First International Conference on
Spoken Language Processing (ICSLP 90)
Kobe, Japan
November 18-22, 1990

STERNOHYOID MUSCLE ACTIVITY AND PITCH CONTROL AT THE ONSET OF UTTERANCES

Zyun'ici B. Simada
Department of Physiology, School of Medicine, Kitasato University
Sagamihara, Kanagawa, 228 Japan
Satoshi Horiguchi, Seiji Niimi, Hajime Hirose*
Research Institute of Logopedics and Phoniatrics, University of Tokyo
Bunkyoo, Tokyo, 113 Japan

ABSTRACT

Sternohyoid activity involving accentual contrast was examined in a male speaker of Tokyo Japanese. The contrast of interest was in the initial position of sentence frames; thus, unaccented vs. accented pairs of two-mora meaningful words were compared in terms of pitch movements and electromyographic activity. The utterance types all showed such activity preceding the audio onset. For words with accent, this activity became still for a short period and was followed by consistent reactivation; moreover, the mean level of this pre-audio activity was lower for one type of utterances with initial accent. On the other hand, words with no accent showed a different timing for the offset of the pre-audio activity. These findings lend support to the role played by the sternohyoid muscle in pitch lowering, and specifically in the accentual distinction in Japanese.

1. INTRODUCTION

The quest for the pitch-lowering mechanism used during speech has been both intriguing and frustrating. In the larynx, it has been difficult to find anatomical structures that can strengthen and weaken the longitudinal tension of the vocal folds antagonistically.

Electromyographic (EMG) experiments and other studies to date have clarified that the cricothyroid (CTh) muscle is primarily responsible for tightening the folds. But, on one hand, although the pitch once raised falls with a relaxation of the CTh, this process of falling still needs to be assisted by the active control of a group of infrahyoid muscles. According to Sonninen [10], this muscle group adjusts the framework composing the larynx as a whole, thereby having an effect on the effective mass of the vocal folds or the forces exerted externally on them. In this group, the sternohyoid (SH) muscle has been most studied and is, in fact, reported to be active in pitch falling in languages such as Chinese [5], Thai [3], English [1], Dutch [2] and the Kinki dialect of Japanese [6, 11]. However, the SH activity in question does not occur for every speaker; we were unable to find any relevant activity in previous EMG experiments on Tokyo Japanese [8, 9]. What has made the problem more complicated is that the SH muscle is also involved in segmental articulatory gestures, e.g. jaw opening, tongue lowering and so forth [7].

Here, we report on a speaker of Tokyo Japanese who used a consistent patterning of the SH muscle for the accentual contrast in sentence-initial syllables. The pattern was not so simple as that observed in Osaka Japanese by Sugito and Hirose [11] or very recently by Kori et al. [6]. Sugito and Hirose describe one female speaker who showed a burst of activity or nothing depending on whether utterances began with a low- or high-pitched accent. A similar alternation of SH activity was replicated by our subject, but we feel that it is possible to interpret this correctly only by taking its precise timing into account.

*Current affiliation: Department of Neurootology, Tokyo Metropolitan Neurological Hospital, Hutyuu, Tokyo, 183 Japan.

2. METHOD

2.1. Subject

The subject in the present study was one of the authors, a male speaker of Tokyo Japanese, who lived for eight years as a child in Itoigawa, a city in Niigata Prefecture, where a dialect similar in word accent to Tokyo Japanese is spoken, and then moved to a city on the outskirts of Tokyo.

2.2. Speech material

The EMG experiment was carried out twice on different days. For both sessions, we prepared pairs of sentences which were contrastive as to whether an accent appeared in their initial position or not. The test words used in Experiment 1 were *umi* "giving birth" and *u^mi* "sea", the former being unaccented and the latter accented on the first syllable (we will use the symbol ^ hereafter to indicate the accent kernel; for detailed Japanese phonology, see Vance [13]). This pair of words was compared in three sentence frames:

... ni iku "(it) goes to ..."
ari wa ... ni iku "an ant goes to ..."
ka^me wa ... ni iku "a turtle goes to ..."

In Experiment 2, the word pairs

imi "mourning, abstinence" vs. *i^mi* "meaning"
ame "candy" vs. *a^me* "rain"
kimi "you" vs. *ki^ne* "pestle"
kani "crab" vs. *ka^mi* "god"
himo "string" vs. *hi^me* "princess"
hane "wing" vs. *ha^mu* "ham"

were prepared and pronounced in the sentence frame *... o miyo^o* "let's look at ...". We chose these stimuli to look into the segmental influence of the sounds [a, k, h]. As a control of Experiment 1, the utterances *umi / u^mi ni iku* were also included.

2.3. Data recording and analysis

In both experimental sessions, bipolar hooked-wire electrodes were used to derive the action potential from the SH muscle (the CTh muscle also was recorded, but here we do not discuss its activity). The subject was asked to produce 13 to 15 repetitions of each pair of sentences alternately, with a short pause between the tokens, and recorded on a PCM tape recorder together with the EMG signals. Also, all signals were digitized on-line with 12-bit resolution using a newly devised multichannel signal processor connected to an NEC desktop computer. The EMG signals, after being full-wave rectified, integrated over a period of 5 ms, and finally low-pass filtered, were sampled at a rate of 1 kHz. The audio signal was sampled at a rate of 5 kHz.

The data of interest were transferred to a Hewlett-Packard computer for analysis (from Experiment 1, the on-line recorded data were submitted to the analysis; the data processed from the PCM recordings were used from Experiment 2). Pitch intervals and the onsets or offsets of certain EMG events were demarcated on a display terminal with the aid of a movable cursor, and sequences of pitch periods were computed.

3. RESULTS

Figure 1 compares the integrated traces obtained in Experiment 1. Trace A represents a token beginning with accented [ka], trace B with accented [u], trace C with unaccented [u] and trace D with unaccented [a]. Also shown in the figure are the fundamental frequency contours of the tokens. Examining these examples, we note two main findings. First, all of the utterance types showed activity preceding the audio onset. This "pre-audio" activity becomes still for a short period during the accented words (A and B) and then shows reactivation. On the other hand, during the unaccented words (C and D), the pre-audio activity lasts after the onset of voicing and the reactivation is delayed. In

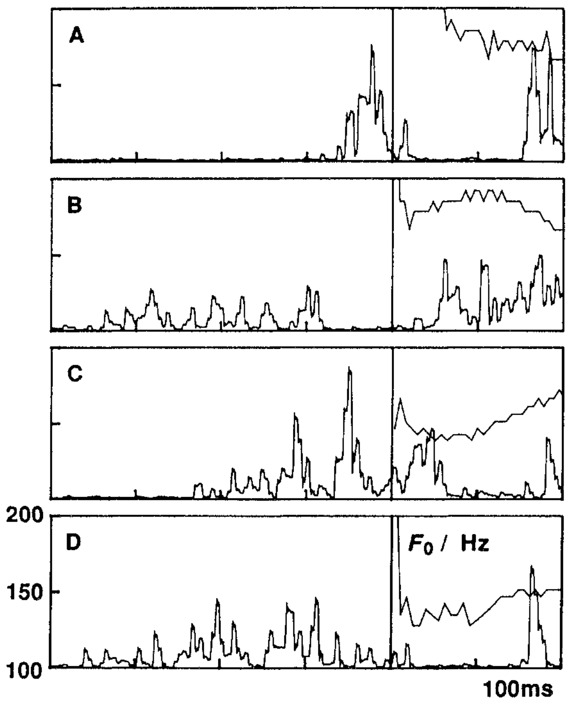

Figure 1. Traces of the full-wave rectified, integrated action potential from the SH during the phase preceding each token of *ka^me wa u^mi ni iku* (A), *u^mi ni iku* (B), *umi ni iku* (C) and *ari wa umi ni iku* (D). Each trace is lined up at the instant of the release of [k] (A) or the onset of the voicing of the tokens (B, C, D), as determined from the acoustic waveform. This is marked in the figure by a vertical line. Fundamental frequency contours are also shown.

addition, the pre-audio activity appears to be lower for the accented token *u^mi* than for the unaccented tokens *umi* and *ari*.

We first made measurements of the onsets and offsets of the pre-audio activity for the tokens of each utterance type (we treated all of the tokens beginning with *ka^me* and those with *ari* as belonging to one utterance type, respectively). The reference point of the measurements is described in the legend of Figure 1. We determined the boundaries of the EMG onsets or offsets by visual inspection, but we sometimes found difficulties in pinpointing their exact locations (in particular, for some tokens of the unaccented *umi* because the EMG activity occurred sporadically up to the part [mi]). Figure 2 and Table I present the median and the hinge distributions of the onset and the offset times of the pre-audio activity, respectively.

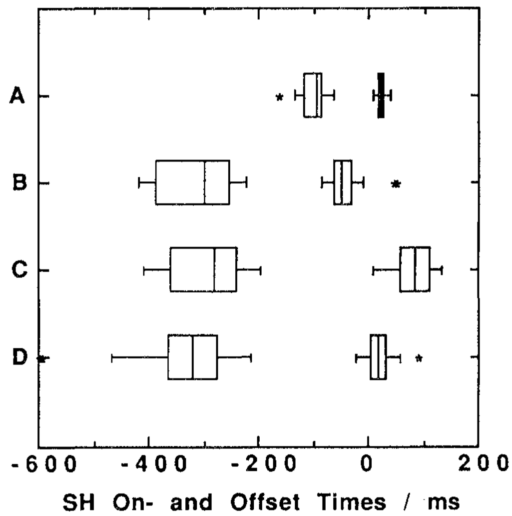

Figure 2. SH onset and offset times for the utterances beginning with *ka^me* (A), *u^mi* (B), *umi* (C) and *ari* (D), respectively. Each box with whiskers shows the median and the hinge distributions of the onset (left) and the offset (right) times [12]. Data points outside the inner fences are plotted with asterisks. Time 0 is the time indicated in Figure 1 by a vertical line. n = 26 for A and D; 13 for B and C.

Table I. The median values of the onset and the offset times of the pre-audio activity as measured from Experiment 1.

	onset time / ms	offset time / ms
A: ka^me	-94.3	22.8
B: u^mi	-299.2	-48.2
C: umi	-281.2	83.4
D: ari	-322.1	14.5

3.1. EMG timing for accented words

The utterance type beginning with an initial accent in *ka^me* was characterized by a burst of activity that occurred 94.3 ms earlier than the release of [k] and lasted for a period of 122.0 ms (Figure 2, A). This first burst was followed by an inactive phase and 112.5 ms later by a second active phase. This pattern was consistently observed for all the tokens. The first burst was concomitant with the articulation of the sequence [ka] which usually calls for SH activity. The second activity was assumed to pertain to the sharp fall in pitch for the accented syllable [ka].

For accented *u^mi* (Figure 2, B), the pre-audio activity was suppressed earlier before the onset of voicing (offset time: -48.2 ms) than for unaccented words *umi* and *ari* (Figure 2, C and D), although there were no significant differences in EMG onset between these utterance types. The SH muscle became active 43.2 ms after the voicing started (81.0 ms from the EMG offset). Here we stress that the same pattern of the suppression and the reactivation of the SH was observable during the accented *ka^me*. In Japanese, when utterances begin with an accented syllable, the pitch starts at a high level and displays a sharp fall toward the next syllable. Consequently, it is reasonable to conclude that the second reactivation of SH is closely relevant to this pitch fall.

3.2. EMG timing for unaccented words

In contrast to accented *u^mi*, unaccented *umi* and *ari* showed a pre-audio activity which continued after the onset of voicing (Figure 2, C and D). The offset times were 83.4 ms for *umi* and 14.5 ms for *ari*; the longer offset time for *umi* is worthy of lengthy discussion. As mentioned above, the SH for *umi* sometimes showed activity sporadically after the offset of the pre-audio activity, whereas the muscle became inactive after the offset for the tokens *ari* (see Figure 1). As to the latter, we view this inactive phase as being due to the pitch which gradually rises from the first toward the second syllable, since this rise is characteristic of Tokyo Japanese.

Why does the pre-audio activity continue after the onset of voicing for the utterance *umi* in particular? In Tokyo Japanese, the pitch rises as it moves from the first to the second syllable if an utterance begins with an unaccented short syllable. However, the pitch contours measured actually from the tokens uttered by the subject never showed a rise immediately after the audio started. They started at a rather high level followed by a sharp fall and turned to a rising direction at a pitch level of about 139 Hz. The turning point was located at a period 77.4 ms from the onset of voicing. This time point appeared to bear some correlation with the

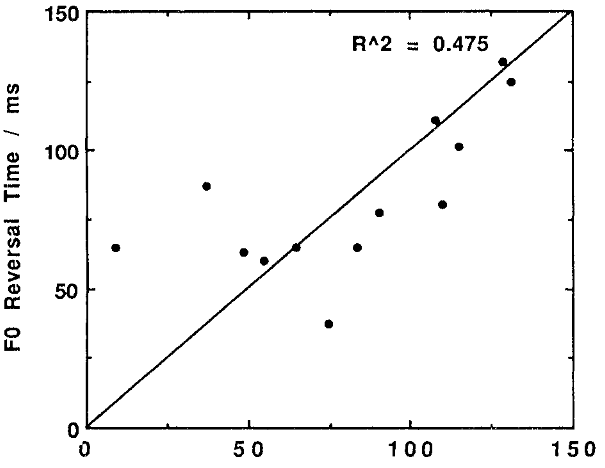

Figure 3. Correlation between the offset times of the SH for tokens of *umi* ... and the reversal times of *F*o (i.e., the periods from the voicing onset to the time when *F*o started to rise). R^2 is the coefficient of determination between the two variables.

offset of the pre-audio activity; indeed, 0.475 was the coefficient of determination between these two variables (Figure 3).

On the other hand, a similar abrupt descent in pitch was observed for most of the *ari* tokens, but the pitch after the descent was a rather lumpy, not smooth, contour to the extent that it was impossible to exactly specify the reversal point of the pitch contour. The large variability of pitch in *ari* might be related to the earlier stopping of the pre-audio activity (offset time: 14.5 ms).

3.3. Mean levels of the pre-audio activity

Figure 1 also illustrates the lower level of pre-audio activity for the accented token *u^mi*. This is further clarified quantitatively in Figure 4, which, for each utterance type, plots the duration of the pre-audio activity (i.e., the offset subtracted from the corresponding onset time) against the mean activity level over the duration. Evidently, we can see three separate groups following the different accentual structure in the initial position of the sentence frames, namely A, B and C including D, respectively. Here, it is noteworthy that the EMG level is significantly lower for accented *u^mi* (B) than for unaccented *umi* (C) and *ari* (D).

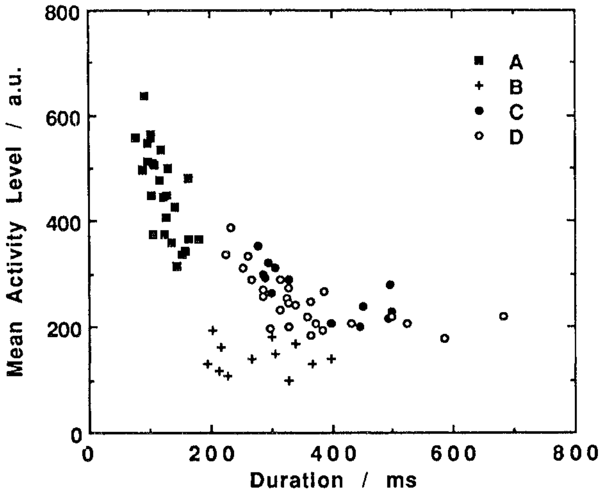

Figure 4. The pre-audio activity of the SH as represented in terms of both duration and the mean activity level over the duration ("a.u." stands for arbitrary units). The letters A, B, C and D correspond to utterances beginning with *ka^me* (A), *u^mi* (B), *umi* (C) and *ari* (D), respectively.

3.4. Results from Experiment 2

Briefly, the SH muscle displayed a patterning similar to that found in Experiment 1. The pre-audio activity lasted after the onset of voicing for words with no accent, whereas it ceased before the onset and was followed by an inactive phase for words with an accent. There were no prominent differences in patterning between words beginning with the vowel [i] and those with the vowel [a], which had been expected to involve a larger SH activity.

4. DISCUSSION

In this paper, we would like to argue two points. The first concerns the SH activity preceding the audio onset. In our experiments, this appeared even when utterances had an accent in initial position. The EMG level of *u^mi* was lower compared to that of unaccented words, but it differed from the results of one subject in Sugito and Hirose [11], who never showed any activity for words beginning with a high accent. This implies that our subject uses the SH muscle further to control other--either segmental or prosodic--speech gestures than pitch lowering. Examining the results of previous studies, we can note that the SH muscle is often activated prior to audio. For example, Gårding et al. [4] observed a similar activity in Swedish and mentioned it as preparatory for the speech mode. It is likely that the activity initiated before audio tunes the vocal folds so finely under a synergistic interaction with other laryngeal muscles that the folds are set ready to vibrate smoothly and in a stable manner.

Second, in interpreting the function of SH concerning pitch control, we must distinguish between the transitory control of lowering pitch from high to low and the steady control of keeping a fallen pitch at that low level. Our patterns did not show any activity during the high-pitched part of the accented test words and was again initiated when the pitch showed a sharp fall after the accent. However, as can be seen in Figure 1, the onset of the reactivation was not time-locked at certain pitch locations. The muscle initiated its activity for *u^mi* shortly before the pitch started falling (B), whereas it became active for *ka^me* after the pitch had already begun to fall (A). Erickson et al. [3] pointed out the timing delay in infrahyoid muscle activity in two speakers of Thai, a language that has one falling tone. According to these authors, it was only for the case of the mid-to-low fall that the SH muscle preceded the pitch peak so as to cause the initiation of the fall. On the grounds of a similar inconsistency of SH patterns, Collier [2] has held the opinion that they do not account well for the pitch lowering in Dutch. There certainly exist some contradictions, yet a good correlation has been shown for an English speaker between SH activity and rising or falling pitch in the mid range of the fundamental frequency [1]. More importantly, the SH muscle has been reported to remain active, usually with CTh activity suppressed, while pitch is kept at a low level. From these and our observations we conclude that the sternohyoid muscle helps at least to keep a fallen pitch at a low level during utterances.

ACKNOWLEDGEMENTS

This work was supported in part by Grant-in-Aid for Scientific Research No. 01440071 from the Japanese Ministry of Education, Science and Culture. We are grateful to Takashi Katakura for technical assistance in transmitting data files to the Hewlett-Packard computer.

REFERENCES

[1] J. E. Atkinson, "Correlation analysis of the physiological factors controlling fundamental voice frequency," J. Acoust. Soc. Am., Vol. 63, pp. 211-222, 1978.

[2] R. Collier, "Physiological correlates of intonation patterns," J. Acoust. Soc. Am., Vol. 58, pp. 249-255, 1975.

[3] D. Erickson, T. Baer, K. S. Harris, "The role of the strap muscles in pitch lowering," in *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, ed. by D. M. Bless, J. H. Abbs, pp. 279-285, San Diego, California: College-Hill Press, 1983.

[4] E. Gårding, O. Fujimura, H. Hirose, "Laryngeal control of Swedish word tones: A preliminary report on an EMG study," Annu. Bull. RILP (University of Tokyo), No. 4, pp. 45-54, 1970.

[5] P. A. Halle, S. Niimi, S. Imaizumi, H. Hirose, "Modern Standard Chinese four tones: Electromyographic and acoustic patterns revisited," Annu. Bull. RILP (University of Tokyo), No. 24, pp. 41-58, 1990.

[6] S. Kori, M. Sugito, H. Hirose, S. Niimi, "Participation of the sternohyoid muscle in pitch lowering: Evidence from Osaka Japanese," Annu. Bull. RILP (University of Tokyo), No. 24, pp. 65-75, 1990.

[7] J. Ohala, H. Hirose, "The function of the sternohyoid muscle in speech," in *Reports of the 1969 Autumn Meeting of the Acoustical Society of Japan*, pp. 359-360, 1969.

[8] Z. Simada, H. Hirose, "The function of the laryngeal muscles in respect to the word accent distinction," Annu. Bull. RILP (University of Tokyo), No. 4, pp. 27-40, 1970.

[9] Z. Simada, H. Hirose, "Physiological correlates of Japanese accent patterns," Annu. Bull. RILP (University of Tokyo), No. 5, pp. 41-49, 1971.

[10] A. Sonninen, "The external frame function in the control of pitch in the human voice," Ann. N. Y. Acad. Sci., Vol. 155, pp. 68-90, 1968.

[11] M. Sugito, H. Hirose, "An electromyographic study of the Kinki accent," Annu. Bull. RILP (University of Tokyo), No. 12, pp. 35-51, 1978.

[12] J. W. Tukey, *Exploratory Data Analysis*, Reading, Massachusetts; Menlo Park, California; London; Amsterdam; Don Mills, Ontario; Sydney: Addison-Wesley Publishing Company, 1977.

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