EVIDENCE FOR TONE-SPECIFIC ACTIVITY OF THE STERNOHYOID MUSCLE IN MODERN STANDARD CHINESE *

PIERRE A. HALLÉ

Laboratoire de Psychologie Expérimentale
CNRS and Paris V (Paris)

The role of the cricothyroid muscle (CT) in raising F_0 is well understood, but the activity of F_0-lowering strap muscles such as the sternohyoid (SH) has been less thoroughly investigated, especially in speech. This study focused on the active participation of the SH in the production of tones 2 (mid-rising) and 4 (high-falling) in Modern Standard Chinese. The other tones, however, together with the role of the CT and vocalis muscles, were also investigated in order to replicate earlier findings and to provide a more comprehensive picture of the production of Chinese tones. EMG data recorded from two male speakers show that the SH is consistently utilized to reset F_0 to a mid-low value at the onset of tone 2. Based on a comparison with earlier results for Thai speakers, we argue that this is a mandatory manoeuvre for producing rising F_0 contours in most contexts. The SH muscle also participates in the F_0 fall of tone 4, but less consistently. We argue that the latter manoeuvre may not be obligatory, especially in the case of speakers with a high-pitched voice.

Key Words: *Mandarin, tone production, electromyography, laryngeal muscles*

INTRODUCTION

Although Modern Standard Chinese (Mandarin, for short) is one of the most extensively studied tone languages, little attention has been paid to the articulatory processes involved in the production of its tones. The author knows of only one published study that has addressed this issue directly by collecting electromyographic

---

This work was supported in part by a grant to the author from the Japanese Society for the Promotion of Science (1989-1990). I thank the Research Institute of Logopedics and Phoniatrics of Tokyo University for gracious hospitality; I am especially indebted to Prof. Hajime Hirose and Dr. Seiji Niimi for helpful advice and discussion, and to Dr. Satoshi Imaizumi for assistance. For their invaluable contribution, I give my thanks to Drs. Kohichi Tsunoda and Kiyoshi Ohshima who inserted the electrodes. Special thanks are also due to the subjects who bravely performed while stoically enduring the discomfort of the experiment.

Correspondence concerning this manuscript should be sent to P. A. Hallé, Laboratoire de Psychologie Expérimentale, 28 rue Serpente, 75006 Paris (France). E-mail: labexp@frmop.cnusc.fr FAX: (33) (1) 40 51 70 85

(EMG) data (Sagart, Hallé, Boysson-Bardies, and Arabia-Guidet, 1986). The study was limited to the cricothyroid (CT) and sternohyoid (SH) muscles; it used two female speakers who read a small corpus of syllables embedded in a carrier sentence. In contrast, the tones of Central Thai have been more thoroughly investigated by Erickson (1976). Her study bore on CT, vocalis, and strap muscles, whose activities were recorded from four subjects. The four tones of Mandarin have counterparts in terms of pitch contour among the five tones of Thai. Not surprisingly, then, the EMG activity patterns in Mandarin tones observed by Sagart *et al.* generally bear a strong resemblance to those observed in Thai by Erickson. There were, however, noticeable differences in the activity patterns of the SH muscle between Mandarin tones 2 and 4 and their Thai counterparts.

Mandarin tone 2 and the Thai rising tone are both characterized by a mid-to-high rising F_0 contour, with a slight trough after tonal onset. Both Erickson's and Sagart *et al.*'s studies found a burst of CT activity preceding the rising part of the tonal contour. In Erickson's data, the initial trough of the tonal contour clearly resulted from a burst of activity of all three strap muscles, especially the thyrohyoid but also the SH and the sternothyroid. Evidence for a similar pattern in Mandarin tone 2 is very dim in Sagart *et al.*'s data, mainly because segment-related activity of SH almost completely blurred F_0 related activity, while, in Erickson's data, segment-related activity of strap muscles was negligible.

Mandarin tone 4 and the Thai falling tone are both characterized by a high-to-low falling F_0 contour. Both studies found the high initial F_0 level to result from a peak of CT activity. In Thai, the F_0 fall was clearly assisted, in the second half of the tonal contour, by a substantial burst of activity of all three strap muscles. In Mandarin, no increase of SH activity during the second half of tone 4 was found, except for the utterance-final syllable /zi4/ ([tsɿ] in tone 4)[1], an activity which was probably related to utterance-final downdrift.

Sagart *et al.* speculated that the difference between Mandarin tone 4 and the Thai falling tone may reflect "different acoustic characteristics". No obvious difference can be seen, however, in the shape or in the height of their F_0 contours that could result from such a radical difference in strap muscle activity. Since acoustically similar outputs can be produced by different means, the discrepancy may simply reflect different individual strategies. But the difference between the Thai and the Mandarin data may also be related to duration or stress[2]: Although Sagart *et al.* indicate that target syllables in their material always received "strong stress", the duration of the tone-carrying part in tone 4 was only

---

1. The 'pinyin' transcription of Mandarin is used here as a phonological transcription; digits 1 to 4 are appended to tonal syllables to denote tone. Phonetic transcriptions include symbols that are traditionally used in Chinese studies (e.g., [ɿ], a high-front apical vowel found after [s], [ʦ], and [ʦʰ]).
2. In Mandarin, local stress and global tempo are the primary factors determining tone duration (Coster and Kratochvil, 1984). Tone duration also depends on the segmental structure, and on the tone itself for syllables in citation form (Howie, 1976), where tone 3 is found to be longer; in running speech, however, tone duration depends little on the tone (Coster and Kratochvil, 1984).

about 200 msec; in Erickson's study, where long vowels were used[3], tonal contour duration ranged from 300 msec to 425 msec for the falling tone – a sizeable difference. The active participation of strap muscles in lowering F_0 may emerge only at such long durations. Long durations may also be necessary for a burst of strap muscle activity to emerge at the onset of Mandarin tone 2. This view is supported by the "variable norms" proposed by Kratochvil which express lawful relationships between duration and tonal contour (Kratochvil, 1985). Kratochvil used schematic representations of tone shapes F_0 values at six equally spaced points of a syllable's tone-carrying part) measured in a corpus of spontaneous speech. He found that the tone shapes were largely determined by tone duration for tonal syllables (though not for non-tonal syllables), and were best modelled by a series of linear relationships between F_0 (at each point of the schematic tone shape) and tone duration. Figure 1 shows the idealized tone shapes for each of the four tones at different durations, as modelled by these linear relationships. Kratochvil's findings strongly suggest that, for tone durations longer than 150—200 msec, some active F_0-lowering device is at work before the onset of tone 2 and in the second half of tone 4. The F_0 contour of non-tonal syllables (not shown in Figure 1) was largely flat, around the 210 Hz level. This F_0 level may be considered as a baseline or 'neutral' F_0 level in Kratochvil's data: F_0 levels departing from the neutral level conceivably result from an active manoeuvre of raising or lowering F_0. Such is the case for the initial F_0 dip in tone 2 and for the tonal offset of tone 4: They both become proportionally lower with increased duration and, from about 150—200 msec on, are both below the neutral F_0 level. These patterns suggest that for sufficiently long durations, an F_0-lowering activity (e.g., SH contraction) is at work in tones 2 and 4.

Kratochvil's findings for Mandarin tones 2 and 4 fit well with the EMG data for the Thai rising and falling tones. They suggest that EMG patterns similar to those observed for the Thai rising and falling tones should be found for Mandarin tones 2 and 4, provided that they are of long duration. In addition, they suggest that increased duration or stress coincides with increased intensity of laryngeal activity. The reason why Sagart *et al.* did not find F_0-related activity of SH in tone 4 may be an insufficient degree of stress of the target syllables. In the case of tone 2, an additional problem was the participation of the SH in segmental articulation, especially around syllable onset, which obscured its role in F_0 control. There was some indication of F_0-related activity of the SH muscle in tone 2 only in the syllable /bi/ ([pi]), where segment-related SH activity was the weakest. In contrast, the syllable /buu/ ([bu:]), which was used in Erickson's study, induced almost no segment-related strap muscle activity.

This study was primarily designed to re-examine the role of strap muscles in the production of Mandarin tones 2 and 4. In addition, tones 1 and 3 and the role of the CT and VOC muscles were also investigated, in order to replicate, and if possible to complement, earlier findings on Mandarin tone production. The SH muscle participates in segmental articulation by lowering or fixing the hyoid bone during jaw opening and

---

Thai has a phonemic vowel length distinction, Mandarin does not. Erickson chose CV syllables with a long vowel because they may carry any of the five tones. CV syllables with a short vowel can only carry 'static' tones: high tone or low tone.

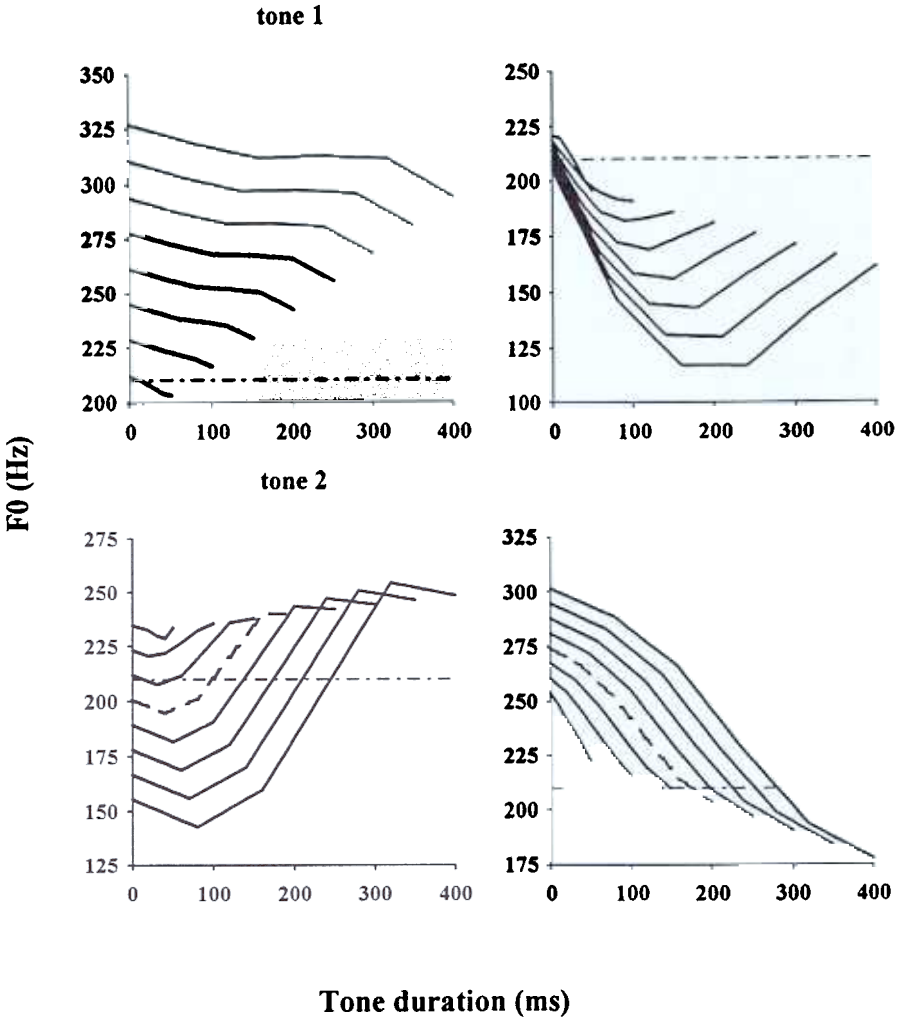

Fig. Tone contour as a function of tone duration, in the four tones (source: Kratochvil, 1985) The baseline F_0 level of 210 Hz (see text) is indicated by the horizontal dashed-dotted lines. The dashed contours in tones 2 and 4 correspond to 200 msec duration.

during tongue lowering and backing gestures (Collier, 1975). Therefore, we used target syllables whose segmental structure required these gestures only to a limited extent. We used more refined techniques for time alignment and time normalization than in the two aforementioned studies so as to reduce temporal distortions due to variability in segment duration. Finally, attempts were made to quantitatively assess differences in F_0-related

EMG activities.

We relied on the most widely accepted accounts of F_0-related laryngeal muscle activities: The CT is the main determinant of F_0 rises in all F_0 registers (Hirose, Simada, and Fujimura, 1970; Ohala and Hirose, 1970; Gårding, Fujimura, and Hirose, 1970; Hirose and Gay, 1972; Collier, 1975; Erickson, 1976; Atkinson, 1978; Harris, 1981). The vocalis (VOC) muscle's activity has also been observed to correlate with F_0 (Erickson, 1976; Atkinson, 1978), although less consistently than CT activity (e.g., Sawashima, Gay, and Harris, 1969). The role of VOC is possibly limited to counterbalancing CT tension in static configurations of the folds, as is suggested by Erickson's (1976) data: VOC activity correlates with F_0 in Thai 'static' tones, not in 'dynamic' tones.

The strap muscles are often regarded as synergic (Erickson, Liberman, and Niimi, 1977; Atkinson, 1978); they show a negative correlation with F_0 and seem to actively contribute to lowering F_0 only below an F_0 'threshold' level close to the F_0 midrange (Erickson, 1976; Erickson and Atkinson, 1976), similar to the F_0 level that we described above as 'neutral'. We assume, then, that strap muscles may cause or assist F_0 falls in medium and low F_0 ranges (see also Ohala and Hirose, 1970; Simada and Hirose, 1970, 1971; Atkinson, 1973, 1978; Collier, 1975; Sagart *et al.*, 1986), whatever the exact nature of the mechanisms involved (for a discussion, see Ohala, 1972; Erickson, Baer, and Harris, 1983). Some strap muscles also seem to be active in the high F_0 range, especially in singing (Sonninen, 1956, Faaborg-Andersen and Sonninen, 1960, Niimi, Horiguchi, and Kobayashi, 1991). Niimi *et al.* (1991) reasoned that the sternothyroid should play the same role as the CT, since it also helps tilting the thyroid cartilage downward, and found supporting evidence in the high F_0 range for trained singers producing high-pitched 'covered voice'. The SH may also be active in the high F_0 range (Roubaut, 1993); a plausible explanation is that SH co-contracts with the geniohyoid (GH): This co-contraction pulls the hyoid bone forward and downward, and conceivably helps tilt the thyroid cartilage forward and raise F_0 (Honda, personal communication; see also Yoshida, Honda, and Kakita, 1993); however, it is usually observed only in extreme gestures for F_0 raising, that is, not in normal speech.

We did not investigate other muscles that may also contribute to F_0 control: for example, the lateral cricoarytenoid, found to parallel the CT (Atkinson, 1978); the GH, found to act as "an extra boost" to raise F_0 (Erickson *et al.*, 1977; see also Honda, 1983), and the cricopharyngeal muscle, active in lowering F_0 (Honda, 1988; Honda and Fujimura, 1991). Finally, subglottal pressure (Ps) is generally regarded as playing a secondary role in F_0 control (Ohala, 1978; but see Atkinson, 1973, 1978). Rose (1984) has shown how both Ps and vocal fold tension contribute to the production of tones in the Chinese dialect of Zhenhai, a northern Wu dialect. It seems, however, that the domain of tone in (northern) Wu dialects is, unlike Mandarin, wider than the syllable: Tone-spreading often occurs as, for example, in Shanghai dialect (Zee and Maddieson, 1980) where the domain of tone may be a whole breath group. Since Ps participation in F_0 control is more likely to occur at the breath group level than within syllable-sized domains (Atkinson, 1978; Collier, 1975), we may assume that the role of Ps is secondary, not primary, in the production of Mandarin syllabic tones.

We thus limited ourselves to studying the role of three laryngeal muscles in the production of Mandarin tones: CT, VOC, and SH.

METHOD

Speech material

Like Sagart *et al.* (1986) and Erickson (1976), we used syllables embedded in a carrier sentence to avoid contamination by non-speech muscular activity. The carrier sentence was /yi2ge *S* zi4/ ("a character *S* "). The target syllable *S* under scrutiny was one of four syllables produced with each of the four tones. In order to minimize SH contribution to segmental articulation, the syllables used were /bi/, /mi/, /yi/, and /hu/ ([pi], [mi], [ji], and [xu]): High vowels [i] or [u] following either a bilabial closure or an homorganic approximant minimize jaw opening and tongue lowering. Tongue backing is expected to occur for [xu], but early in the utterance. (Recall that /buu/ in Erickson's data induced very little strap muscle activity.) The target syllable *S* was not in prepausal position, was stressed, and was preceded and followed by unstressed syllables. This was to avoid strong tonal context effects as well as utterance-final intonation downdrift on the target syllable.

Subjects

Two male subjects, L and Z, were successfully tested. Two additional subjects participated in the experiment, but their data could not be used due to displacement of EMG electrodes or to contamination by unwanted muscular activity. Both retained subjects were native speakers of Mandarin, born and raised in Beijing, aged 26 and 39 respectively, with no known speech pathology.

Experimental procedure

Hooked wire EMG electrodes were inserted in the CT, VOC (only for L), and SH muscles, using the long-established technique of the Research Institute of Logopedics and Phoniatrics at the University of Tokyo. Correct insertion was controlled with various non-speech manoeuvres before and after the experiment, and periodically during its course. The subjects were asked to pronounce the 16 sentences (4 syllables x 4 tones) at a comfortable speech rate; there were ten separate blocks, so that each sentence was repeated ten times. Electrode checking was performed every three blocks. The audio and EMG signals were recorded by means of a U-matic video recorder for subject L, by means of a multi-channel DAT recorder for subject Z. The raw signals were then replayed, digitized, and stored in computer files.

Data analysis

Signal processing. Each of the original three- or four-channel interleaved signal files was first split into single-channel files so as to take advantage of the many available single-channel signal processing programs. F_0, amplitude, and rate of spectral change[4] were computed from the audio signal every 10 msec using time frames of 31.2 msec or 35.7 msec for F_0 (adapting to the speaker's lowest pitch), 20 msec for amplitude,

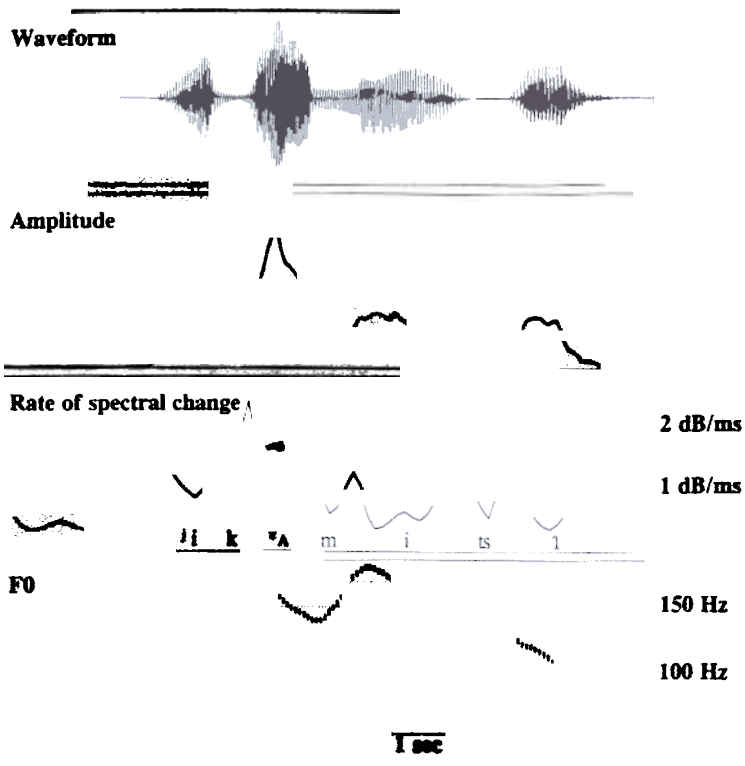

Fig. 2 Processing of the speech signal: one utterance of /mi4/ (subject L

and 32 msec for rate of spectral change. (An example is shown in Figure 2.) The latter function is useful to locate the major speech events. We used it for time alignment and time normalization, as discussed below.

The raw EMG signals were first low-passed filtered (1 kHz cut-off frequency). Their amplitude was then computed every 10 msec with a time frame of 20 msec; this processing is equivalent to 'rectification and integration'.

As a rule, the SH and CT signals were rather intense and clean, but the VOC signal from subject L was somewhat weaker: Although the VOC electrode had been correctly inserted, the electric signal had been less amplified than for CT or SH. However, subsequent analyses proved the recorded activity of the VOC to be meaningful.

The spectral distance between two *adjacent* time frames divided by the frame duration was taken as the rate of spectral change. The spectral distance between two frames was defined as the RMS difference in dB between the two corresponding short-term Bark-scaled energy spectra (dB per msec is the unit used in Figure 2).

Averaging method. The usual method of averaging a physiological function, such as integrated EMG activity, across repetitions of a sentence consists in first locating a specific acoustic event in each utterance, the 'line -up' point, then aligning all utterances on that point, and finally averaging the function across the lined-up utterances, within a domain of interest. This 'ensemble averaging' procedure is quite valid as long as the domain of interest lies close to the line-up point and the fluctuations in articulation rate are small across repetitions. A novel method of non-linear time alignment has recently been proposed to cope with articulation rate differences (Strik and Boves, 1991): Basically, dynamic programming is used to optimally time-warp each realization of the same sentence so as to minimize its acoustic distance from a reference realization. This method can be understood as using a series of many line-up points instead of just one: It may be needed for long, complex sentences. We used an alternative method using only two line-up points between which lies the domain of interest, that is the target syllable. The release burst of [k] in the syllable /ge/ and the vocalic onset in the syllable /zi4/, which precede and follow the target syllable and can always be easily located (at a clear peak of the rate of spectral change curve for /zi4/) were chosen as the two line-up points. For each sentence, the reference utterance was the one whose time distance between line-up points was the median of all such distances. In order to get them aligned with the reference utterance, the other utterances were linearly compressed or expanded so that line-up points coincided. As a result, both time alignment and time normalization (to the time scale of the reference utterance) were performed. As expected, this method greatly reduced the timing differences between repetitions of each sentence. The vowel of the target syllable was of special interest since it bore the tonal contour under examination[5]: The variability of its onset and offset (relative to the first line-up point) and of its duration was computed for each sentence before and after time normalization. Standard deviations dropped considerably, down to 10—15 msec. (10 msec was the sampling rate of all functions derived from the audio or EMG signals.) The mean tone duration of the target syllable ranged from 170 to 245 msec according to syllable type. Average durations of both tones 2 and 4 were about 215 msec.

RESULTS

EMG patterns in the four tones

The timing of CT and SH activity related to the production of the target syllable tone was stable and consistent across syllable-type. This is illustrated in Figure 3 where /bi4/ and /mi4/ are superimposed. Note that the timing of EMG activities is stable relative

---

Traditionally, phonological descriptions of Mandarin (e.g., Chao, 1969) alleged that the domain of tone is the entire voiced part of the syllable. Howie's (1974) phonetic studies showed, however, that the domain of tone in Mandarin is confined to the rhyme: the syllabic vowel and any voiced segment that may follow it. In particular, F_0 movements in syllable-initial sonorants are irrelevant to the tonal contour *per se*. In the material we used, all target syllables were CV syllables. Hence, their tone-carrying part was V, the syllabic vowel.

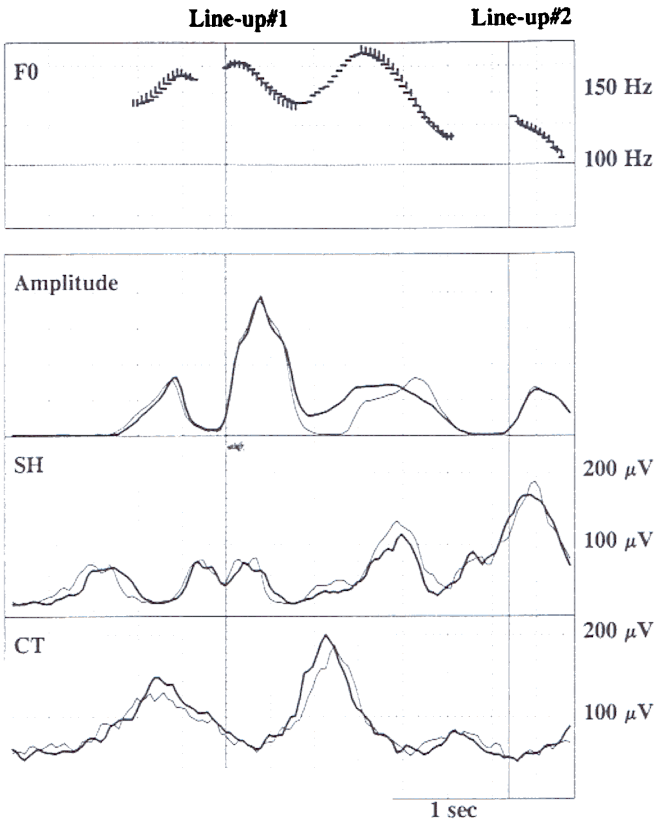

Fig. 3. Superimposed curves for averaged /bi4/ and /mi4/ utterances. Vertical bars for F_0 and thin lines for amplitude are for /bi4/, horizontal bars and thick lines for /mi4/ (subject L).

to the vowel of the target syllable. It may be thought of as largely invariant relative to the *whole* syllable only if, for example, the occlusion is considered part of the /bi/ syllable. In any case, the timing is such that similar contours are obtained *on the vowel*, whatever the initial portion of the syllable. The F_0 movement observed in the initial sonorant /m/, for example, reflects the continuous regulation of F_0 and is usually not taken to characterize the tone of the /mi/ syllable (see Note 5).

The patterns of CT and SH activity were found to be largely invariant across segmental variations and across the two subjects. The EMG activity patterns shown in Figures 4 (L's data) and 5 (Z's data) are averaged across all four syllables.

The results regarding target syllables obtained by Sagart *et al.* (1986) were largely replicated. F_0 rises (preceding tonal onset in tones 1 or 4, in the second half of the tonal

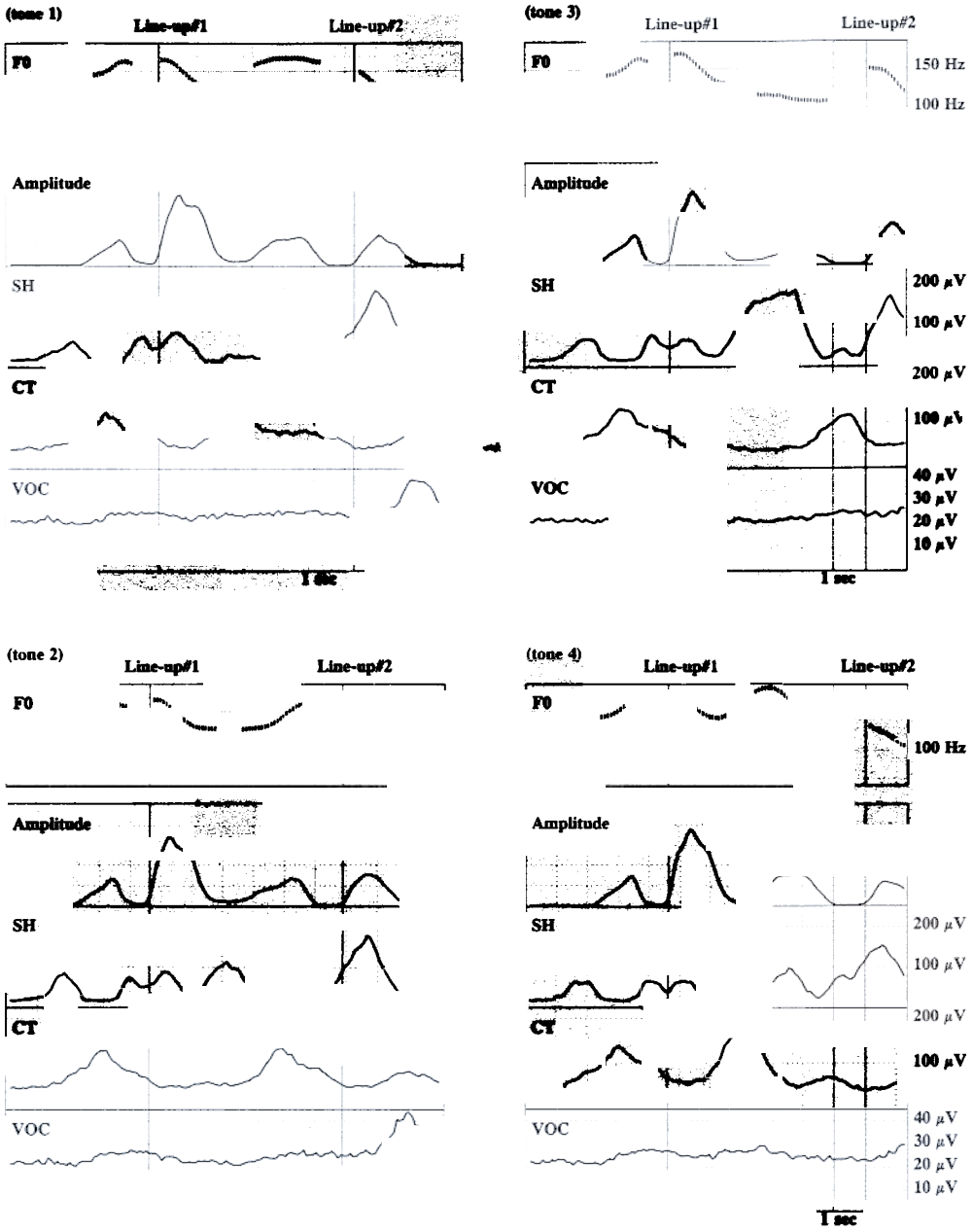

Fig. 4. F_0, amplitude, and EMG patterns for the four tones averaged across repetitions and syllables (subject L).

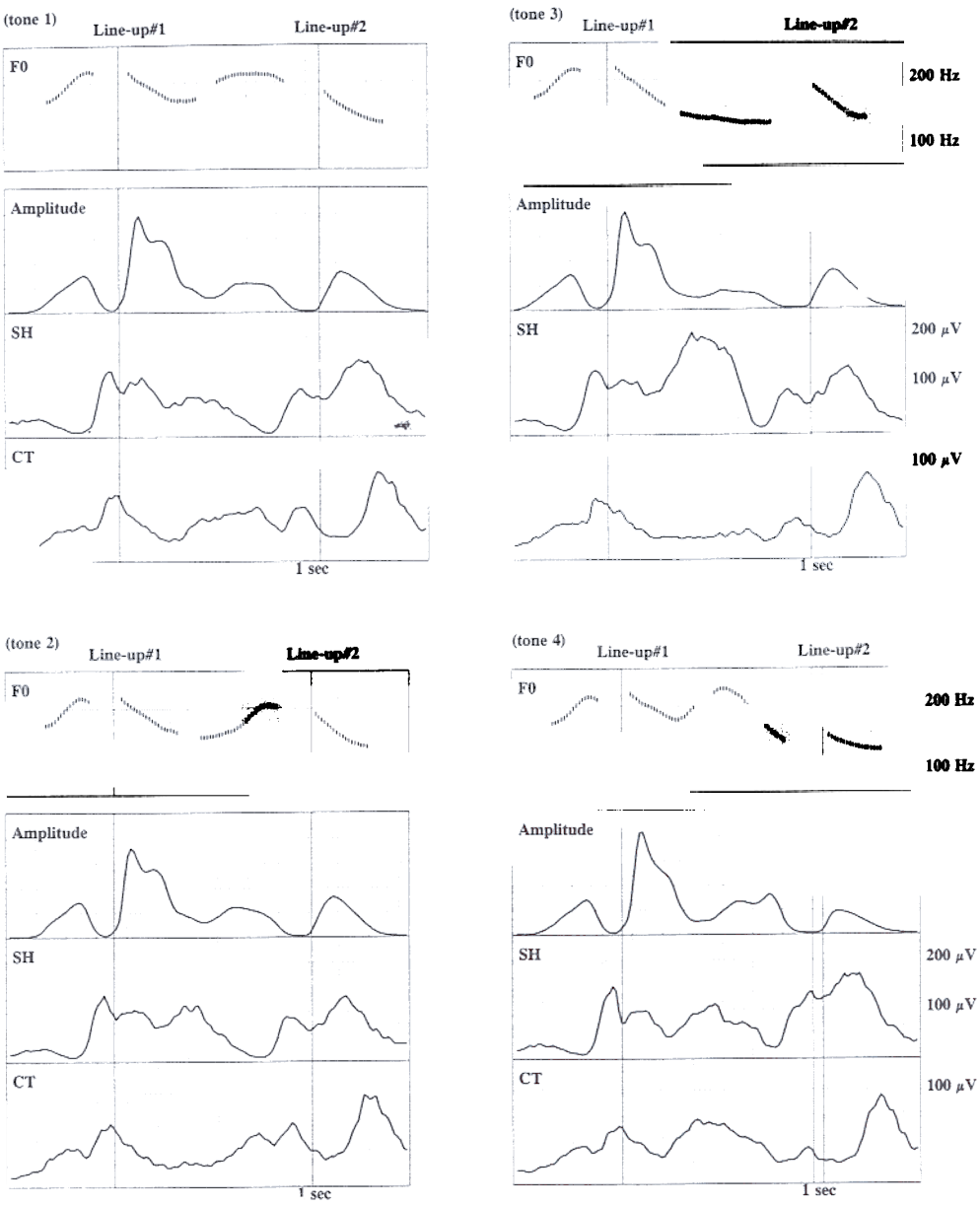

Fig. 5. F_0, amplitude, and EMG patterns for the four tones averaged across repetitions and syllables (subject Z).

contour in tone 2) were preceded by a burst of CT activity. A moderately high level of CT activity was maintained throughout tone 1 whose tonal contour was high and level. CT activity otherwise remained at a minimal, 'rest' level. In particular, CT activity was at this rest level throughout tone 3. Extremely high SH activity was found in tone 3; it is presumably related to the low F_0 level in this tone. Finally, we also found an increase of SH activity in the initial part of all target syllables; this increase was only moderate in tone 1 where it was presumably mainly related to segmental articulation: It was lowest for /yi/ (L) or /hu/ (Z) and somewhat larger for /mi/ and /bi/ (both subjects).

Some new patterns, however, emerged in our data. First, a moderate burst of SH activity occurred in the second part of tone 4, whatever the segmental structure for subject L, only for syllables /bi4/ and /hu4/ in the case of subject Z. Second, a similar burst of SH activity was very clear preceding the initial part of tone 2 for subject L in all syllables, for subject Z in /hu2/ and, to a lesser extent, in /bi2/. Finally, VOC activity (L) was much weaker than that of other muscles. In consequence, VOC activity patterns, as shown in Figure 4, remain unclear. Consistent patterns were revealed, however, by a comparison method exposed in the following section: VOC activity roughly paralleled CT activity; it was the weakest during tone 3, and reached a higher level during tone 1, in the beginning of tone 4, and in the second half of tone 2. The reliability of these SH and VOC patterns of activity is examined in the following sections.

Aside from EMG patterns related to target syllables, we observed a consistent pattern of EMG activities at the ends of all utterances for both subjects: a very intense burst of SH activity, centred on the offset of the utterance-final syllable /zi4/, followed by a rather large increase of activity of the CT (and VOC for L). This intense burst of SH activity may be responsible for the intonation downdrift that terminates breath groups. The subsequent increase in CT and VOC activity presumably increased fold stiffness and thickness, which, combined with vocal fold abduction resulted in voice termination.

SH activity in tones 2, 3, and 4

Since the SH is generally involved in both F_0 control and segmental articulation (Collier, 1975), the question arises of whether the SH activity found in tones 2, 3, and 4 is at all related to F_0 control. As can be seen in Figures 4 and 5, the SH activity related to the target syllable is the weakest in tone 1, where it precedes syllable onset. This activity, if it was related to F_0 control, could only contribute to the initial F_0-raising since tone 1 is essentially high-level. Such SH contribution can be observed in extreme gestures for raising F_0 but is unlikely in the present case. Hence, we may assume that SH activity in tone 1 is mainly segment-related. The SH activity profile in tone 1 may thus serve as a 'baseline' for estimating F_0-related SH activity in other tones. For example, in order to estimate F_0-related SH activity in tone 2, two sets of utterances were compared: the same targets syllables (to minimize differences in segment-related SH activity), produced with tone 1 and with tone 2. The two sets of utterances were first time-aligned and time-normalized together; the difference in SH activity levels between the two sets was then assessed by means of Student's t values computed at each point in time. The plot of t values along time indicates where differences are significant, hence, where SH activity in the tone compared with tone 1 is F_0-related. Figures 6 and 7 are an illustration of this

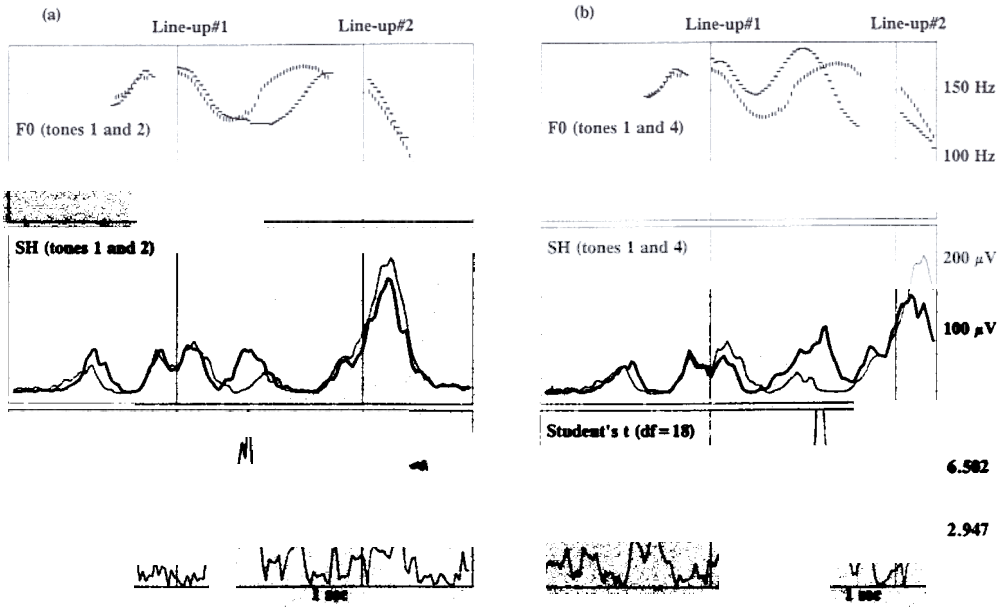

Fig. 6. Comparison of SH activity between (a) tone 2 and tone 1, and (b) tone 4 and tone 1 (subject L, syllable /yi/). The two labelled levels of t values shown in this figure and in Figures 7 and 8 correspond to $p = 0.01$ and $p = 0.00001$. Vertical bars (F_0) and thin lines (EMG) are for tone 1, horizontal bars and thick lines are for tones 2 or 4.

procedure for /yi/ (L) and /hu/ (Z) respectively. They show that SH activity in the initial part of tone 2, and in the second part of tone 4 was significantly larger than in the corresponding portions of tone 1. This was especially clear in L's data. In Z's data, F_0-related SH activity in tone 4 was more spread out; for this subject, an increase of SH activity also occurred at the onset of utterance-final /zi4/ much more markedly after tone 4 than after tone 1, 2, or 3, as can be seen in Figure 5 (all syllables), or in Figure 7 (/hu/ syllable). The intense SH activity observed in tone 3 for both subjects (see Figures 4 and 5) leaves little doubt as to its F_0-related nature: Indeed, this was confirmed by t-test comparisons.

CT and VOC activity in tones 1, 2, and 4

Just like the SH in tone 1, the CT remains at a minimal activity level in tone 3: Indeed, tone 3, the low-falling tone, does not require F_0-raising activity. The CT activity profile in tone 3 may thus serve as a baseline for estimating F_0-related CT activity in other tones. It is already fairly well known, however, that CT activity is mainly F_0-related. Not surprisingly, then, this was confirmed by t-test comparisons similar to those conducted for the SH. As to the VOC muscle, since its activity profile in tone 3 was the lowest and unlikely to be F_0-related, it was used as a baseline in t-test comparisons

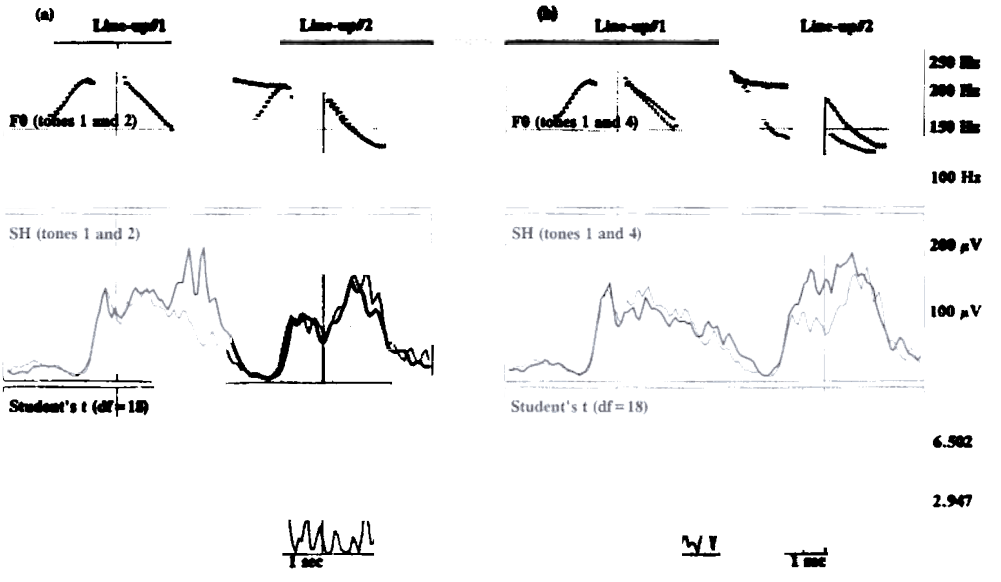

Fig. 7 Comparison of SH activity between (a) tone 2 and tone 1, and (b) tone 4 and tone 1 (subject Z, syllable /hu/). Vertical bars (F_0) and thin lines (EMG) are for tone 1, horizontal bars and thick lines are for tones 2 or 4.

between tone 3 and the other tones. VOC activity in tones 1, 2, and 4 strikingly paralleled CT activity: It was significantly higher than the baseline almost in the same regions as for the CT, as illustrated in Figure 8 showing L's data for /yi1/ and /yi4/ compared to /yi3/.

Overall differences in EMG activity

In the previous sections, tone-specific patterns of EMG activity were isolated by visual inspection of EMG and F_0 curves as in Figures 4 and 5, and by local comparisons with baseline EMG profiles that were assumed to reflect minimally F_0-related EMG activity. Another means for assessing tone-related differences and for 'factoring out' segment-related differences consists in comparing the variability induced by tone variation alone (keeping the segments constant) to that induced by segmental variation alone (keeping the tones constant). In other words, we may compare inter-tone variability to inter-segmental variability. A simple measure of the overall difference between two activity profiles of a given muscle is given by the following distance:

$$d(x,y) = \left\{ \sum_{t=t_0}^{t_f} |x(t) - y(t)| \right\} \Bigg/ \left\{ \sum_{t=t_0}^{t_f} (x(t) + y(t)) \right\}$$

where $x(t)$ and $y(t)$ stand for the integrated EMG activity along time of a given muscle

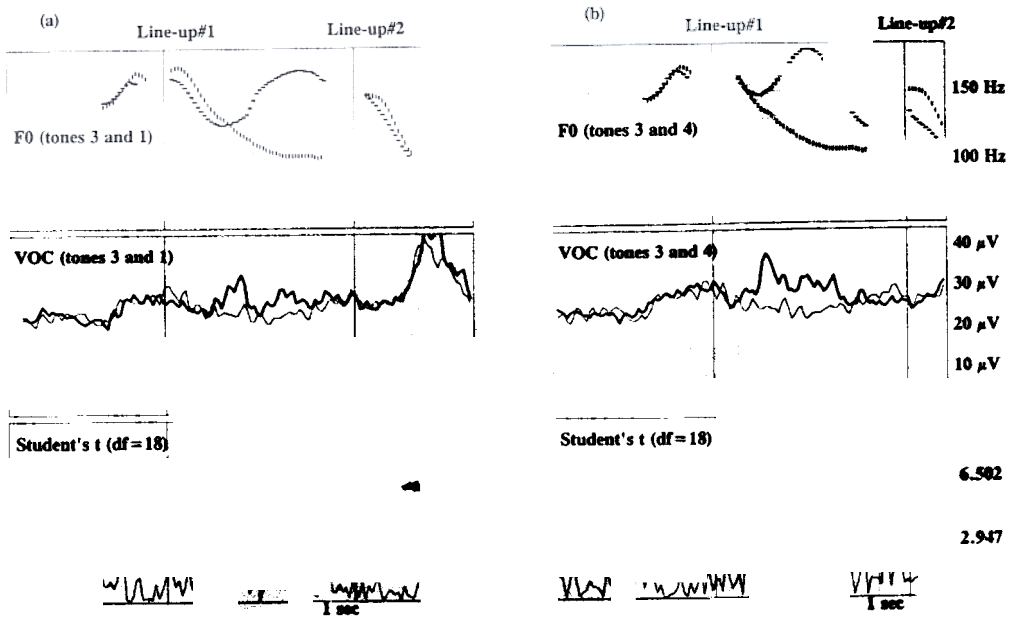

Fig. 8. Comparison of VOC activity between (a) tone 1 and tone 3, and (b) tone 4 and tone 3 (subject L, syllable /yi/). Vertical bars (F_0) and thin lines (EMG) are for tone 3, horizontal bars and thick lines are for tones 1 or 4.

(e.g., the SH) in two utterances whose target syllables differ with respect to either tone or segments. The two utterances must first be time-aligned and time-normalized together. The time interval $[t_0, t_f]$ corresponds here to the time domain common to both utterances. The second sum serves to normalize $d(x,y)$ with respect to duration and mean integrated EMG activity. It can be assumed that this distance mainly reflects differences in target syllables, since the carrier sentence basically does not change. Inter-tone distances were larger than inter-segmental distances for both subjects, whatever the muscle under scrutiny [L: $t(46) = 10.9$, $p < 0.0001$ for CT; $t(46) = 9.2$, $p < 0.0001$ for SH; $t(46) = 3.4$, $p < 0.002$ for VOC; Z: $t(46) = 4.8$, $p < 0.0001$ for CT; $t(46) = 4.6$, $p < 0.0001$ for SH]. Put another way, the effect of the 'tone' factor overrode the effect of the 'segmental' factor.

Inter-tone distances for SH are of particular relevance in this study: Distances between tone 2 and tone 1, or between tone 4 and tone 1 were both significantly larger than inter-segmental distances for subject L; the same trend was observed for subject Z but did not reach significance level. This result, shown in Table 1, confirms – at least for subject L – that SH activity in tones 2 and 4 is tone-specific.

Correlation between F_0 and EMG activities

The examination of positive or negative correlations between F_0 and the activity profile of a given muscle is yet another means of assessing its F_0-related activity. We essentially followed the cross-correlation technique proposed by Atkinson (1978): The

TABLE 1

Inter-tone distances for SH profiles between tones 1 and 2 or tones 1 and 4 *versus* inter-segmental distances (distances are averaged across syllables or across tones)

Subject	inter-segmental	between tones 1 and 2	between tones 1 and 4
L	0.098	0.135 *	0.166 **
Z	0.128	0.137	0.166

* $p < 0.005$ ** $p < 0.0001$

correlation measure is the maximum Pearson correlation coefficient (in absolute value) obtained by gradually time-shifting the F_0 contour relative to the EMG profile. At the same time, the time shift for which the correlation is maximal is an estimation of muscle response time or latency. When the correlation is computed over a whole utterance, it may be taken as a gross indication of the involvement in F_0 control of the muscle under scrutiny. The time leads of CT, VOC, and SH activity relative to F_0 movements obtained by the cross-correlation method are compared in Table 2 to visual estimations made from F_0 and EMG curves[6]. The latencies we find are in good agreement with the response times estimated by Baer (1981). Baer measured time lags between EMG single firings and related local F_0 perturbations in otherwise flat F_0 contours (about 100 Hz) produced by a male speaker. He found response times of about 80 msec for the CT, and of about 100 msec for strap muscles. The correlation values indicate that, globally, CT and VOC correlate positively with F_0, whereas SH correlates negatively with F_0 (see Table 2). A more detailed analysis was conducted within a region limited to the target syllable and its immediately preceding context[7]: SH correlated negatively with F_0 for all target syllables in tone 2, 3, or 4, as is shown in Table 3. In tone 1, the correlation was much weaker, and

---

[6] For CT and VOC, we measured the interval between the peak of CT activity at the onset of tone 4 and the corresponding peak of the F_0 contour. For SH, we measured the interval between the initial peak of SH activity in /yi2/ (L's data) or /hu2/ (Z's data) and the initial trough in the contour of tone 2.

Latencies seem to vary within an utterance, perhaps according to within-utterance position, or to stress. This could be seen for the utterance-initial syllable /yi2/: The time lag between the peak of CT activity and the peak of the tonal contour was about 40—50 msec, while it was about 90 msec for the target syllable /yi2/. Hence, a more accurate estimation of muscle latency – and thereby of correlation – by means of cross-correlation is obtained when restricting the domain of computation to the time interval where EMG activities may affect a given tonal contour.

TABLE 2

Muscle latencies (msec) estimated by measuring the distance between related events and by the cross-correlation method (with the corresponding correlation value)

Subject	muscle	related events	cross-correlation	correlation
L	CT	96	87	0.882
	VOC	45	52	0.615
	SH	113	110	-0.769
Z	CT	56	86	0.526
	SH	143	128	-0.652

TABLE 3

Local correlation – in the target syllable domain –between F_0 and SH in each of the four tones (averaged across syllables)

Subject	tone	tone 2	tone 3	tone 4
L	-0.375	-0.867	-0.967	-0.867
Z	-0.235	-0.967	-0.877	-0.794

probably not related to F_0: In this tone, the F_0 contour begins with a slight rise and then remains high-level, while SH activity exhibited a moderate increase at syllable onset and then returned to a rest level. The weak negative correlation observed for tone 1 is thus due to the moderate increase of SH activity at syllable onset, which is believed to be related to segmental articulation rather than to F_0. The correlation was found to be no less strong with tones 2 or 4 than with tone 3, where the tone-related role of the SH is clear. Hence, the negative correlation between SH and F_0 in the case of tones 2 and 4 reflects – at least partly – the role of the SH in the production of these tones.

GENERAL DISCUSSION

The patterns of EMG activity that we observed are very similar to those observed by Erickson (1976) in Thai high, rising, low, and falling tones, the 'sister' tones, so to speak, of Mandarin tones 1 to 4. Interestingly, the similarity also holds for the patterns of SH activity in tones 2 and 4, which were not clearly found or not found at all by

Sagart *et al.* (1986). Otherwise, the results of Sagart *et al.* (patterns of SH activity in tone 3 and of CT activity in all four tones) are essentially confirmed by Erickson's data as well as by our data: Here, we seem to be on stable ground.

The doubtful points, then, arise from the SH activity patterns found in tones 2 and 4. In the data reported here, the tone-related nature of SH activity in tones 2 and 4 was supported by multiple sources of evidence. The discrepancies between the earlier results of Sagart *et al.* (1986) and Erickson (1976) on the one hand and our results on the other hand, may simply be due to individual variations. Yet, before we surrender to this 'random variation' explanation, we should consider alternative accounts that point to systematic sources of variation. In the case of tone 2, the reason why Sagart *et al.* did not find clear evidence for SH participation in F_0 control may have been the strong involvement of the SH in segmental articulation of syllables such as /fa/, /ge/ ([kˠʌ]), etc. For /bi/, where segment-related SH activity was the weakest, they found some indication of F_0-related SH activity in tone 2. But SH involvement in segmental articulation cannot explain the absolute lack of SH activity in the second half of tone 4 that they have reported. In the Introduction it was tentatively proposed that the target syllables of Sagart *et al.* were not sufficiently stressed for an F_0-lowering activity to be observed: In /bi4/, the vowel was about 200 msec long, as opposed to about 350 msec in the Thai data of Erickson (1976). In the new Mandarin data reported here, the vowels of target syllables were only slightly longer: In /bi4/, the vowel was about 230 msec long *for both speakers* (233 msec for L, 231 msec for Z). Since the tone-related SH activity in tone 4 is much clearer in subject L than in subject Z, it now seems unlikely that vowel duration explains the discrepancy between the present results and those of Sagart *et al.*

Another explanation is suggested by the differences in speakers' pitch range. There seems to be a general trend for speakers with a lower pitch range to use SH more consistently in high-to-low steep F_0 falls, as in Mandarin tone 4 or the Thai falling tone. In our data, L has a lower pitch than Z (by about 50 Hz) and produces clearer SH patterns of tone-related activity in tone 2 or 4; the one subject whose SH activity was recorded by Sagart *et al.*, a female speaker, had a higher pitch and produced no SH pattern at all in tone 4. In Erickson's study, the contribution of strap muscles to F_0 fall in the falling tone and to F_0 resetting in the rising tone is clear for all three male subjects, but is more confused and noisy for the female subject (Erickson, 1976, pp. 57—58 and 117). Taken together, these data suggest that speakers with a high-pitched voice can produce rapid high-to-low F_0 falls by simply relaxing F_0-raising activity. Speakers with a lower-pitched voice additionally utilize an F_0-lowering device. This may be related to biomechanical properties of the laryngeal system. (A smaller/lighter larynx may return faster to rest position.) In sum, there may be some systematicity in the inter-individual variations of F_0-related SH activity for Mandarin tone 4 or the Thai falling tone. In contrast, the production of Mandarin tone 2 or the Thai rising tone seems to result from a more universally used manoeuvre: In order to produce a mid-to-high rising tonal contour, speakers must first reset F_0 to a low-mid level (active participation of the strap muscles), then raise F_0 to a high level. Hence, the lesser inter-individual variations observed in the production of Mandarin tone 2 and the Thai rising tone.

F0-related participation of the SH in the initial portion of tone 2 or in the second

half of tone 4 is consistent with the somewhat puzzling findings of Kratochvil (1985): For example, the longer a tone 2, the lower its onset F_0 *below a neutral* F_0 *level* (see Figure 1). Indeed, only the action of some F_0-lowering device can explain this result. Our data strongly suggest that the SH or, more generally, the strap muscles are such a device. But how does increased duration result in *proportionally* lower minima of F_0 in tones 2 and 4? When tones are longer (or syllables more stressed), SH activity may be maintained longer or may be more intense. Even if SH activity remains largely unchanged in longer tones (in that case, the notion of "variable norms" would not apply to articulation patterns), its effects are simply unopposed over a longer time interval and can modify F_0 movements to a larger extent. Because target syllables were not sufficiently varied in stress, our data do not permit observation of a possible correlation between tone duration and intensity (or duration) of related EMG activities. Further investigations are needed to clarify this point.

Finally, we may ask to what extent the patterns of EMG activities that were found here can remain unaffected by tonal context. The patterns isolated here should be thought of as ideal patterns that apply to stressed syllables carefully embedded in a carrier sentence designed so as to minimize contextual and phrase-intonation effects. Indeed, only controlled material can avoid the "ubiquitous variability" of motor command patterns (MacNeilage and deClerk, 1969; MacNeilage, 1970), and, given this restriction, we found a relative invariance in the patterns relevant to the target syllables. What happens, however, when tonal context effects become substantial, that is, when tonal coarticulation occurs? Do the ideal patterns still hold (at least qualitatively), do they change in unpredictable ways, or do they change in some systematic way? Again, further investigations are needed. In our data, however, we already find some indications that the patterns of EMG activity related to the utterance-final syllable /zi4/ are not invariant: In some instances, different EMG patterns were observed in different tonal contexts for a similar tone contour on /zi4/; in other instances, different EMG patterns were observed in similar tonal contexts for a different tone contour on /zi4/[8].

To sum up, evidence gathered from controlled speech material strongly suggests that there are *ideal* patterns of EMG activity associated with each tone, and that they

For subject L, the burst of CT activity preceding /zi4/ (causing the high onset F_0 in tone 4) was much larger after target syllables in tone 3 than after targets in other tones, in particular tone 4. As a result, the tone shape of /zi4/ was preserved after tone 3 but not after tone 4: Conceivably, the lowering effect of a preceding tone with a low endpoint F_0 (such as tone 3 or 4) was compensated by a larger CT activity in the case of tone 3, and was not in the case of tone 4. For subject Z, a larger SH activity together with a smaller CT activity was observed before /zi4/ in tone 4 utterances. The low endpoint of target syllables in tone 3 did not affect the tone shape of /zi4/ (maybe because the SH activity related to tone 3 production was released earlier than for subject L), whereas the somewhat different EMG pattern after target syllables in tone 4 caused a lower/flatter tone shape in /zi4/. Both subjects produced a lower/flatter tone contour on /zi4/ after tone 4 than after other tones, but they did so by using different strategies of F_0 control, that is, with different EMG patterns.

involve SH activity in tones 2 and 4 as well as in tone 3. This is not to say that these patterns are invariant. Future research will have to address the important question of how patterns are modified under stress variation, tonal context variation, and speaker variation.

(Received April 29, 1993; accepted March 4, 1994)

REFERENCES

ATKINSON, J.E. (1973). Physiological factors controlling F_0 results of a correlation analysis. *Journal of the Acoustical Society of America*, **54**, 319 (Abstract).

ATKINSON, J.E. (1978). Correlation analysis of the physiological factors controlling voice frequency. *Journal of the Acoustical Society of America*, **63**, 211—222.

BAER, T. (1981). Investigation of the phonatory mechanism. In C.L. Ludlow and M. O'Connell Hart (Eds.), *Proceedings of the Conference on the Assessment of Vocal Pathology (ASHA Reports 11)* (pp. 38—47). Rockville, MD: American Speech-Language-Hearing Association.

CHAO, Y.R. (1968). *A Grammar of Spoken Chinese*. Berkeley, CA: University of California Press.

COLLIER, R. (1975). Physiological correlates of intonation patterns. *Journal of the Acoustical Society of America*, **58**, 249—255.

COSTER, D.C., and KRATOCHVIL, P. (1984). Tone and stress discrimination in normal Peking dialect speech. In B. Hong (Ed.), *New Papers in Chinese Linguistics* (pp. 119—132). Canberra: Australian National University Press.

ERICKSON, D. (1976). *A Physiological Analysis of the Tones of Thai*. Ph.D. Dissertation, University of Connecticut.

ERICKSON, D., and ATKINSON, J.E. (1976). The function of strap muscles in speech. *Haskins Laboratories Status Report on Speech Research*, **SR-45/46**, 205—210.

ERICKSON, D., BAER, T., and HARRIS, K. (1983). The role of the strap muscles in pitch lowering. In D.M. Bless and J.H. Abbs (Eds.), *Vocal Fold Physiology: Contemporary Research and Clinical Issues* (pp. 279—285). San Diego, CA: College-Hill Press.

ERICKSON, D., LIBERMAN, M., and NIIMI, S. (1977). The geniohyoid and the role of the strap muscles. *Haskins Laboratories Status Report on Speech Research*, **SR-49**, 97—102.

FAABORG-ANDERSEN, K., and SONNINEN, A. (1960). The function of the extrinsic laryngeal muscles at different pitch: An electromyographic and roentgenologic investigation. *Acta Otolaryngologica*, **51**, 89—93.

GÅRDING, E., FUJIMURA O., and HIROSE, H. (1970). Laryngeal control of Swedish word tones. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics (University of Tokyo)*, **4**, 45—54.

HARRIS, K. (1981). Electromyography as a technique for laryngeal investigation. In C.L. Ludlow and M. O'Connell Hart (Eds.), *Proceedings of the Conference on the Assessment of Vocal Pathology (ASHA Reports 11)* (pp. 70—86). Rockville, MD: American Speech-Language-Hearing Association.

HIROSE, H., and GAY, T. (1972). The activity of the intrinsic laryngeal muscles in voicing control: an electromyographic study. *Phonetica*, **25**, 140—164.

HIROSE, H., SIMADA, Z., and FUJIMURA, O. (1970). An electromyographic study of the activity of the laryngeal muscles during speech utterances. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics (University of Tokyo)*, **4**, 9—25.

STRIK, H., and BOVES, L. (1991). A dynamic programming algorithm for time-aligning and averaging physiological signals related to speech. *Journal of Phonetics*, **19**, 367—378.

YOSHIDA, Y., HONDA, K., and KAKITA, Y. (1993). Non-invasive EMG measurement of laryngeal muscles and physiological mechanism of prosody control. *ATR Technical Report*, **TR-A-0168**, 1—13 (in Japanese).

ZEE, E., and MADDIESON, I. (1980). Tones and tone sandhi in Shanghai: phonetic evidence and phonological analysis. *Glossa*, **14**, 45—88.

HONDA, K. (1983). Relationship between pitch control and vowel articulation. In D.M. Bless and J.H. Abbs (Eds.), *Vocal Fold Physiology: Contemporary Research and Clinical Issues* (pp. 279—285). San Diego, CA: College-Hill Press.

HONDA, K. (1988). Action of the cricopharyngeus muscle in lowering voice fundamental frequency. *Proceedings of the '88 Acoustical Society of Japan Spring Meeting*, 169—170 (in Japanese).

HONDA, K., and FUJIMURA, O. (1991). Intrinsic vowel F_0 and phrase-final F_0 lowering: phonological *vs.* biological explanations. In J. Gauffin and B. Hammarberg (Eds.), *Vocal Fold Physiology: Acoustic, Perceptual and Physiological Aspects of Voice Mechanisms* (pp. 149—158). San Diego, CA: Singular Publishing Group.

HOWIE, J.M. (1974). On the domain of tone in Mandarin. *Phonetica*, **30**, 129—148.

HOWIE, J.M. (1976). *Acoustical Studies of Mandarin Vowel and Tones*. Cambridge, U.K.: Cambridge University Press.

KRATOCHVIL, P. (1985). Variable norms of tones in Beijing prosody. *Cahiers de Linguistique Asie Orientale*, **14**, 135—174.

MACNEILAGE, P. (1970). Motor control of serial ordering of speech. *Psychological Review*, **77**, 183—196.

MACNEILAGE, P., and DECLERK, J. (1969). On the motor control of coarticulation in CVC monosyllables. *Journal of the Acoustical Society of America*, **45**, 1217—1233.

NIIMI, S., HORIGUCHI, S., and KOBAYASHI, N. (1991). F_0 raising role of the sternothyroid muscle - An electromyographic study of two tenors. In J. Gauffin and B. Hammarberg (Eds.), *Vocal Fold Physiology: Acoustic, Perceptual and Physiological Aspects of Voice Mechanisms* (pp. 183—188). San Diego, CA: Singular Publishing Group.

OHALA, J. (1972). How is pitch lowered ? *Journal of the Acoustical Society of America*, **52**, 124 (Abstract).

OHALA, J. (1978). Production of tone. In V. Fromkin (Ed.), *Tone: A Linguistic Survey* (pp. 5—40). New York: Academic Press.

OHALA, J., and HIROSE, H. (1970) The function of the sternohyoid muscle in speech. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics (University of Tokyo)*, **4**, 41—44.

ROSE, J.P. (1984). The role of subglottal pressure and vocal cord tension in the production of tones in a Chinese dialect. In B. Hong (Ed.), *New Papers in Chinese Linguistics* (pp. 133—168). Canberra: Australian National University Press.

ROUBAUT, B. (1993). *Mécanismes vibratoires laryngés et contrôle neuro-musculaire de la fréquence fondamentale*. Doctoral Dissertation, University of Paris XI.

SAGART, L., HALLÉ, P., BOYSSON-BARDIES, B., and ARABIA-GUIDET C. (1986). Tone production in modern standard Chinese: an electromyographic investigation. *Cahiers de Linguistique Asie Orientale*, **15**, 153—174.

SAWASHIMA, M., GAY, T., and HARRIS, K. (1969). Laryngeal muscle activity during vocal pitch and intensity changes. *Haskins Laboratories Status Report on Speech Research*, **SR-19/20**, 211—220.

SIMADA, Z., and HIROSE, H. (1970). The function of the laryngeal muscles in respect to the word accent distinction. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics (University of Tokyo)*, **4**, 27—40.

SIMADA, Z., and HIROSE, H. (1971). Physiological correlates of Japanese accent patterns. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics (University of Tokyo)*, **5**, 41—49.

SONNINEN, A. (1956). The role of the extrinsic laryngeal muscles in length adjustment of the vocal cords in singing. *Acta Otolaryngologica*, **Suppl. 130.**