Ann. Bull. RILP, No. 15, 17 - 30 (1981)

# LARYNGEAL ADJUSTMENTS IN JAPANESE WORD ACCENT*

Hirohide Yoshioka and Hajime Hirose

## Abstract

The aim of the present study is to clarify the role of intrinsic laryngeal muscles in accomplishing various accent types in the Tokyo dialect of Japanese. Specifically, the activity patterns of the thyroarytenoid (VOC) and the posterior cricoarytenoid (PCA) were investigated in addition to that of the cricothyroid (CT). Supplemental cine filming of the glottal image was also made using laryngeal fiberoptic endoscopy. The test utterances contained various meaningful words composed of /VCV/ sequences with or without an accent kernel (C = single obstruent or geminate, V = high front vowel /i/). It was revealed that the VOC showed activity patterns similar to those of the CT, peaking for the accent kernel in general. For the accented first mora in /V⌉CV/ sequence, however, the peak timing of the VOC activity tended to be earlier than that of the CT, although it was approximately coincident with that for the accented second mora in /VCV⌉/ sequence. Furthermore, the peak value of the VOC for the accented first mora was much higher than that for the accented second mora, while the peak CT activities did not show any significant difference between these two types of word accent. These findings suggest that the VOC activity is related not only to the realization of Japanese accent but also to segmental phonetic features such as glottalization in word-initial vowel production in Japanese. As for the PCA, the activity for the voiceless stops or affricates following an accented first mora was smaller than for those in the second mora not preceded by an accent kernel, regardless of whether or not this second mora has an accent kernel. The results were in good agreement with those of fiberoptic observation in that the temporal patterns of glottal width always showed larger openings for the latter group.

---

* A version of this paper was presented at the 96th Meeting of the Acoustical Society of America in Honolulu, 27 November - 1 December, 1978.

## Introduction

The Japanese word accent, typically in the Tokyo dialect, is characterized by the sequence of subjective binary pitch pattern—high or low— for each mora. More specifically in terms of its distinctive feature, the accent type for an individual word can be defined by the following, introducing the term "accent kernel" which indicates the location of an abrupt drop of (phonological) pitch. First, does the word possess an accent kernel or not? Secondly, to which mora is the single kernel, if any, attached? (e.g., Hattori, 1960). In addition, as a non-distinctive (configurational) feature, the first mora of any word, at least typically in isolated pronunciation, always starts on a lower pitch compared with the second mora, unless the accent kernel is attached on the first mora. Accordingly, a $/V_1CV_2/$ sequence for example can form three distinct words, $/V_1\urcorner CV_2/$, $/V_1CV_2\urcorner/$ and $/V_1CV_2/$, which are introspectively recognized as [high-low], [low-high], and [low-high], respectively. Although the acoustic correlates with manifestation of various accent types of course include the intensity and durational changes of each mora as well as the perturbation of format frequencies, the contour of the $F_0$ is considered most substantially involved in Japanese. Therefore, the relationship between the continuous temporal pattern of pitch contour and the discrete linguistic information of accent type has been a major project in the acoustic study of speech (e.g., Fujisaki and Sugito, 1977).

Apart from these acoustic viewpoints, many studies in speech production research, mostly using electromyographic technique (EMG), have been carried out to clarify the relationship between the laryngeal muscle activity and the pitch control, and proved that the cricothyroid (CT) is a principal pitch raiser both in speech and phonation (e.g., Hirano et al., 1969; Sawashima et al., 1969; Shimada and Hirose, 1971; Atkinson, 1978; and Sugito and Hirose, 1978). The purpose of this study is to further investigate the participation of laryngeal muscles in accomplishing various (pitch) accent types in the Tokyo dialect. Specifically, the activity patterns of the thyroarytenoid muscle and the posterior cricoarytenoid muscle were investigated along with that of the cricothyroid muscle.

## Method and procedure

The electromyographic (EMG) data were obtained by use of bipolar hooked-wire electrode techniques (Basmajian and Stecko, 1962). The electrodes, consisting of a pair of platinum-tungsten alloy wires (50 microns in diameter with isonel coating), were inserted perorally into the posterior cricoarytenoid muscle (PCA) under indirect laryngoscopy with the aid of a specially designed curved probe (Hirose et al., 1971). Before the insertion, topical anaesthetic was applied to the mucous membrane of the hypopharynx using a small amount of 4% lidocaine spray (Xylocaine). For placement into the thyroarytenoid muscle (VOC) and the cricothyroid muscle (CT), a percutaneous approach was adopted using hypodermic needles (26 gauze and 1-1/2 inches in length) as the guides (Hirano and Ohala, 1969).

The interference voltages of the EMG signals were recorded on an FM multichannel data recorder together with the acoustic signal. The action potentials were then fed into a digital computer system and sampled at a rate of 200/sec, after being rectified and integrated over a 5 msec time window, for further processing to obtain the muscle activity patterns for ensemble-averaged tokens (Kewley-Port, 1977). The figures to be presented in this paper represent activity patterns aligned with reference to the onset of intervocalic consonants and smoothed with a time constant of 35 msec before ensemble-averaging.

As for the fiberoptic data, the glottal view through a flexible laryngeal fiberscope (Olympus VF-O type, 4.5 mm in outer diameter) was photographed with a cine camera at a rate of 60 frames/sec simultaneously with the EMG and speech signals for some selected tokens of each utterance type (Sawashima and Hirose, 1968; Sawashima, 1976). On each frame of the films, the distance between the vocal processes, which is considered one of the best indicators of glottal width, was measured during pertinent intervocalic consonantal segments. The plots of the glottal aperture as a function of time on an arbitrary scale will be shown below, in comparison with the EMG activity patterns of the abductor (PCA).

Along with the EMG and fiberoptic data processing, acoustic analyses were made, including the $F_0$ contours by use of an adaptive autocorrelation method (Lukatela, 1973), as well as the audio envelope curves, which serve as a gross identification of each acoustic event.

| 「それは ___ です」 | "sorewa ___ desu" |
|---|---|
| 意志 | /i⌉si/ |
| 石 | /isi⌉/ |
| 逸史 | /iQsi/ |
| 一子 | /i⌉Qsi/ |
| 維持 | /i⌉zi/ |
| 意地 | /izi⌉/ |
| 位置 | /i⌉ci/ |
| 一 | /ici⌉/ |
| 粋 | /iki/ |
| 息 | /i⌉ki/ |
| 生き | /iki⌉/ |
| 一気 | /i⌉Qki/ |
| 意義 | /i⌉gi/ |

Table 1. List of test words and their carrier sentence.

Two native adult male speakers of the Tokyo dialect served as the subjects. Among the possible phoneme sequences composed of /VCV/ (C = single obstruent or geminate, V = high front vowel /i/) with or without an accent kernel, only meaningful words in Japanese were selected as the test words. As shown in Table 1, some of these phoneme sequences may form two distinct meaningful words depending on the presence or absence of the accent kernel and on its location, if any. The position of each accent kernel is indicated by the diacritic symbol "⌉". The subjects were asked to pronounce each of the test words embedded in a frame sentence "sorewa ----- desu," at least 15 times, in random order. No particular instruction was given to the subjects about the vocal intensity, the pitch, or the speaking rate.

## Results

Figure 1 contains the averaged electromyographic (EMG) activity patterns of the thyroarytenoid muscle (VOC) and the cricothyroid muscle (CT) for two different words composed of the same phoneme sequence /isi/, where the location of an accent kernel varies. In addition, the $F_0$contours and the corresponding audio envelopes of representative single tokens for the production of both utterances embedded in the frame sentence "sorewa ----- desu" are included at the bottom. The vertical line across the time axis in each graph corresponds to the voicing offset of the vowel [i] preceding intervocal [ʃ] segment, which served as the line-up point for the EMG averaging. It may be clear, at a glance, that there is a single peak of the EMG activity corresponding to each accent kernel not only in the CT but also in the VOC. That is, for the first mora accented word, the VOC activity starts to abruptly increase around at -200 msec before the line-up point, reaching its maximum at -120 msec. The CT also begins to be rapidly activated at -200 msec, and it peaks approximately at -60 msec. On the other hand, for the second mora accented word, the activity

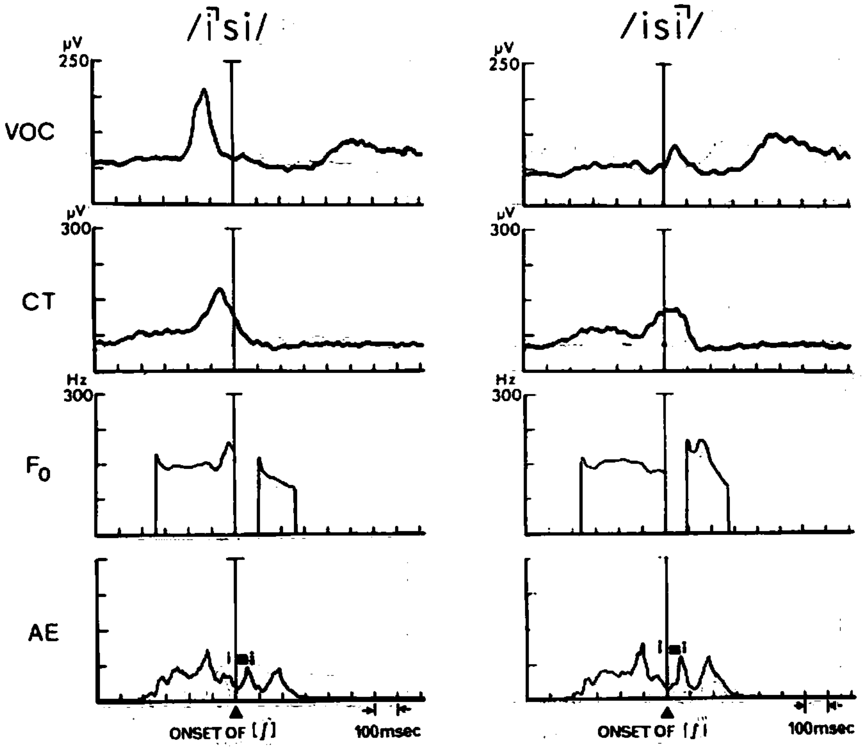

Fig. 1. Averaged EMG curves of VOC and CT, representative Fo contours, and corresponding audio envelope curves for /i¬si/ and /isi¬/.

pattern of the VOC as well as that of the CT shows a slight temporal suppression before the line-up, then both peaking around +50 msec. Incidentally, the VOC activity for either utterance type reveals a moderate reactivation initiated at the end of pronunciation, while the CT does not. This particular activity of the VOC seems to be correlated to the so-called glottal stop-like gesture for the termination of phonation, which is often reported in the literature (e.g., Hirose, 1971).

Figure 2 contains the test words, each having a voiced fricative /z/ instead of the voiceless one in the intervocalic position with accent distinction. Despite the difference in voicing feature between these intervocalic fricatives, the outlines of Figure 2 are considerably similar to those of Figure 1. It also implies that the difference of EMG activity directly correlated to the accent distinction is well preserved notwithstanding the replacement by the voiced fricative /z/; curves of the VOC and the CT are equally characterized as single peaked activity patterns regardless of the accent type distinction. The peak values of the VOC, however, are quite different between these two accent types in that the

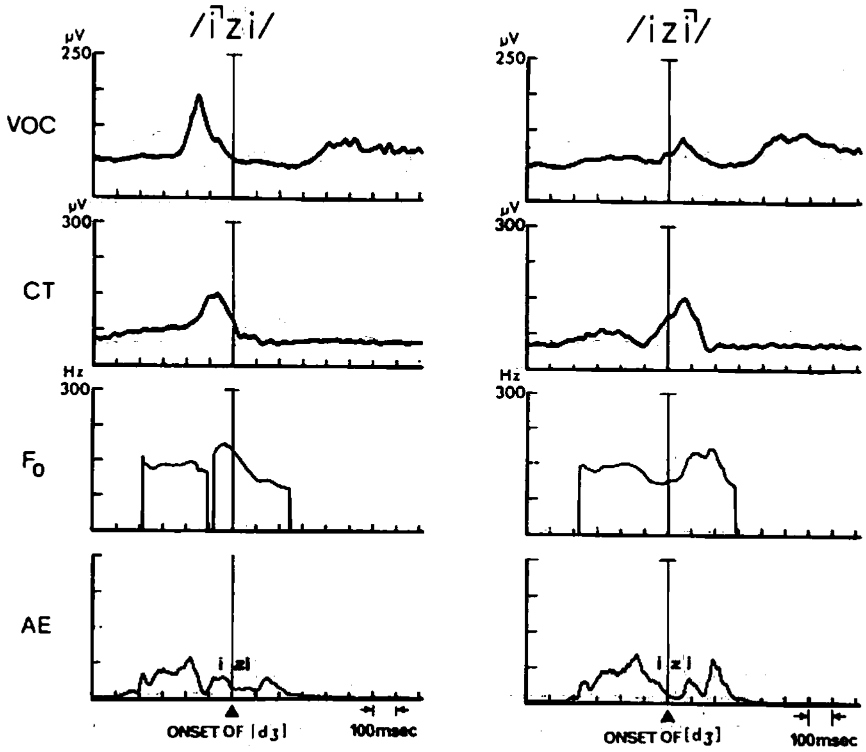

Fig. 2. Averaged EMG curves of VOC and CT, representative Fo contours, and corresponding audio envelope curves for /i⌝zi/ and /izi⌝/.

peak for the accented first mora is far larger than that for the accented second mora. In contrast, the CT does not show such noticeable difference in the peak value attributed to the location of the accent kernel.

Figure 3 shows the EMG patterns for a so-called "non-accent" type of word, in comparison with those for the same but first mora accented phoneme sequence. As implied by the $F_0$ contour, the "non-accent" word is accompanied by a relatively higher pitch during the second mora, compared to the preceding first mora. In this connection, it is worth noting here that the CT activity shows a

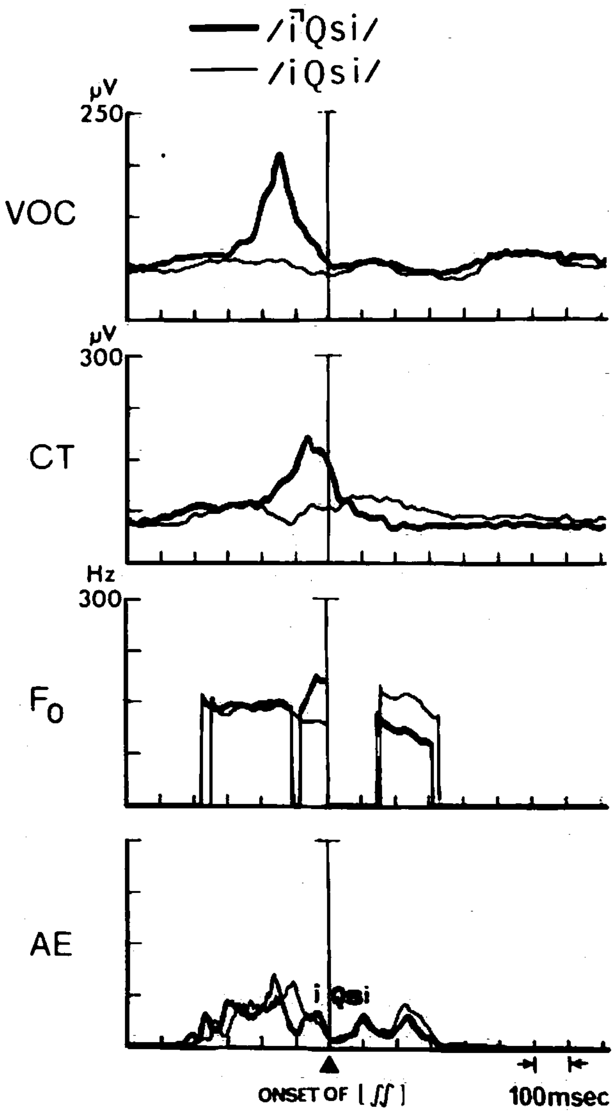

Fig. 3. Averaged EMG curves of VOC and CT, representative Fo contours, and corresponding audio envelope curves for /i⌉Qsi/ and /iQsi/.

slight suppression prior to the first mora production, followed by a gradual reactivation with a small but broad plateau-like peak. On the other hand, the VOC activity for the "non-accented" word does not reveal any prominent peaks; rather it may be characterized as having a tiny suppression across the line-up. In contrast, the EMG patterns for the first mora accented word have large distinct peaks both in the VOC and the CT, which comparably resemble those for other first mora accented words shown in the previous graphs. In addition, it is further interesting to note that the peak timings of the VOC and the CT for this first mora accented word are different; the peak of the VOC is earlier than that of the CT. Consequently, the VOC activity is almost down at the noise level around the line-up, while the CT still shows some activity during the period. Similar findings can also be retrospectively pointed out from the other first mora accented word shown in the previous figure.

Figure 4 contains the optimal phoneme sequence /iki/, which exists among meaningful words in Japanese with three different accent types. Here, it is ascertained that what has been mentioned above with respect to the EMG activity related to accent distinction can be generalized so far as /VCV/ sequence is concerned. Namely, the peak activity corresponding to each accent kernel can be identified both in the VOC and the CT, the peak value difference of the VOC based on the location of accent kernel is clear, and the peak timing difference for the first mora accented word between the VOC and the CT is also obvious. Incidentally, for the "non-accented" type, these two muscles show approximately similar patterns until the line-up to those of the second mora accented one, then revealing a slight and broad peak like a plateau, particularly in the CT. This finding may be of interest especially in view of the phonological fact that such two mora words as /VCV/ and /VCV˥/ are equally considered as [low-high] in the Tokyo dialect.

Figure 5 contains the EMG activity patterns of the posterior cricoarytenoid (PCA) and representative plots of the glottal width during the intervocalic voiceless consonantal segments for the other subject, together with single patterns of the $F_0$ contours and the audio envelope curves at the bottom. The glottal opening curves during the intervocalic [ʃ] segments do not seem notably different, particularly in terms of the degree of the maximum

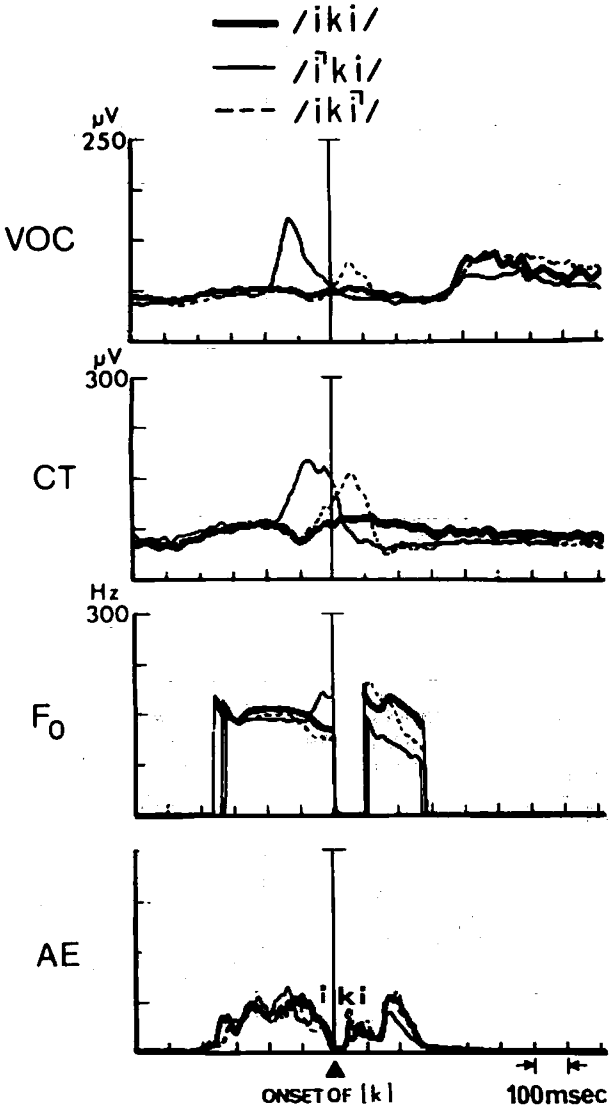

Fig. 4. Averaged EMG curves of VOC and CT, representative Fo contours, and corresponding audio envelope curves for /iki/, /i⌉ki/ and /iki⌉/.

opening. In accordance with this finding, the PCA activity patterns are also similar to each other. Therefore, it may be that the glottal opening gesture for the voiceless single fricative production is not substantially effected by whether or not the affiliated mora is attached with an accent kernel.

Figure 6 compares a geminate combination /Qs/ in intervocalic environment, where the accent type differs. From this graph henceforth, the activity patterns of the PCA will be presented with the standard deviation bracket curves in addition to the averaged ones.

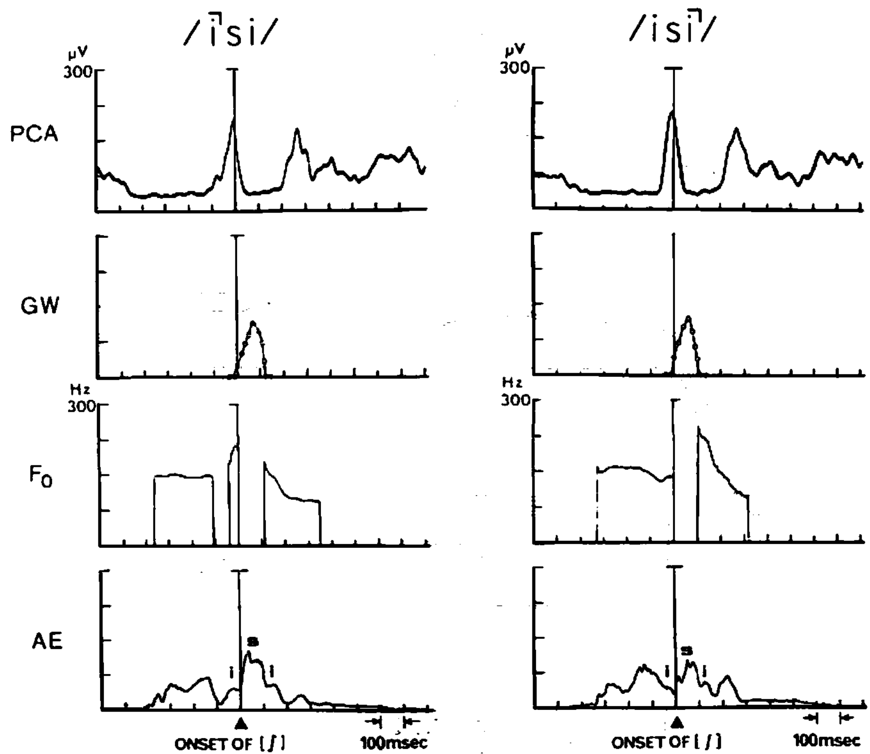

Fig. 5. Averaged EMG curves of PCA, representative glottal width curves and corresponding audio signals for /i˥si/ and /isi˥/.

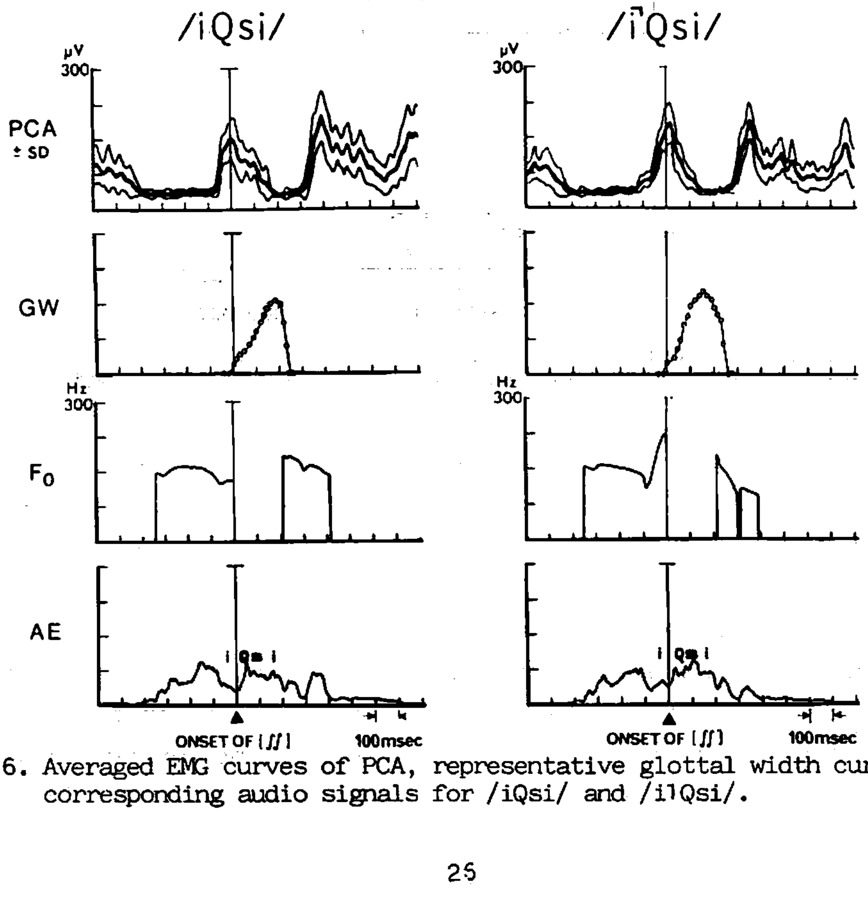

Fig. 6. Averaged EMG curves of PCA, representative glottal width curves and corresponding audio signals for /iQsi/ and /i˥Qsi/.

It may be obvious here that there is a single peak in the PCA activity curve as well as in the glottal aperture plots for either utterance type, and their peak values are approximately similar to each other despite the accent type difference. The timing of the maximum opening, however, seems slightly different; the peak for the "non-accented" word is a little late, although the duration of the glottal opening itself is comparable to that of the first mora accented word. Nevertheless, the PCA patterns look contradictory, in that the PCA curve after the peak, i.e. the line-up for the "non-accent" word, shows a gradual decrease, while that for the first mora accented one comes back to the low level as quickly as the preceding increasing phase. Of course, as is implied by the standard deviation bracket curves, it is doubtful whether such a minute difference is consistent or statistically significant.

Figure 7 contains a pair of words, each having a voiceless affricate phoneme /c/ in the intervocalic position. Here, it is clear that the peak activity of the PCA as well as the peak opening of the glottis for the voiceless affricate is smaller in a post-accented position, compared to the pre-accented position. The PCA activity patterns in this graph further reveal another point of interest: The activation of the PCA for the post-accented voiceless affricate starts approximately at -250 msec, while it does not begin until -70 msec before the line-up for the pre-accented one. The activation for the former case should be considered as occurring quite early, particularly in view of the estimation that the time lag of mechanical contraction from the electrical activity during voluntary movement is unlikely to be far beyond 100 msec. In this connection, the audio envelope curve for this first mora accented word is suggestive; the preceding accented vowel /i⌉/ is produced with a glottal attack, while it is not the case for the initial "non-accented" vowel in /ici⌉/. Taken together, it is conceivable that the early initiation of the PCA activity is not directly correlated to the following opening gesture for the voiceless affricate, but rather to the release of the tight closure of the glottis due to the glottal attack gesture for the word-initial accented vowel (e.g., Hirose and Gay, 1973).

Figure 8 further compares three different words composed of a same phoneme combination /iki/, where the accent type differs.

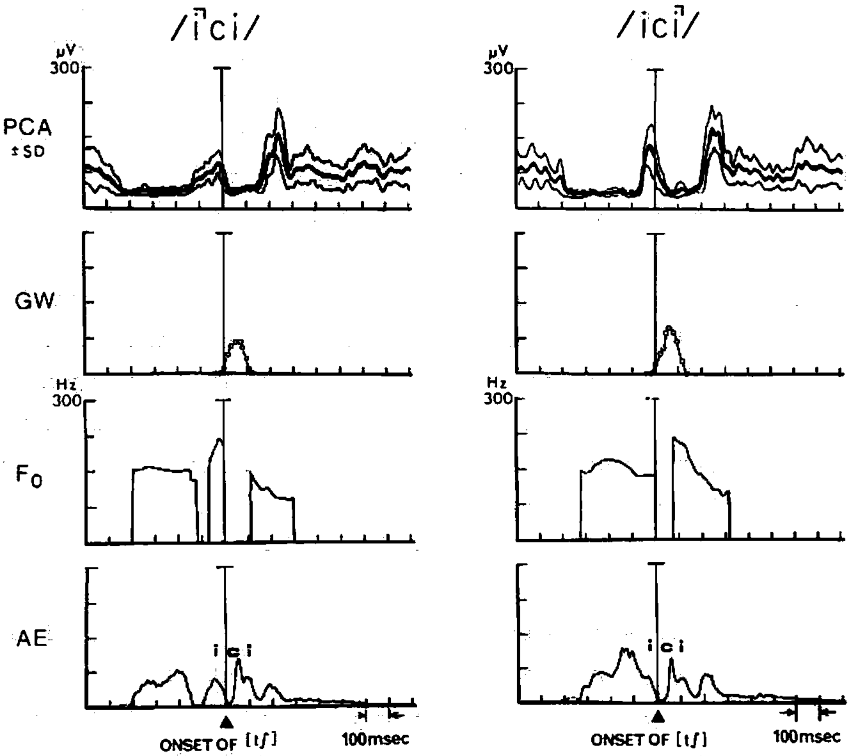

Fig. 7. Averaged EMG curves of PCA, representative glottal width curves and corresponding audio signals for /i˥ci/ and /ici˥/.

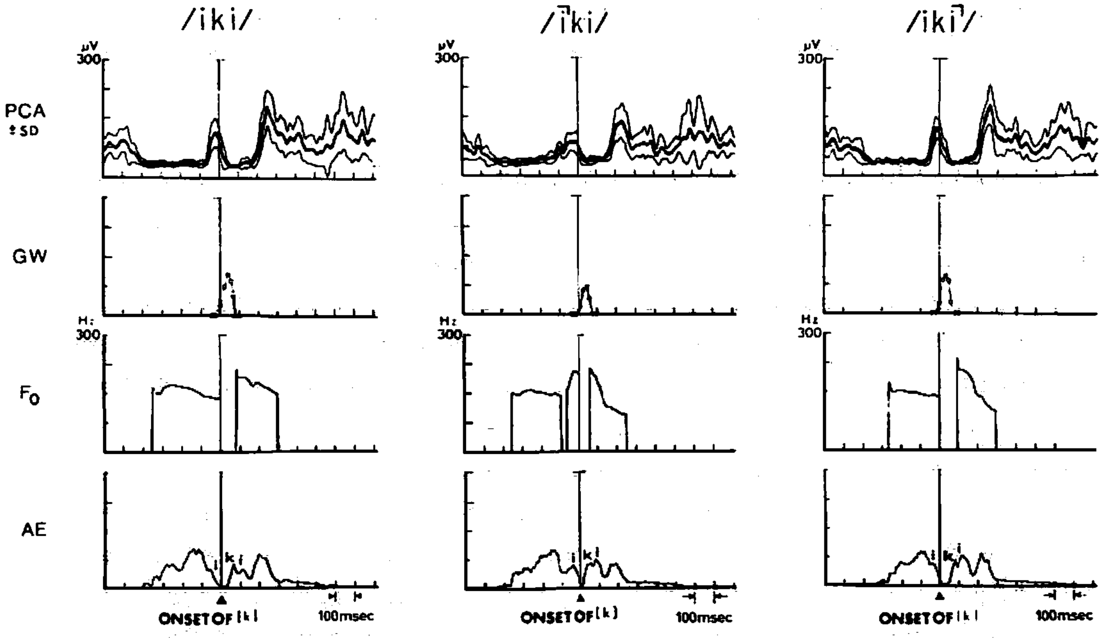

Fig. 8. Averaged EMG curves of PCA, representative glottal width curves and corresponding audio signals for /iki/, /i˥ki/ and /iki˥/.

Here, what has been pointed out in the previous graph is again clear despite the difference of intervocalic obstruent. The peak glottal opening for pre-accented voiceless stop at the right is larger than that for the post-accented one in the middle of the graph. In addition, the early initiation of PCA activity prior to the post-accented stop production is quite obvious, although the audio envelope curve at the bottom does not clearly show a completely silent period preceding the glottal attack. As for the "non-accented" word at the left, the maximum glottal opening during the intervocalic voiceless stop is, unlike the post-accented stop production mentioned above, as large as that for the pre-accented stop. Furthermore, the temporal patterns of the PCA for these two "non-accented" and pre-accented stop productions are strikingly similar to each other and clearly different from that for the post-accented one. These findings on the glottal opening gesture and its neural control related to the accent distinction may be of particular interest, especially in consideration of the fact that the second mora of a "non-accented" word is usually produced rather higher compared to the preceding first mora, and consequently the pitch contour resembles that for the second mora accented word as far as the anterior portion before the accent kernel is concerned.

## Acknowledgment

This work was supported in part by NINCDS Grant NS-13870 and BRS Grant RR-05596 to Haskins Laboratories, New Haven CT, U.S.A.

## References

Atkinson, J.E. (1978); Correlation analysis of the physiological factors controlling fundamental voice frequency. Journal of the Acoustical Society of America, 63, 211-222.

Basmajian, J., and G. Stecko (1962); A new bipolar indwelling electrode for electromyography. Journal of Applied Physiology, 17, 849.

Fujisaki, H., and M. Sugito (1977); Onsei no butsuriteki seishitsu. In S. Ohno, and T. Shibata, (Eds.) Nihongo. Tokyo: Iwanamishoten, 5, pp 63-106.

Hattori, S. (1960); Onseigaku no hoohoo. Tokyo: Iwanamishoten.

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

Hirano, M., J. Ohala, and M. Vennard (1969); The function of laryngeal muscles in regulating fundamental frequency and intensity of phonation. Journal of Speech Hearing Research, 12, 616-628.

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

Hirose, H., and T. Gay (1973); Laryngeal control in vocal attack: An electromyographic study. Folia Phoniatrica, 25, 203-213.

Hirose, H., T. Gay, M. Strome, and M. Sawashima (1971); Electrode insertion technique for laryngeal electromyography. Journal of the Acoustical Society of America, 50, 1449-1450.

Kewley-Port, D. (1977); EMG signal processing for speech research. Status Report on Speech Research (Haskins Laboratories), SR-50, 123-146.

Lukatela, G. (1973); Pitch determination by adaptive autocorrelation method. Status Report on Speech Research (Haskins Laboratories), SR-33, 185-193.

Sawashima, M. (1976); Current instrumentation and techniques for observing speech organs. Technocrat, 9-4, 19-26.

Sawashima, M., and H. Hirose (1968); A new laryngoscopic technique by use of fiberoptics. Journal of the Acoustical Society of America, 43, 168-169.

Sawashima, M., T.J. Gay, and K.S. Harris (1969); Laryngeal muscle activity during vocal pitch and intensity changes. Status Report on Speech Research (Haskins Laboratories), SR-19/20, 211-220.

Shimada, Z., and H. Hirose (1971); Physiological correlates of Japanese accent patterns. Annual Bulletin (Research Institute of Logopedics and Phoniatrics, University of Tokyo), 5, 41-50.

Sugito, M., and H. Hirose (1978); An electromyographic study of the Kinki accent. Annual Bulletin (Research Institute of Logopedics and Phoniatrics, University of Tokyo), 12, 35-52.