed previously, Japanese voiceless obstruents are thought to be underlyingly specified for [+spread glottis] rather than [-voice], just as English voiceless obstruents are in Iverson & Salmons (1995) (see §2.2.4 for discussion).

2.5.2 WHISPERED VOWELS

The third type of vowel specified above is a vowel that has been devoiced by a phonological (i.e. feature-changing) rule. This would involve a straight-forward delinking of the vowel's [+voice], the association of the preceding consonant's [+spread glottis]¹⁴ to the vowel's Laryngeal Node, and finally deletion of the delinked [+voice] via Stray Erasure¹⁵.

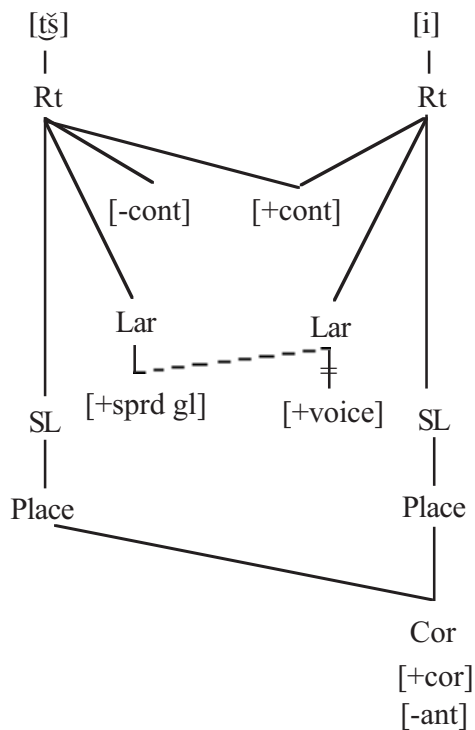


Figure 2.18 Geometric representation of a whispered vowel.

The vowel would then be specified for [+spread glottis], and would surface as a voiceless (i.e. whispered vowel). It is not expected that this type of production will be found in normal speech due to the time and articulatory effort it takes to produce

¹⁴Linking the [+spread glottis] of either the preceding or following voiceless obstruent would achieve the required representation; however, spectrographic evidence presented above and the fact that devoicing occurs before a pause (assumed to carry no features) speaks to a spread from the preceding obstruent.

¹⁵See Blevins (1995: 223-224) for discussion of Stray Erasure.

it; careful articulation of vowels while whispering simply requires too much effort to be of use during normal, fast speech. Among the 3600 vowels produced by the participants in the current study, no production seemed to be truly whispered vowel. However, see Tsuchida (1977) §3.4.6 for example spectrograms of whispered vowels.

2.5.3 FRICATIVIZED VOWELS

The most often produced vowel type in the current data set, and indeed possibly in everyday speech, are the fricativized vowels Hahn (1991) calls ‘obstruentized’. Since fricativized vowels appear to be quite rare among the world’s languages and have not been documented to any length, it seems beneficial to review the articulatory characteristics of this type of production and what type of modifications to the geometry given in Figure 2.17 would be required.

Based on close observation of the productions collected for this study and other native speaker’s speech, the following modifications to the geometry of a voiced vowel are posited.

- vocalicity ([-consonantal]) is reduced or lost
- place of articulation ([+coronal]) and continuity ([+continuant]) are retained
- voicing may or may not be lost

The first of these considerations, the loss or reduction of vocalicity, could be achieved by manipulation of the phonological feature [consonantal].

As mentioned in §1.4.2, Kaisse (1992) places the feature [consonantal] to a daughter position of the Root node, as opposed to it being an intrinsic part of the Root node. This is because there are several cases where it appears that [consonantal] and only [consonantal] is spreading, mandating its autonomy in the geometric hierarchy. Ample evidence is presented for this (see Kaisse 1992 §2 for detailed discussion): obstruentization of /y/ in Cypriot Greek (Newton 1972); glide hardening in Bergüner Romansch (Kamprath 1986); and consonant dissimilation in the Athabaskan language of Ahtna (Kari 1990). Especially of note is the high vowel devoicing followed by obstruentization in Uyghur detailed in Hahn (1991) that was discussed earlier.

Kaisse (1992) attributes the high vowel fricativization occurring in Uyghur to the spread of [+consonantal] from the following consonant, since the intermediately derived, contour segment that begins as a voiceless vowel and ends as a voiceless fricative is [+consonantal] on its right edge. In contrast, the direction of spread of [consonantal] happening with Japanese HVD is necessarily from the preceding consonant, since the formed fricative shares the place of articulation of the preceding consonant, not the following: $ki\check{s}i \rightarrow k\check{j}si \rightarrow kx\check{j}si$ ‘riverbank’ (cf. * $k\check{s}i\check{s}i$).

One possible derivation would then be a spreading of the feature [consonantal] as proposed by Kaisse (1992)—the feature [-cons] will be delinked from the vowel, followed by the subsequent spreading of the preceding consonant’s [+cons]. This representation is given below for a voiced, fricativized vowel.

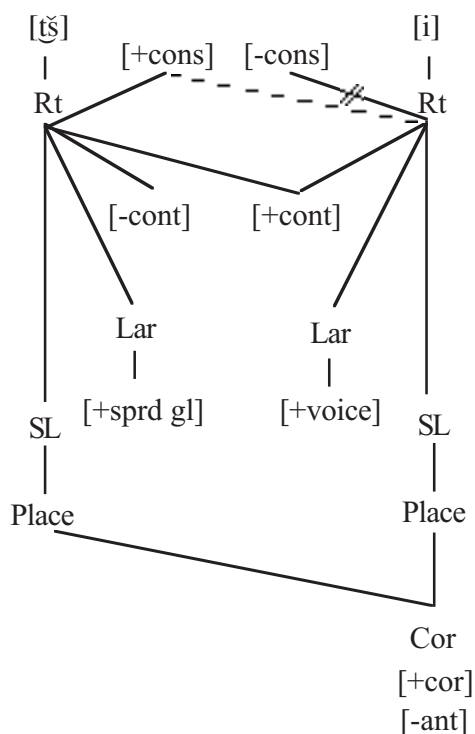


Figure 2.19 Geometric representation of a fricativized vowel.

The above manipulation of the feature [+cons] would also explain the affricatization of the velar stop seen in the discussion of fricativization above (e.g. the 2nd mora of Figure 2.11). In those cases as well, the redundant [+cont] specification of the vowel,

linked to the [-cont] specification of the velar stop, would result in the required velar affricate (i.e. [kx]).

The problem with this analysis is that it does not predict the gradience of frication that is observed in the production of vowels where the voicing has been lost and the formant information is carried by the frication. The correlation of the presence of frication and voicing—i.e. voicing is reduced as frication is increased—indicates that this is a phonetic process, not a phonological one. It may well be more appropriate to simply posit a phonetic process of overlap of the oral closure gesture (i.e. a specification for [+consonantal]) similar to the overlap of the specification for voicing that was discussed earlier in this chapter.

For now, the representation above will be taken as given, with the exact nature of this process, as well as the strength of the tie between oral closure and devoicing, being left for further research.

2.5.4 DEVOICED VOWELS

As noted above, fricativized vowels may be either voiced or devoiced; hence the reservation of the term ‘devoiced’ in this thesis for only those fricativized vowels that are also fully devoiced. Many of the vowels observed in the current study fall in this category.

The geometric representation given above can handle this case quite easily by simply spreading the feature [+spread glottis] from the preceding consonant so that voicing is lost as per the discussion in §2.2.4. This representation is given below.

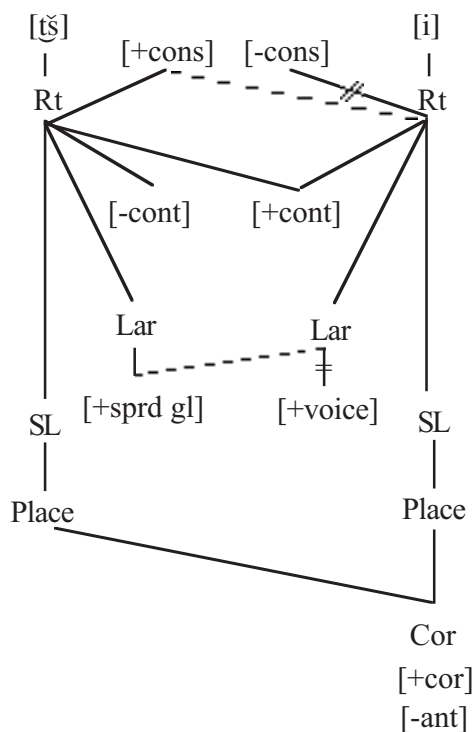


Figure 2.20 Geometric representation of a devoiced vowel.

This representation reflects the fact that the primary articulation of a fully devoiced (and fricativized) vowel has changed considerably. As an extension of the closure forwarded above for fricativized vowels, the portion of the production associated with the vowel has now become a full fricative due to the spread of [+cons].¹⁶ Still, the configuration of the vowel articulation is sufficiently maintained to allow the formation of formant frequencies (although not the same frequencies as a fully open vowel); the frication at the close oral closure results in a heavily vowel-colored fricative. This vowel-colored fricative can be fully sustained; productions as long as two full seconds have been observed by the current author when someone was asked to repeat themselves slowly several times.

¹⁶In the framework of Clements & Hume (1995), this would also entail a collapse of the V-Pl node; in the framework of Keyser & Stevens (1994) a shift of the dominant node from the Root node to the Supralaryngeal node.

It is important to stress here that while these productions cannot be considered to be vowels in the strictest sense, enough of the formant frequency information is maintained to allow identification of the target production.¹⁷

2.5.5 DELETED VOWELS

The final stage of vowel reduction would be deletion, the complete loss of vocalic information. This would correspond to the case where no vowel information has been preserved; the frication of the preceding consonant is not colored by the deleted vowel in any manner. This would correspond phonologically to a vowel timing slot that has had all specifications for segmental features deleted from it. The stranded timing slot could then be either preserved through association to the preceding consonant's timing slot via Compensatory Lengthening (Hayes 1989), or lost through Stray Erasure.

In regard to duration of the mora the vowel is deleted from, if the timing slot were lost along with the segmental information associated with it, it is expected that a shortened mora would be produced. Alternatively, the timing slot may be preserved through association to the preceding consonant via Compensatory Lengthening after the segmental information is lost. In this case it is expected that the duration of the mora will be approximately the same as the duration of a mora where the vowel is not deleted, consistent with the findings of Kondo (1997, Ch. 7). See also Han (1994: 79) for discussion and data.

A geometric representation of a deleted vowel according to this definition is of little use, since no information would be retained under the Root Node. Instead, a representation of the timing tier where the vowel's timing slot has been preserved through Compensatory Lengthening is given in the following figure.

¹⁷Interestingly, native Japanese speakers attending a seminar at the author's university could not identify the most highly reduced-vowel tokens in isolation, but could identify them in context.



Figure 2.21 Geometric representation of a deleted vowel, with Compensatory Lengthening.

It will be noted that full deletion of the vowel—i.e. a full loss of vocalic information—was not a strategy utilized by most participants in this study. In almost all reduced productions there is still enough coloring of the frication to identify the vowel that was underlyingly present.

An apparent instance of deletion of the vowel's timing slot can be seen in the following production. The 2nd mora vowel appears to have been completely deleted, resulting in a shortened mora.

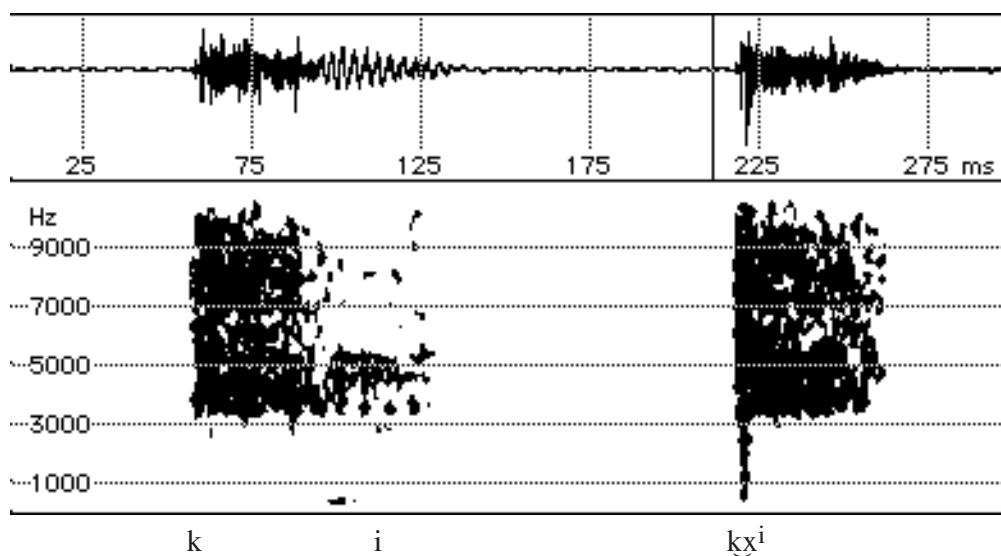


Figure 2.22 Production of *kiki* 'anxiety' by participant ANa showing deletion of the vowel timing slot.

As can be seen, the only thing present in the 2nd mora is the short duration frication associated with the release of the velar stop. However, although the formant structure of the frication is not well-developed in the spectrogram, it is audible enough to identify the vowel that was present. Although the entire vocalic segment

appears to have been deleted, the vowel is still perceptually present similar to the preservation of vowel information in Kaisse (1985, 1990).

The geometric representations presented above are perhaps best thought of as a continuum of vowel reduction—full production (i.e. specification for all vowel features), whispering (i.e. deletion of only the feature [voice]), fricativization (i.e. adjustment of the feature [-cons]), devoicing (i.e. deletion of [voice] and [cons]), and deletion (i.e. the loss of all segmental information of the vowel).

2.6 SUMMARY OF CHAPTER 2

In the earliest characterizations of HVD (e.g. McCawley 1968), devoicing was viewed as a strictly phonological change of the feature [+voice] to [-voice]. However, observations of glottal spreading gestures and the electric potentials of the muscles causing glottal spreading (Sawashima 1971; Tsuchida 1997) has not upheld this. Based on the one speaker tested, Tsuchida (1997) posited 2 loss of voicing processes for Japanese. The 1st is phonological in nature, involving devoicing by spreading of the feature [spread glottis] from the preceding stop. The sharing of the feature by the stop and the vowel thereby results in the temporal realignment of the spreading gesture to the midpoint of their productions (Kingston 1990; Iverson 1997). The 2nd process is phonetic in nature, operating between voiceless fricatives, and involves a gestural overlap of the vowel voicing gesture by the glottal spreading gesture of a preceding and following fricatives. This results in gradient amounts of vowel voicing. In this model phonological devoicing occurs between plosives, but phonetic loss of voicing occurs between voiceless fricatives.

It was noted that the data from the current study supported gradient amounts of voicing between the allophonic affricates occurring in the tokens used. It seems reasonable to assume that the fricative portion of affricates affect vowel voicing in the same way that voiceless fricatives do. However, it was also noted that devoicing between velar plosives shows the same gradiency that devoicing between fricatives and affricates do, and further that there appears to be only one glottal spread after productions of affricates as well. Tsuchida (1997) is therefore supported by this

study—i.e. there are both phonological devoicing and phonetic loss of voicing at work in Japanese—with the modification that both processes occur in all environments. That is to say, both processes occur between all voiceless segments.

Attention was also given to the mode of production of Japanese devoiced vowels. Specifically, it was maintained that many productions are actually fricativized vowels (Ladefoged 1996), usually but not always accompanied by devoicing. This is quite similar to the ‘obstruentization’ in Uyghur, a Turkic language, discussed in Hahn (1991). Examples of both the frication observable in spectrograms and the number of zero crossings were presented as evidence.

An attempt toward clarifying the terminology surrounding HVD was made, as well as a step toward appropriate geometric representations of the various stages of vowel reduction observed in devoicing data. The various stages posited were *fully voiced vowels*, characterized by voicing lasting the duration of the oral cavity configuration with accompanying well-defined formant structure; *partially voiced vowels*, characterized by voicing lasting only a portion of the time formant structure does; *fricativized vowels* which may or may not be devoiced, characterized by heavy amounts of frication attributed to an increased oral closure; *devoiced vowels*; characterized by heavy frication which serves as the vehicle for the formation of the formant frequencies of the original vowel; and *deleted vowels*, characterized by both a loss of voicing and formant activity. It was noted that in all productions checked some evidence of vowel coloration was observed even at the most heavily reduced vowel sites. In addition, very few productions seemed to have a deletion of the timing slot (as evidenced by a short period of frication) associated with devoicing; almost always a longer period of vowel-colored frication was produced.

Phonological devoicing is thought to consist of a spread of the feature [+spread glottis] from the preceding obstruent. This is based upon the place of articulation of the affricate derived from the preceding stop ([kx^hʲi], cf. *[kʃ^hʲi]) or the continued frication of the preceding affricate or fricative, as well as the shift of the frication from the preceding obstruent toward the vowel that is observable in spectrograms). Fricativization is thought to possibly consist of a spread of the feature [+consonantal] from the preceding obstruent. However, it was noted that this characterization of fricativization is not satisfactory due to the gradient nature of

frication observable both on the vowel site and the following stop's closure, and will need to be further explored in detail.