Ann. Bull. RILP. No. 10, 61-79 (1976)

DEVOICED AND WHISPERED VOWELS IN JAPANESE

Raymond S. Weitzman*, Masayuki Sawashima,
Hajime Hirose and Tatsujiro Ushijima.

Devoiced vowels[1] can and have been defined as those vowels produced with no vocal fold vibration and with the glottis abducted.[2] It has sometimes been stated that a devoiced vowel is phonetically identical to a whispered vowel.[3] In the literature whispered vowels are classically defined as those produced with the supraglottal cavities assuming the shape for the particular vowel with a narrowing (or even closing) of the membranous glottis while the cartilaginous glottis is open. Air flowing through this narrow constriction is made turbulent, producing the hissing sound of whisper.[4]

This identification of devoiced vowels and whispered vowels is certainly not agreed upon by all scholars. For example, Abercrombie has differentiated devoiced and whispered vowels in terms of an open glottal state versus a narrow glottal state.[5] Abercrombie's view has gained added weight by the recent fiberscopic studies of Sawashima, who found that the maximum glottal opening during the pronunciation of Japanese [i̥] in the environment [k_t], [k_k], and [k_s] was greater than the width associated with the obstruents [t, k] and [tt] and [kk] and comparable to the widths associated with [s] and [ss] in voiced vowel environments.[6]

Herein is a preliminary progress report on a study to gather more data on laryngeal activity during the production of devoiced vowels in Japanese and to explore further the relationship between devoiced vowels and whispered vowels. Of particular concern here was to examine glottal and supraglottal laryngeal gestures during whispered speech and to see whether a vowel devoicing gesture occurs even in whispered speech when the vowels /i/ and /u/ occur between voiceless consonants.

Experimental Procedures

A series of test words--some meaningful, but most nonsense--having the canonical forms C̥V̥tee, C̥Vdee, Vtee and Vdee were collected together (see Table 1).[7] V̥ was either [i̥] or [u̥], V either [i] or [u], and C̥ was one of the

KITEE	TSUTEE	KIDEE	TSUDEE
KUTEE	CHITEE	KUDEE	CHIDEE
SHITEE	CHUTEE	SHIDEE	CHUDEE
SHUTEE	PITEE	SHUDEE	PIDEE
SUTEE	ITEE	SUTEE	IDEE
FUTEE	UTEE	FUDEE	UDEE
HITEE		HIDEE	

* California State University, Fresno

following voiceless obstruents: [p, k, c, č, ʦ, s, š, ç]--i.e., all those that can precede either /i/ or /u/ or both in Japanese. The words were placed in the sentence frame "sore o _____ to yuu" (That we call _____.). The words were read with no accent kernel and in the sentence frame by two male Japanese subjects (Subject H and Subject S), both speakers of the Tokyo dialect. The sentences were read in three ways. First, they were read in a normal speech mode. Secondly, they were read in a soft whisper mode--i.e., the kind of whisper one ordinarily uses when whispering into someone's ear. Thirdly, they were read in a stage whisper mode--i.e., the kind of whisper one would use if the listener was quite a distance away from the speaker. The sentences in each mode were repeated twice (a few three or four times) by Subject H and three times (a few four times) by Subject S. The utterances were recorded using a Sanken CL-201 Close Talking Microphone and a 2-track Sony TC-707S Stereo Tape Recorder. An Olympus Fiberscope was inserted through the subject's nose and positioned so as to obtain a good view of glottis and supraglottal laryngeal area. During the recording session black-and-white motion pictures were made of laryngeal activity at the rate of 50 frames/second, using a Beaulieu 16mm cinecamera. The speech recording and the film recording were synchronized by time marks so that a correlation could later be made of the speech wave and laryngeal events depicted on the film. Those film frames covering the relevant portions of each utterance token were examined visually and measurements taken of the glottal width between the vocal processes.[8] Since the fiberscope technique does not allow absolute measurements to be made, the measurements taken do not reflect the actual dimensions of the glottis. However, the measurements do reflect the relative changes in glottal width during the pronunciation of the utterances. The measurements were made to an accuracy of 0.5 millimeters on the magnified film image.

Results

Normal Speech: Figures 1 and 2 show the time course of glottal width for representative samples of devoiced vowels in normal speech. In normal speech the glottis moves from an adducted, voicing position to an abducted, voiceless position and then back to an adducted voicing position when the test words have the canonical form ç̥V̥tee. The maximum point of abduction was consistently reached during the ç̥V̥ interval. When ç̥ was a stop, the maximum glottal opening usually occurred during the devoiced vowel interval. These findings generally agree with Sawashima's earlier findings. In the cases where ç̥ was a fricative or an affricate, it could not be ascertained exactly in which segment the point of maximum abduction was reached, since it was not possible in these cases to determine the boundary between the devoiced vowel and its preceding voiceless consonant.[9]

In the case of Subject H it appeared that the arytenoid masses were not fully apart from each other, remaining partially in contact, so that the posterior portion of the cartilaginous glottis was not visible. In the case of Subject S the arytenoid masses appeared almost completely, if not completely, apart from each other so that a good portion of the cartilaginous glottis was visible.[10] For both subjects the vocal folds were apart from each other along their entire length, the epiglottis was well forward away from the glottis, and the false vocal folds were well back from the glottis.

In Figure 3 are still photographs taken from the films showing the typical laryngeal gesture at the point of maximum abduction for each subject.

Soft Whispered Speech: As can be seen from Figures 4 and 5, the time course of glottal width during the C̥V̥t interval in soft whispered speech bears some similarities as well as differences to that in normal speech. There is clearly an abducting gesture in which the glottal width increases from a relatively narrow base. Just prior to the take-off point of the abducting movement, the glottal width is not only small but there is a radical constriction of the larynx above the glottis involving the arytenoid masses, false vocal folds, and epiglottis (see discussion below for /i/ and /u/ in the words /itee/ and /utee/).

At the point of maximum opening, the arytenoid masses seemed to be more closely approximated and cover a greater portion of the open cartilaginous glottis than in normal speech in the case of Subject H. Usually, however, one or both of the vocal processes were visible. The anterior portion of the membranous glottis is closed--a phenomena not observed in normal speech. Furthermore, the false vocal folds seemed to be more closely drawn in toward the glottis and appeared more prominent than they were in normal speech. In some cases if not most, the epiglottis seemed to be drawn back somewhat toward the glottis. In general, there was a greater degree of laryngeal constriction than found associated with devoiced vowels in normal speech (see Figure 6A).

In the case of Subject S, the arytenoid masses were drawn apart from each other at the point of maximum opening, so that the posterior part of the cartilaginous glottis was clearly visible. There did not seem to be any closure of the anterior portion of the membranous glottis, as with Subject H. The area of the glottis seemed at times similar to that found in normal speech and at times somewhat narrower and pinched slightly inward at the vocal processes. As with Subject H, the false vocal folds were drawn inward toward the glottis, and the epiglottis drawn back further than in normal speech (see Figure 6B).

A comparison of /i/ and /u/ in /itee/ and /utee/ in soft whispered speech with the same vowels surrounded by voiceless consonants reveal some similarities but also some very striking differences. For both subjects the shape of the glottis was much the same in the two environments, but the glottal width was much narrower (see Figure 7). Abduction did occur but usually during the latter half of the vowel in anticipation of the following voiceless stop. The most prominent difference, however, was the extreme degree of constriction above the glottis, which was nearly identical to that found just prior to the abducting movement in those cases where the vowel is between voiceless consonants. For both subjects the false vocal folds were drawn radically in toward the glottis and the epiglottis was drawn back considerably. In the case of Subject H the top of the epiglottis was visible occasionally. For both subjects the arytenoid masses appeared to be in contact over a greater area (see Figure 8).

Stage Whispered Speech: Figures 9 and 10 show the time course of glottal width during the representative samples of words with the canonical form C̥V̥tee in stage whispered speech. As in normal and soft whispered speech, abduction of the glottis was observed to begin about the beginning or just prior to the C̥V̥t sequence and reached a point of maximum opening

during the $\overset{\circ}{C}\overset{\circ}{V}$ interval. The laryngeal state just prior to the abduction movement was almost identical to that for /i/ and /u/ in the words /itee/ and /utee/, which will be discussed below.

In the case of Subject H the glottis at the peak of abduction was triangular in shape and the anterior portion of the membranous glottis closed, as in soft whispered speech. The arytenoid masses appeared partially in contact, and thus covered the posterior portion of the cartilaginous glottis. One or both of the vocal processes were usually visible. As in soft whisper, the false vocal folds were drawn inward toward the glottis, and the epiglottis drawn back somewhat. The total effect gave the larynx above the glottis a rather constricted appearance (see Figure 11A).

In the case of Subject S the glottis had a somewhat bulbous or key-hole shape, and there appeared to be no closure at the anterior portion of the membranous glottis, although it was difficult to tell since the epiglottis covered the anterior portion. Unlike Subject H, the arytenoids were drawn well apart from each other so that the posterior of the cartilaginous glottis was visible, as in soft whispered speech. The false vocal folds were drawn inward and sometimes appeared to cover the vocal folds. The general laryngeal appearance was that of constriction in the anterior half and abduction in the posterior half (see Figure 11B).

For both subjects the glottis at times appeared very large. Frequently the maximum relative glottal width was greater than or nearly equal to that found in normal speech. There is some evidence to suggest that in whispered speech, particularly stage whispered speech, the larynx is raised appreciably.[11] This may account for the large appearance of the glottis in stage whispered speech. Since no attempt was made in this experiment to measure the changes in laryngeal height and since absolute measurements of glottal width are not possible with current fiberscopic techniques, however, it is not possible to say definitely if laryngeal raising or further glottal abduction or a combination of both were responsible for the apparent large size of the glottis.

An examination of the glottis and larynx during the pronunciation of /i/ and /u/ in the words /itee/ and /utee/ during stage whispered speech again, as in soft whispered speech, revealed some remarkable differences from these same vowels when surrounded by voiceless consonants. The degree of constriction of the larynx is even more radical than that found for the same vowels in the same environment in soft whispered speech. Generally the constriction is so great that glottal width could not be measured with any accuracy, particularly in the case of Subject S (see Figure 12). In the case of Subject H, the arytenoid masses, false vocal folds, and epiglottis all contributed to constricting the larynx to the extent that only a very small portion of the glottis in the vicinity of the vocal processes was visible (see Figure 13A). The anterior portion of the membranous glottis seemed to be closed and the area of closure seemed to be somewhat greater than observed in the other cases. The false vocal folds were in close proximity and at times were even in contact with each other. In viewing the running films, vibrations of the false vocal folds could be observed occasionally.

The laryngeal constriction was so radical in the case of Subject S that only a very small chink of the glottis was barely visible (see Figure 13B).

Conclusions

The results presented above confirm the position that devoiced vowels and whispered vowels are not phonetically identical, at least in their laryngeal gestures. Also the results clearly indicate that the vowel devoicing process occurs in whispered speech as well as in normal speech. In normal speech, the take-off point of devoicing is from an adducted, voicing state, while in both types of whispered speech the take-off point is from an open but highly narrowed glottal state and a highly constricted supraglottal laryngeal state. The devoicing gesture--i.e., the abduction of the glottis --seems to be basically the same, however.

The differences observed between the devoicing of the vowel in normal speech and the two types of whispered speech in terms of unconstricted larynx versus somewhat constricted larynx can perhaps best be accounted for in terms of the differences between the two modes of speech. Laryngeal constriction is a basic component in whispered speech and is maintained to a certain degree even during the devoicing gesture.

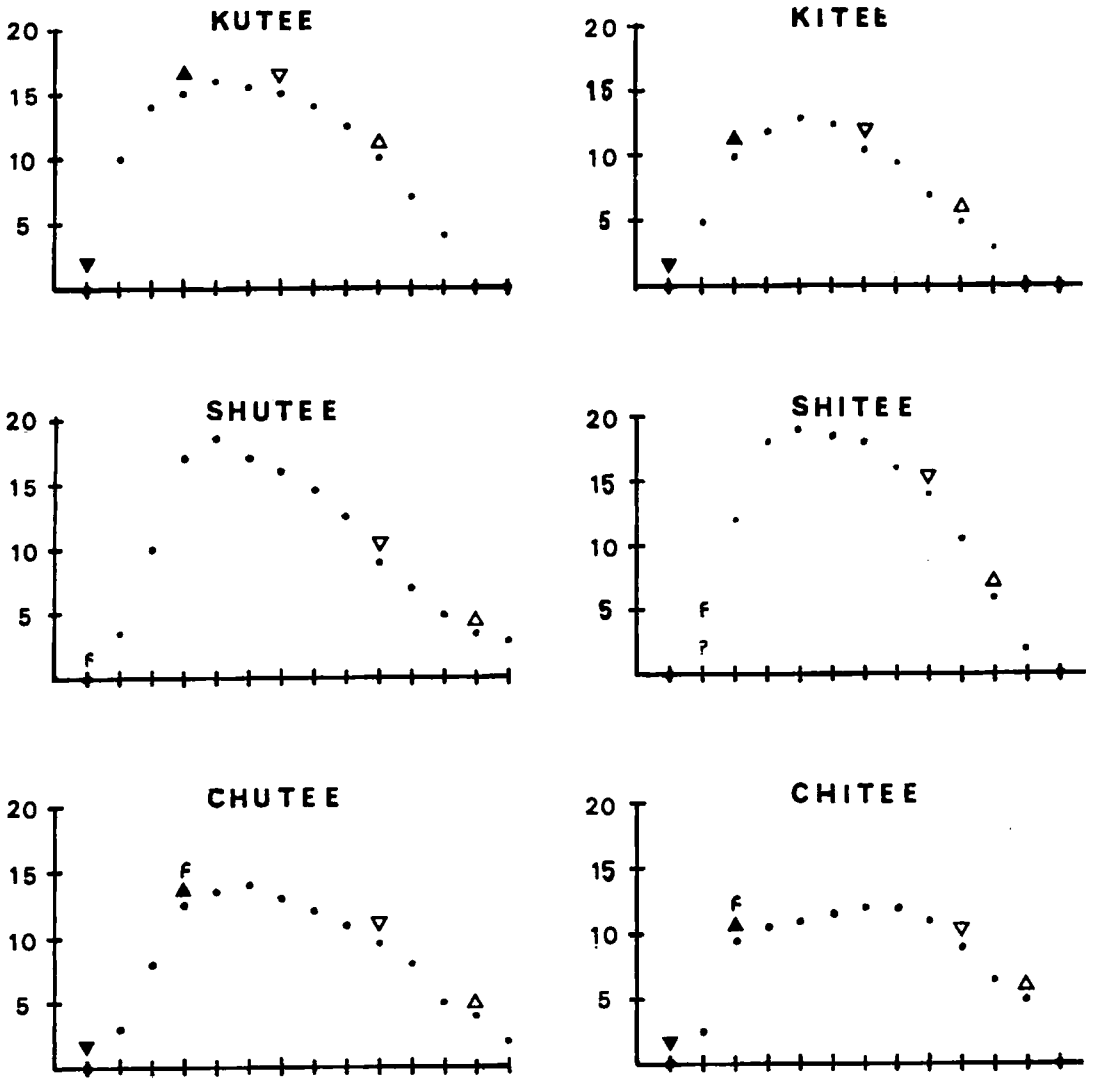


Figure 1. Time course of glottal width for representative samples of devoiced vowels during normal speech (Subject H). The abscissa represents the frame number, and the ordinate represents an arbitrary scale of glottal width in millimeters. Legend: ▼ indicates the frame that occurs at the point or just before the closure for an initial stop; ▲ indicates the release of the initial stop; ▽ indicates the closure for the stop [t] following the devoiced vowel, and △ indicates the frame that occurs at the point or just before the release of [t]; "f" indicates the frame at the point or just before the onset of frication.

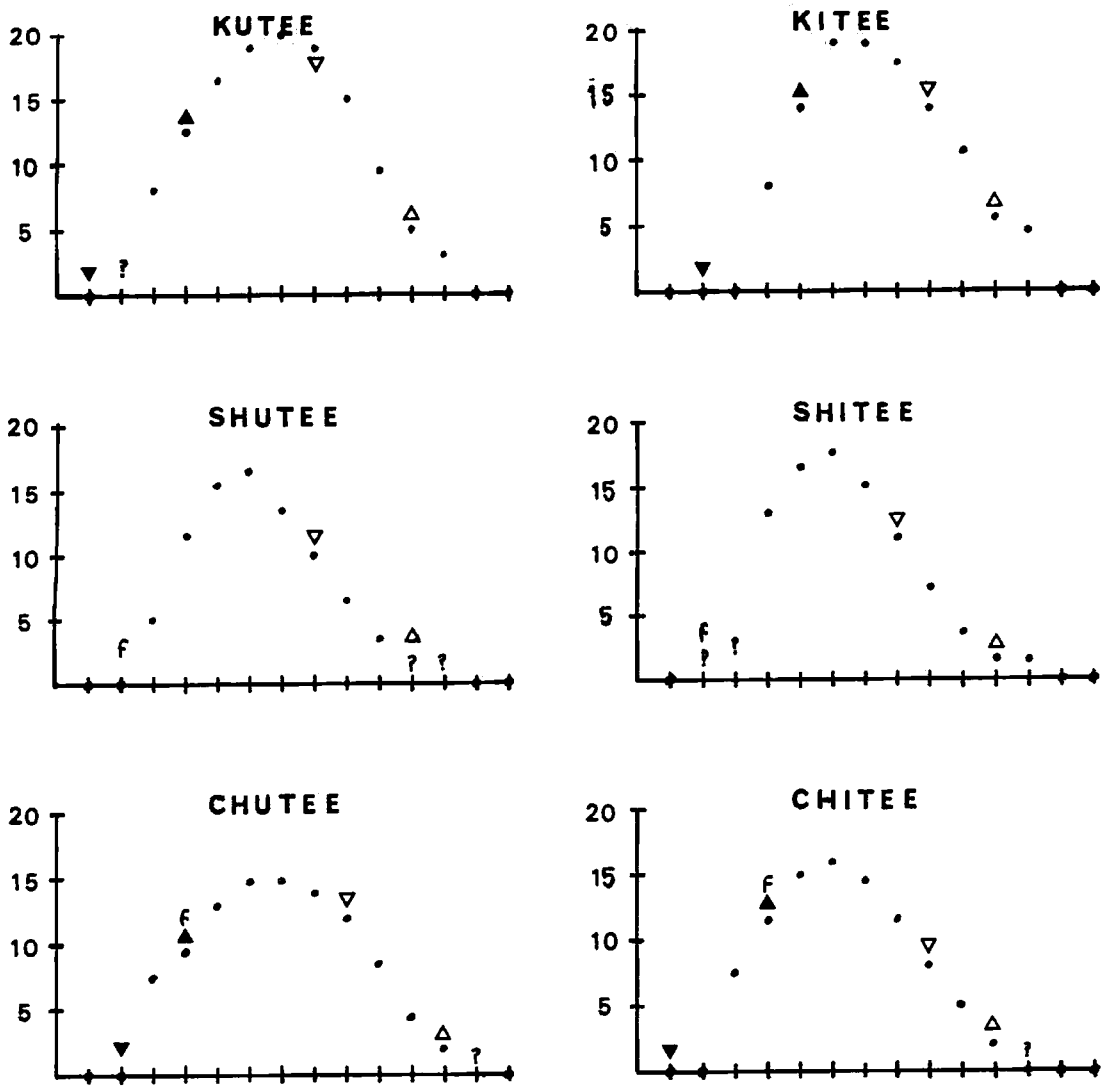
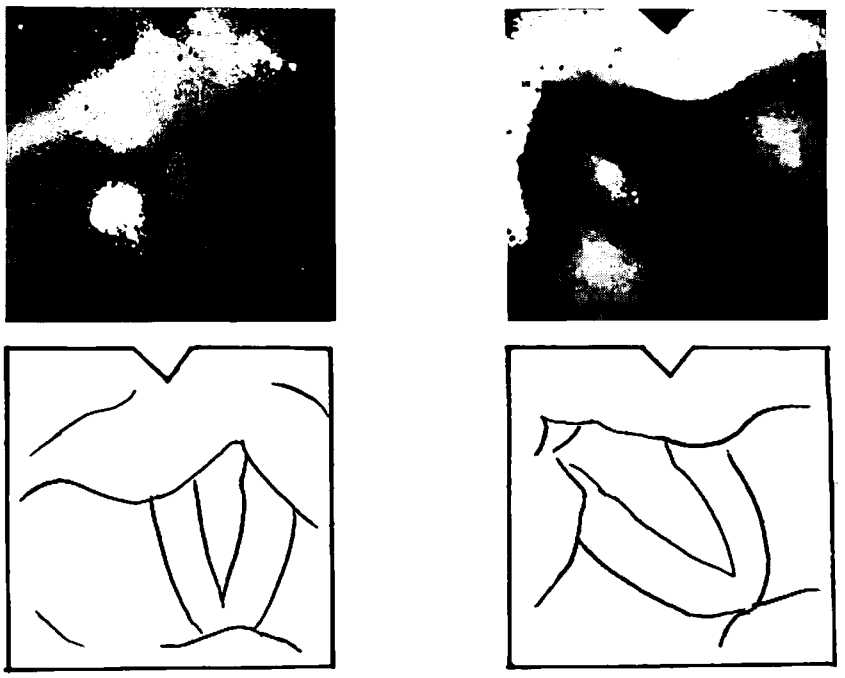


Figure 2. Time course of glottal width for representative samples of devoiced vowels during normal speech (Subject S). For explanation see Fig. 1.



A: SUBJECT H

B: SUBJECT S

Figure 3. Photographs showing the typical laryngeal gesture at the point of maximum abduction in normal speech.
(Test word: /kitee/)

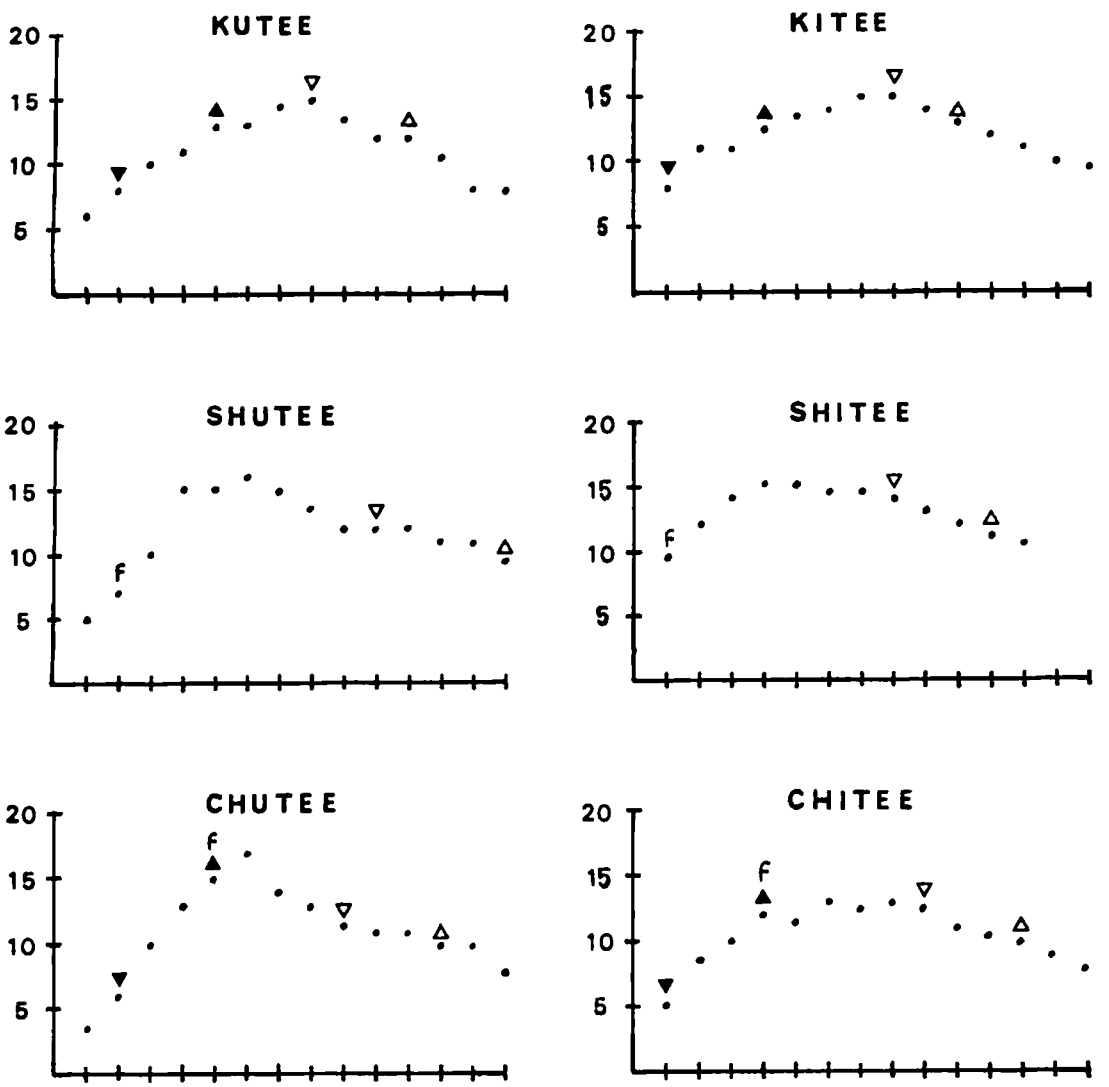


Figure 4. Time course of glottal width for representative samples of devoiced vowels during soft whispered speech (Subject H). For explanation see Figure 1.

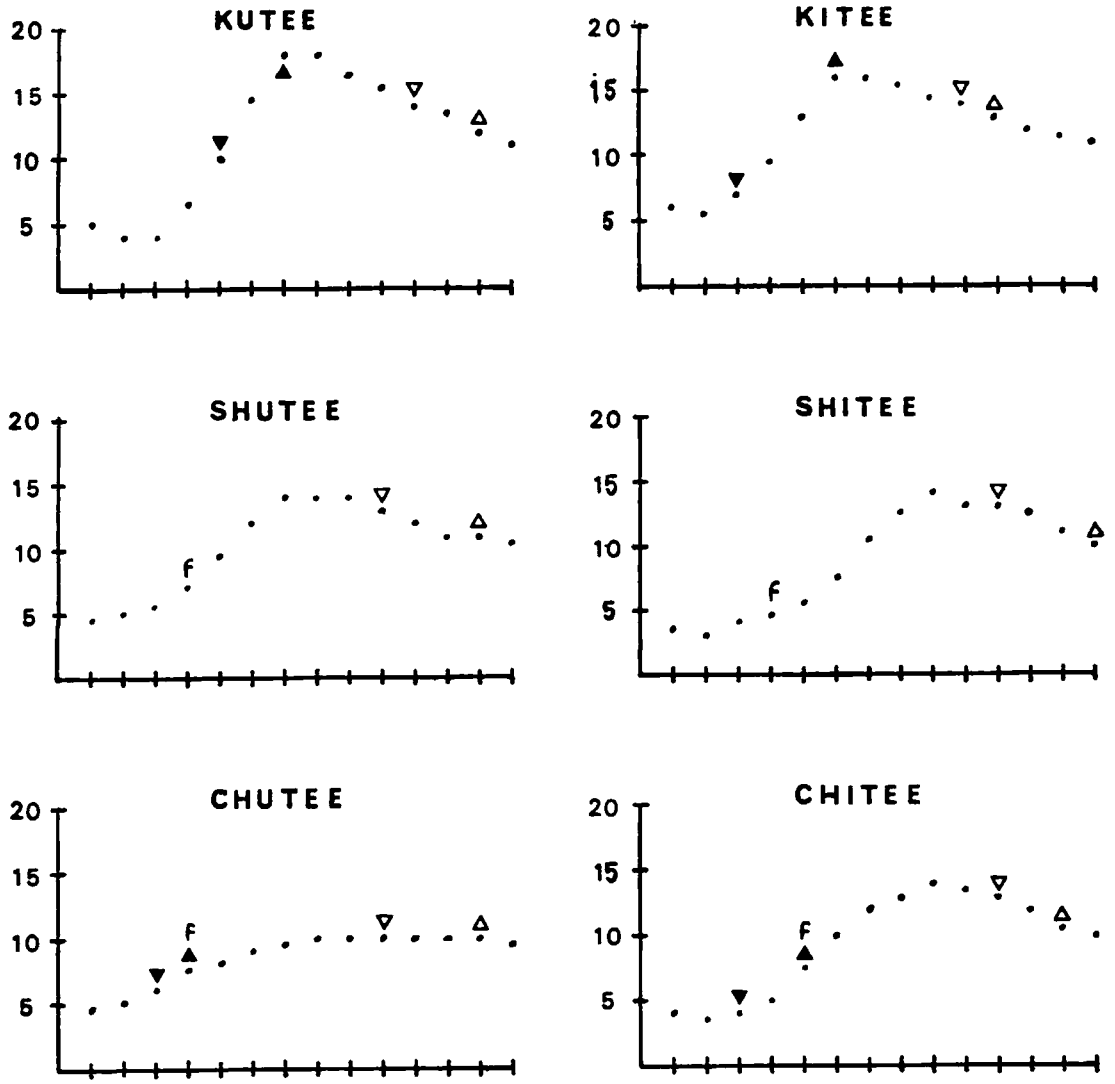
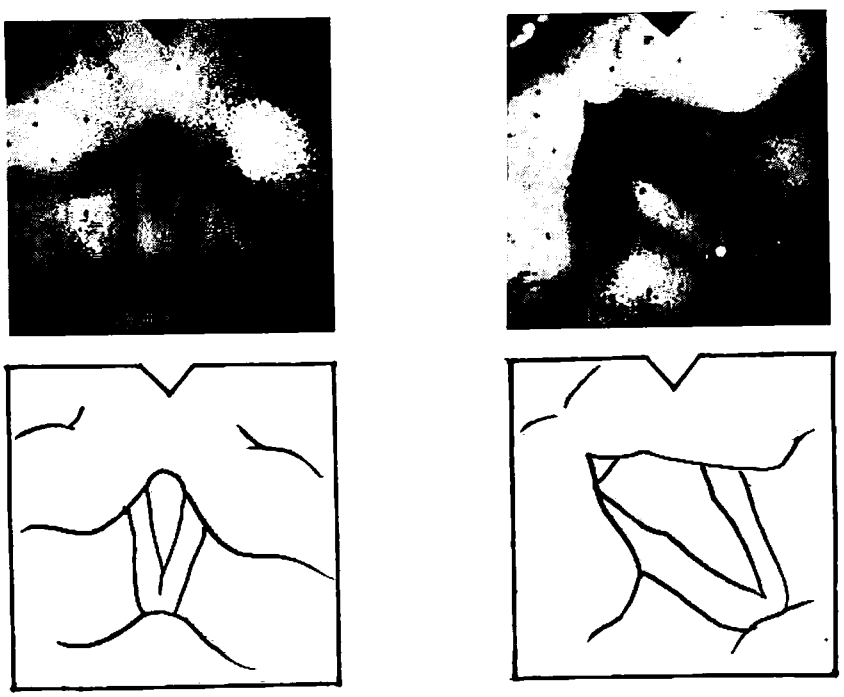


Figure 5. Time course of glottal width for representative samples of devoiced vowels during soft whispered speech (Subject S). For explanation see Figure 1.



A: SUBJECT H

B: SUBJECT S

Figure 6. Photographs showing the typical laryngeal gesture at the point of maximum abduction in soft whispered speech.
(Test word: /kitee/)

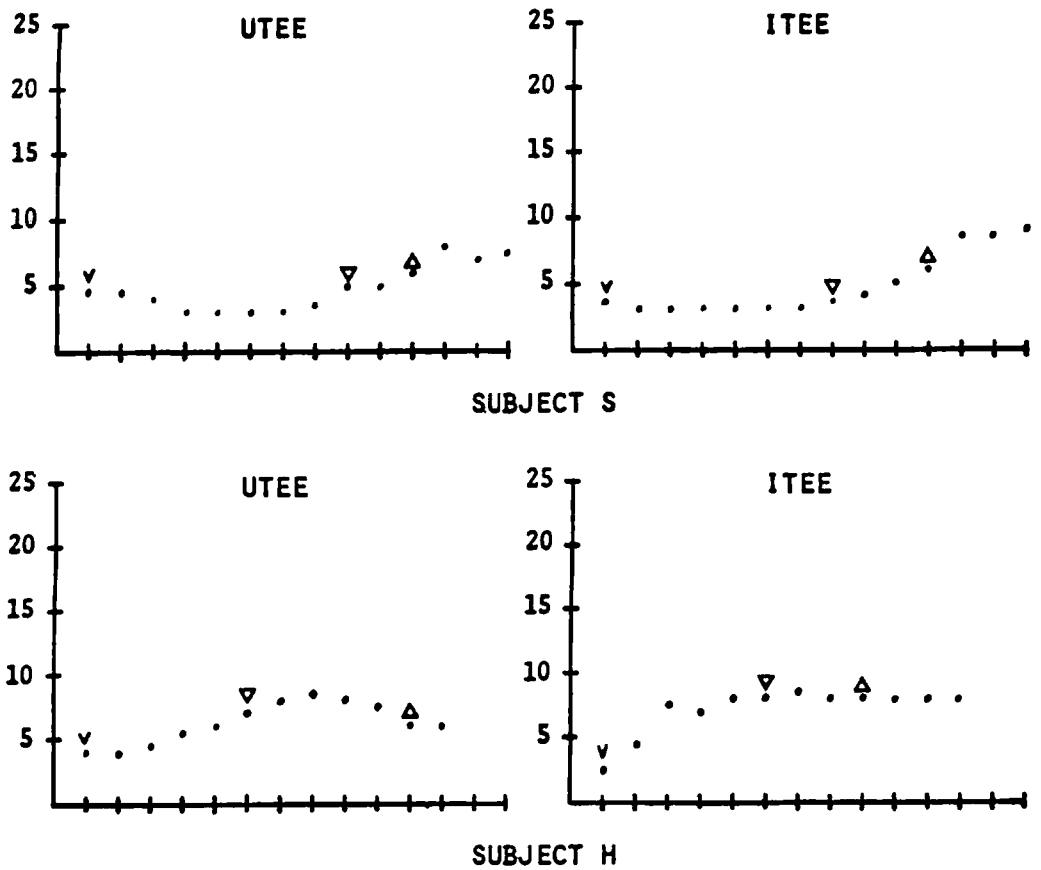


Figure 7. Time course of glottal width for representative samples of the words /utee/ and /itee/ during soft whispered speech. Legend: "v" indicates the frame that occurs at the point or just before the onset of the initial vowel; for explanation of other symbols see Figure 1.

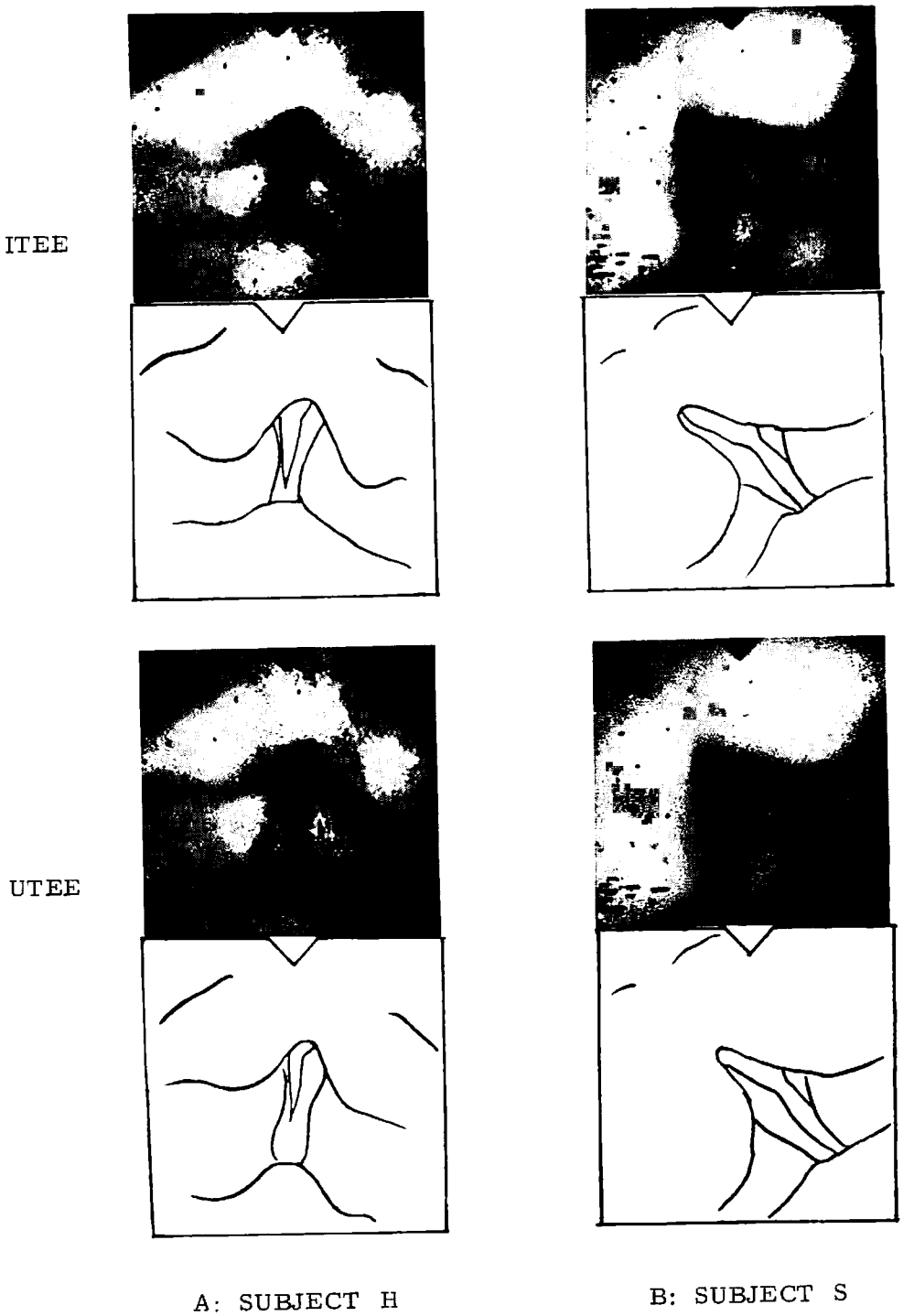


Figure 8. Typical laryngeal state during the pronunciation of /i/ and /u/ in /itee/ and /utee/ in soft whispered speech.

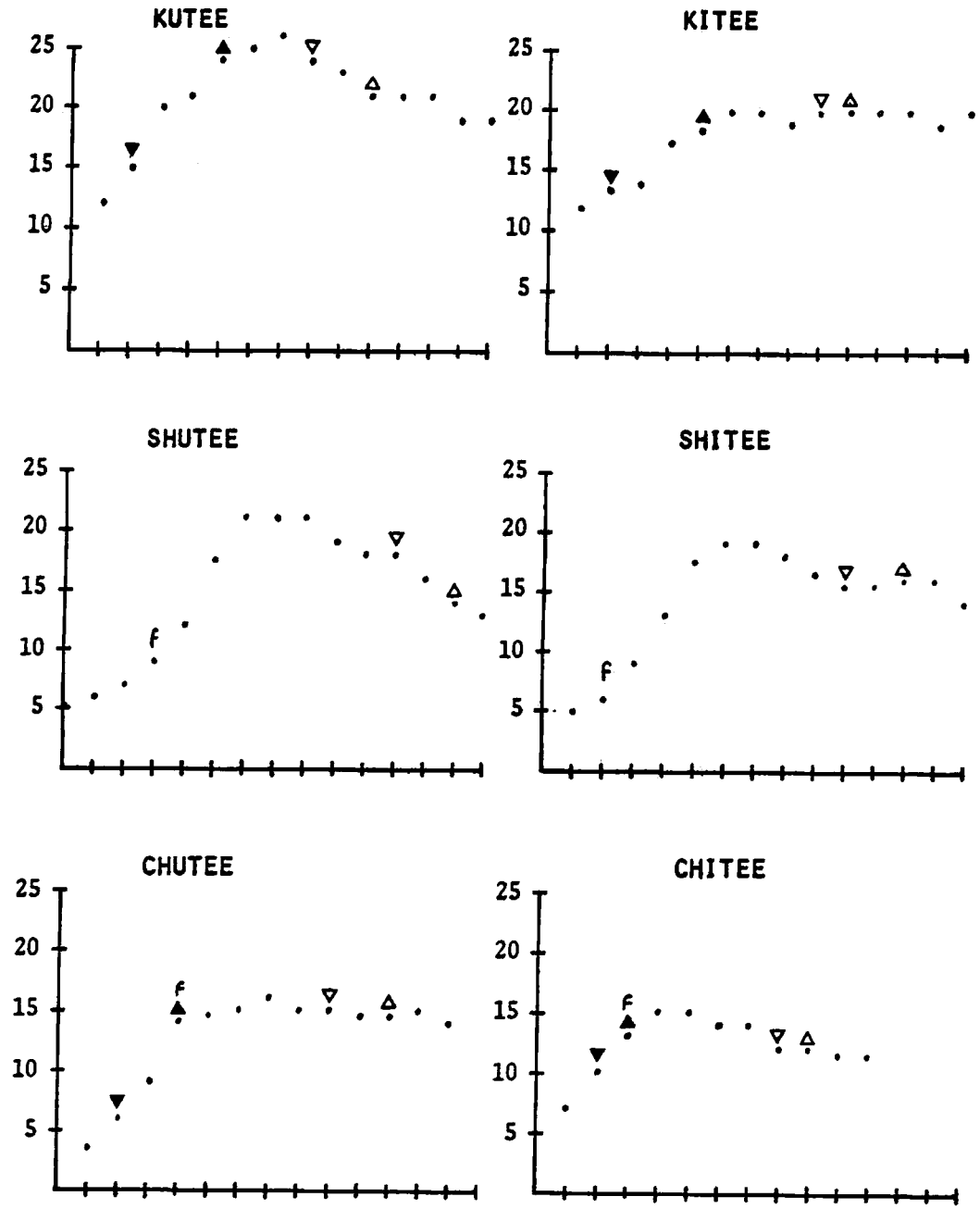


Figure 9. Time course of glottal width for representative samples of devoiced vowels during stage whispered speech (Subject H). For explanation see Figure 1 .

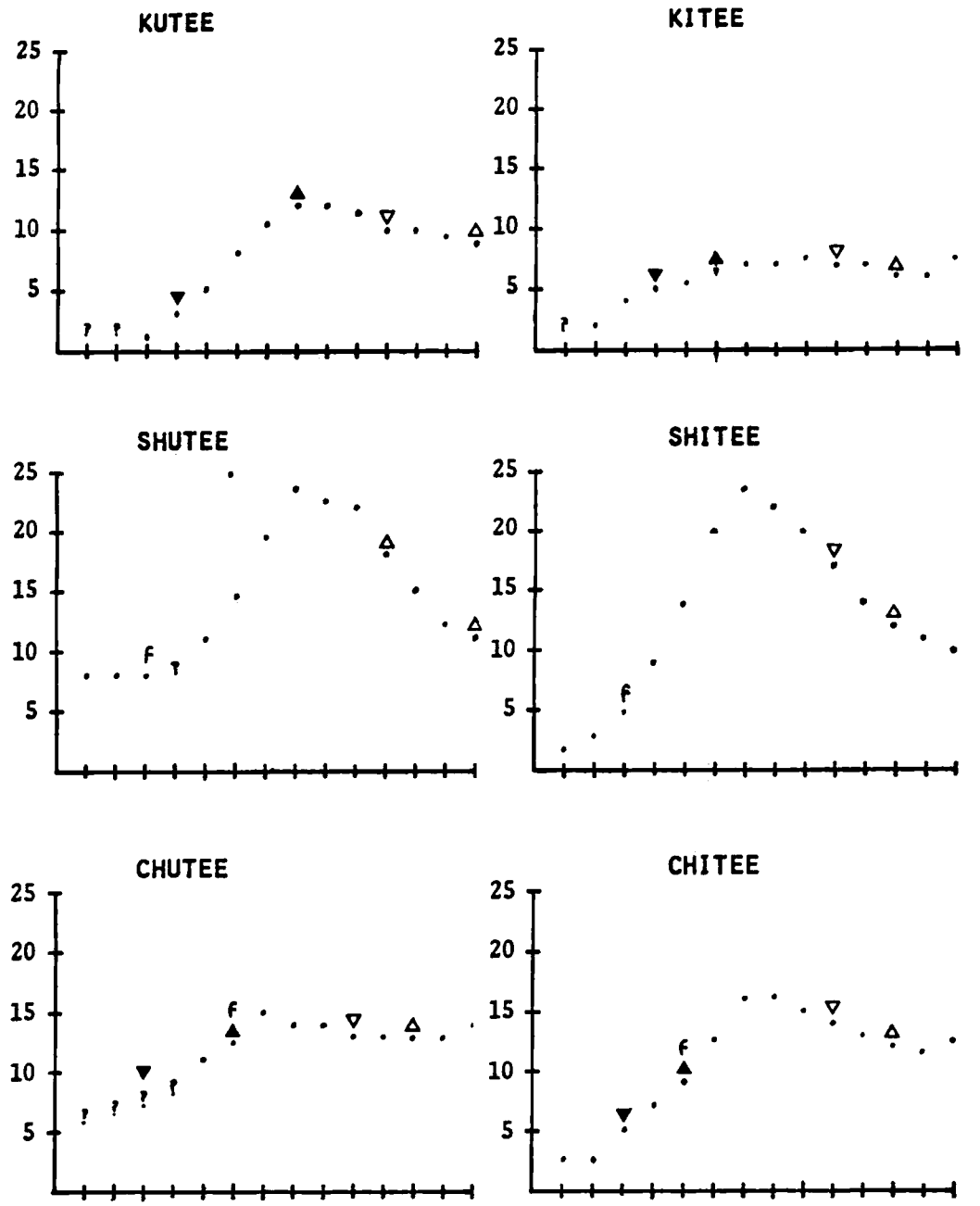
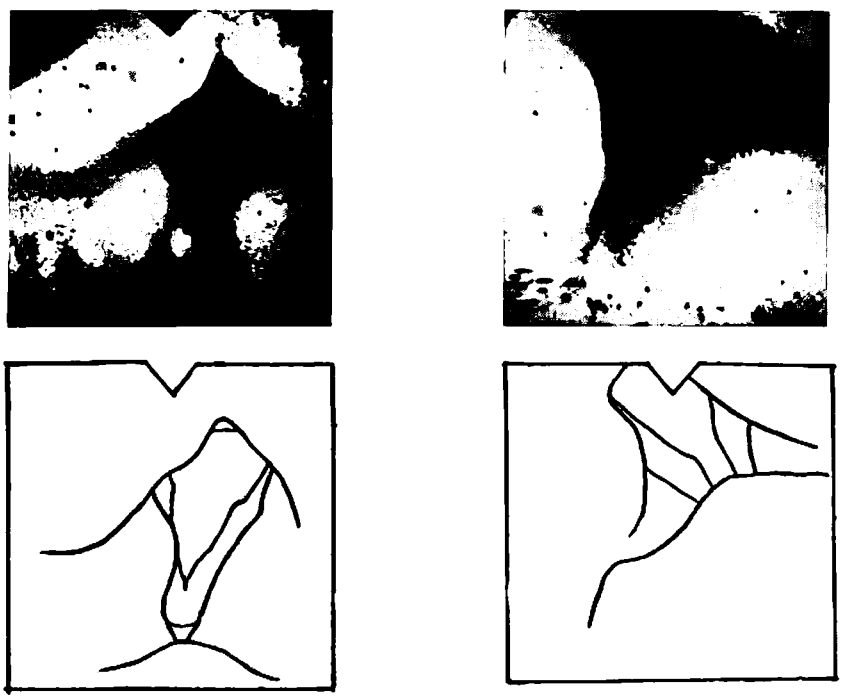


Figure 10. Time course of glottal width for representative samples of devoiced vowels during stage whispered speech (Subject S). For explanation see Figure 1.



A: SUBJECT H

B: SUBJECT S

Figure 11. Photographs showing the typical laryngeal gesture at the point of maximum abduction in stage whispered speech.(Test word: /kitee/).

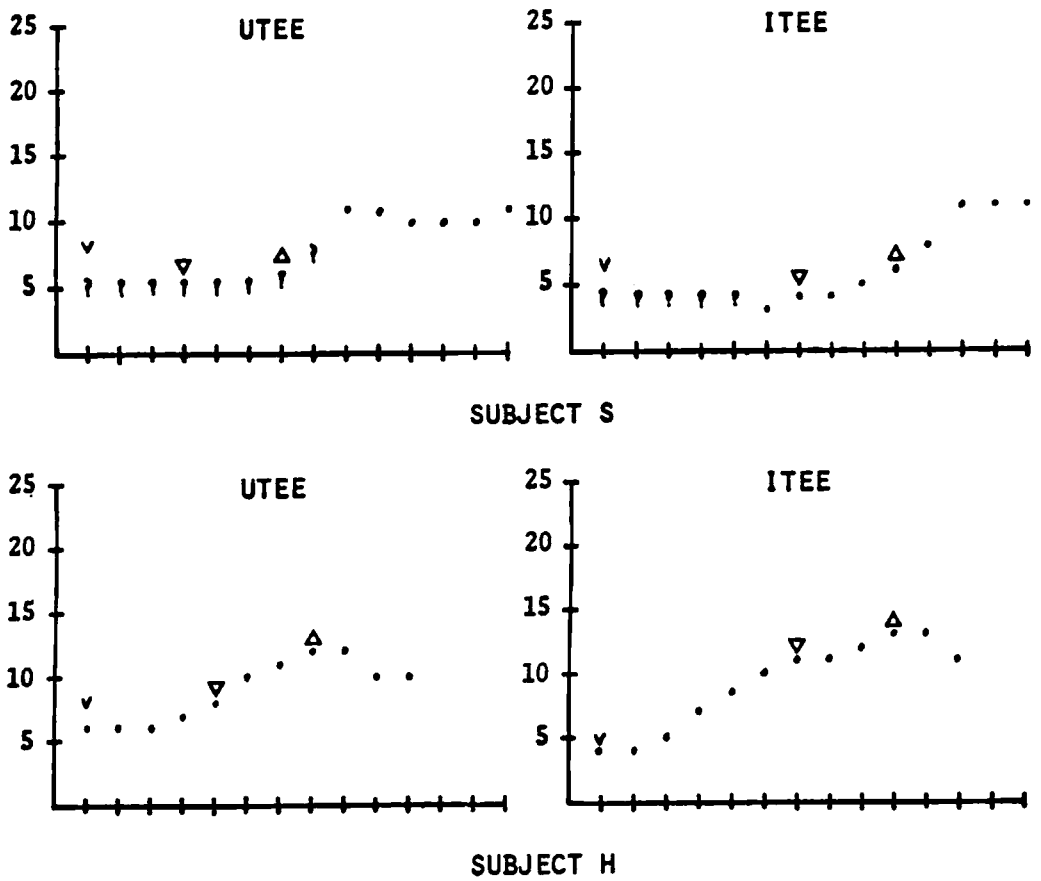


Figure 12. Time course of glottal width for representative samples of the words /utee/ and /itee/ during stage whispered speech. For explanation see Figures 1 and 8.

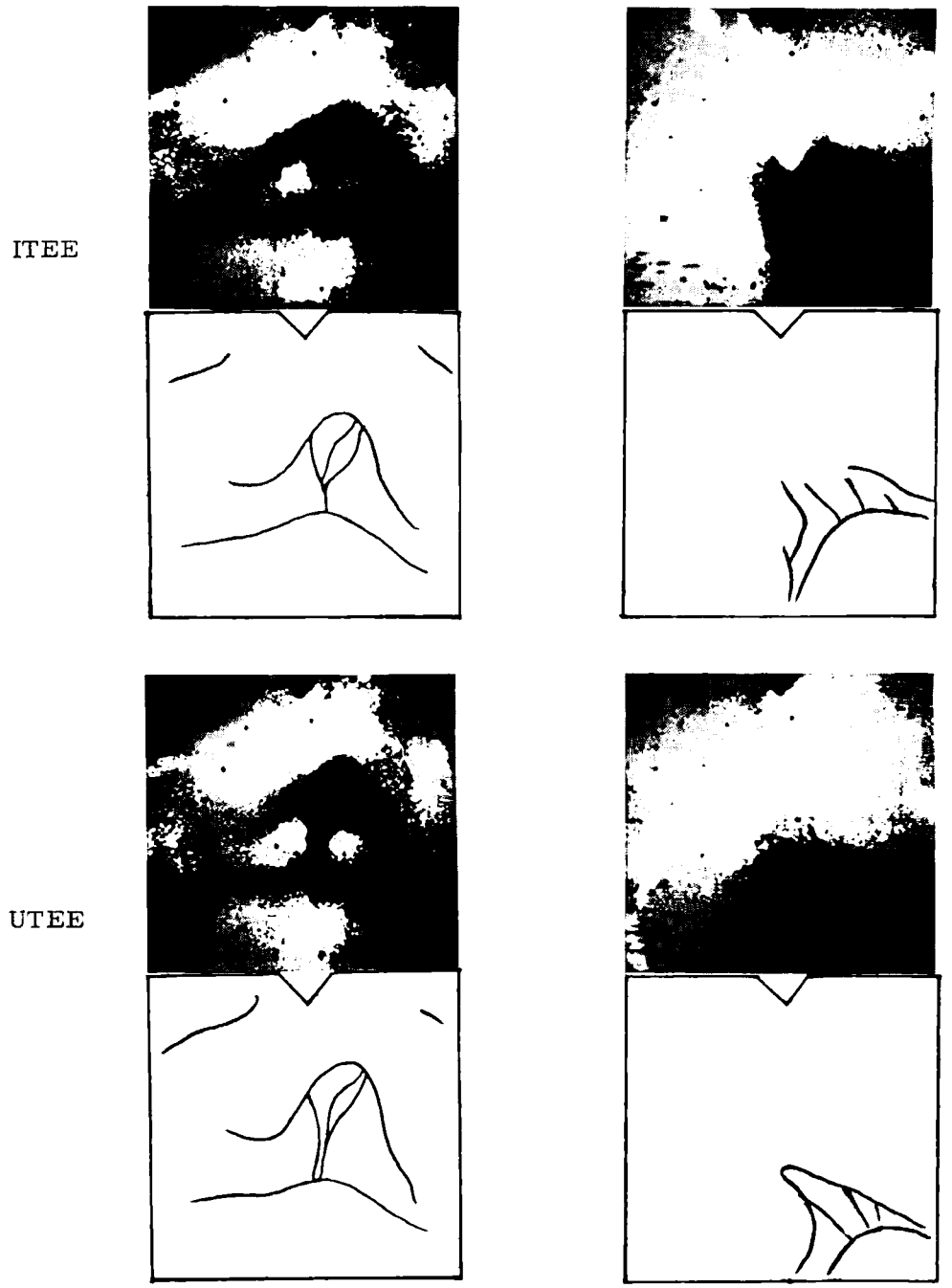


Figure 13. Typical laryngeal state during the pronunciation of /i/ and /u/ in the words /itee/ and /utee/ in stage whispered speech.

Notes

1. The term 'devoiced vowel' as used here is equivalent to the term 'voiceless vowel.'
2. See Han, Mieko. Japanese Phonology: An Analysis Based upon Sound Spectrograms, Kenkyusha, Tokyo, 1962; Abercrombie, David. Elements of General Phonetics, Aldine Publishing Co., Chicago, 1967.
3. See Heffner, R-M. S. General Phonetics, The University of Wisconsin Press, Madison, 1964; Edwards, E. R. Nihongo no Onseigakuteki Kenkyuu, Koseikaku-kan, Tokyo, 1935.
4. See Brosnahan, L. F. and Malmberg, B. Introduction to Phonetics, W. Heffer & Sons, Ltd., Cambridge, 1970.
5. Abercrombie, David (pp. 58-60).
6. Sawashima, M. "Devoicing of Vowels," in Annual Bulletin, Research Institute of Logopedics and Phoniatrics, University of Tokyo, No. 5, pp. 7-14 (1971).
7. Only the forms C$\overset{\circ}{V}$tee and Vtee will be discussed here. Analysis of the others has not been completed.
8. The vocal processes were not always readily identifiable nor even visible at times. When such cases occurred, care was taken to measure the glottal width across the same point of the glottis consistently.
9. There is a strong possibility that in such cases no vowel, devoiced or otherwise, actually is present. This has been suggested by others before and will be explored in detail in a later report.
10. The cartilaginous glottis being partially hidden by the arytenoid masses was also due to the position of the fiberscope lens with respect to the glottis.
11. Dr. T. Ushijima, personal communication.