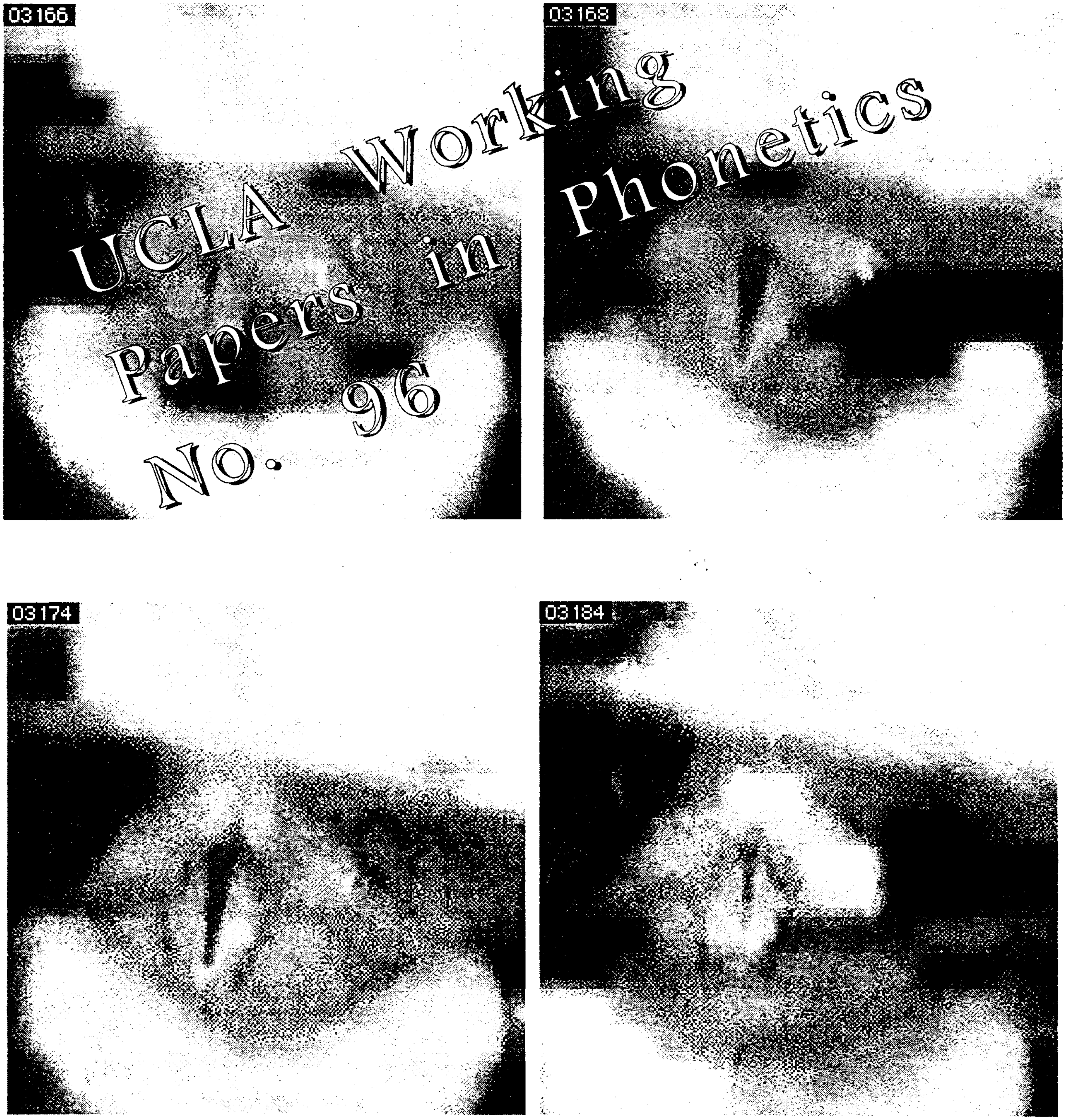
UCLA Working Papers in Phonetics

No. 96



October 1998

The UCLA Phonetics Laboratory Group

Soo-Woong Ahn
Adam Albright
Victoria Anderson
Marco Baroni
Barbara Blankenship
Taehong Cho
Sandy Disner
Gorka Elordieta
Melissa Epstein
Heidi Fleischhacker
Christina Foreman
Susan Meyers Fosnot
Cécile Fougeron
Sean Fulop
Matt Gordon
Narineh Hacopian
Bruce Hayes
Chai-Shune Hsu
Sun-Ah Jun
Pat Keating
Robert Kirchner
Jenny Ladefoged
Peter Ladefoged
Hyuck-Koon Lee
Ian Maddieson
Elizabeth Menard
Amy J. Schafer
Aaron Shryock
Donca Steriade
Sumiko Takayanagi
Henry Tehrani
Kimberly Thomas
Siri G. Tuttle
Moto Ueyama
Jie Zhang

As on previous occasions, the material which is presented in this volume is simply a record for our own use, a report as required by the funding agencies which support the Phonetics Laboratory, and a preliminary account of research in progress for our colleagues in the field.

Funds for the UCLA Phonetics Laboratory are provided through:
USPHS grant 5 T32 DC00029
NSF grant SBR 9319705
NSF grant SBR 9511118
and the University of California

Correspondence concerning UCLA Working Papers in Phonetics should be addressed to:

Phonetics Laboratory
Department of Linguistics
UCLA
Los Angeles, CA 90095-1543

The General Editor of UCLA Working Papers in Phonetics is Ian Maddieson.

UCLA Working Papers in Phonetics

Number 96

October 1998

Table of Contents

Electromyographic Evidence for a Gestural-Overlap Analysis of Vowel Devoicing in Korean*

Sun-Ah Jun[1], Mary E. Beckman[2], Seiji Niimi[3], and Mark Tiede[4]

[1])Phonetics Laboratory, Department of Linguistics, University of California, Los Angeles
[2])Linguistics Department, Ohio State University [permanent address] &
ATR Interpreting Telecommunications Laboratories, Kyoto [visiting researcher, 1994-95]
[3])Research Institute of Logopedics & Phoniatrics, Faculty of Medicine, Tokyo University
[4])ATR Human Information Processing Laboratories, Kyoto

1. Introduction

In languages such as Japanese, where the specification of a peripheral vowel quality is not inherently associated with the longer duration of a heavy (or "stressed") syllable, it is common to observe productions of syllables with short peripheral vowels which are completely voiceless. While this phenomenon of "devoicing" has been noted for Montreal French (Cedergren & Simoneau, 1985), Shanghai Chinese (Zee, 1988), Greek (Dauer, 1980; Arvaniti, 1994), Korean (Jun & Beckman, 1993, 1994; Kim, Niimi, & Hirose, 1993), and Turkish (Jannedy, 1995), it has been studied most extensively in standard (Tokyo) Japanese, where traditionally it has been described in terms of a symbolic rule that either categorically deletes the vowel or (the more typical characterization) categorically changes the vowel's phonological feature specification from the default value of [+voice] to the marked value of [-voice]. Close examination of the phonetic evidence, however, shows that neither of these phonological characterizations is tenable. Indeed, all of the evidence is more compatible with an alternative account which casts the phenomenon as "undershoot" — that is, as a failure to achieve an interval of audible voicing due to overlap between the glottal adduction target in the vowel and the glottal abduction target for a neighboring voiceless consonant. By this account, vowel devoicing is not a phonological process at all, but another example of the kind of gradient phonetic effect resulting from more or less subtle adjustments to the timing of otherwise invariant gestural specifications, as in Lindblom's (1963) account of vowel reduction in Swedish, or a very simple version of Browman and Goldstein's (1990) account of casual speech processes in English.

Such a gestural overlap account is supported over the symbolic account by the spectral patterns typically observed in devoiced vowels. In acoustic studies of the phenomenon, waveforms and spectrograms of target syllables do not show two clearly differentiated types. Rather, there is a continuum of degrees of devoicing, from tokens with one or two reduced glottal pulses to tokens having no trace even of the vowel's lingual gesture. This is true of the studies cited above of Montreal French, Turkish, and Korean, as well as of most studies of Japanese (e.g., Maekawa, 1990; Kondo, 1994).

A study of the perception of such voiceless syllables in Japanese by Beckman & Shoji (1984) also supports the gestural overlap account over the category-changing phonological rule. Minimal pairs such as /kaʃi/ 'sweets' versus /kaʃu/ 'poet' are distinguished much less accurately when the target syllable is voiceless, but still considerably better than at chance level, at about the same rate as *she* and *shoe* are distinguished in English when the fricatives are excised from context in experiments on coarticulation (Yeni-Komshian & Soli, 1981).

Further evidence for gestural overlap and undershoot comes from examining the distributional characteristics of the phenomenon. That is, the account is supported by studies showing that devoicing occurs to variable extents in different contexts, and that it occurs more commonly in just those conditions where a neighboring consonant's devoicing gesture covers a greater proportion of the vowel's gestural activation interval — e.g., because the vowel is particularly short. Thus, in Japanese, devoicing is limited to phonemically short vowels, and it most commonly affects the two high vowels, /i/ and /u/, which are phonetically very short even when they are not devoiced. Devoicing does occur sometimes with mid and low vowels (Maekawa, 1990; Kondo, 1993), but here it is relatively more rare. In other languages, also, devoicing is usually described as affecting the high vowels of the language, and as occurring more frequently in prosodic positions associated with shorter vowel durations. In Montreal French, for example, devoicing is observed for syllables containing /i/, /y/, or /u/, but it is much less likely to affect the second-to-last syllable in a rhythm group, and it never affects the absolute rhythm-group final syllable (Cedergren & Simoneau, 1985; Cedergren & Perrault, 1994). Both of these are positions where the vowel is prosodically lengthened and less likely to be completely covered over by the preceding consonant's gesture (Levac, Cedergren, & Perreault, 1993). In Turkish, similarly, devoicing is observed in syllables containing /i/, /y/, /ɯ/, or /u/, and it is more commonly observed at faster speech rates and in open syllables — where the vowel tends to be shorter than in closed syllables (Jannedy, 1995).

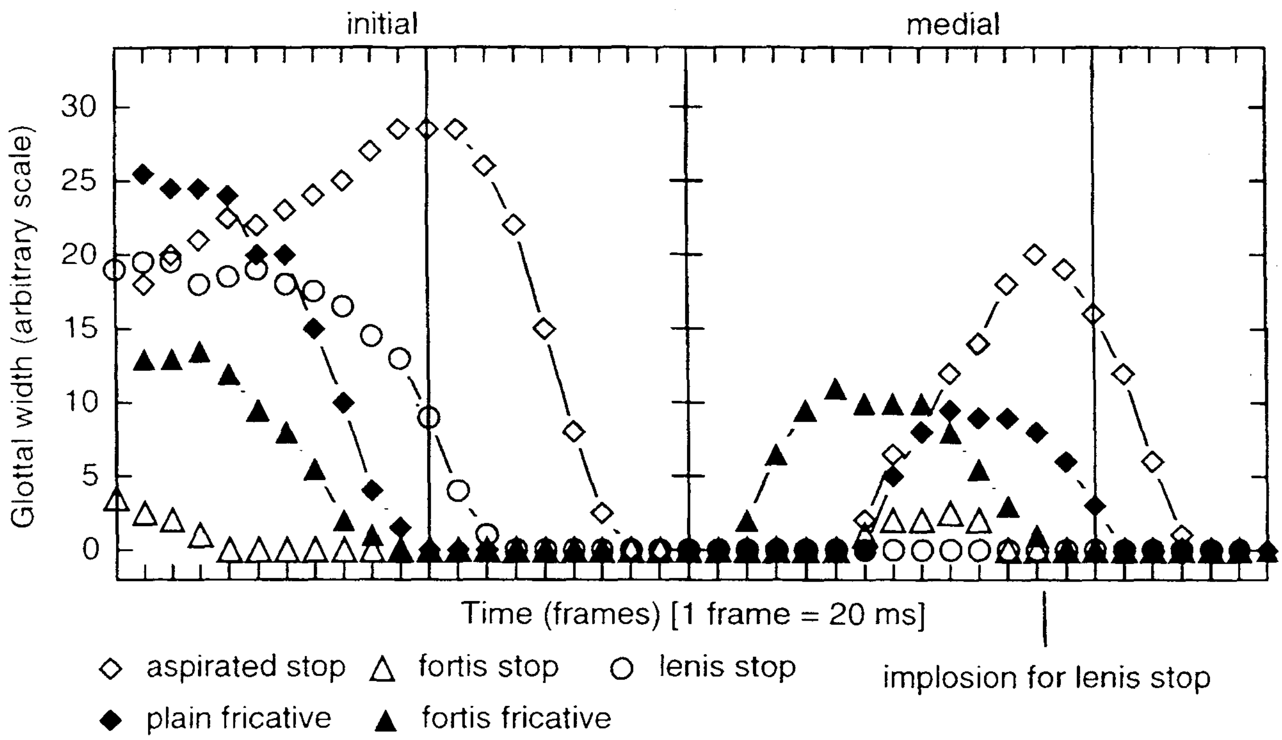


Figure 1. Glottal aperture data for representative utterance-initial consonants (left) and word-medial consonants (right) from Kagaya (1974). The data for the five traces in each panel are estimated from Kagaya's Figures 2 and 3, and then replotted on one graph by aligning at oral release in the aspirated versus fortis versus lenis stops, and at the offset of frication noise in the plain (or lenis) versus fortis fricatives.

In Korean, too, devoicing is observed for syllables containing /i/, /ɨ/, or /u/. The language is also unique in having a particularly revealing set of contrasting voiceless consonants, which are differentiated specifically by their patterns of relative phasing between the stop-source and vowel-filter shapes, and thus can provide inherently different patterns of gestural overlap with the vowel. The different patterns are illustrated in Figure 1a, adapted from Kagaya (1974). For example, the three time traces with open plotting symbols in the left-hand panel are glottal aperture for representative utterance-initial aspirated versus fortis versus lenis stops, aligned at the oral release of the stop. The maximum glottal aperture value is largest in the aspirated stop, and the timing of this local peak is latest relative to the aspirated stop's oral release. In the fortis stop, by contrast, the glottis is completely adducted well before the stop release, to provide the characteristic pressed-voice quality at the onset of phonation in the following vowel. Given these contrasting patterns of overlap, the undershoot account predicts that devoicing should be most frequently observed in a high vowel after an aspirated stop, and it should be least frequently observed in a high vowel after a fortis stop, at least in utterance- or phrase-initial position. These were the relative frequencies documented by Jun & Beckman (1994).

Thus, converging evidence from several types of studies support an account of devoicing across languages that attributes the phenomenon to variation in the relative timing of the gestures for the target vowel and for a neighboring consonant. Devoicing is observed when the degree of gestural overlap affects a proportionally long enough interval of the vowel to cause undershoot of the aperture necessary for voicing.

At the same time, there is also evidence against any very simple undershoot account of devoicing, at least for Japanese. For example, Imaizumi, Hayashi, & Deguchi (1995) examined the likelihood of devoicing in elementary school teachers' productions of children's' names, each of which contained one or two high vowels in a consonantal context conducive to devoicing. The teachers produced the names in three different production modes — a fast lab-speech reading of a list, and two different spontaneous speech modes in which the teacher instructed a normally hearing child or a hearing-impaired child to point at cartoon figures labeled with gender-appropriate names from the list. As predicted by the gestural overlap account, there were many more instances of devoicing in the fast list readings, and many fewer instances of devoicing in the slow, careful speech addressed to the hearing-impaired listeners, as predicted by the simple undershoot model. However, when Imaizumi and colleagues modeled the likelihood of devoicing directly as a function of observed vowel duration, they found a substantial residue of variance across the three task modes, which they attributed to a task-appropriate adjustment to the magnitude of the vowel's voicing gesture as well as to its timing relative to the consonant. Couching these results in terms of Lindblom's current understanding of the possible forms of non-categorical variation (Lindblom, Brownlee, Davis & Moon, 1992), we might say that the "transform" for the hyperarticulated clear speech directed at the hearing-impaired child does not simply adjust timing to inhibit undershoot. Rather, there is also some other articulatory adjustment to provide aerodynamic conditions more conducive to vocal fold vibration during the still rather short period of the phonemically short, unstressed high vowel. To substantiate this interpretation, we need a more direct examination of the articulation, at all levels that are accessible to examination.

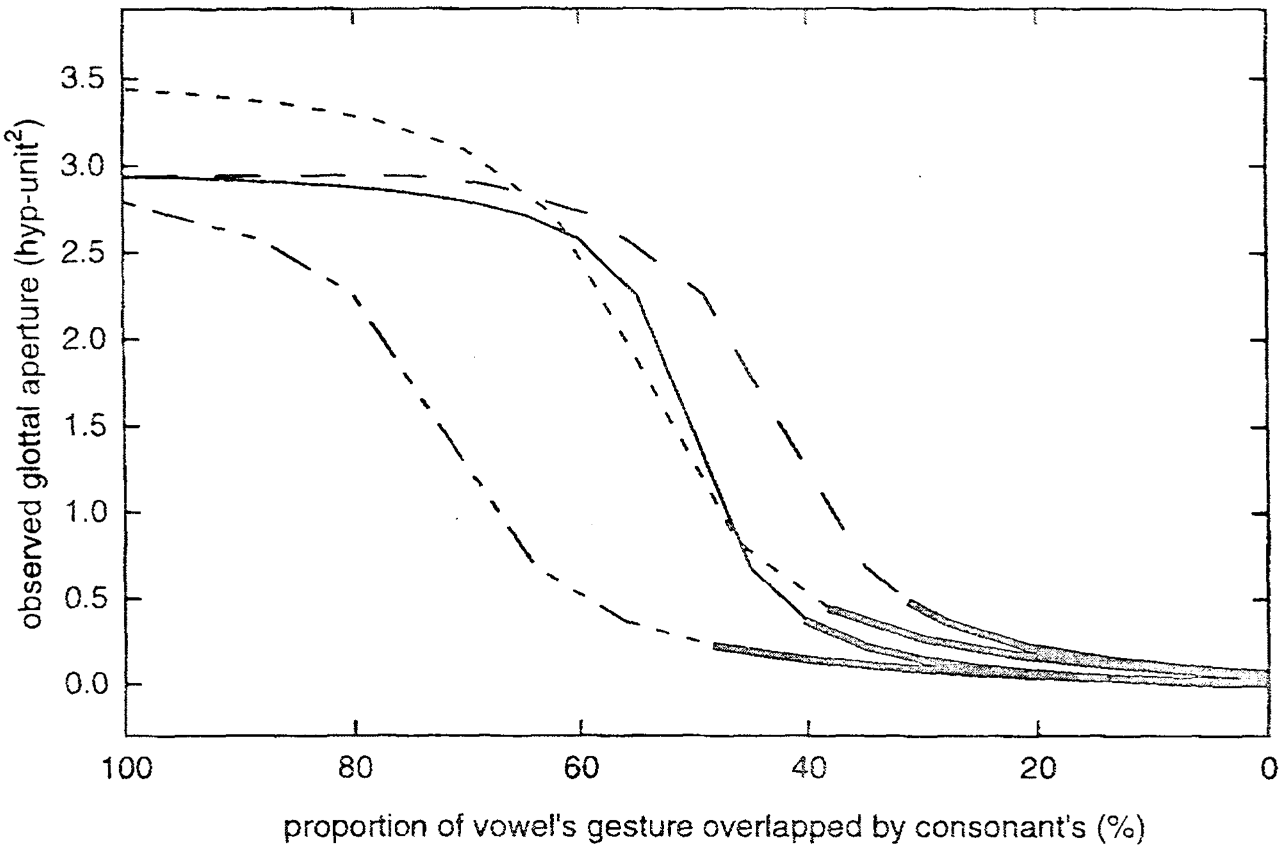


Figure 2. Hypothetical mapping functions for predicting observed glottal aperture from the specification of different degrees of overlap between an invariant vocalic glottal adduction gesture and an invariant consonantal adduction gesture when oral constriction and subglottal pressure are held constant. Values plotted with thicker line segments are mean glottal aperture over all phases of the source wave period, during conditions of voicing. The different broken-line curves suggest the effect on the mapping function of orthogonal changes to some other concommitant specification. The short dashed line shows the hypothetical effect of increasing the abduction target for the consonant at the beginning of an intonational phrase, the long dashed line shows the hypothetical effect of increasing the activation interval duration for the vowel's adduction target in careful speech, and the mixed (short-short-long) line shows the hypothetical effect of decreasing subglottal pressure near the end of a turn-final sentence.

What articulatory data that there are do not unambiguously distinguish simple undershoot from a more sophisticated transform account of Japanese vowel devoicing. For example, Yoshioka (1981) looked at glottal aperture traces measured from successive frames of a cinegraphic record photographed through a laryngeal fiberscope during repeated productions of words such as /hise:/ and /ʃihe:/. Of the 28 productions of /hise:/, four had voiced [i] and the remaining 24 were judged to have devoiced [i̥]. As in earlier studies from the same laboratory (e.g., Sawashima, 1971; Hirose, 1971; Weitzman, Sawashima, Hirose & Ushijima, 1976), the tokens with devoiced [i̥] typically showed a qualitatively different pattern from that observed in the tokens in which the target vowel was voiced. In tokens containing a voiced [i], there were two glottal aperture peaks for the surrounding voiceless fricatives, separated by a clear dip to full adduction in the vowel. In tokens judged to contain a voiceless [i̥], by contrast, glottal aperture typically peaked only once, in the word-initial fricative, and then slowly declined over the entire interval up to the glottal adduction for voice onset in the /e:/ of the second syllable. However, the lack of any clear minimum for the vocalic segment in the representative glottal trace could still be attributed to undershoot rather than to any real difference in the adduction gesture itself, because there was no token-by-token control of the degree of between-gesture overlap. Indeed, in the one pair of individual tokens for which Yoshioka shows spectrographs, the CVC sequence containing the devoiced vowel is shorter than the sequence containing the voiced vowel, as would be predicted by a simple undershoot account. While the difference is not large absolutely, it is large relative to the length of the interval of voicing in the token with the voiced [i]. And the same is true also of the pair of spectrograms in Hirose's (1971) earlier report.

In first reporting such contrasting patterns in the average glottal aperture traces, Hirose acknowledged that the glottal aperture differences are not unequivocal evidence for a difference in gesture per se: "It might still, at this point, be argued that there might be no difference in the motor command for the laryngeal movement between voiced and devoiced [i], but depending on difference in some physical conditions near the level of the glottis, the resulting glottal state would vary as a result of 'passive' but highly non-linear effect" (Hirose, 1971, p. 162). The "physical conditions" might be the result of specifying other concurrent gestures. For example, suppose the talker targets a larger backdrop exhalation rate to produce a high boundary tone (cf. Herman, Beckman & Honda, 1996), or suppose the talker targets a less forceful oral seal for a phrase-medial lingual stop (cf. Fougeron & Keating, to appear; Fougeron 1996; Hsu & Jun 1996). Either of these subtle adjustments in the magnitude of other gestures would affect the pressure differential across the glottis and change the observed glottal aperture for an invariant aperture target. Or the "physical conditions" could result from different degrees of gestural overlap. That is, suppose that the observed glottal aperture during the activation interval for the vowel's oral filter shape is not a linear function of targeted glottal aperture, but rather is a consequence of "blending" the vowel's adduction gesture with the overlapped abduction gesture of the preceding voiceless consonant. The blending function for different degrees of overlap could be nonlinear (as shown by the hypothetical curves in Figure 2), producing the appearance of two qualitatively distinct programs even when the speaker exercises continuously variable control over the phasing between the adduction gesture for the vowel and the glottal abduction gesture for the voiceless consonant.

Because of the possibility of such nonlinear mapping functions, Hirose (1971) felt it imperative also to examine articulation at a level of the speech chain closer to the gesture (or input "motor command"). He did so by looking at the smoothed electromyographic (EMG) activity level recorded from hooked-wire electrodes inserted into the fibers of the thyroarytenoid (TA) and of the lateral cricoarytenoid (LCA) during citation-form productions of words containing the high vowel /i/. Contracting either the TA or LCA might pull the arytenoid cartilages forward in a way to rotate them medially, and thus help to draw the vocal processes together. Comparing the EMG trace averaged over tokens of /ʃise:/ containing a voiced [i] with the averaged trace for tokens containing a devoiced [i̥] Hirose found qualititatively different patterns. For both talkers in the study, the average TA trace had an early peak about 100 ms before the onset of the voiced high vowel which was not observed in the average trace for the tokens with devoiced [i̥]. For one of the talkers, there was a comparable difference between the two average traces for the LCA. Hirose concluded that devoicing is "a matter concerning the neural process that determines the motor command" and "that the choice between different gestures is made as an optional application of a phonological rule" (Hirose, 1971, p. 166).

In his later study using similar word types embedded in a carrier sentence, Yoshioka (1981) also looked at smoothed EMG traces for an adductor muscle — the interarytenoid (IA) — and also for an abductor muscle — the posterior cricoarytenoid (PCA). Again, the averaged traces showed qualitatively different patterns. In the average trace for the tokens with voiced [i] there were two clear PCA peaks separated by a local minimum, and a clear small IA peak in that interval. In the average trace for the tokens with devoiced vowel, by contrast, the PCA had a single peak (albeit with a broad following skirt) and an uninterrupted IA trough. Yoshioka summarizes the results as being "generally in good agreement" with the earlier studies inasmuch as "the abductor and adductor muscles show EMG activity patterns clearly different from those for voiced tokens." At the same time, he also noted "considerable variations at the EMG level, particularly in the PCA curves, within the same devoiced group" which made him hesitate to ascribe the variants to a phonological rule ("there still remains the question whether these auditorily explicit two-way allophones are based on two different articulatory programming patterns at a certain neural level, or on the mere nonlinear effect of rather wide fluctuations of the EMG potentials"). He concludes on the cautionary note that "the significance of the averaged curve, especially for the devoiced group, should be reconsidered.... The alternative is to observe the activity patterns for a considerable number of single tokens of the same utterance types, paying special attention to the critical cases" (Yoshioka, 1981, p. 248). (Kiritiani, 1994, later came to a similar conclusion after reanalysing some examples from Sawashima et al., 1978.) In this paper, therefore, we will take Yoshioka's lesson to heart, and look at all the cases that we can of sets of tokens in a database of physiological recordings of Korean words in which a high vowel was variably voiced or devoiced.

2. Method

The database was recorded from one female native speaker of Chonnam Korean (the first author) as she produced the same speech materials used in Jun & Beckman (1994). This was a corpus of disyllabic CVCV target words embedded in the dialogue illustrated in Table 1. Since the dialogue contained two pairs of question and answer sentences, it provided tokens of two different target words in each of two different discourse-focus positions. In the Korean prosodic system, the most salient indication of attentional focus is the phrasing. A word with narrow focus is always initial to an Accentual Phrase, a tonally defined prosodic constituent that is associated with many phonetic processes such as voicing of phrase-medial lenis stops (Jun, 1989, 1996). All words after the focused word until the next major intonational phrase boundary are then grouped together into the same Accentual Phrase. Thus, in our corpus of dialogues, the target word is produced first under narrow focus, making its first consonant initial to an Accentual Phrase, and then is repeated with narrow focus on the preceding deictic pronoun, so that the initial consonant of the target word is medial to its phrase. There were 25 different target words, listed in Table 2. (One target word — [s'uk^{h}i] — is a nonsense word, but the speaker had no difficulty pronouncing it.) The list was designed to provide a balanced set of preceding and following voiceless consonant types for the first vowel, which was the segment of interest. This vowel was any one of the three high vowels of Korean, and each of two surrounding consonants was one of the three stop types or two fricative types in the language.

Table 1. Sample dialogue.

(1) Q: ike kutu-nja?	"Is this 'SHOE'?"
(2) A: ani, tʃəke kutu-ja.	"No, THAT is 'shoe'."
(3) Q: kɨləm, ike kipʰi-nja?	"Then, is this 'EVASION'?"
(4) A: ɨŋ, kɨke kipʰi-ja.	"Yes, THAT is 'evasion'."

The underlined parts are focused, and /kutu/ 'shoe' and /kipʰi/ 'evasion' are the two target words. In sentences (1) and (3), the target word is focused, and hence initial to the Accentual Phrase, whereas in (2) and (4), the target word is post-focal and hence medial to the focused Accentual Phrase starting at /tʃəke/ or /kɨke/ 'that'.

Table 2. Corpus of target word types.

Initial C type	Medial consonant type									
	lenis stop		aspirated stop		fortis stop		lenis fricative		fortis fricative	
lenis	kutu	'shoes'	kipʰi	'evasion'	tuk'e	'thickness'	kisa	'knight'	kus'o[1]	'hardens'
aspirated	tʰuki	'venture'	kʰukʰi	'cookie'	pʰup'e[2]	'pear'	tʰusu	'pitcher'	pʰus'i[3]	'seed'
fortis	k'ɨtʃʰi[4]	'end'-NOM	t'upu[5]	'tofu'	t'ɨk'i[6]	'tearing'	t'ɨsi	'will' NOM	t'is'o[7]	'tears'
lenis s	sikje	'a watch'	sikʰje	'rice-drink'	sik'i[8]	'tableset'	suso	'hydrogen'	sus'o[9]	'bull'
fortis s'	s'ukʰi	NONCE	s'ɨta	'write'	s'it'a[10]	'to wash'	s'usi	'pricker'	s'is'o[11]	'washes'

[1]From /kut/ 'to harden'+ /so/ (polite ending). The fortis fricative here results from phonological word-formation constraints affecting codas or coda-onset sequences at morpheme boundaries within compound words and inflected forms. In careful speech, [kut s'o] is also a possible pronunciation of this word.

[2]From /pʰus/ 'unripe' + /pe/ 'pear' (cf. fn 1). [3]From /pʰus/ 'unripe'+ /s'i/ 'seed'

[4]From /k'ɨtʰ/ 'end' + /i/ (nominative ending). [5]A nonstandard (but acceptable) variant of /tupu/.

[6]From /t'ɨt/ 'tear' + /ki/ (nominalizer). [7]From /t'ɨt/ 'tear'+ /so/ (polite ending).

[8]From /sik/ 'food' + /ki/ 'equipment'. [9]From /sus/ 'male'+ /so/ 'cattle'.

[10]From /s'is'/ 'wash'+ /ta/ (infinitive ending). [11]From /s'is'/ 'wash'+ /so/ (polite ending).

The data were recorded at the Research Institute of Logopedics and Phoniatrics, University of Tokyo, under the direction of the third author. The EMG activity level was recorded from hooked-wire electrodes (Hirano & Ohala, 1969) inserted perorally into the PCA and percutaneously into the TA. These two signals were stored on two channels of a multi-channel analog FM tape recorder. Two of the other channels were used to record the audio signal (from a directional microphone placed about 50 cm in front of the speaker's mouth) and an intra-oral air pressure signal (from a pressure transducer mounted on a catheter inserted into the pharynx through the left nostril). A video view of the glottis was recorded separately at 60 fields (30 frames) per second using a laryngeal fiberscope inserted into the pharynx through the right nostril. During productions of the target dialogues, time stamps were recorded on the video fields, and a corresponding timing pulse train was recorded directly onto the audio track of the videotape. (The RILP physiology system is designed to record the timing pulses also onto a channel of the FM recorder, but no signal was registered during this recording session because of a faulty connection.) The speaker read all of the dialogues once, then read two other corpora before reading these materials a second time, then read the two other corpora a second time, and so on. She was able to read the total list of 25 dialogues three times before it became necessary to remove the electrodes.

Copies of the FM tape and the videotape were sent to ATR Human Information Processing Laboratories for the second and fourth authors to digitally process. The audio, the air pressure, and the two EMG signals were digitized at 24kHz in appropriately-sized chunks so that the resulting files would each contain the records for one dialogue. These files were transferred to a Sparcstation in the ATR Interpreting Telecommunications Laboratories (where the second author was then a visitor), and the four signals were separated using the waves+ demux routine. The two EMG signals were then downsampled to 2400 Hz, rectified, and smoothed by averaging over a moving 70 ms triangular window centered at successive time samples. An Entropic waves+ header was added to each of the four signal files to allow synchronized display, as illustrated in Figure 3 for sentence (3) in the first production of the dialogue in Table 1. The top panel displays the wideband spectrogram, the second and third panels display the processed TA and PCA signals, the fourth panel displays the intra-oral air pressure signal, and the fifth panel displays the fundamental frequency. (The F0 curve was calculated at 100 Hz for each audio file using the ESPS get_f0 program, which uses an auto-correlation based dynamic programming algorithm.) The processed files were then transferred by ftp to the UCLA Phonetics Laboratory Sparcstation for analysis by the first author. Here spectrographic displays of the audio signal were examined to determine whether the vowel was voiced, partially devoiced, or completely devoiced, using the criteria described in Jun & Beckman (1994): i.e., vowels less than 30ms long and having only a low frequency energy were labeled as 'partially devoiced'. EMG events of interest were labeled using the waves+ label function.

The processing of the videotape involved making digitized movie sequences using a Macintosh Quadra 950 to control a Sony EVO-9650 VTR and a RasterOps 24STV video digitizer. There were 25 digitized movie sequences — one for each of the intervals during which timepulses were being recorded for the first reading of one of the 25 dialogues. First the VHS format copy was redubbed to a Hi-8 copy in order to add VTR-legible timecodes from the original timecodes (which were merely superimposed on each image in a format not accessible to the VTR circuitry). Then time offsets were identified for the beginning and end of the interval during which the timecode counter was being incremented during the reading of one target dialogue. The interleaved fields of each video frame from this range were digitized independently, resulting in 60 separate images for each second of video. Because the original experiment timecode counter was incremented at this 60fps (fields per second) rate, the digitized sequences are synchronized exactly with the superimposed timecodes. Each image was sampled in a 24bit RGB color space and compressed using JPEG compression. Because the digitization hardware used introduced an undesirable vertical offset differential between odd and even sampled fields, each image was subsequently resampled using a "jitter" filter to align the top of the superimposed timecode (see Figure 4).

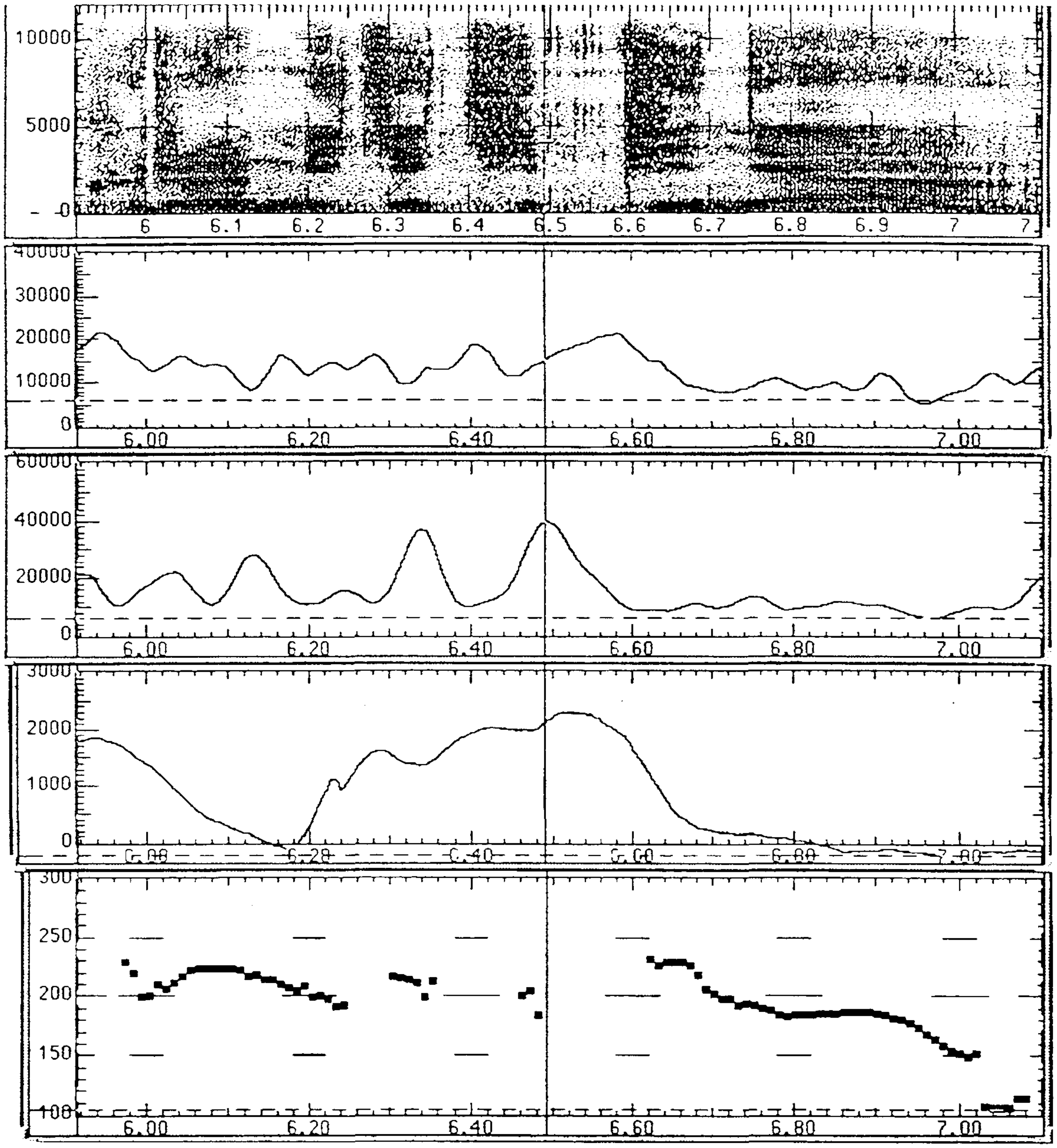


Figure 3. Sample waves+ display for sentence (3) in the first production of the dialogue in Table 1. Starting from the top, the windows are (1) wide-band spectrogram calculated from the audio signal file, (2) smoothed signal recorded from the thyroarytenoid (TA), (3) smoothed signal recorded from the posterior cricoarytenoid (PCA), (4) intra-oral air pressure, and (5) fundamental frequency. The audio signal file and the other four signal files have wave+ headers which allow the displays to be time-locked to each other despite the different sampling rates, and thus provide for a common cursor positioning. The displayed cursor aligns the signals at the implosion into the medial /p^h/ in the target word /kip^hi/.

The original video is visually noisy. A periodic artifact consisting of alternating intensity bands runs diagonally across each image. In an effort to mitigate the effects of this artifact on any subsequent analyses that use automatic edge-detection algorithms, the data were processed using a masking filter applied in the frequency domain. Because this operation is computationally expensive the filter was applied only to a 256 pixel square region framing the glottis (see Figure 5). The frame size was chosen to include the entire glottis and the full inner arc of the epiglottis, in order to be able to normalize any measurement of glottal aperture for the position of the larynx relative to the video focus. (The distance between the glottis and the camera view varies because the larynx moves up and down for different fundamental frequencies, and because the end of the fiberscope moves up and down for different positions of the soft palate.) Although the original superimposed timecode is not visible in these filtered images, a field index synchronized with the original is visible in the upper left corner of each image. The 25 field sequences were recorded as QuickTime Movie files to a compact disk. The CD and the VHS copy of the original videotape were then sent to the UCLA Phonetics Laboratory for analysis by the first author.

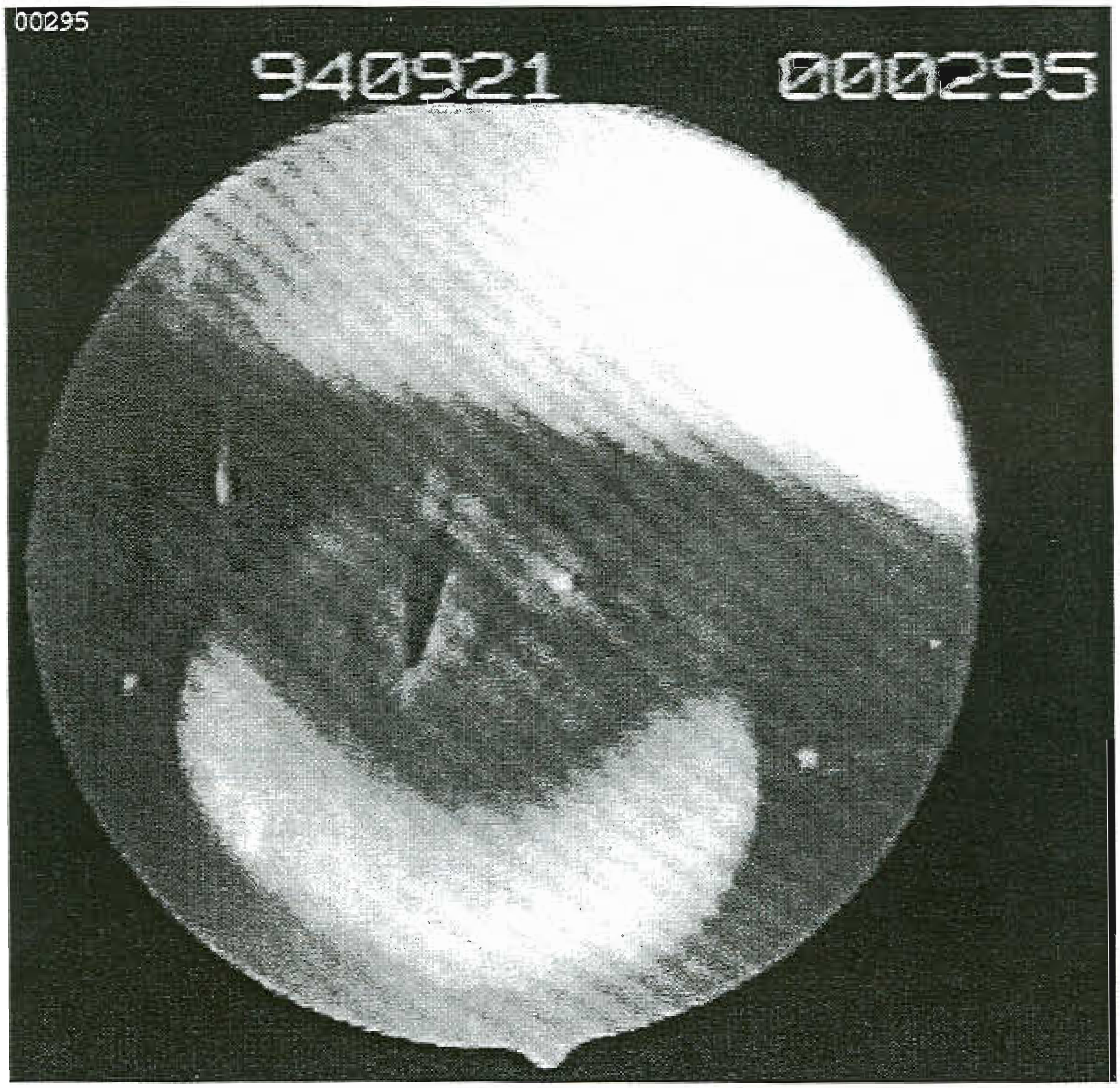


Figure 4. Sample full-sized video view of field number 00295. This field is the one just before the release of the second consonant in the target word /kip^{h}i/ ‘evasion’ in sentence (3) of the sample dialogue. The large “000295” in the upper right corner of the view is the original timecode, and the smaller “00295” in the upper left is the new VTR-legible timecode.

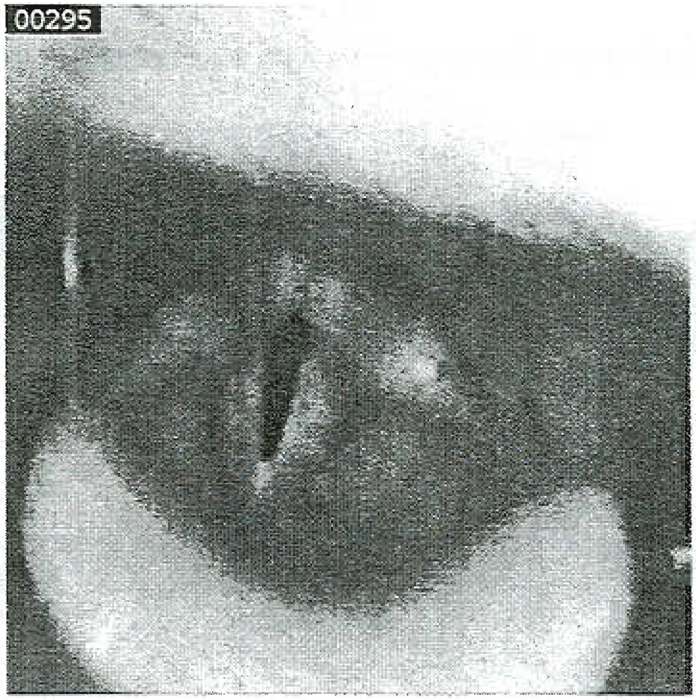


Figure 5. The zoom-window ripple-filtered view of field number 00295, shown in full unfiltered view in Figure 4.

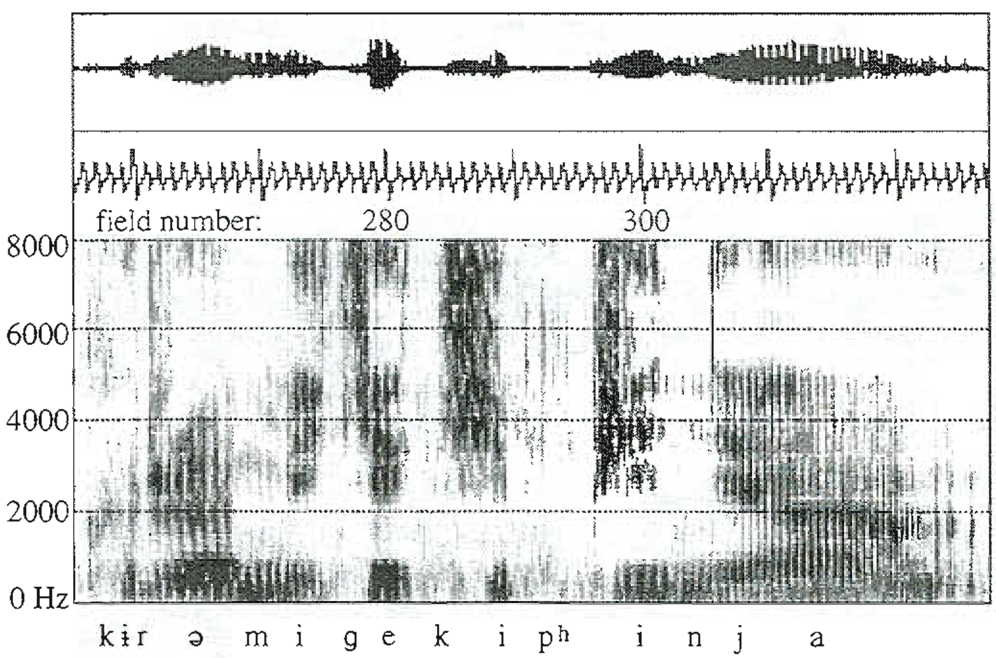
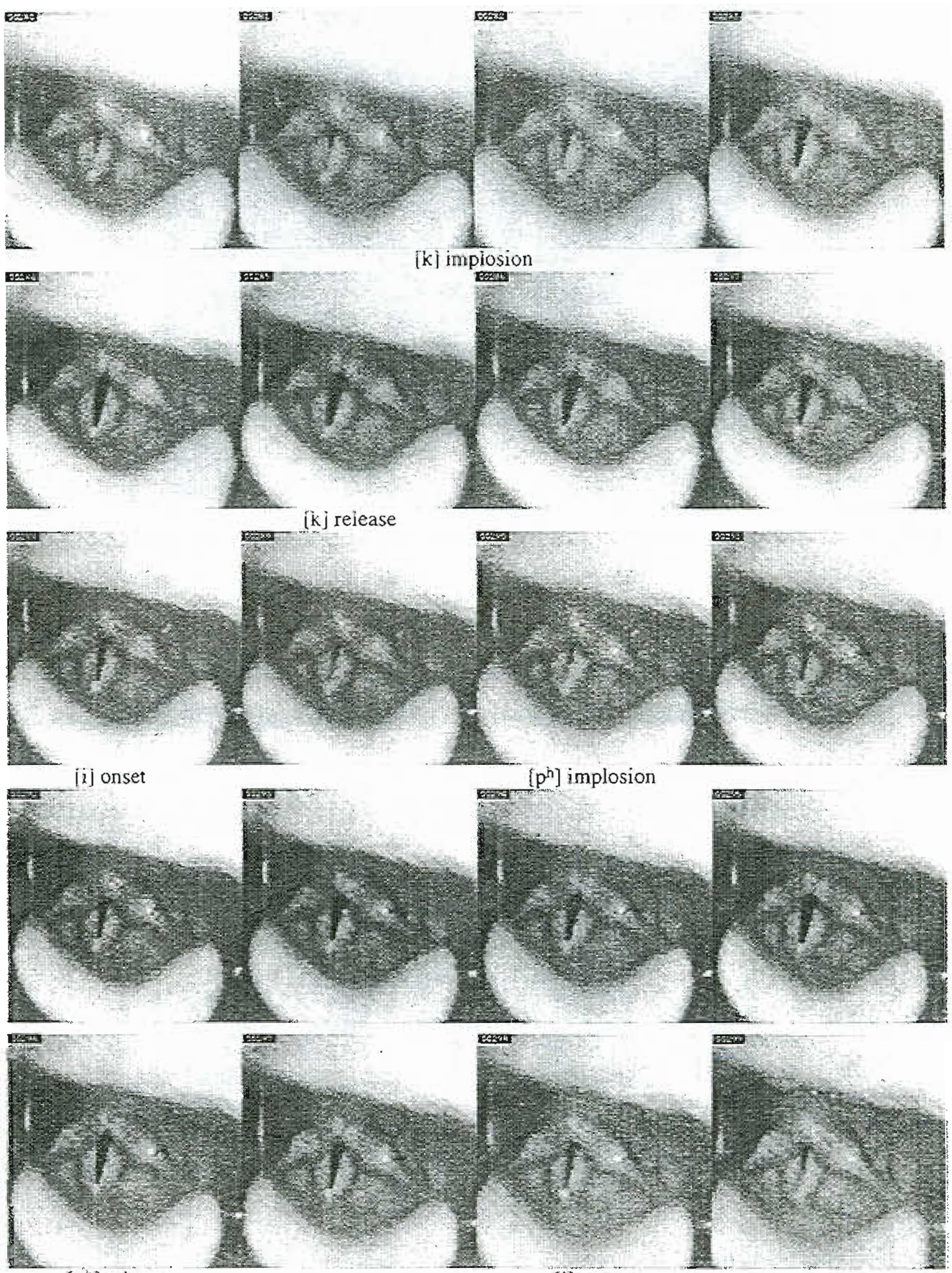


Figure 6. Audio waveform, timepulses, and spectrogram for the utterance shown in Figure 3 (above), and ripple-filtered zoom-window video views of the 20 fields starting about 2 fields (~30ms) before the implosion of the word-initial lenis stop to 1.5 fields (25ms) after the onset of voicing in the second syllable of the target word (see next page).

Fig. 6 (continue) Glottal view of /kip^{h}i/: from the field number 280 until 299.



Word: [kip^{h}i] -- voiced V1 [i], AP initial

At UCLA, the audio track of the videotape was digitized in stereo at 10 KHz per channel onto the hard drive of a PC using the Kay Elemetric CSL hardware. The speech audio signal was thus separated from the simultaneously recorded time pulse train, and could be viewed in synchrony with it using CSL display software. Spectrographic displays of the speech audio track could then be used to identify the timecodes for the video fields nearest to the voice onset time in the target vowel (if it was voiced) and to the closure and release times of each of the surrounding consonants. Figure 6 illustrates this identification process. The first part of the figure shows the audio waveform, timepulses, and spectrograph for the same utterance. The fields associated with the relevant acoustic events are then labeled in the second part of figure, which shows a sequence of 20 video fields starting 2 fields (~30ms) before the implosion of the word-initial /k/ and ending about 1.5 fields (25ms) after the onset of voicing in the vowel following the word-medial /p^h/.

3. Results and Discussion

Figure 7 shows the distribution of devoiced vowels in the three repetitions of the corpus, counting partially devoiced and completely devoiced vowels together as a single category in contradistinction to voiced vowels. Although there were fewer instances of devoicing overall, the relative distribution of devoiced vowels across the different consonantal contexts and different prosodic positions reproduced the results of Jun & Beckman (1994). The effect of the preceding consonant type was larger than the effect of the following consonant type, and the effect of prosodic position was largest for a preceding lenis stop. Comparing the different preceding consonantal contexts, we see the highest percentage of devoiced vowel tokens in the context of a preceding aspirated stop or preceding plain (or lenis) fricative, and almost no devoicing in the context of a preceding fortis stop or fricative. Comparing the different following consonantal contexts, we see more devoicings before an aspirated or fortis stop and fewer devoicings before a fricative, and among stops, we see less devoicing before lenis than others.

As in Jun & Beckman (1994), we can relate the effects of preceding consonant type to the published literature on glottal aperture (Kagaya, 1974; Hirose, Lee, & Ushijima, 1974), and to experiments on the effects of accentual-phrase position on the closure duration and relative length of residual voicing at the beginning of a lenis stop (see, e.g., Jun, 1990, 1996a; Silva, 1992; and compare the traces for the lenis stops in the left and right panels of Figure 1 from Kayaga's data above). That is, the degree of vowel devoicing is correlated with the degree of adjacent consonant's glottal opening. In the same prosodic position, vowels are more likely to be devoiced when the preceding consonant has large glottal aperture than when it has small glottal aperture. For the same consonant type, the following vowels are more likely to be devoiced when they occur accentual phrase-initially than accentual phrase-medially. Since consonants have a longer closure duration at accentual phrase initial position, we assume that the glottal aperture is bigger phrase initially than phrase medially. This result is what is predicted by the simple undershoot account. The differential effect of a following lenis stop as compared to the aspirated and fortis stops can be explained in part by the same prosodic facts. That is, a word-medial lenis stop is also necessarily accentual-phrase-medial, and thus will have a short closure which undergoes the lenis-stop voicing "rule" (see Jun, 1995, 1996a, for the phonological status of this phenomenon).

The differential effect of a following fricative as opposed to a following aspirated or fortis stop is somewhat more difficult to interpret in terms of the literature on glottal aperture, but this effect might be related to the time course of the oral constriction gesture. That is, studies on kinematics of oral articulations often show faster, more ballistic movements into stops than into sibilant fricatives (e.g., Kuehn & Moll, 1976). Given the transglottal airflow requirements for voicing, and the quicker buildup of oral air pressure from the closure into the stop, glottal vibration may cease sooner in a vowel-stop sequence than a vowel-fricative sequence. Alternatively, it may be the case that the glottal abduction gestures for these two stop types differ from the gestures for the fricatives in ways that do not show up on glottal aperture traces. That is, whereas the plain and fortis fricatives only contrast with each other, the aspirated and fortis stops contrast also with the lenis stop. This three-way contrast may provide an impetus to trigger an active increase in vocal fold tension just before the implosion, to prevent residual voicing into the stop closure of the sort that is observed typically in word- and phrase-medial lenis stops. (A study of English stop production shows elevated cricothyroid activity for some speakers in their voiceless stop productions which was explained in this way — see Löfqvist, Baer, McGarr & Seider Story, 1989.) This suggests that vowel devoicing is not mere undershoot of the glottal target opening. That is, it is not a simple linear function of degree of overlap between adductive and abductive glottal gestures. Rather its causes are a combination of the degree of glottal overlap, the aerodynamic conditions resulting from the time course of the overlayed oral constriction gesture, and perhaps even the differential tuning of the gestural "target" in response to the inventory of contrasting types in each particular subsystem of the language's consonant system.

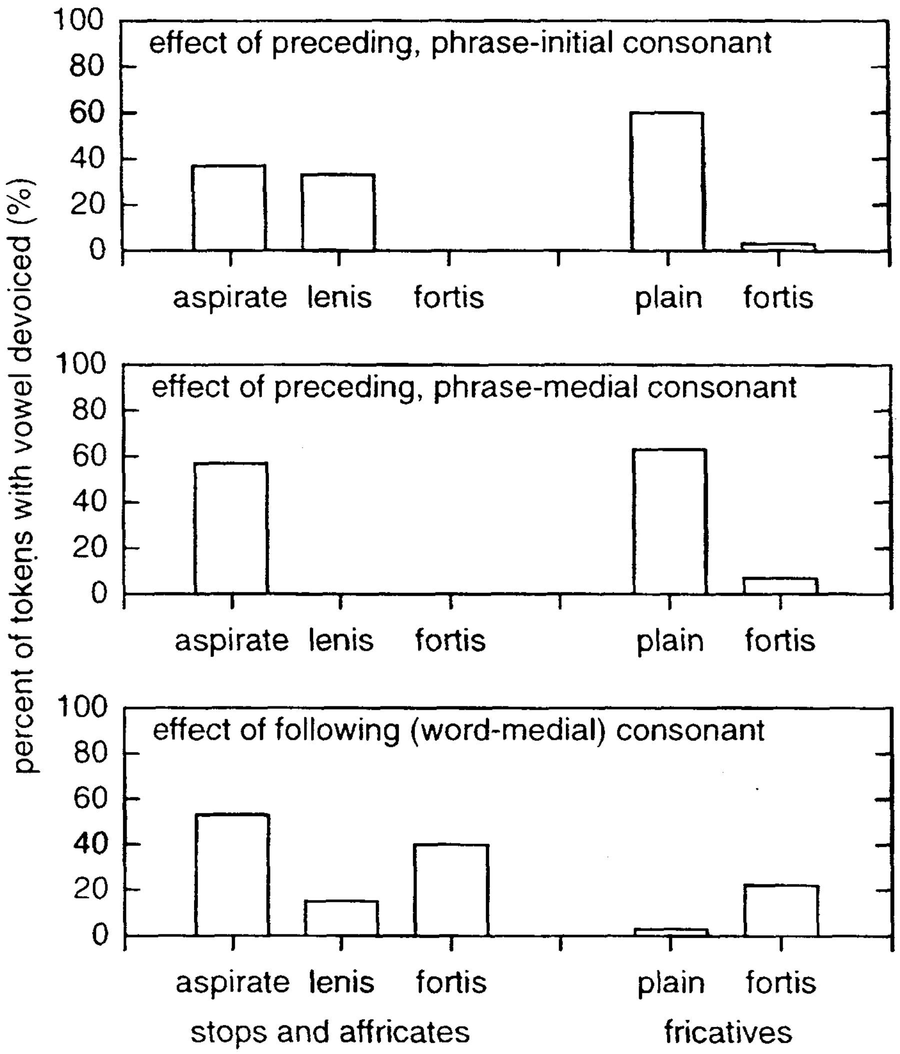


Figure 7. Percentage of tokens in which the vowel is completely or partially devoiced in the different contexts provided by the corpus of dialogues. N is 30 tokens for the top two panels and 60 tokens for the bottom panel.

Observations of the glottal video fields lend further support to the notion that vowel devoicing in Korean is not categorical. Even though we had digitized movie sequences for only the first set of repetitions of the 25 word types, there was considerable variability in the time course of glottal closing and opening over the CVC sequence relative to the associated oral gestures in the vowel and flanking consonants. That is, there were not two qualitatively different patterns for the voiced versus devoiced vowel tokens in these utterances, or even two clearly differentiated regions in any distribution of tokens along some natural continuum of glottal adduction patterns. Figure 6 above and Figures 8-12 illlustrate the range of patterns that we saw.

Comparing Figure 6 with Figure 8 shows one type of difference that we saw between voiced and devoiced tokens of the same word type. Where Figure 6 shows the spectrogram and video field sequence for [kip^{h}i] 'evasion' with a voiced first [i], Figure 10 shows the sequence of video fields for a token of the same word with a partially devoiced [i] in the same prosodic context, at the beginning of the accentual phrase. In both examples, the glottis is maximally open around the release of the [k] and only gradually adducts over the next few fields, up to the field where glottal pulsing was first observed in the [i]. This is the pattern one might expect for an aspirated stop. Indeed, in both of these tokens, the glottal opening pattern at and just after the release of the phrase-initial [k] is very similar to the pattern at and just after the release of the word-medial [p^h]. Earlier studies of Korean stops (e.g., Han & Weitzman, 1970; Dart, 1987) have shown that many speakers do produce a short interval of breathy-voiced "aspiration" in phrase-initial lenis stops; hence the term "slightly aspirated" in many older descriptions of the contrast, such as Martin (1951). The two examples are also similar in the time course of glottal abduction for the postvocalic [p^h]. Beginning one or two fields after the oral implosion, the glottis gradually opens to achieve maximum adduction in the field just before the [p^h] release. The difference between the two tokens is only a very subtle matter of the length of interval between the initiation of glottal adduction from the [k] into the [i] and achievement of full glottal abduction for the [p^h]. In the partially devoiced token, this sequence is accomplished within 9 video fields, whereas in the fully voiced token it covers 11 fields. Table 3(a) shows the duration (measured in number of fields) of the intervals between acoustic landmarks commonly used to measure stop and vowel duration. Given the 60 fps sampling rate of the video, the difference is one of about 30 ms. As a result of this shorter duration, the glottal abduction gesture in the following [p^h] (which is timed relative to the stop's release into the following vowel) begins much earlier relative to the peak glottal abduction in the preceding [k]. The partial devoicing might be explained entirely in terms of the shorter duration of the target vowel and consequently greater overlap between the end of the glottal adduction gesture for the vowel after a preceding voiceless stop and the onset of the glottal abduction gesture for the following stop. That is, the movie sequence here is compatible with an interpretation in terms of undershoot resulting from greater temporal overlap among a sequence of invariant glottal gestures.

The pattern in Figure 9, by contrast, is not so easily explained. This figure compares two tokens of [p^{h}us'i] 'raw seed' in accentual-phrase medial position. Again, one token shows a fully voiced and one a partially devoiced target vowel. However, the difference in duration is not as large. The entire CVC sequence is 18 fields in both cases, and the interval between the maximum glottal opening at the release of the [p^h] and the release of the [s'] into the following vowel differs by only a half a field -- 14 versus 13.5 fields in Table 3(b). Note, however, that the maximum glottal opening in the [s'] is not timed relative to the consonant offset in this fortis fricative. Rather, the glottis opens much more quickly to provide the airflow necessary for turbulent flow at the oral stricture. In both tokens, maximum glottal abduction is reached one or two frames after the onset of frication for the [s']. The relevant interval duration in this word type, therefore, is not the time from [p^h] release to [s'] offset. Nor can it be the time from [p^h] release to maximum glottal opening in the [s'], which hardly differs. Thus, temporal overlap and undershoot cannot be the entire explanation of the partial devoicing here. Note, however, that the first field showing maximal abduction in the [s'] occurs immediately after the frication onset in the token with the fully voiced vowel, but not until one field later in the other token. This suggests a somewhat more gradual abduction out of the partially devoiced vowel. That is, there could be a small difference in "settling time" for the glottal abduction gesture. Suppose that the offset of voicing depends not on the time of maximal glottal opening, but rather on the time when there is a degree of abduction sufficient for airflow through the glottis to make an air pressure differential across the oral constriction and cause turbulence there, rather than an air pressure differential at the glottis that cause vibrations of the vocal folds. This "sufficient degree of abduction" seems to be achieved earlier relative to the [p^h] release in the token with the partially devoiced vowel. The [s'] onset is 3.5 fields after [p^h] release rather than 5 fields later, as in the token with the fully voiced vowel. In other words, while the contrast here cannot be explained in terms of overlap of invariant gestures, it still seems to be a case of undershoot. That is, one of the glottal gestures varies in more than relative onset time, but the relevant gestural reorganization still could be a temporal one -- an increase in "settling time" for the glottal abduction gesture of the following consonant which results in greater overlap and undershoot of the adduction gesture in the target vowel.

Figures 10-11 show another way in which the relative timing of the glottal gestures can vary to produce different degrees of devoicing in the target vowel. The figures show the spectrogram and the relevant sequence of glottal video fields for each of the four tokens of the word [p^hup'e] 'an unripe pear' which together represent four different points in the voicing continuum. The fully 'voiced' token in Figure 10(a) has a full vowel, as indicated by the energy at the first and second formant frequencies. The first 'partially devoiced' token in Figure 10(b) shows three glottal pulses with very weak energy only around the second formant of the vowel. The second 'partially voiced' token in Figure 10(c) shows only one clear glottal pulse with some weak energy around the second formant during the aspiration noise. Finally, the 'completely devoiced' token in Figure 10(d) shows no trace of voicing, but only aspiration noise.

Unlike in Figures 6 and 8 (for [kip^{h}i]) or Figure 9 (for [p^hus'i]), however, the overall duration of the CVC sequence is roughly constant across the different voicing patterns. Moreover, the general characterization of the timecourse of glottal closing and opening does not change across the continuum from the fully voiced token in Figure 11(a) to the fully devoiced token in Figure 11(d). There is always a large glottal opening during the word-initial aspirated stop [p^h], and a gradual narrowing to the smaller opening at the beginning of the word-medial tense stop [p']. The maximum glottal opening at the release of the [p^h] looks to be larger in the token with a completely devoiced vowel in Figure 11(d) than in the first partially devoiced token in Figure 11(b), but it is not larger than in the fully voiced token. Moreover, there is some glottal opening at the voice onset in the vowel for all four tokens, whether the token is judged to be 'partially devoiced' or fully 'voiced', and the peak glottal opening during the preceding [p^h] is very similar between voiced and completely devoiced category. Therefore, it is not possible to predict the acoustic voicing category just from the overall duration and from such qualitative observations of the pattern of glottal opening across the the word-initial CVC sequence. The relevant difference among these four tokens seems rather to be in the relative timing of acoustic events such as consonant closure and release. The occurrence of one or another of the four different voicing categories seem to be related to the durations of the flanking consonants' closure and release gestures. That is, as the duration of the vowel (the number of fields recorded during the vowel's voiced interval) decreases, the duration of the flanking consonants' closure and release intervals increases, as shown in Table 3 (c). As the vowel becomes less and less voiced, the number of fields recorded during the vowel decreases from 2.3 to 0 while the number of fields during the preceding consonant's closure and release increases from 6 to 8. This suggests that the vowel devoicing in the token in Figures 10(d) and 11(d) is a result of the aerodynamic conditions caused by the very large overlap between the glottal adduction gesture for the vowel and the oral closure and release gestures of the flanking consonants. It is not as simple a case of undershoot as in Figure 6 versus Figure 8, but it is nonetheless compatible with an undershoot account, once we acknowledge that there can be variability in the relative durations of the oral gestures independent of the timing of the glottal gestures for the vowel and neighboring consonants.

Table 3. Number of video fields for closure (from implosion to explosion) and release (from consonant release to vowel onset) of the preceding consonant, vowel (V1), and for the closure of the following consonant. For [kip^{h}i] and [p^{h}us'i], two voicing categories (voiced, partially devoiced) are shown, whereas for [p^{h}up'e] data are for four different degrees of devoicing (voiced, partially devoiced-I, partially devoiced-II, completely devoiced). See Figures 6, 8, 9, and 11.

(a) [kip^{h}i]

[kip^{h}i]	voiced	part. devoiced
k closure	3.5	3
k release	3	2.5
vowel	2	1.5
p^h closure	6	5

(b) [p^{h}us'i]

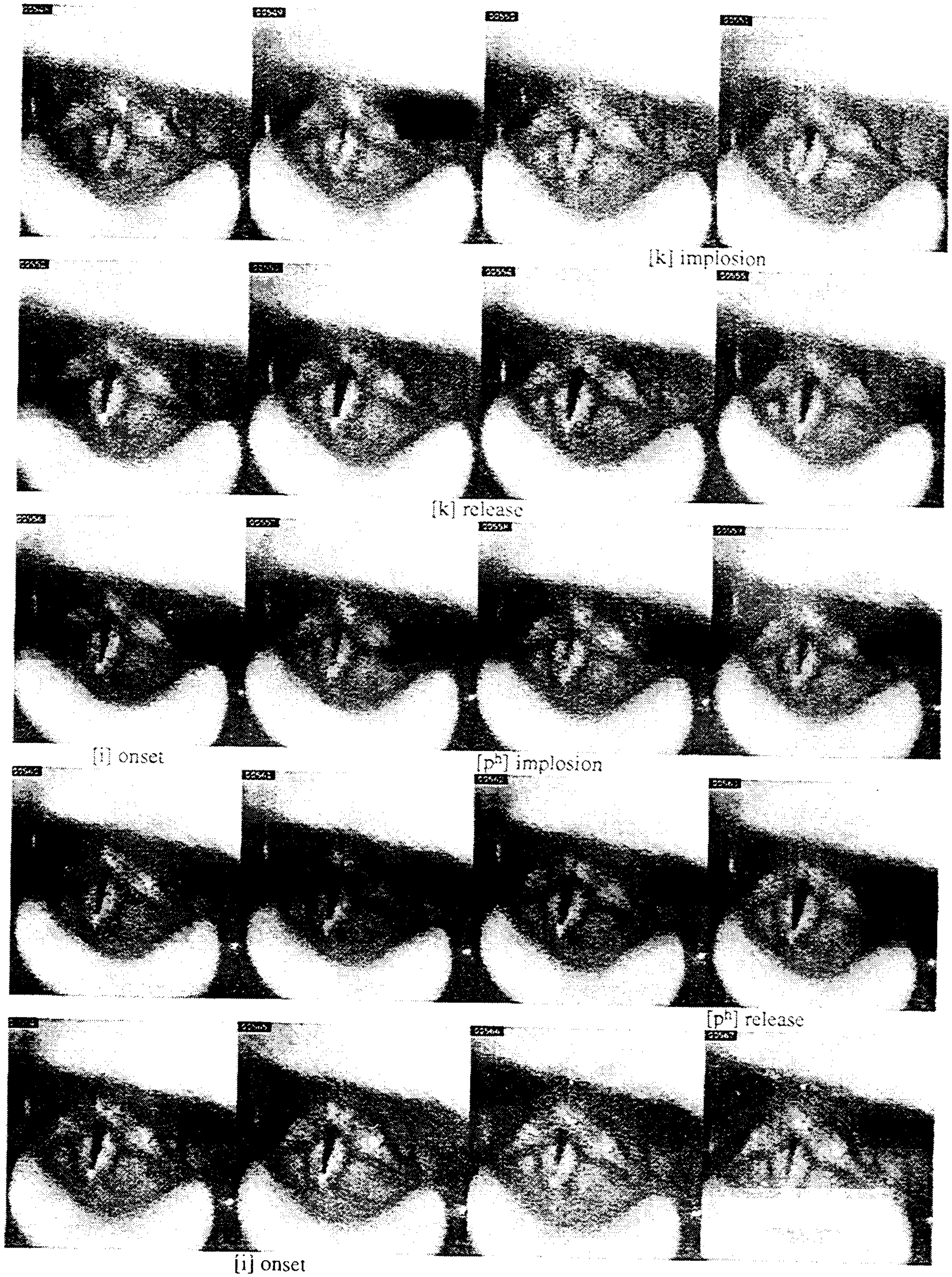
[p^{h}us'i]	voiced	part. devoiced
p^h closure	4	4.5
p^h release	3	2.5
vowel	2	1
s' frication	9	10

(c) [p^{h}up'e]

[p^{h}up'e]	voiced	partially devoiced-I	partially devoiced-II	completely devoiced
p^h closure	4	3.3	3.8	4.5
p^h release	2	2	3	3.5
vowel	2.3	2	0.3	0
p' closure	7	6.5	7.3	7.8

However, not all examples of devoicing can be explained just in terms of gestural timing and undershoot. Figures 12 and 13 show spectrograms and movie sequences for two tokens of [sigje] 'clock'. The phrase-initial token has a fully voiced and the phrase-medial token has a fully devoiced vowel, so much devoiced that the spectrogram shows continuous sibilant energy up to the sudden reduction in the amplitude of the frication noise between 4000 and 5000 Hz at the onset of the rather "leaky" closure of the following voiced lenis stop. The following lenis stop is voiced in both cases, so the difference in the vowel should be due to the gestural configuration of the initial CV sequence. This CV sequence is considerably longer in the token with the fully voiced vowel. However, this cannot explain the difference, since the maximum glottal opening in the preceding [s] is considerably larger in the phrase-initial token, so that full glottal adduction is not achieved any earlier relative to the closure into the following lenis stop. Indeed, the glottal opening during the two fields of the voiced vowel is if anything somewhat larger than the last two fields at the end of the frication noise just before the [gj] implosion in the token with no voicing. Apparently, the difference between these two tokens has to do with the greater air pressure in the phrase-initial position. The larger subglottal airpressure for the high tone at the beginning of the focused phrase causes breathy-voiced vibration for the vowel, even though the glottis is somewhat open here. Figure 14, which shows the intraoral air pressure in the two tokens, supports this interpretatoin of the glottal video fields. In short, observations of the sequence of glottal views during voiced and devoiced vowels in different tokens of the same word support Hirose's (1971) suggestion that the observed variation in glottal state may indicate not a "difference in the motor command" but rather a "'passive' but highly nonlinear effect" of the interaction with varying "physical conditions near the level of the glottis", such as a difference in transglottal flow due to the higher subglottal airpressure associated with the production of the associated high tone at the beginning of the focused accentual phrase.

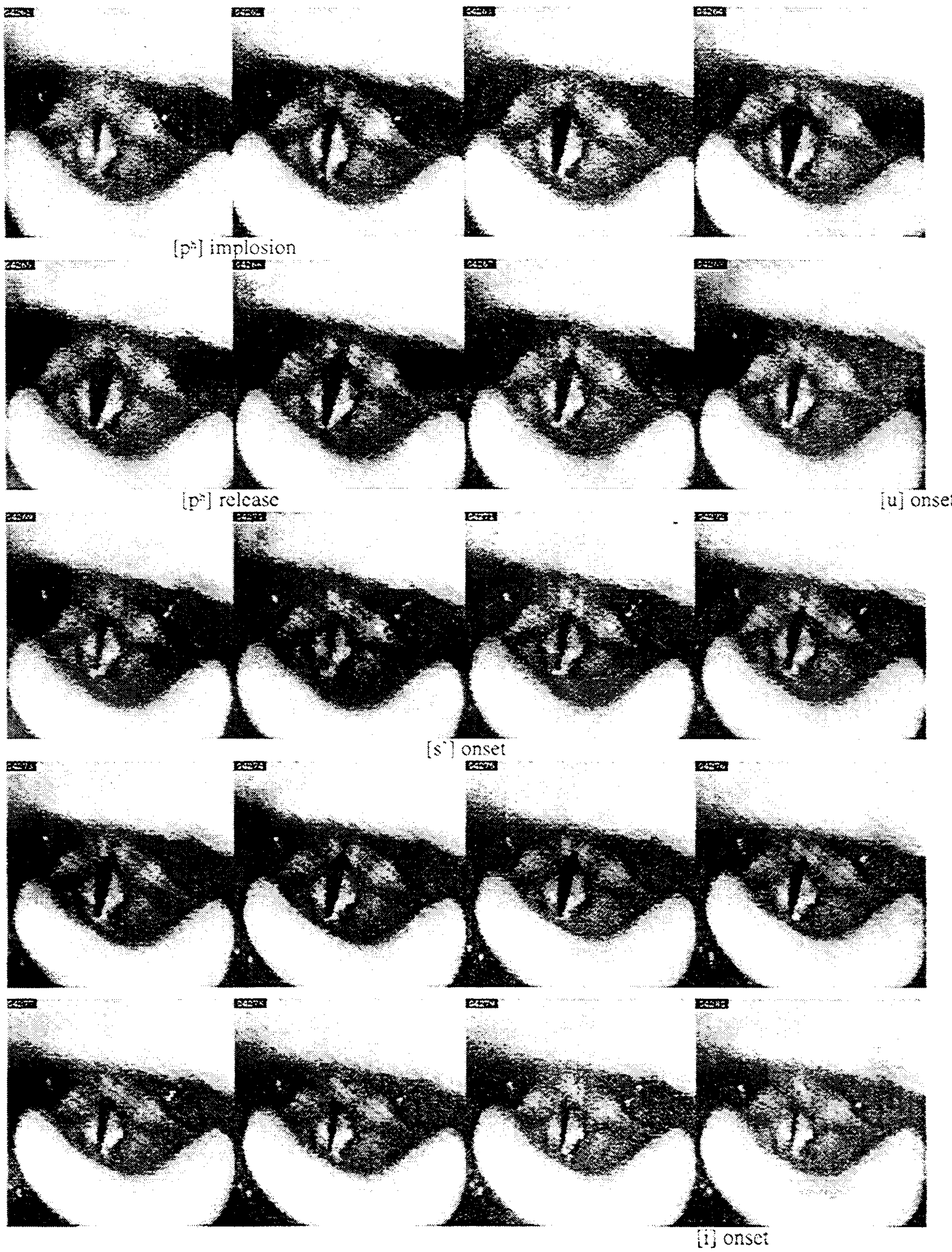
Figure 8. Video fields of glottal width pattern of word [kip^{h}i] 'evasion', when the first vowel in the target word is partially devoiced. This token is also produced in an accentual phrase initial position. (field numbers from 548 until 567)



Word: [kip^{h}i] -- partially devoiced [i]. AP initial

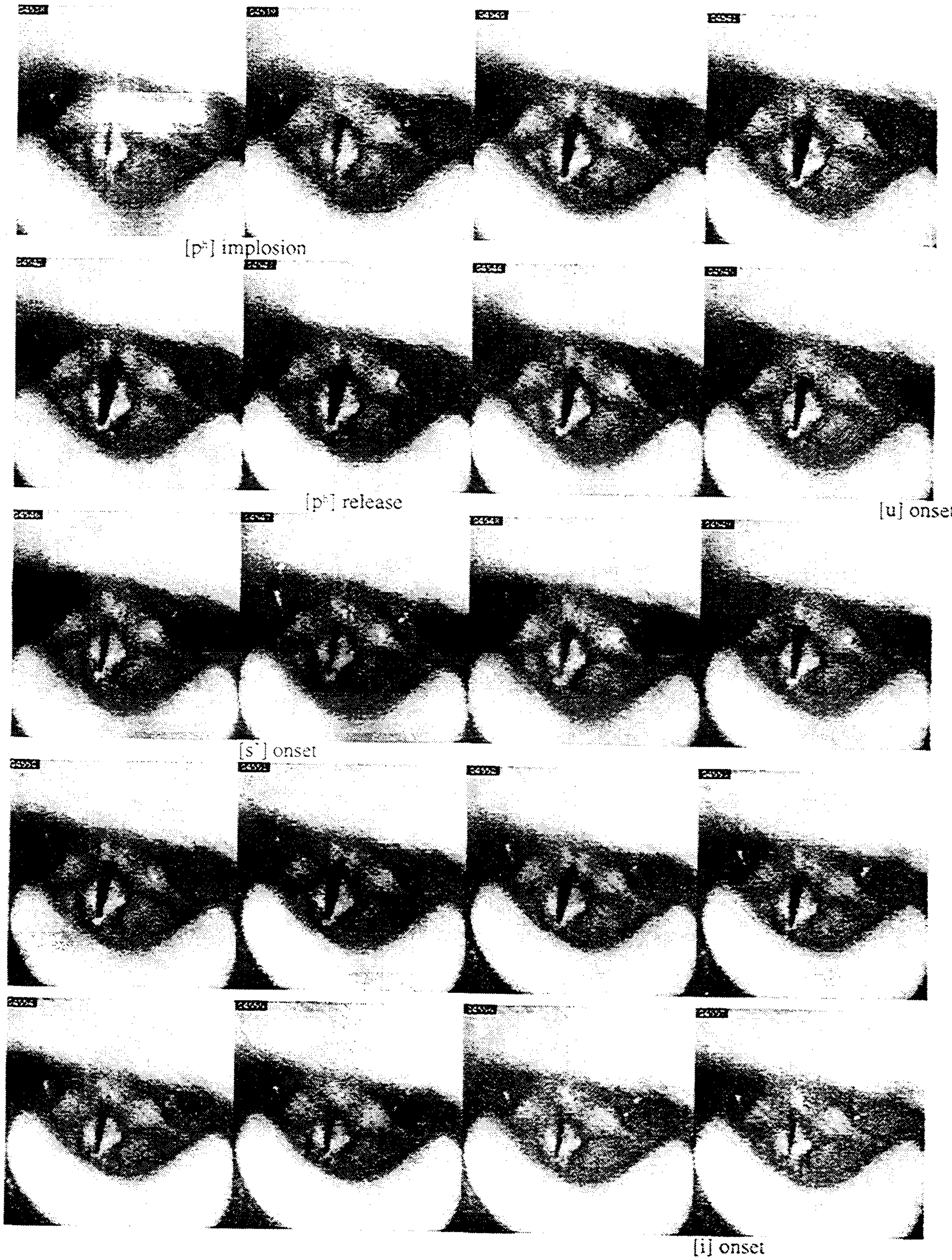
Figure 9. Video fields of glottal width pattern of word [p^{h}us'i] 'raw seed', when the first vowel in the target word is (a) voiced and (b) partially devoiced. Both tokens were produced in an accentual phrase medial position

(a) [p^{h}us'i] with a voiced [u] (field numbers from 4261 until 4280)



Word: [p^{h}us'i] -- voiced [u], AP medial

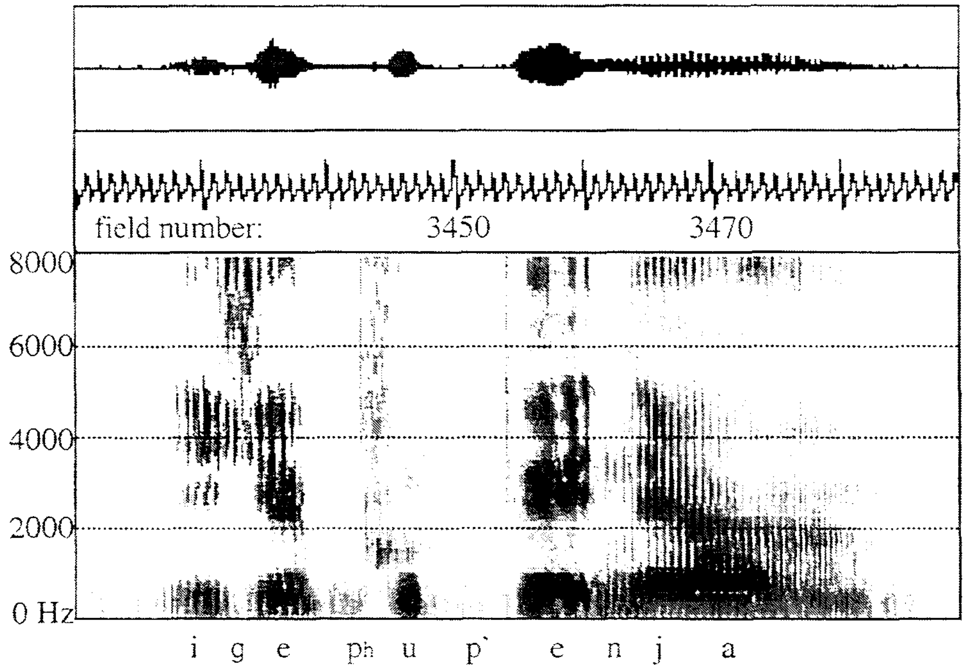
(b) [p^hus'i] with a partially devoiced [u]
(field numbers from 4538 until 4557)



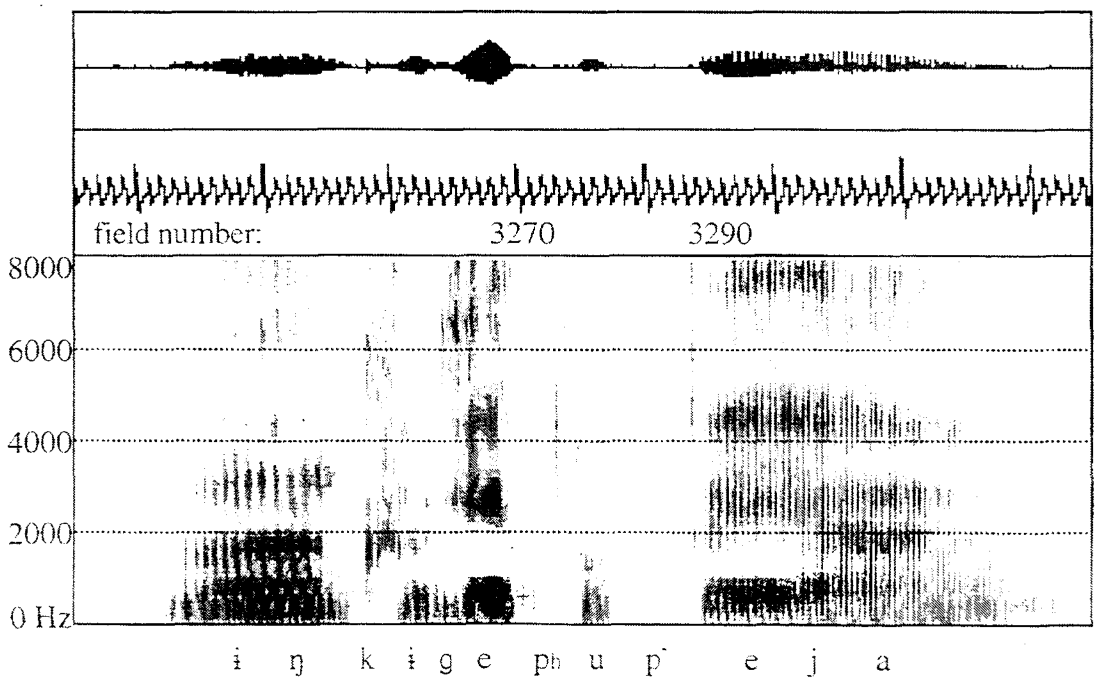
Word: [p^hus'i] -- partially devoiced [u], AP medial

Figure 10. Waveform, timepulses, and spectrogram of [p^hup'e] showing four categories in a continuum: (a) voiced [u], (b) partially devoiced [u] (labeled as partially devoiced-I), (c) partially, but close to completely, devoiced [u] (labeled as partially devoiced-II), and (d) completely devoiced vowel ([u̥]).

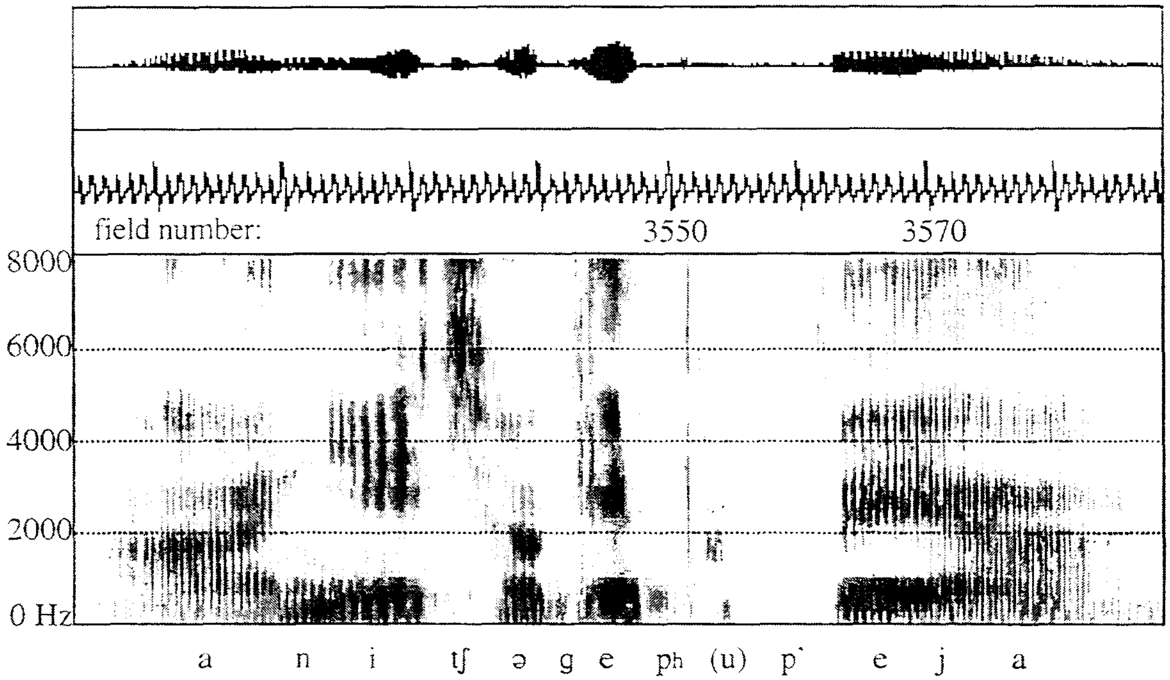
(a) a voiced token of [p^hup'e]



(b) a partially devoiced-I token of [p^hup'e]



(c) a partially devoiced-II token of [pʰup'e]



(d) a completely devoiced token of [pʰup'e]

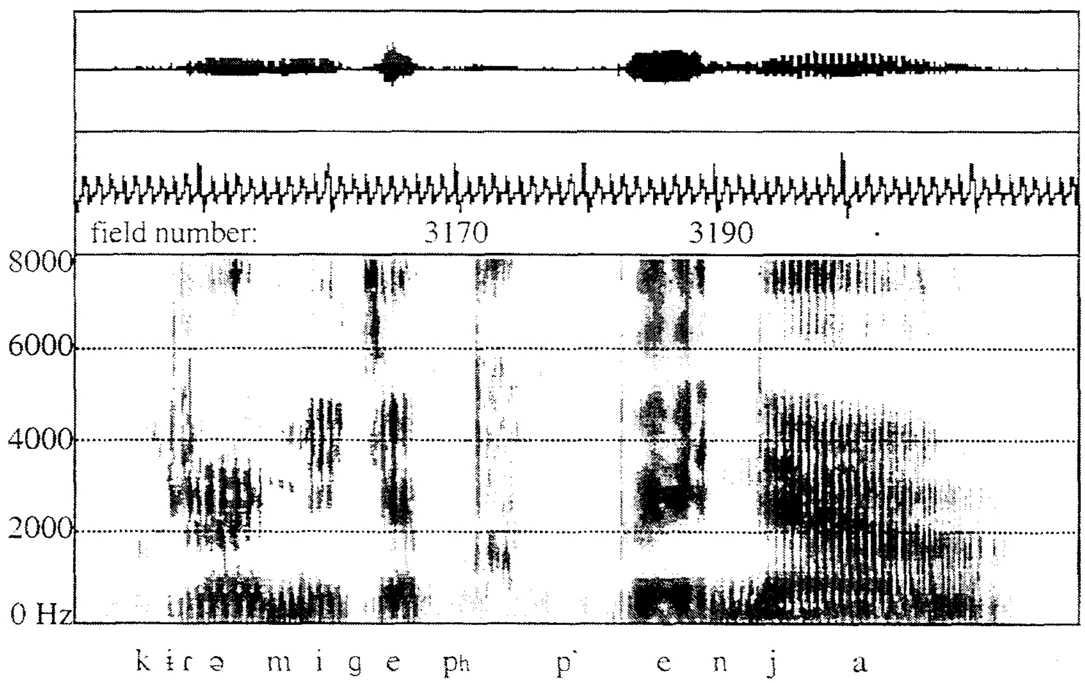
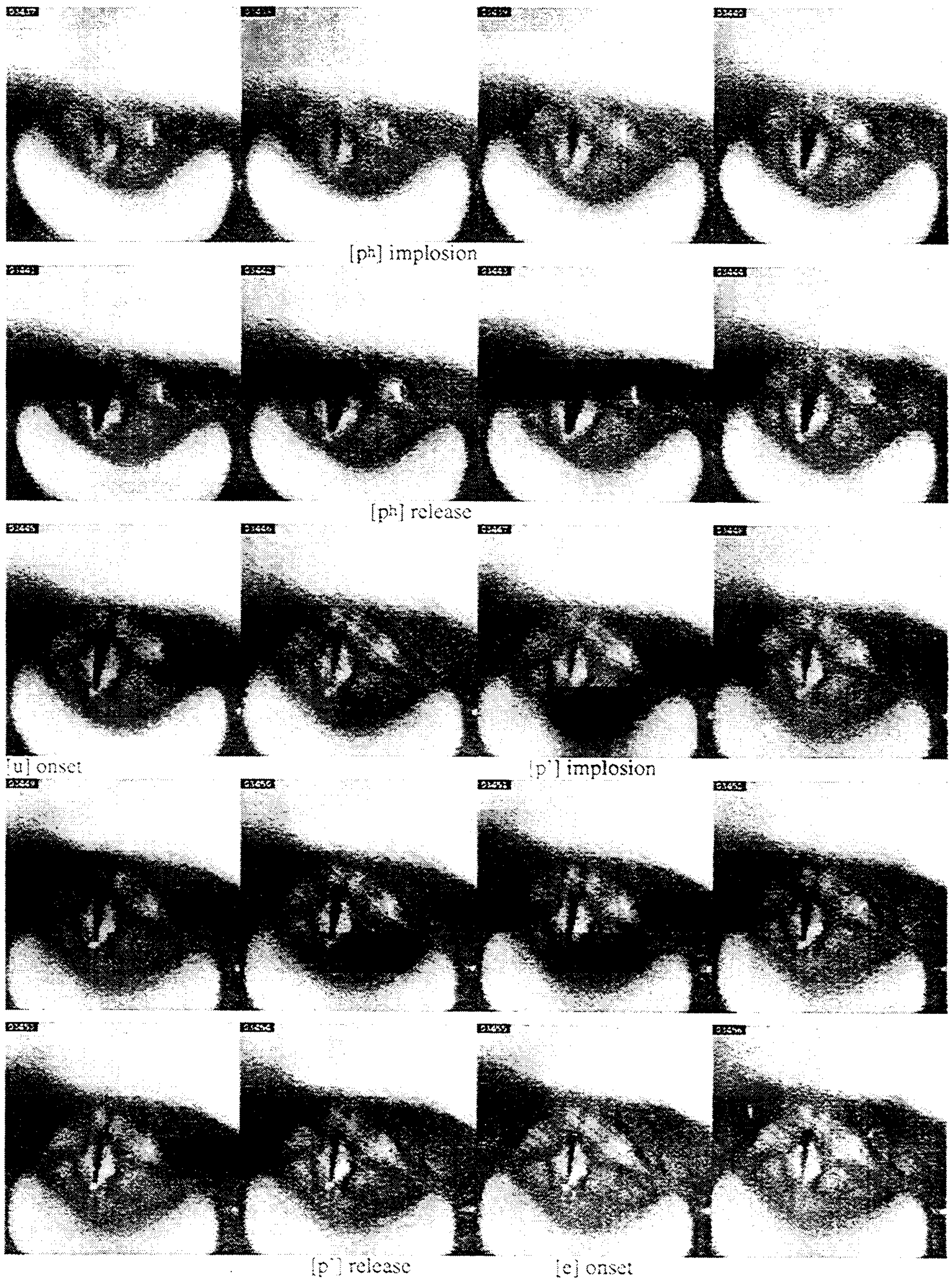


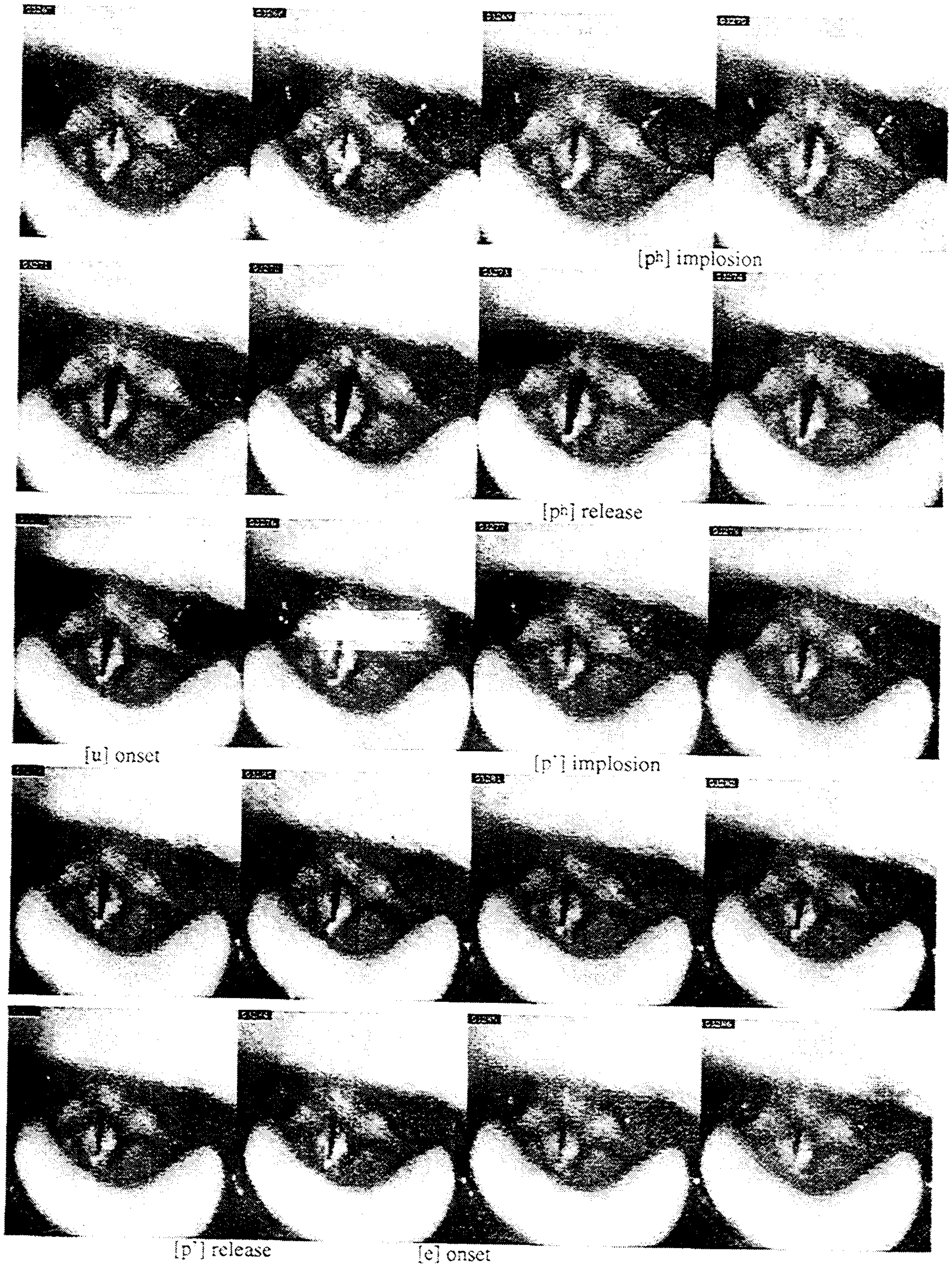
Figure 11. Video fields of glottal width pattern corresponding to each example in Figure 8: (a) voiced, (b) partially devoiced-I, (c) partially devoiced-II, and (d) completely devoiced vowel [u] in [pʰup'e].

(a) voiced (field numbers from 3437 until 3456)



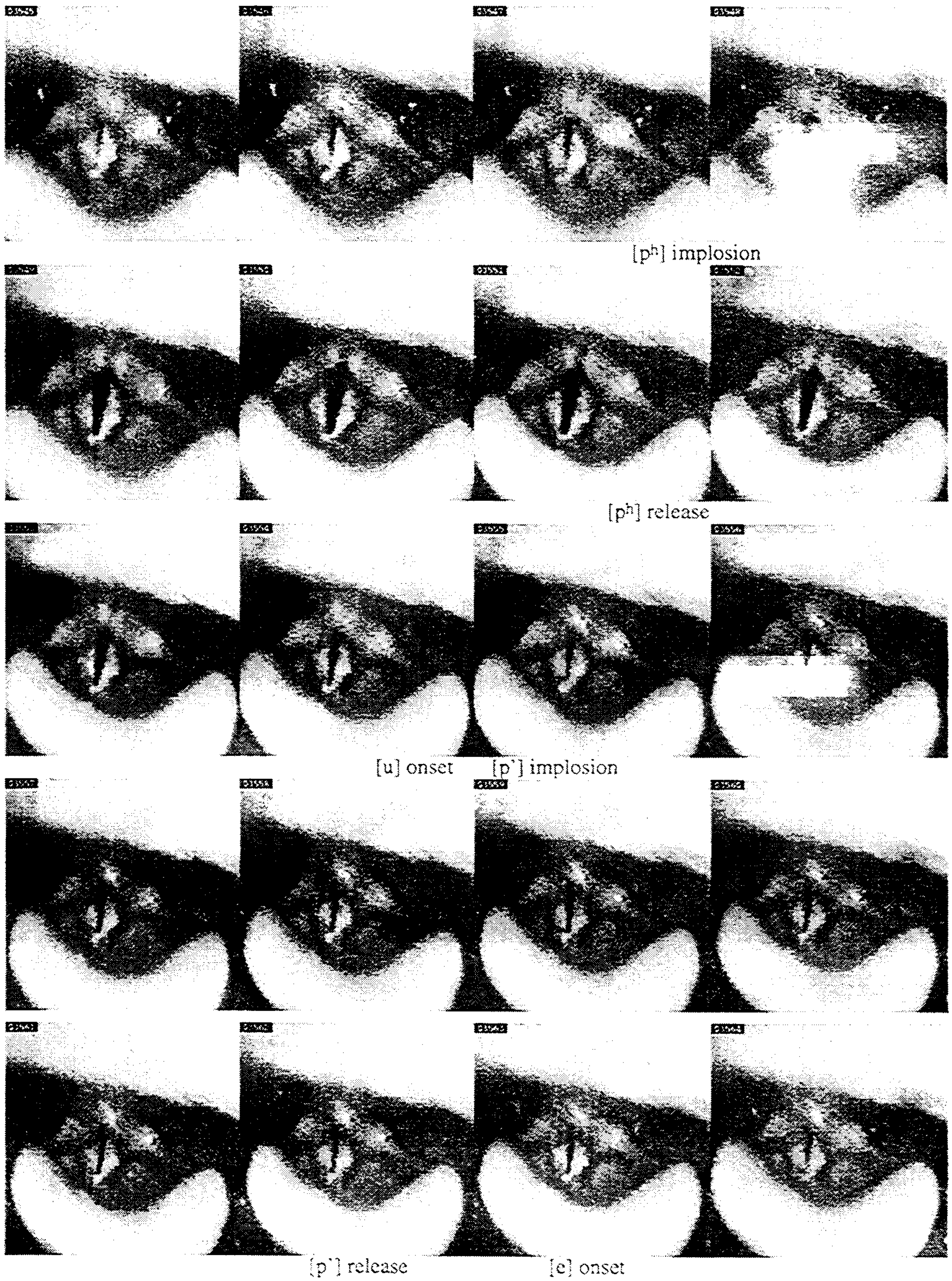
Word: [pʰup'e] -- voiced [u], AP initial

(b) partially devoiced-I (field numbers from 3267 until 3286)



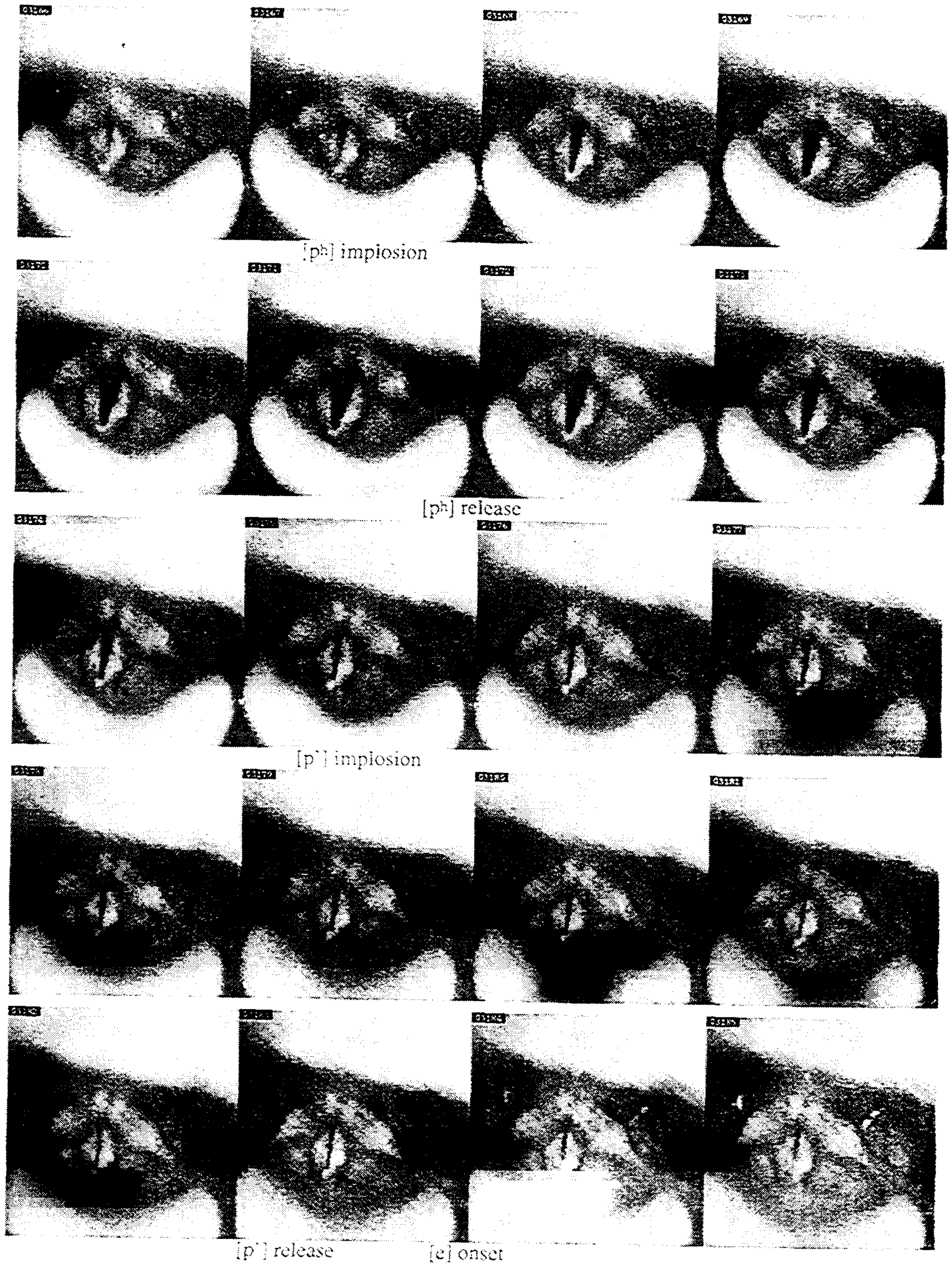
Word: [pʰup'e] -- partially devoiced [u], AP medial

(c) partially devoiced-II (field numbers from 3545 until 3564)



Word: [pʰup'e] -- partially devoiced [u], AP medial

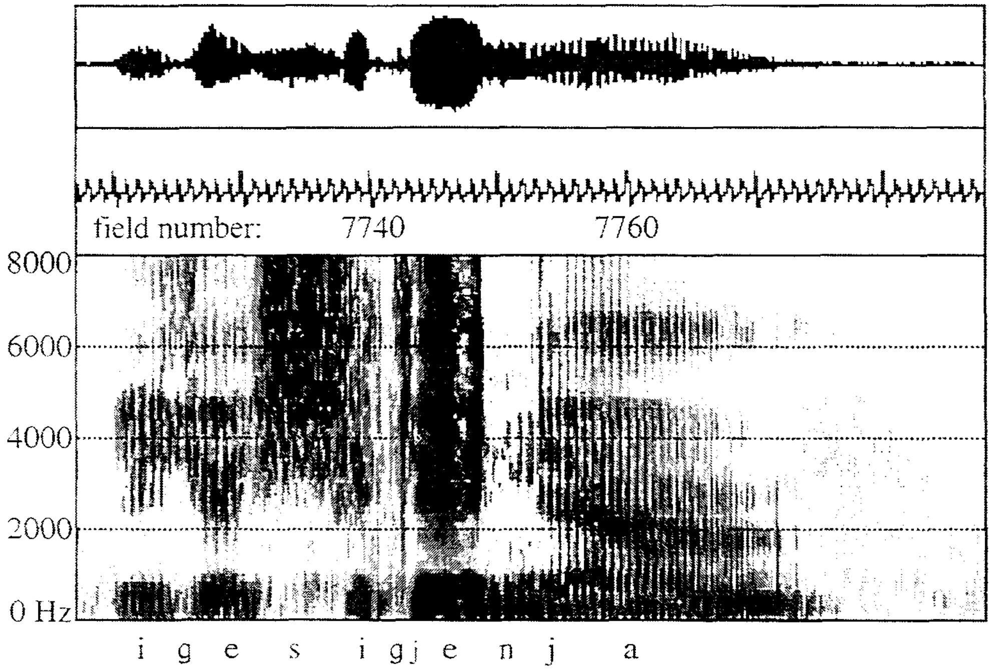
(d) completely devoiced (field numbers from 3166 until 3185)



Word: [pʰup'e] -- completely devoiced [u], AP initial

Figure 12. Waveform, timepulses, and spectrogram of [sigje] ‘clock’ corresponding to each token in Figure 12. (a) when the first vowel in the target word is voiced and (b) completely devoiced.

(a) [sigje] with a voiced [i] — in accentual phrase initial position



(b) [sigje] with a completely devoiced [i] — in accentual phrase medial position

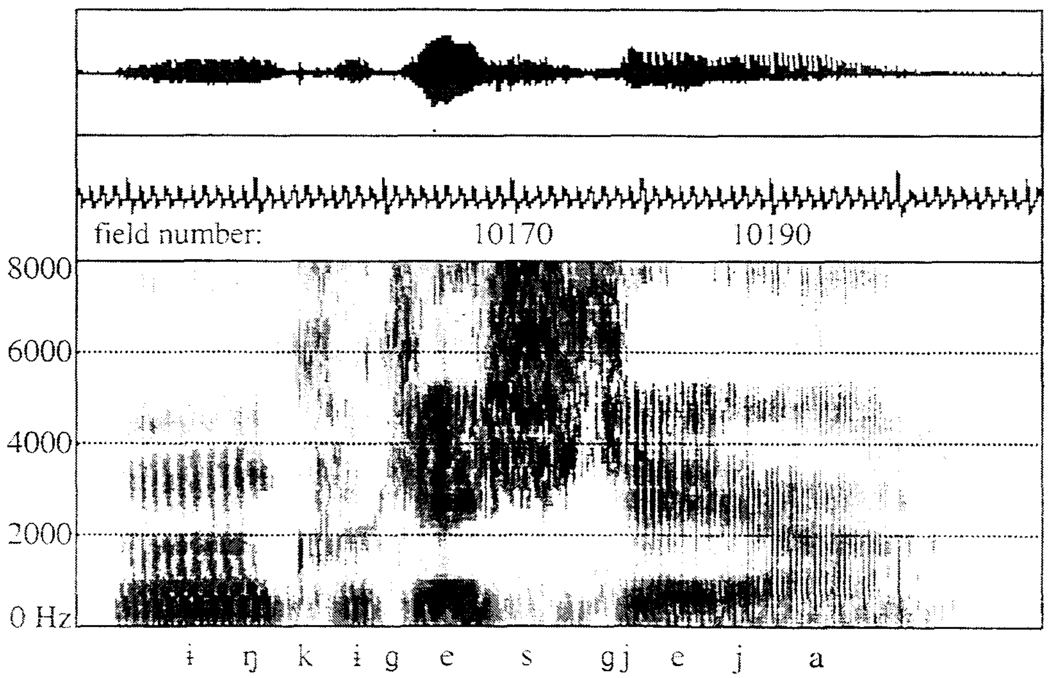
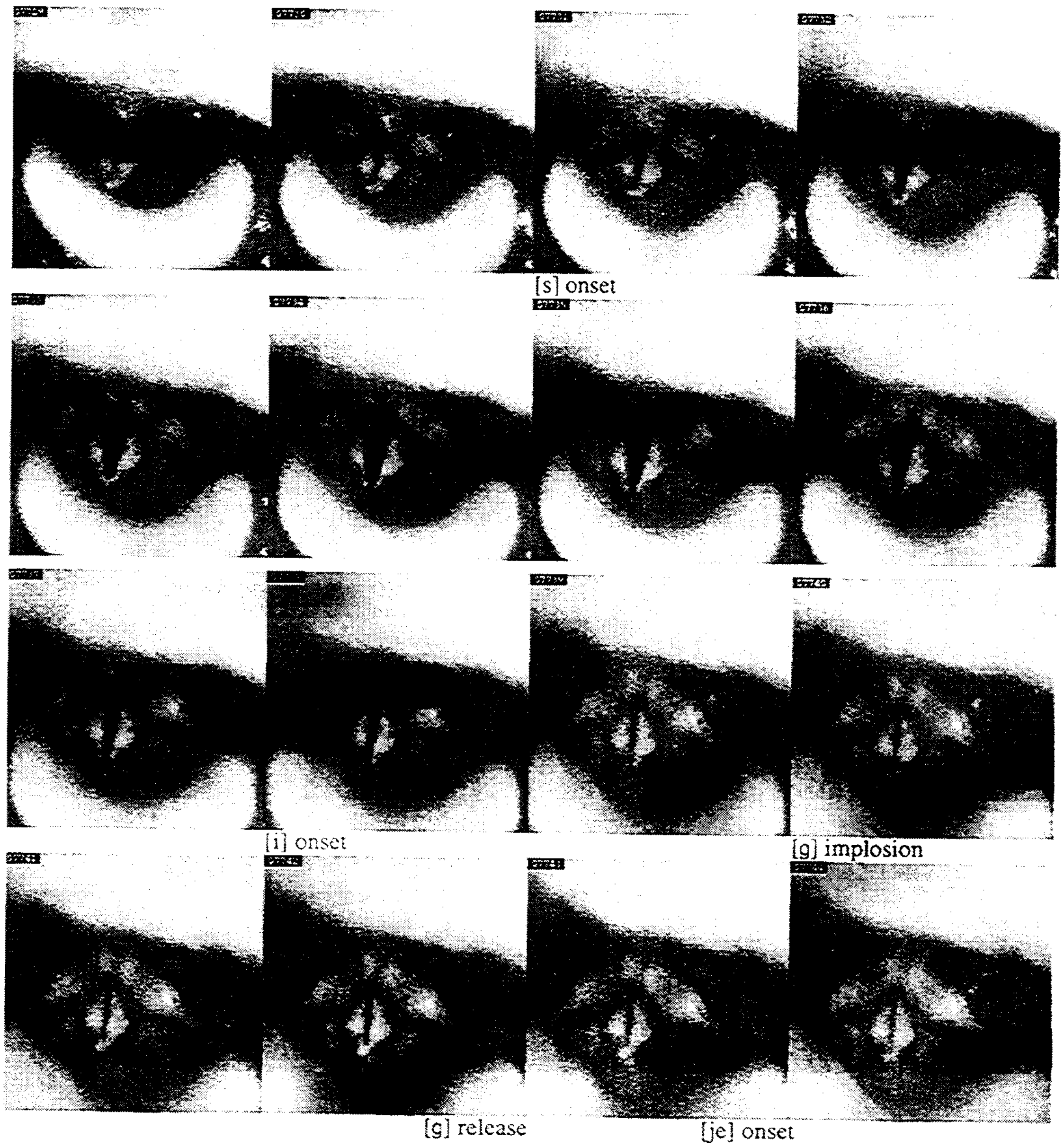


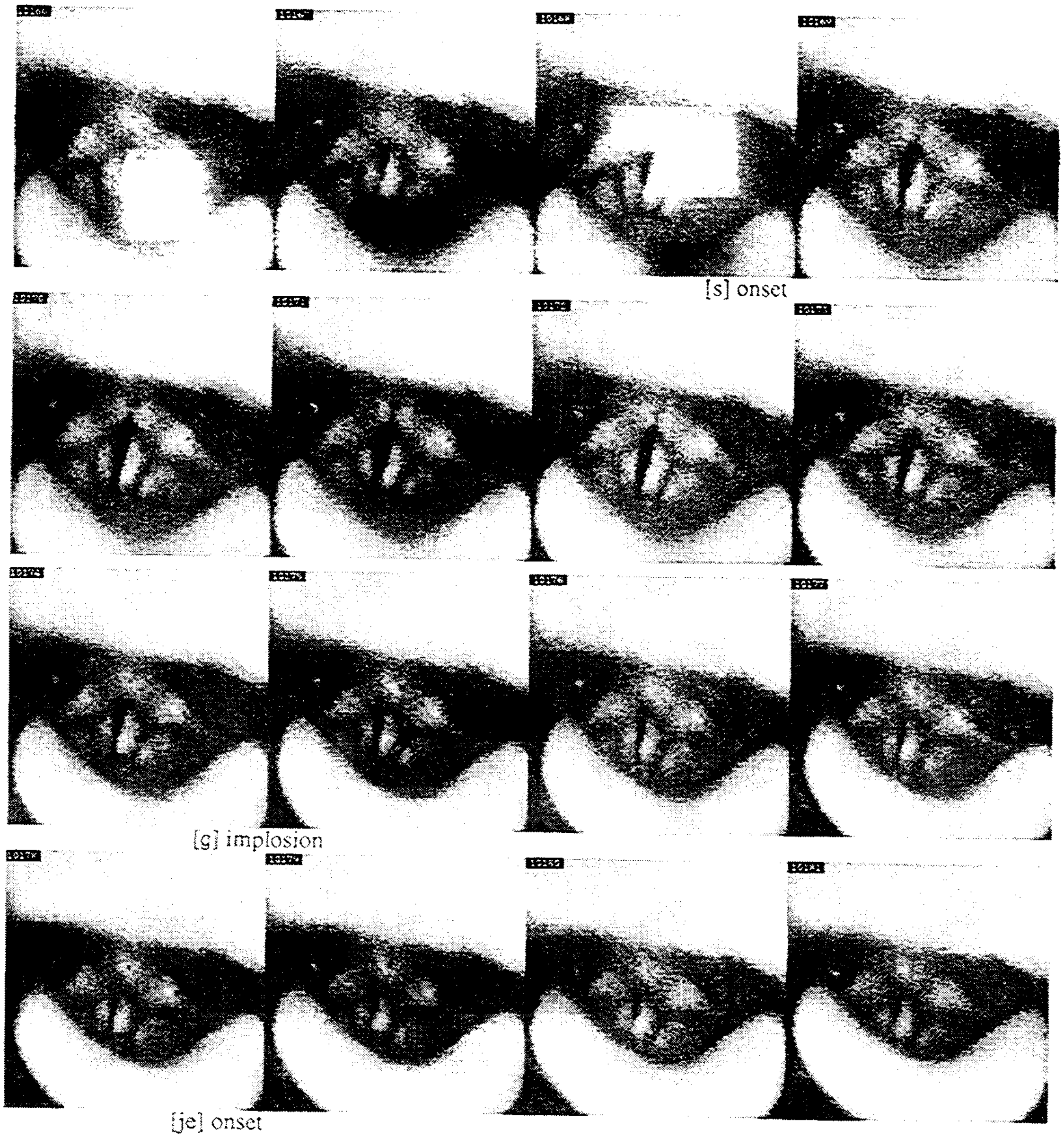
Figure 13. Video fields of glottal width pattern of word [sigje] 'clock', when the first vowel in the target word is (a) voiced and (b) partially devoiced. The target word in (a) is produced accentual phrase-initially while that in (b) is produced accentual phrase-medially.

(a) [sigje] with a voiced [i] — in accentual phrase initial position
(field numbers from 7729 until 7744)



Word: [sigje] -- voiced [i], AP initial

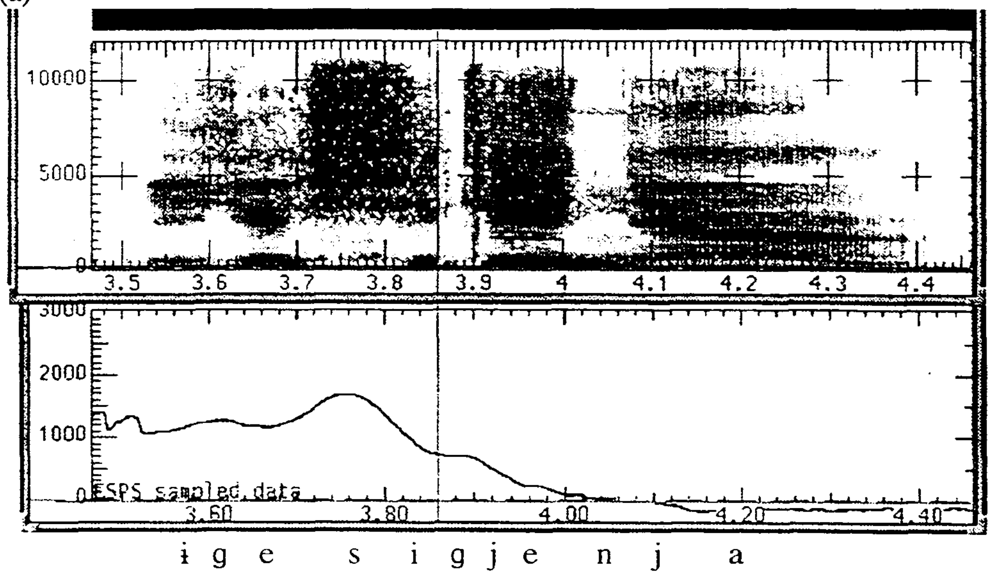
(b) [sigje] with a completely devoiced [i̥] — in accentual phrase medial position
(field numbers from 10166 until 10181)



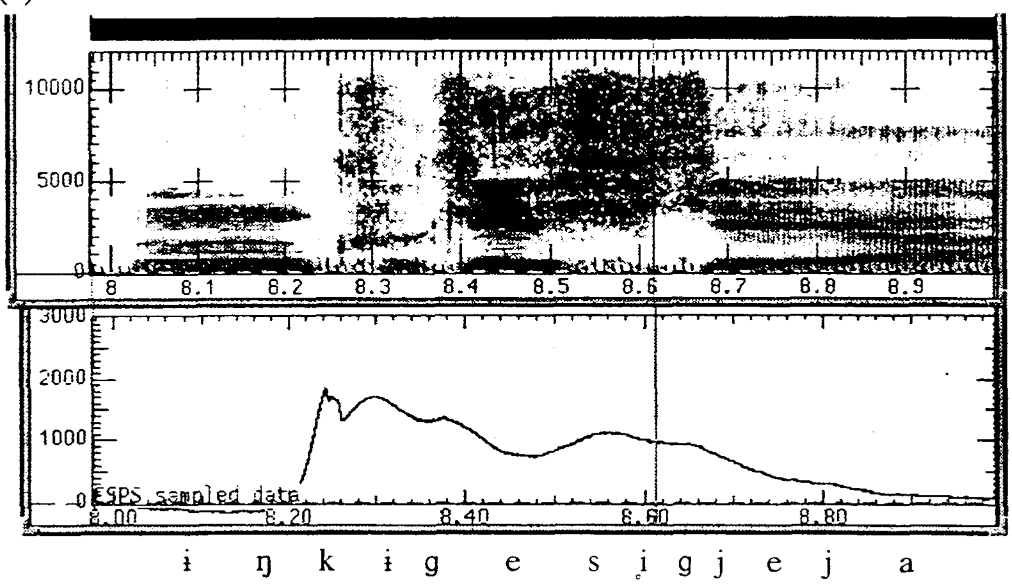
Word: [sigje] -- completely devoiced [i], AP medial

Figure 14. Intra-oral air pressure view of the two tokens of [sigje]. (a) with a voiced [i] and (b) with a completely devoiced [i̥]. A vertical line marks the beginning of [g].

(a)



(b)



Data from EMG muscle activities further support this suggestion that vowel devoicing in Korean might be a 'passive' result of a complex conjunction of more or less subtle variation in gestural timing and magnitude and its interaction with other concurrent gestures for the oral stricture and the associated tonal pattern. Figures 15-18 illustrate this point. Each figure shows the spectrogram, traces of EMG level measured in the CT and TA, intra-oral pressure, and fundamental frequency for pairs of tokens differing in the degree of voicing or devoicing in the target vowel. In general, each consonant in a target CVC sequence shows a single PCA "event". This "event" is consistently a well-defined peak with a high maximum value when the consonant is an aspirated stop or a fricative. The "event" was also a PCA maximum for tense and lenis stops, although it was not always so high or so clearly "peaky". In almost all cases, however, our data differed from the previous findings for Japanese in showing two separate PCA "events" for the two flanking consonants, even when the intervening vowel is devoiced. Also, there was consistently a TA "event" -- a peak -- associated with the offset of the intervening vowel, regardless of whether the vowel was voiced or devoiced. Figure 15 illustrates this for two tokens of the word [p^hup'e], with (a) voiced and (b) completely devoiced vowels repectively. There is a clear TA peak at the cursor marking the end of the vowel, and it separates the two clear PCA "events" for the preceding [p^h] and following [p']. Thus the count of EMG "events" does not distinguish whether the vowel is voiced or devoiced. The speaker did not produce a single large PCA "command" to open the glottis for the [p^h] and keep it open through to the end of the [p'] for the token with the devoiced vowel. Rather, there are two PCA peaks, separated by a TA peak for a vocal-fold tensing gesture, perhaps to inhibit "passive" residual voicing at the beginning of the [p'] closure.

Similar patterns can be observed in Figures 16 and 17, which show the EMG data for the tokens of [kip^hi] and [p^hus'i] described above in Figures 6, 8 and 9. All these examples show a PCA peak for each of the consonant preceding and the consonant following the target vowel, and an intervening TA peak around the offset of the target vowel, whether it is fully voiced or partially devoiced or completely devoiced. Similar EMG patterns can also be seen in the partially devoiced and completely devoiced token of [k^huk^hi] 'cookie' shown in Figure 18. Hirose (1971) suggests that the timing of the TA peak relative to the PCA peaks might occur earlier when the intervening vowel is devoiced, but informal observation of our EMG data suggests that there is no consistent difference. We intend to make more quantitative evaluation of Hirose's claim, to see if there is a relationship between timing of the TA peak and the probability (or degree) of vowel devoicing. We also intend to quantitify our informal observations of the glottal opening area. However, our informal observations to date suggest strongly that maximal glottal opening area, or the duration of glottal opening by themselves will not predict vowel devoicing.

Conclusion

This paper gives a glimpse of the extensive variability that can be observed in different tokens of "devoiced" vowels. There is variability in the "degree of devoicing" in the acoustic signal. There is also considerable variability in the patterns of glottal closing and opening across different devoiced tokens. There seems to be no categorical difference between devoiced tokens and voiced tokens in these views. Nor does there seem to be a categorical distinction in the pattern of EMG activity events. All of these observations support the notion that vowel devoicing in Korean cannot be described as the result of application of a phonological rule that changes a [+voice] specification to [-voice]. Rather, devoicing seems to be a highly variable "phonetic" process, a more or less subtle variation in the specification of such phonetic metrics as degree and timing of glottal opening, or of associated subglottal pressure or intra-oral airflow assoicated with concurrent tone and stricture specifications. Some of the token-pair comparisons are amenable to an explanation in terms of gestural overlap and undershoot. However, the effect of gestural timing on vocal fold state seems to be a highly nonlinear function of the interaction among specifications for the relative timing of glottal adduction and abduction gestures, of the amplitudes of the overlapped gestures, of aerodynamic conditions created by concurrent oral or tonal gestures, and so on. In summary, to understand devoicing, it will be necessary to examine its effect on phonetic representations of events in many parts of the vocal tract, and at many stages of the speech chain between the motor intent and the acoustic signal that reaches the hearer's ear.

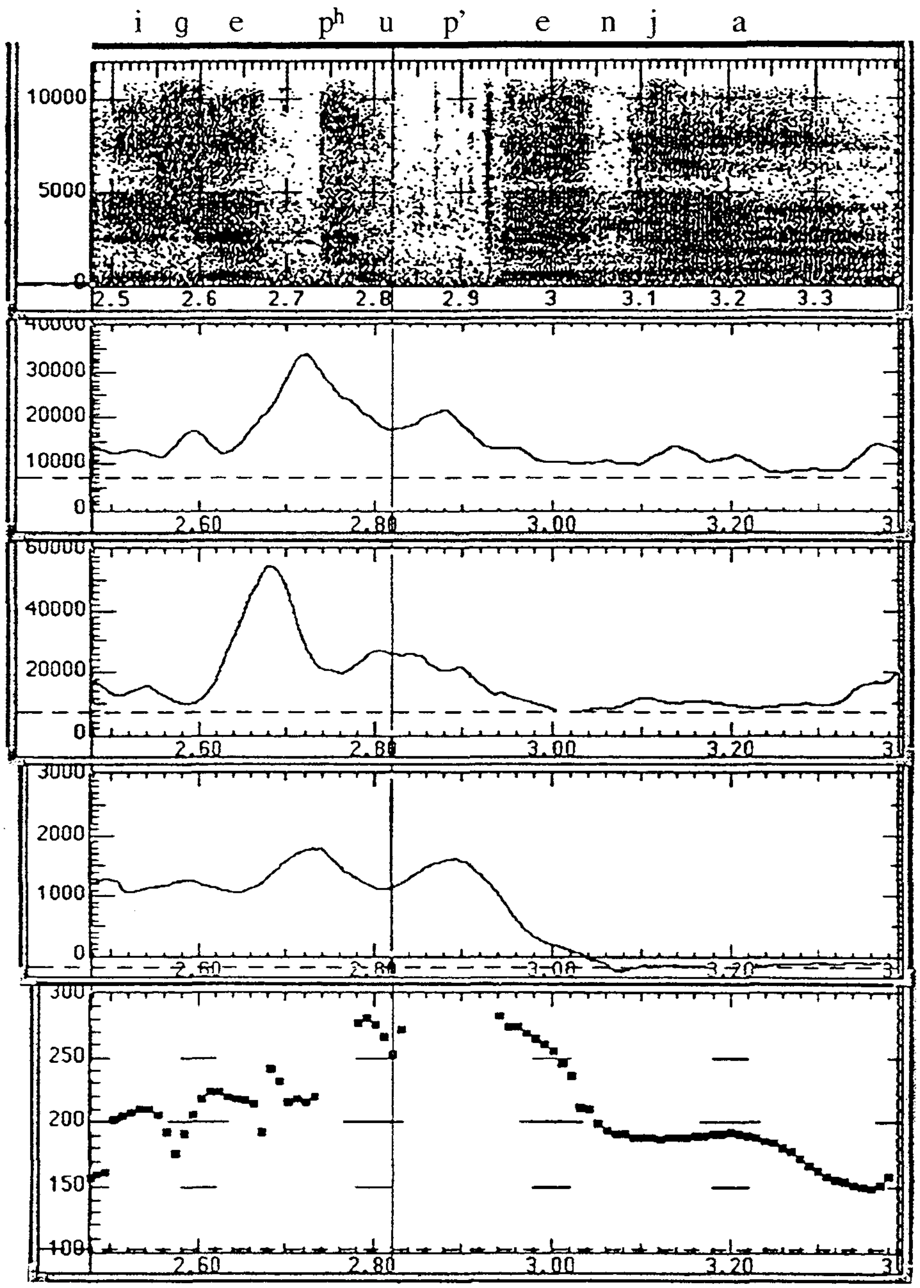
* This paper also appears in *The Jouranl of Speech Sciences*, Vol. 1, pp. 153-200 (1997).

Acknowledgments

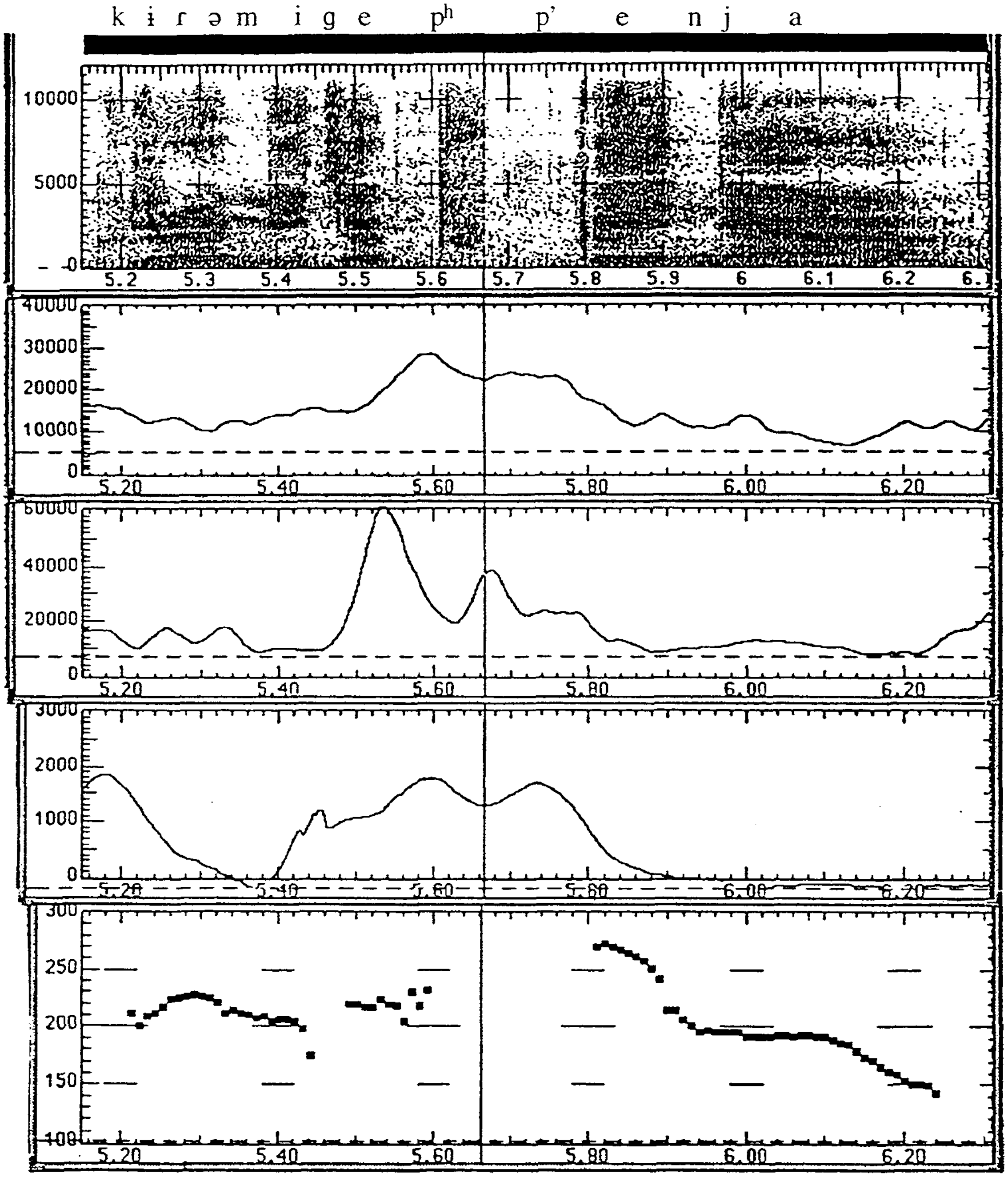
We thank Drs. Emi Z. Murana and Hirotake Iwamoto for their assistance in recording the physiological data at RILP, Hiroyuki Hirai for his generous help when we digitized the physiological data at ATR-HIP, and Norio Higuchi for letting us store and process the very large digitized database on the ATR-ITL machines. The first two authors owe Kiyoshi Honda considerably more than he will let us acknowledge for his advice as we learned how to process and interpret the EMG signals.

Figure 15. EMG activities for (a) voiced [u] and (b) completely devoiced [u̥] in [pʰup'e]. PCA is shown in the second window and TA is shown in the third window. Detailed description of each window is the same as that in Figure 3.

(a) voiced [u]

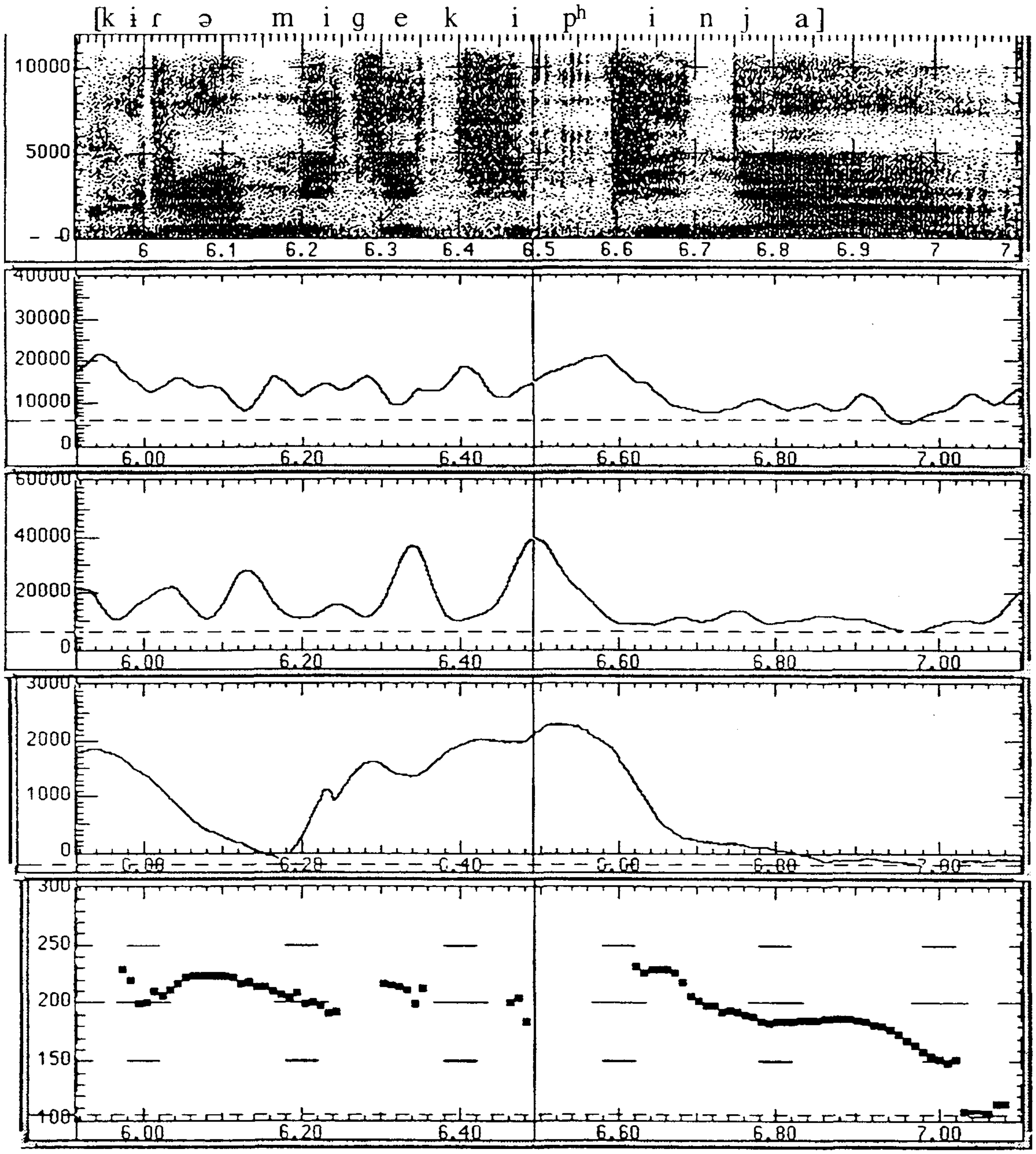


(b) completely devoiced [u̥] in /pʰup'e/.

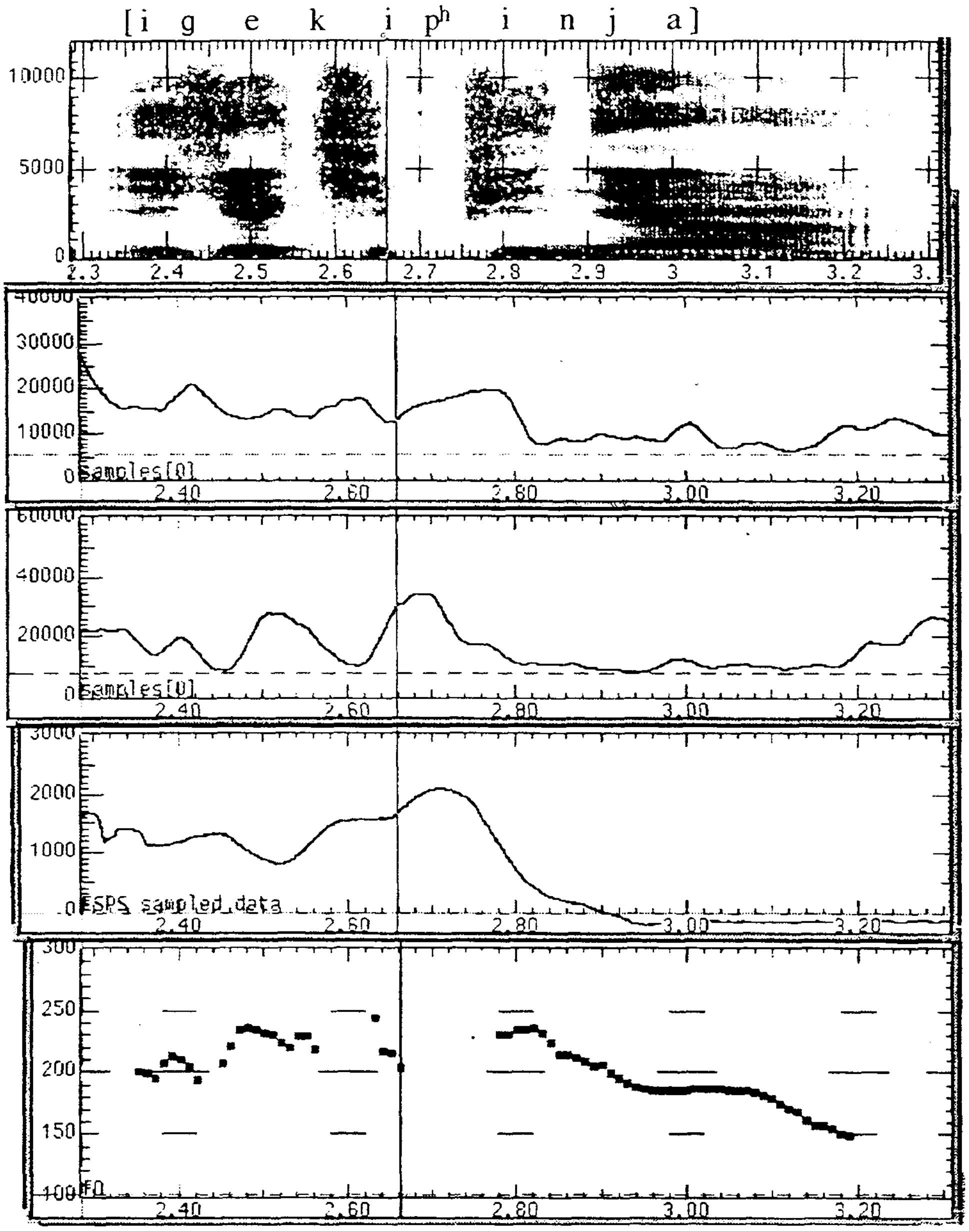


Figures 16. EMG patterns in the (a) voiced and (b) partially devoiced token of [kip^{h}i] 'evasion'. TA peak is shown in the second window, and PCA peak is shown in the third window.

(a) [kip^{h}i] — voiced

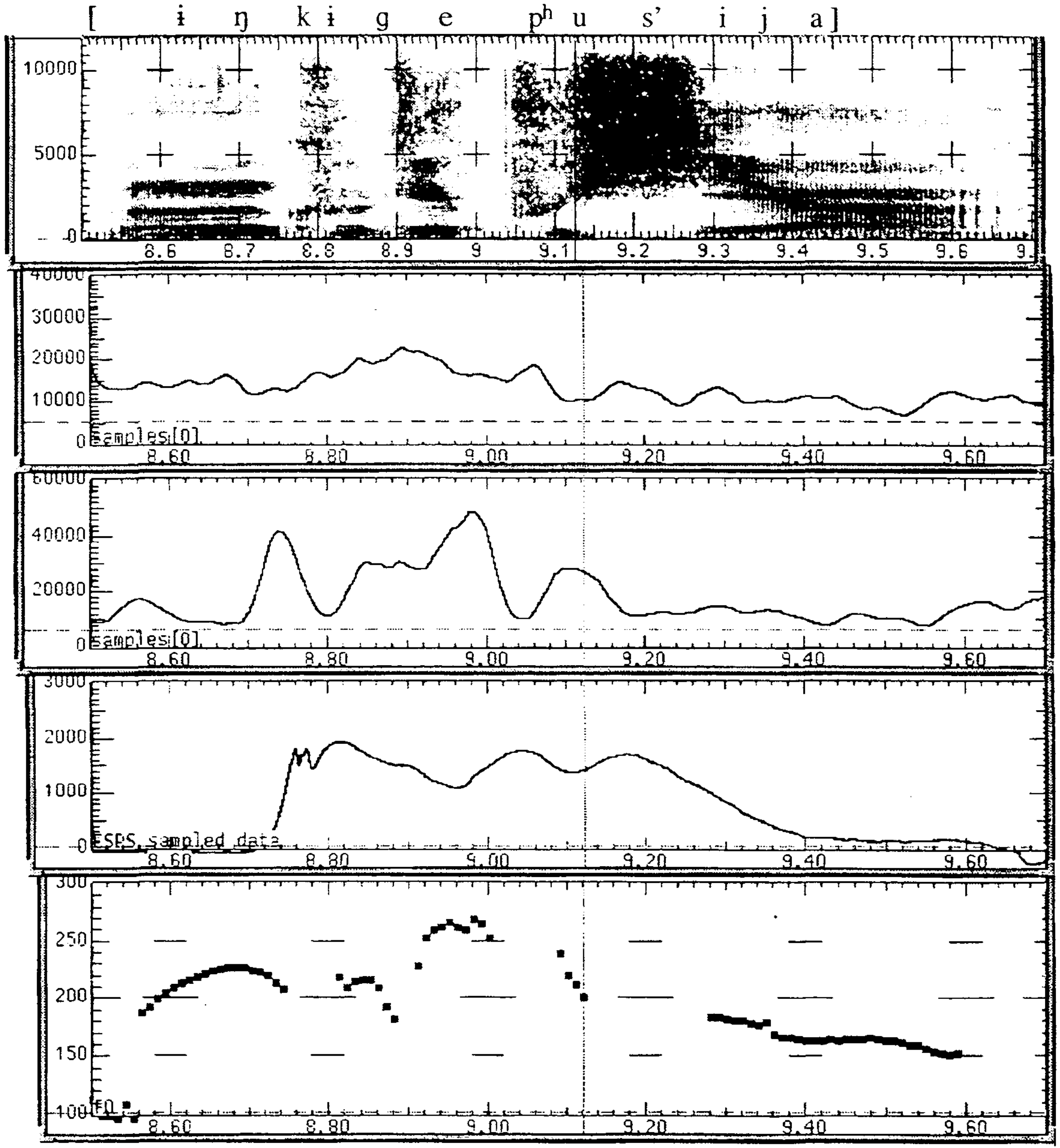


(b) [kip^{h}i] — partially devoiced [i̥].

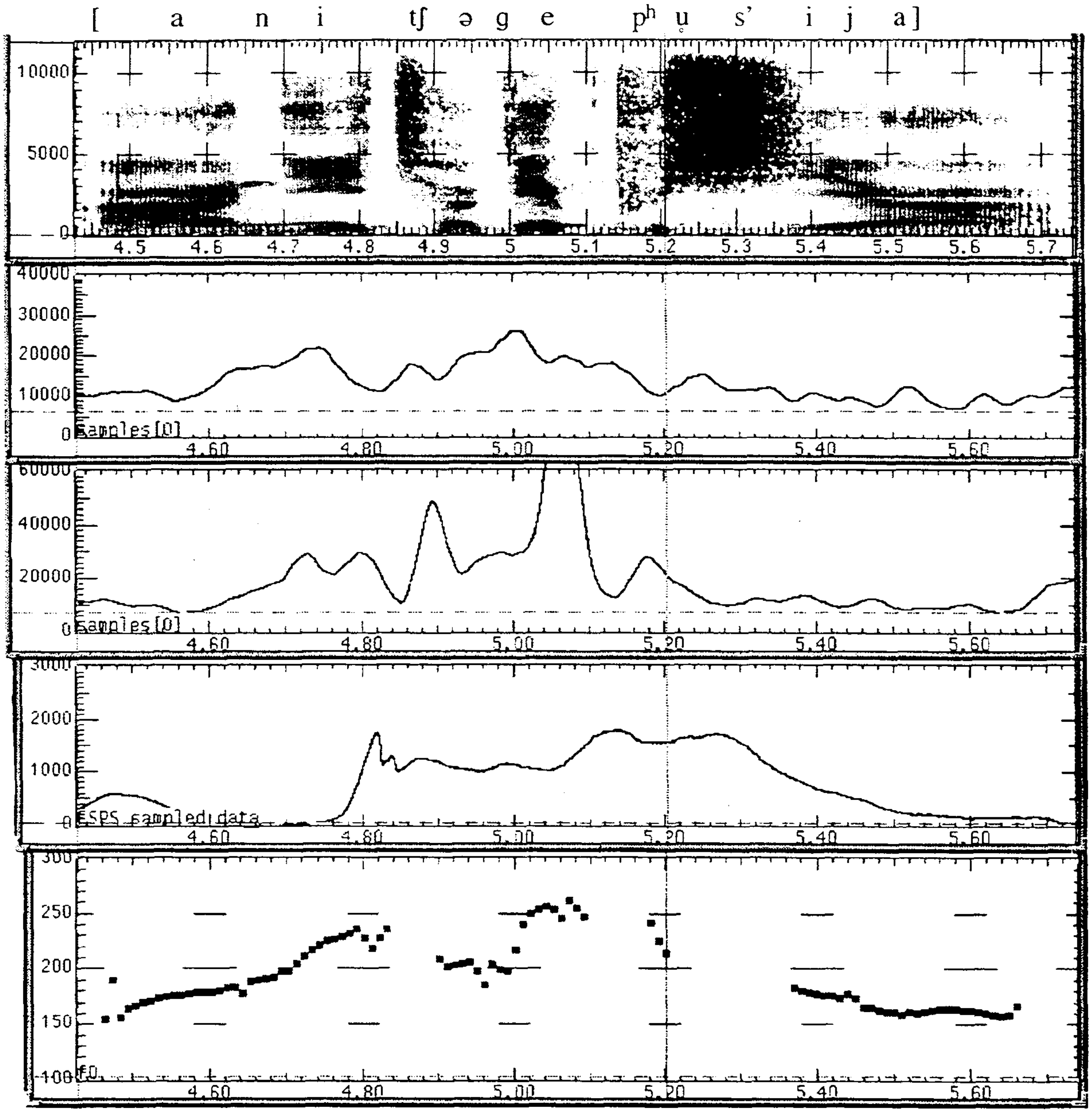


Figures 17. EMG patterns in the (a) voiced and (b) partially devoiced token of [pʰus'i] 'raw seed'.

(a) [pʰus'i] — voiced [u]

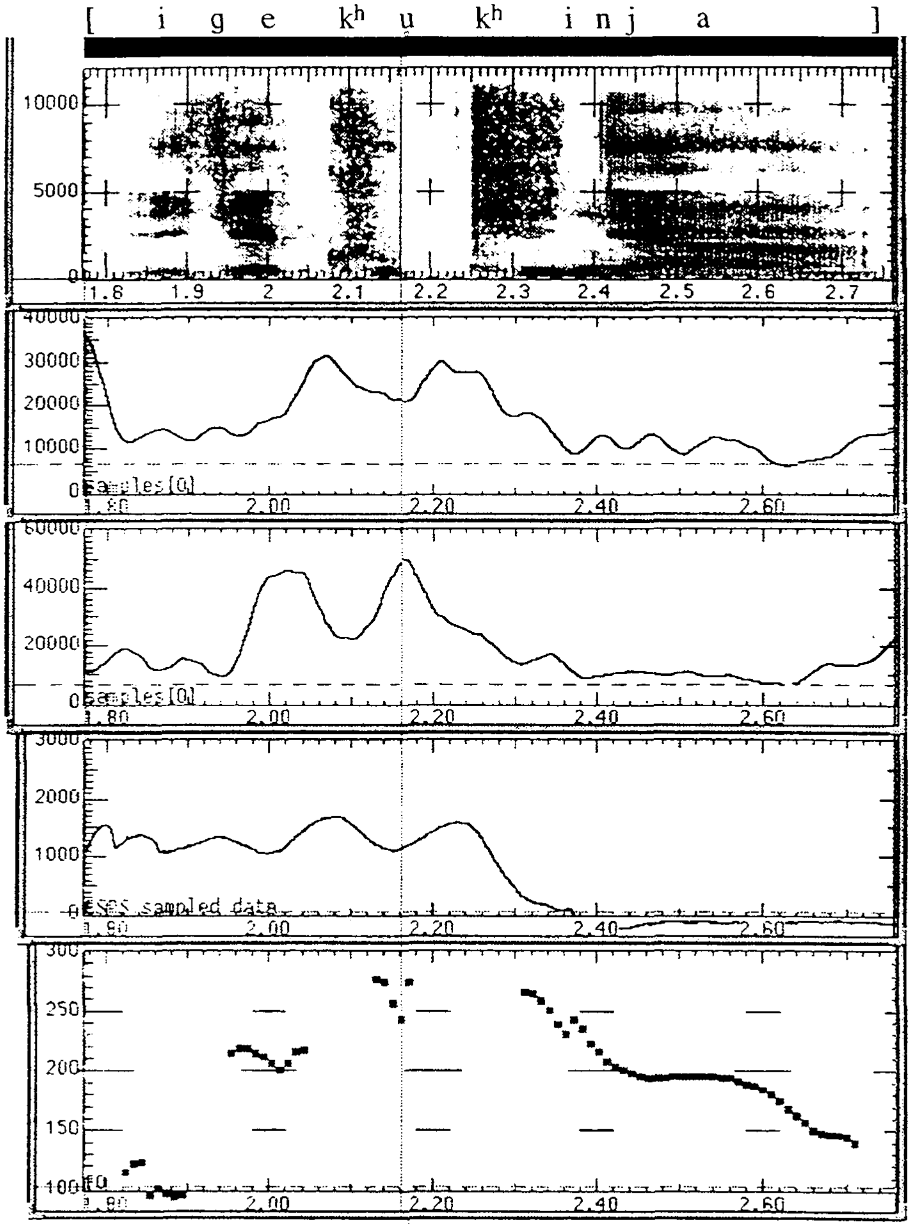


(b) [p^hus'i] — partially devoiced [u̥]

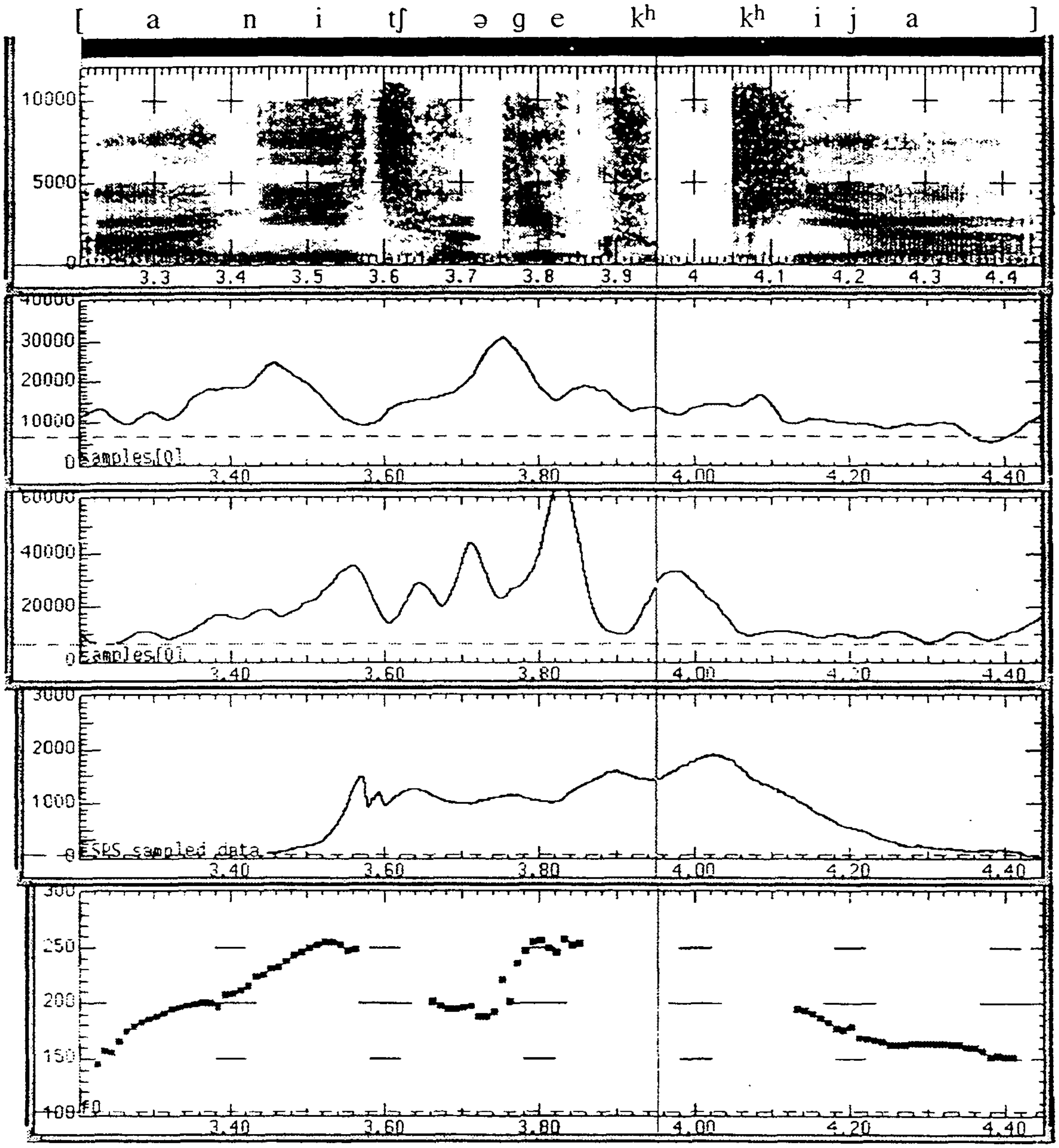


Figures 18. EMG patterns in the (a) partially devoiced and (b) completely devoiced token of [k^huk^hi].

(a) [k^huk^hi] — partially devoiced [u̥]



(b) [kʰukʰi] — completely devoiced [u̥]



References

Arvaniti, Amalia (1994) Acoustic features of Greek rhythmic structure. *Journal of Phonetics*, 22, 239-268.

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

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

Cedergren, H. J., & L. Simoneau (1985) La chute des voyelles hautes en français de Montreal: 'As-tu entendu la belle syncope?' In WHAT INITIAL Lemieux & H. J. Cedergren, eds., *Les tendencies dynamiques du français parle a Montreal*, pp. 57-144.

Cedergren, H. J., & Perreault, H. (1994). Speech rate and syllable timing in spontaneous speech. *Proceedings of the 1994 International Conference on Spoken Language Processing*, vol. 3, pp. 1087-1090.

Dart, S. (1986). An aerodynamic study of Korean stop consonants: Measurements and modeling. *J. Acoust. Soc. Am.* 81(1), 138-147

Dauer, R. M. (1980) The reduction of unstressed high vowels in Modern Greek. *Journal of the International Phonetic Association*, 10, 17-27.

Fougeron, Cécile (1996) "Variation de débit nasal en fonction de la position prosodique de [n] et [ɑ̃] en Français", in *the Proceedings of XXIth Journée d'Etudes sur la Parole*. pp. 215-218.

Fougeron, Cécile & Pat Keating (to appear) Articulatory strengthening at edges of prosodic domains. *JASA*

Han, M. S., & Weitzman, R. (1970). Acoustic features of Korean /P, T, K/, /p, t, k/ and /pʰ, tʰ, kʰ/. *Phonetica* 22, 112-128

Herman, Rebecca, Mary Beckman, & Kiyoshi Honda (1996). Subglottal pressure and final lowering in English. *Proceedings of the 1996 International Conference on Spoken Language Processing*, vol. 1, pp. 145-148.

Hirano, M. & J. Ohala (1969) Use of hooked-wire electrodes for electromyography of the intrinsic laryngeal muscles. *Journal of Speech and Hearing Research*, 12. pp. 362-373.

Hirose, Hajime (1971) The activity of the adductor laryngeal muscles in respect to vowel devoicing in Japanese. *Phonetica*, 23, 156-170.

Hirose, H., C. Y. Lee, & T. Ushijima (1974) Laryngeal control in Korean stop production. *Journal of Phonetics*, 2, 145-152.

Hsu, Chai-Shune & Sun-Ah Jun (1996) Is Tone Sandhi Group part of the Prosodic Hierarchy in Taiwanese?. Paper presented at the 3rd joint meeting of ASA and ASJ. Honolulu, Hawaii.

Imaizumi, Satoshi, Akiko Hayashi, & Toshisada Deguchi (1995). Listener adaptive characteristics of vowel devoicing in Japanese dialogue. *Journal of the Acoustical Society of America*, 98, 768-778.

Jannedy, Stefanie (1995) Gestural phasing as an explanation for vowel devoicing in Turkish. *OSU Working Papers in Linguistics*, no. 45, pp. 56-84.

Jun, Sun-Ah (1989) The accentual pattern and prosody of Chonnam Dialect of Korean. In S. Kuno et al., eds., *Harvard Studies in Korean Linguistics*, no. 3, pp. 89-100.

Jun, Sun-Ah (1990) The prosodic structure of Korean — in terms of voicing. In E-J. Baek, ed., *Proceedings of the Seventh International Conference on Korean Linguistics* . vol. 7, pp. 87-104.

Jun, Sun-Ah (1995) Asymmetrical prosodic effects on the laryngeal gesture in Korean. In B. Connell and A. Arvaniti (eds.) *Phonology and phonetic evidence: Papers in Laboratory Phonology* IV: 235-253, Cambridge University Press, England.

Jun, Sun-Ah (1996a) *The Phonetics and Phonology of Korean Prosody: Intonational Phonology and Prosody*, Garland. [Ph.D dissertation, Ohio State University (1993)]

Jun, Sun-Ah (1996b) Influence of microprosody on macroprosody: a case of phrase initial strengthening. *UCLA Working Papers in Phonetics* 92: 97-116

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

Jun, Sun-Ah, & Mary E. Beckman (1994) Distribution of devoiced high vowels in Korean. *Proceedings of the 1994 International Conference on Spoken Language Processing*, vol. 1, pp. 479-152.

Kagaya, Ryohei (1974) A Fiberscopic and Acoustic study of the Korean stops, affricates and fricatives. *Journal of Phonetics* 2, 161-180.

Kim, Hyun-Gi, Seiji Niimi, & Hajime Hirose (1993) Devoicing of vowel in Korean. *Ann. Bull. Research Institute of Logopedics and Phoniatrics*, no. 27, pp. 151-157.

Kiritani, Shigeru (1994) Articulatory movement and units in speech. *Dokkyo International Review*, no. 7, pp. 403-412.

Kondo, Mariko (1994) Mechanisms of vowel devoicing in Japanese. *Proceedings of the 1994 International Conference on Spoken Language Processing*, vol. 1, pp. 61-64.

Kuehn, D. P., & K. L. Moll (1976) A cineradiographic study of VC and CV articulatory velocities. *Journal of Phonetics*, 4, 303-320.

Levac, Louise, Henrietta J. Cedergren, & Helene Perreault (1993) Phonetic evidence of narrow and wide temporal scope for prosodic constituents in French. *Proceedings of an ESCA Workshop on Prosody*, Working Papers, Department of Linguistics and Phonetics, Lund, No. 41, pp. 54-57.

Lindblom, Björn (1963) Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America*, 35 (11), 1773-1781.

Lindblom, B., S. Brownlee, B. Davis & S-J. Moon (1992). Speech transforms. *Speech Communication*, 11, 357-368.

Löfqvist, Anders, Thomas Baer, Nancy S. McGarr & Robin Seider Story (1989) The cricothyroid muscle in voicing control. *Journal of the Acoustical Society of America* 85, 1314-1321.

Maekawa, Kikuo (1990) Hatsuwa sokudo ni yoru yuusei kukan no hendoo. *Technical Report* SP89-148, Densi Zyoohoo Tuusin Gakkai.

Martin, S. E. (1951). Korean phonemics. *Language* 27, 519-533.

Munhall, Kevin, & Löfqvist, Anders (1992) Gestural aggregation in speech: laryngeal gestures. *Journal of Phonetics* 20, 111-126.

Sawashima, M. (1971) Devoicing of vowels. *Ann. Bull. RILP* 5. 7-13.

Sawashima, M., H. Hirose & H. Yoshioka (1978) Abductor (PCA) and adductor (INT) muscles of the larynx in voiceless sound production. *Ann. Bull. Research Institute of Logopedics and Phoniatrics*, no. 12, pp. 53-60.

Silva, David J. (1992) *The Phonetics and Phonology of Stop Lenition in Korean*. Ph.D. dissertation, Cornell University.

Weitzman, Raymond S., Masayuji Sawashima, Hajime Hirose & Tatsujiro Ushijima (1976) Devoiced and whispered vowels in Japanese. *Ann. Bull. Research Institute of Logopedics and Phoniatrics*, no. 10, pp. 61-79.

Yeni-Komshian, G. H. & S. D. Soli (1981) Recognition of vowels from information in fricatives: perceptual evidence of fricative-vowel coarticulation. *Journal of the Acoustical Society of America*, 70, 966-975.

Yoshioka, Hirohide (1981) Laryngeal adjustments in the production of the fricative consonants and devoiced vowels in Japanese, *Phonetica* 38, 236-251.

Zee, Eric (1989) Vowel devoicing in Shanghai. Paper presented at the Northeast Conference on Chinese Linguistics, Ohio State University.