ANN.Bull.RILP(1990)No.24, 41-58

# MODERN STANDARD CHINESE FOUR TONES: ELECTROMYOGRAPHIC AND ACOUSTIC PATTERNS REVISITED.

P.A. Hallé*, Seiji Niimi, Satoshi Imaizumi, and Hajime Hirose

## Introduction

Modern Standard Chinese (henceforth MSC) tones are, like in many South East Asian languages, contour rather than level tones. The traditional account is: tone T1 high level, T2 mid rising, T3 low dipping and T4 high falling. Some disagreement arises about tone T3. According to Chao (1968) it is rising in its second half when in citation form or in prepausal position; it becomes low level ("half third tone") before any of the 3 other tones, and becomes similar to tone T2 before another tone T3. The latter mutation - a case of paradigmatic sandhi - is still a little controversial (see Kratochvil 1987 for a review).

Fo control in MSC tone production seems to be mainly achieved through the regulation of vocal fold tension. A weak correlation can be found between Fo and amplitude contours in general, and in MSC as well (Kratochvil 1981). Since amplitude reflects subglottic pressure Ps, one could argue that Ps and vocal fold tension are used conjointly to achieve Fo control, or that increase of Ps results from increase of tension, or vice versa (see Ohala 1978, for a discussion). However, there is no evidence that Ps alone is responsible for any Fo movement in MSC and we assume here that the primary mechanism of Fo control is the regulation of vocal fold tension through laryngeal muscles.

EMG studies of laryngeal muscles in many languages, e.g. Japanese (Hirose, Simada & Fujimura 1970; Ohala & Hirose 1970), Swedish (Garding, Fujimura & Hirose 1970), English (Hirose & Gay 1972), Dutch (Collier 1975), Thai (Erickson 1976), Chinese (Sagart & Hallé 1986), have shown several laryngeal muscles to be involved in speech Fo control. Among those, CT has long been recognized as responsible for upward Fo movements. Vocalis seems to work in synergy with CT (Erickson 1976). The role of other muscles is more controversial. Many authors (Ohala & Hirose 1970, Atkinson 1978, Collier 1975, Erickson 1976, Sagart & Hallé 1986) have shown strap muscles to be active in lowering Fo. Honda (1988) finds the Cricopharyngeus active too. Various mechanisms have been suggested to explain how extrinsic laryngeal muscles can lower Fo. Ohala (1972) claims that strap muscles control the vertical position of the larynx, modifying the vertical tension of vocal folds. Among strap muscles, SH has been most studied. It participates in supralaryngeal articulation too, lowering or fixing the hyoid bone during jaw opening, tongue lowering and backing gestures (see Collier 1975 for a review).

In the present study, we have investigated CT, Vocalis, and SH activities, as we expected that those muscles were the most

---

* Laboratoire de Psychologie Expérimentale, CNRS (Paris).

likely to contribute in Fo control for MSC tone production. We have already shown that CT was responsible for raising Fo in high tones, and that tone T3, uttered as low level, was associated to an intense and consistent SH activity (Sagart & Hallé 1986). However, active Fo lowering by SH was not demonstrated by us for other tones, although it could be expected for the high falling tone T4, as has been shown in the falling tone of central Thai (Erickson 1976). This may have been due to the ambiguous role of SH, involved in both laryngeal and supralaryngeal articulation. So, the view that Fo fall in T4 is due to a passive relaxation of muscles like CT, cannot yet be disregarded. In the speech material used here, care has been taken to minimize SH involvement in segmental articulation in order to give an answer to this question.

Tone contours in Chinese can be affected by various linguistic factors. One is citation form versus running speech, but citation form is roughly equivalent to prepausal position in running speech. Within running speech, contours are affected mainly by 2 factors: tonal context, and stress. Contextual effects may be merely mechanical, but some are linguistic. Stress on syllables has various degrees, ranging from highly stressed to totally unstressed . In the latter case, the tone contour is affected and becomes highly dependent on the tonal context (see M. Yip 1984, for a tentative rationalization). Thus a special tone called neutral or light tone has been posited. Acoustic correlates of stress are primarily duration and pitch, and to a lesser extent, amplitude (see the "syllabic volume" proposed by Kratochvil 1981). A recent study by Kratochvil (1985), based on a large corpus of running speech has described the relationship between tone contour and syllable duration - which correlates with stress: high tones T1, T4 and also T2 reach higher Fo values for longer durations, and the converse outcome holds for the low tone T3 (see Fig. 1). These acoustic data suggest that longer, i.e. more stressed syllables are produced with more intense CT activity for high tones, whereas they are produced with more intense SH activity for T3. However, some aspects of Kratochvil's data are quite intriguing, like tone T2 lower Fo onset for longer durations, and call for more elaborate explanations in terms of underlying laryngeal gestures.

## Method

### Speech material, subject, and experimental setting

As in our previous experiment (1986), we used syllables embedded in a frame sentence, to avoid contamination by non-speech muscular activity. The frame sentence was /yi2ge X zi4/ (a character X), X being the syllable under scrutiny, belonging to minimal series sharing the same segmentals at the 4 tones. In order to minimize SH contribution to supralaryngeal articulation, that had blurred its role in Fo control in the previous experiment, we have used here target syllables consisting of a bilabial initial [p] and [m], or a glide initial [ʲ] followed by the high front vowel [i] (/bi/, /mi/, and /yi/). Such segmentals minimize jaw opening (high vowel, bilabial closure) and both tongue lowering and backing (high front vowel). In addition, for sake of comparison, we have used the segmentals /hu/ ([xu]): vowel [u] is high back rounded and the fricative initial [x] is velar. Consequently, we would expect little jaw opening or tongue lowering, but a fair amount of tongue backing. It is important to note that the target syllable X is not in prepausal position, is stressed, and preceded and followed by unstressed syllables. This is to avoid strong tonal context effects as well as intonation downdrift on the last syllable of the breath group. The subject was a male native speaker of MSC, born and raised in Beijing, aged 26, with no known speech pathology. The hooked wire EMG electrodes were inserted in CT, Vocalis, and SH, using the long established technique of the RILP Institute, by Dr. Niimi. Correct insertion was controlled with various non-speech maneuvers before and after the experiment, and periodically during its course. The subject was asked to pronounce - at a comfortable speech rate, and distinctly - the 16 utterances (4 segmentals x 4 tones) in 10 separate blocks. Electrode checking was performed every 3 blocks. The Audio and 3 EMG signals were recorded by means of a U-matic video recorder. It appeared later that although the Vocalis electrode had been correctly inserted, the corresponding electric signal had been less amplified than other EMG signals.

Both EMG and phonetic data, are of interest in this paper, in particular timing relationships between them. We will present first the method used to analyze and average across repetitions the EMG data, then present and discuss phonetic measurements, and then EMG activities.

Data analysis

A preliminary analysis of the data, using the RILP Institute PDP 11 system, proved that the data were of interest (Hallé, Niimi & Imaizumi 1989). However, in order to allow further analyses on standard PC computers, the recorded signals have been replayed, digitized an stored into 4-channel signal files (one file per utterance, thus a total of 160 files). The sampling rate was 28 KHz (7 KHz per channel), the resolution 12 bits, and each channel was first low-pass filtered at 3 KHz. Files were then transferred to a standard PC computer. Figures 2-4 show the different stages of signal processing: interleaved 4-channel signal files are first split into single channel files so as to take advantage of existing one-channel signal processing routines (Fig. 2). From the Audio signal, Fo, Amplitude, and Spectral Derivative curves are computed. The latter is used for time alignment and time normalization, which are discussed later, as well as for phonetic measurements (Fig. 3). Raw EMG signals are low-passed filtered at 1 KHz before amplitude is computed (zero offset has been controlled and no high-pass filtering has been used) (Fig. 4). For all curves, the frame period was 10 ms. As a rule, SH and CT signals are rather powerful and clean, but Vocalis signal is weak and probably suffers from a low signal to noise

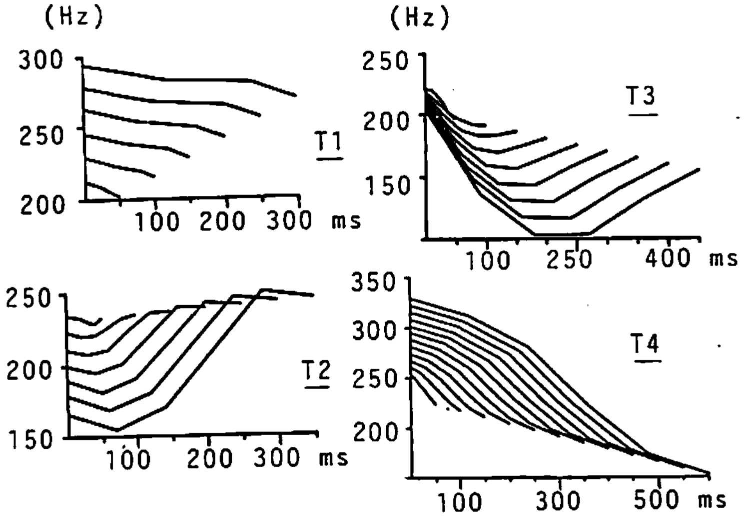


Fig. 1. Relation between tone contour and tone duration. (From Kratochvil 1985)

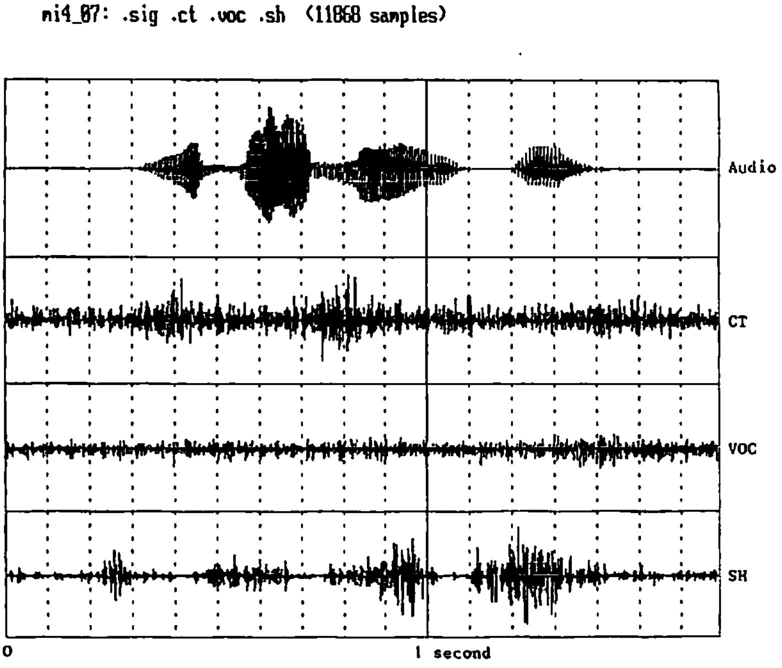


Fig. 2. Audio and EMG raw signals: /yige mi4 zi/ repetition #7.

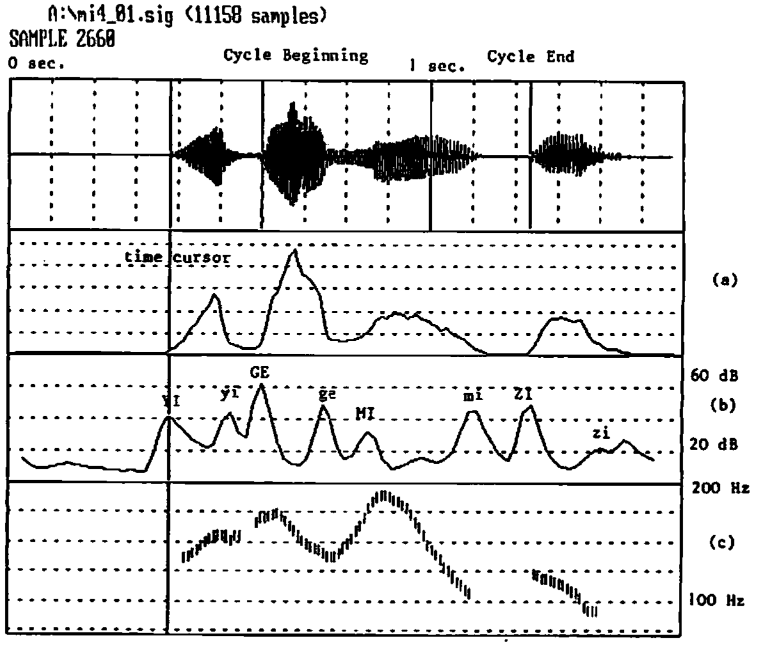


Fig. 3. Audio signal, Fo (c), amplitude (a), and spectral derivative (b). Letters on spectral derivative peaks correspond to vocalic onsets (upper case) or offsets (lower case).

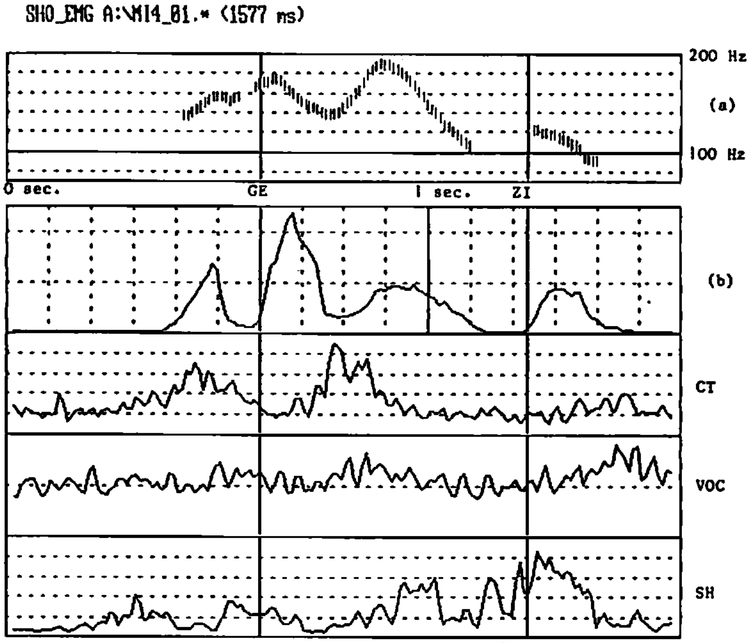


Fig. 4. Audio signal Fo (a) & amplitude (b), EMG amplitudes after low-pass filtering of 1 KHz. The grid scale is 25 microvolts between 2 dashed lines.

ratio. However, subsequent analyses seem to show that the recorded activity of Vocalis is not random.

### Averaging method

In order to line-up and average occurrences of a given utterance, investigators usually locate a specific acoustic event in each occurrence and perform the averaging around it (within a certain averaging range). Indeed, there is no problem when the domain of interest is very close to the line-up event, and when articulation rate fluctuations are small across repetitions. We propose here an alternative method that makes use of 2 line-up references. Occurrences of a given utterance are both time aligned with respect to 2 separate events and time normalized so that the interval between the 2 events is made constant. The underlying idea is to reduce the variability of time locations induced by articulation rate fluctuations, and to widen the domain where averaging is valid. The 2 line-up events chosen here are the vocalic onsets of syllables /ge/ and /zi/. We use the terms line-up or reference "cycle" for the domain extending from one line-up event to the next, that is [/ge/ -> /zi/] in our case. For small variations of the cycle duration, it is reasonable to assume that relative locations are preserved, in other words, that everything is scaled to the cycle duration, so that all events within the cycle are correctly aligned after time scaling has been completed.

Since this assumption had to be checked, we have measured the position of the events related to the target syllable within the line-up cycle: syllabic onset and offset and vocalic onset. The latter measurement was sometimes impossible to make for /yi/ syllables where the boundary between the initial glide and the vowel may be fuzzy. However, the offset of the syllable (coinciding with the offset of the vowel) could always be determined precisely. This event is of primary interest since it coincides with the tone contour endpoint. All the locations, including those of the line-up events, have been estimated in a way that is equivalent to spectrogram visual inspection: they have been made from the spectral derivative curve whose peaks signal the main spectral changes in the speech flow. The process is illustrated in Fig. 3: spectral derivative peaks signal clearly vocalic onsets and offsets. The results show that all locations are roughly scaled in proportion to cycle duration. This is particularly true for vocalic offsets: For example, the linear regressions computed for /bi/ vocalic offset location against cycle duration at each of the 4 tones all yield a very small constant coefficient (mean=41 ms, sd=70 ms), so that /bi/ tone contour endpoint location can be considered, for each tone, as very close to proportional to cycle duration and will occupy within the time normalized cycle a location rather invariant across repetitions. This is found for all syllables. Fig. 5 illustrates the relation of vocalic onset, offset, and duration, to cycle duration for one target syllable. Table 1 summarizes what has been gained by performing the line-up cycle time normalization: standard deviations of vowel onset, offset, and duration decrease dramatically after they have been time-scaled to cycle duration (a specific "reference cycle" is used for each target syllable: it is chosen among the 10 available occurrences as the closest to the mean). Systematic differences between segmentals or between tones, in vocalic onset location and vowel duration, will be discussed in the next section. These differences would entail line-up errors when line-up is tried between different tones or different segmentals. Therefore, we have limited ourselves to lining-up and averaging across repetitions only.

## Results of phonetic measurements

For each target syllable, we report here the mean durations of phonetic segments and their position in the "reference cycle". Segmental durations per syllable type and tone are summarized in Table 2. The vowel durations are rather stable, ranging from 170 ms to 245 ms. However, they are a little shorter for /yi/ and /hu/ syllables. Whereas /yi/ is traditionally transcribed [i], we find a substantial glide initial component (about 80 ms) for the target syllable /yi/, which follows another opened syllable (/ge/) in the frame sentence that has been used. In a similar context, Howie (1976) does not consider there may be any glide. For the sentence initial /yi/, we do not observe any glide.

Vocalic durations are always longer under the tones T3 and T4 than under tones T1 and T2. There is a trend for vocalic onsets to occur earliest for tone T4 (when the initial is not the plosive initial [p]): at the same time, Fo at vocalic onset is always highest for tone T4. These two facts may be related. Fo contours found for each tone are in agreement with the accounts given in literature (Howie 1976; Kratochvil 1981, 1985). However tone T3 has no rising end, it is consistently low and level. The contours are very similar across segmentals, when the vocalic portion only is considered. Fo movements in voiced initials appear as rather smooth transitions between /ge/ and the following syllable.

## Discussion

Howie (1974) showed that the domain of tone in MSC is not the entire voiced part of the syllable, or the entire vocalic part, but is limited to the syllabic vowel and any voiced segment that may follow it. In other words it is the rhyme. This means that Fo contour in voiced initial sonorant like [m] and glides that may precede the syllabic vowel are irrelevant to the tone contour specificity. From a phonological point of view, the "domain of tone" should be a time interval delimiting the tone contour attached to the syllable, 1) minimal but sufficient for contrasting tone contours and 2) ensuring a maximal independence of tone contours from the underlying segmentals and the context. In that meaning, Howie's view is well supported by our data in 2 different ways. First, examination of durations indicate that rhyme (i.e. the vowels in our data) durations are rather

Table 1. Time measurements for each target syllable, before and after "reference cycle" time normalization.

| TARGET SYLLABLE | | CYCLE | Mean, SD (before) & *(after)* normalization for: | | |
|---|---|---|---|---|---|
| seg. | tone | duration (ms) | vowel onset | vowel offset | vowel duration |
| | T1 | 583 | 295 (16),*(11)* | 465 (29),*(10)* | 170 (17),*(10)* |
| /yi/ | T2 | 601 | 319 (19),*(16)* | 497 (17),*(7)* | 180 (11),*(13)* |
| | T3 | 598 | 305 (19),*(16)* | 489 (25),*(11)* | 198 (9),*(9)* |
| | T4 | 609 | 268 (17),*(12)* | 488 (22),*(10)* | 222 (15),*(10)* |
| | T1 | 638 | 295 (12),*(7)* | 517 (18),*(7)* | 222 (9),*(7)* |
| /bi/ | T2 | 628 | 299 (27),*(16)* | 528 (32),*(13)* | 228 (11),*(11)* |
| | T3 | 646 | 295 (24),*(16)* | 531 (27),*(11)* | 236 (15),*(15)* |
| | T4 | 643 | 286 (16),*(11)* | 517 (24),*(11)* | 233 (16),*(13)* |
| | T1 | 629 | 297 (21),*(13)* | 512 (29),*(11)* | 214 (12),*(8)* |
| /mi/ | T2 | 624 | 295 (14),*(10)* | 514 (20),*(15)* | 219 (9),*(8)* |
| | T3 | 636 | 296 (18),*(13)* | 529 (16),*(9)* | 233 (11),*(12)* |
| | T4 | 643 | 274 (13),*(12)* | 519 (25),*(9)* | 245 (19),*(11)* |
| | T1 | 628 | 337 (16),*(11)* | 526 (22),*(11)* | 189 (13),*(10)* |
| /hu/ | T2 | 634 | 340 (10),*(7)* | 534 (14),*(8)* | 195 (10),*(8)* |
| | T3 | 648 | 342 (12),*(4)* | 542 (19),*(6)* | 200 (11),*(10)* |
| | T4 | 623 | 311 (9),*(9)* | 497 (17),*(11)* | 200 (12),*(9)* |
| SDs' summary | | | (16),*(11)* | (22),*(10)* | (12),*(10)* |

Table 2. Syllable initial component and rhyme durations, rhyme onset location per tone and per segmentals (in ms).

| TARGET SYLLABLE | | initial seg. | rhyme | rhyme onset |
|---|---|---|---|---|
| seg. | tone | duration | duration | location |
| | T1 | 103 | 170 | 295 |
| /yi/ | T2 | 131 | 180 | 319 |
| | T3 | 108 | 198 | 305 |
| | T4 | 89 | 222 | 268 |
| | T1 | 120 | 222 | 295 |
| /bi/ | T2 | 128 | 228 | 299 |
| | T3 | 127 | 236 | 295 |
| | T4 | 127 | 233 | 286 |
| | T1 | 119 | 214 | 297 |
| /mi/ | T2 | 128 | 219 | 295 |
| | T3 | 124 | 233 | 296 |
| | T4 | 111 | 245 | 274 |
| | T1 | 157 | 189 | 337 |
| /hu/ | T2 | 170 | 195 | 340 |
| | T3 | 173 | 200 | 342 |
| | T4 | 148 | 200 | 311 |

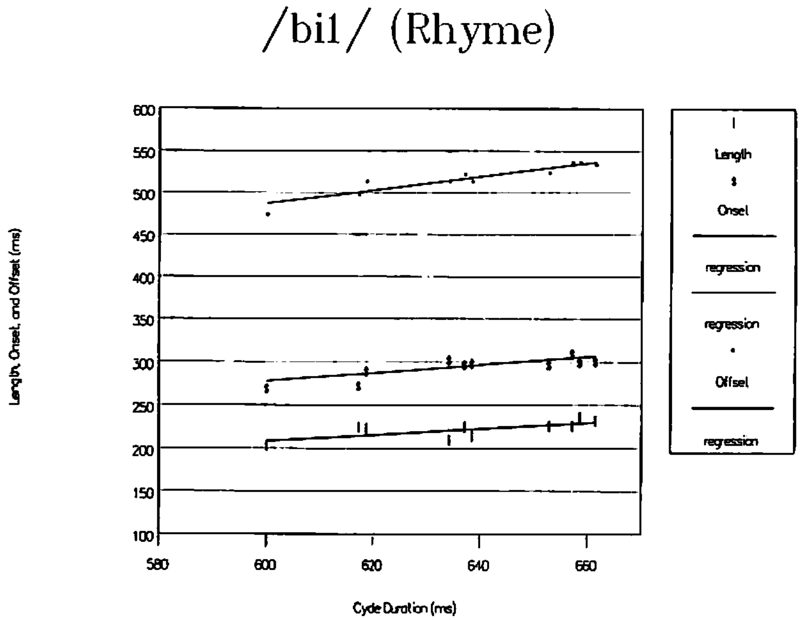


Fig. 5. Vocalic onset, duration, and offset against line-up cycle duration for the 10 repetitions of /bil/ (onset & offset are relative to the cycle beginning).

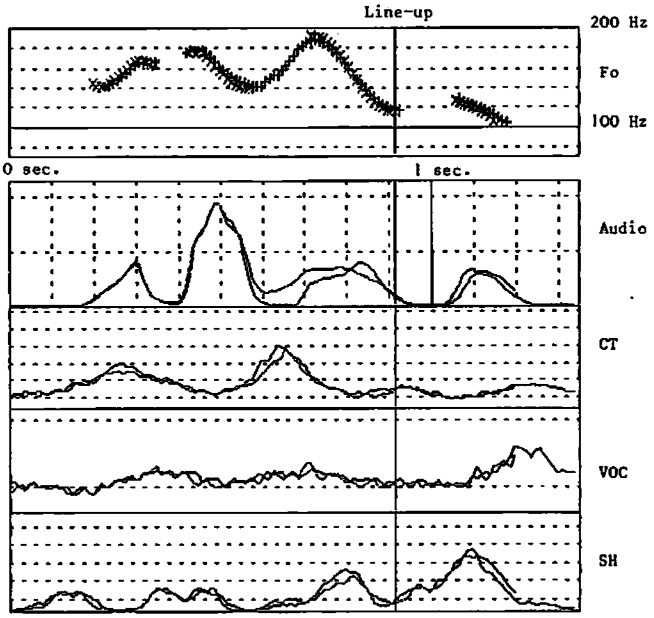


Fig. 6. Superimposed Fo and EMG curves for averaged /mi4/ (+) and /bi4/ (x), lined-up from [i]'s offset: the rhyme is the interval where Fo contours and relative timing with EMG patterns are maximally similar across segmentals.

invariant, whereas syllable entire voiced part durations are not. Second, if we superimpose tone contours, aligning them from their offset, it appears clearly that the common part between, say, /bi/ and /mi/ is the rhyme, whatever the tone may be. The Fo contour in the initial [m] of /mi/ is foreign to the tone contour per se. This is most apparent in the case of tone T4, where the top of the contour, which can be thought as a reference event of T4 contour, occurs at a quite invariant position relative to the rhyme, but certainly not relative to the entire voiced part of the syllable. From a motor control standpoint, the domain of tone should be a time interval maximizing, for each of the 4 tones, the invariance of EMG pattern timing relationships with it. The next section will confirm that, from this standpoint, the domain of tone should be - again - the rhyme and not the entire voiced part of the syllable. This is illustrated in Fig. 6.

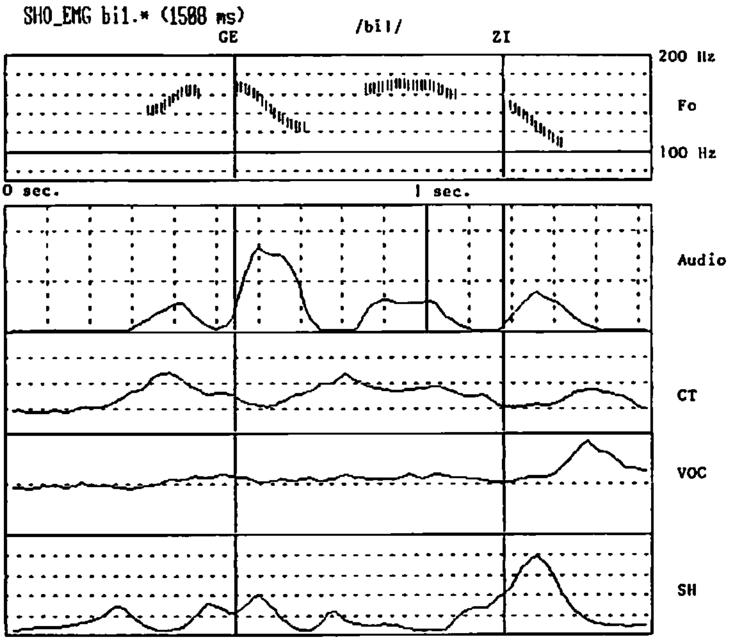
Tone seems to affect rhyme duration, but authors report conflicting results about it. Our data is in agreement with the findings of Howie (1976), who also used carrier sentences, apart from tone T3 which we find slightly shorter than T4. However, T3 appears consistently low level with no rising end in our data, contrarily to Howie's. Kratochvil (1981), who used a corpus of running speech, reports a somehow different ordering of durations, with tinier differences between tones. Such discrepancies reflect, we believe, the nature of the corpus used. In Howie's data or ours, the syllables analyzed were all stressed syllables, while in Kratochvil's, all syllables were considered.

## Results of EMG measurements

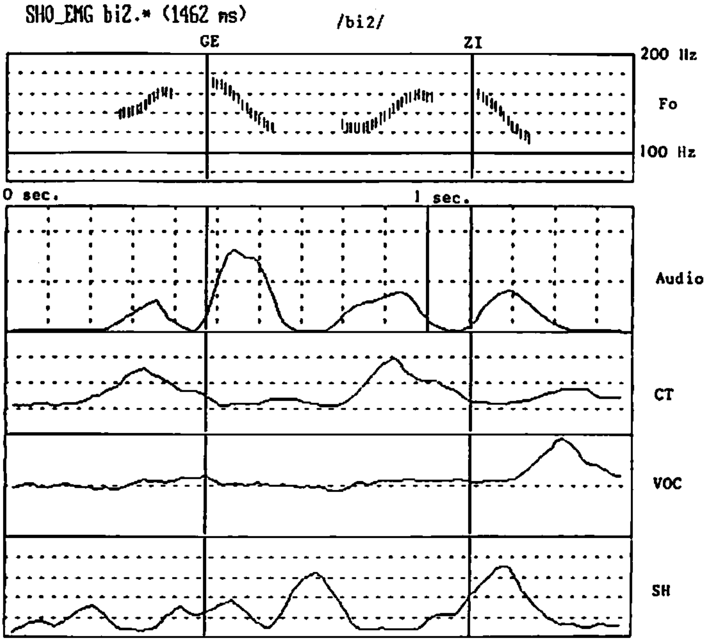
### Canonical patterns for tones

The patterns of CT and SH activities are found to be very stable across segmental variations. Figs. 7.a-d show the patterns for /bi/. The timing relationships of the patterns are found to be stable and consistent with respect to the rhyme but not the entire voiced part of the syllable, confirming the view that rhyme is the proper domain of tone. The patterns can be described as follows:

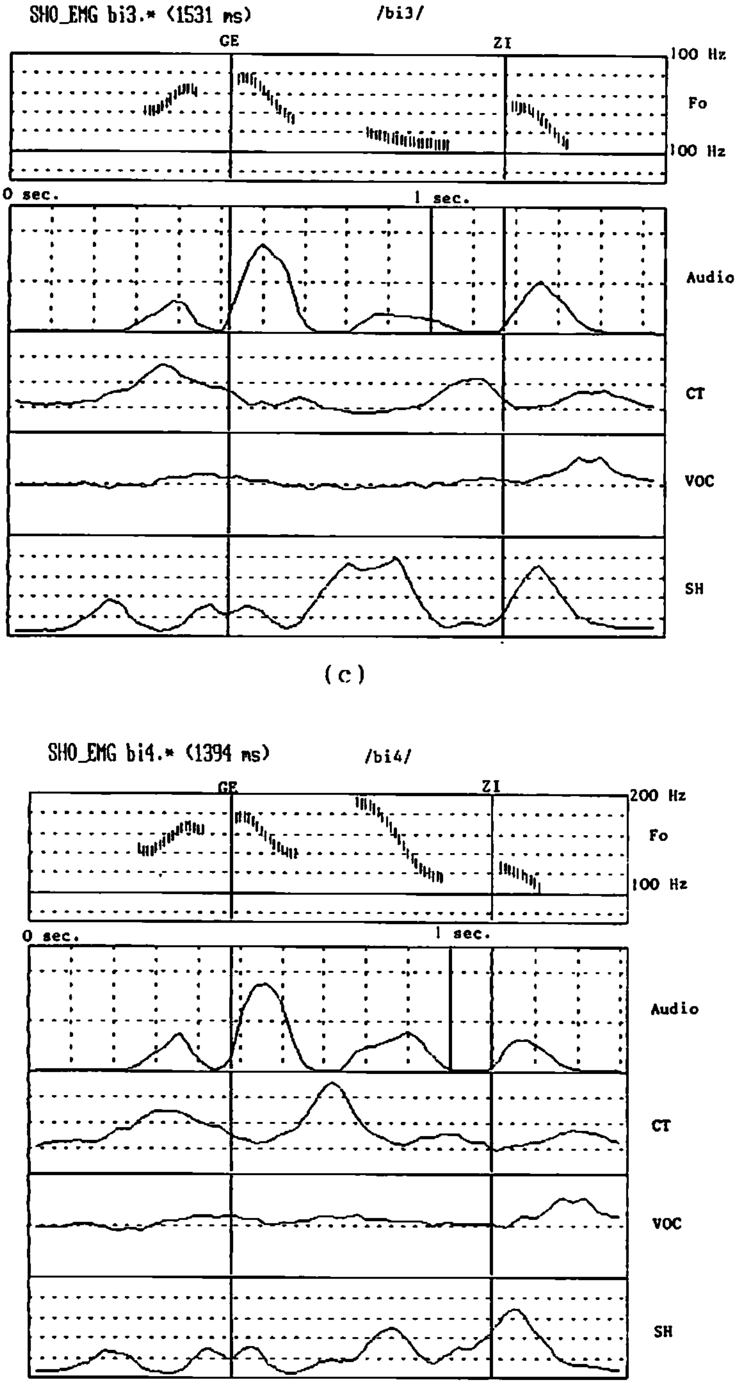
- Tone T1: CT activity begins to increase at about 200 ms before rhyme onset, reaches a peak of moderate intensity at 75-80 ms before rhyme onset, and finally decreases to a steady level that is maintained until the end of the rhyme. SH shows a very small increase of activity, centered on the syllable initial component, that seems to be related to segmental articulation. It is the lowest for /mi1/ and /yi1/, and the highest for /bi1/: jaw opening for [i] is indeed very moderate but more marked with [p] initial. As for /hu1/, the increase of SH activity is a little smaller than in the case of /bi1/, but occurs earlier and is more spread. This might be explained by a greater latency for the tongue backing gestures in [xu].
- Tone T2: A noticeable SH activity reaches a peak value 70-80 ms before rhyme onset. It is adding onto the SH activity found for T1, larger for /bi2/ and more spread, starting earlier for /hu2/.


SHO_EMG bi1.* (1588 ms)
/bi1/
GE
ZI
200 Hz
Fo
100 Hz
0 sec.
1 sec.
Audio
CT
VOC
SH

(a)


SHO_EMG bi2.* (1462 ms)
/bi2/
GE
ZI
200 Hz
Fo
100 Hz
0 sec.
1 sec.
Audio
CT
VOC
SH

(b)



(c)

(d)

Fig. 7.a-d Tone patterns for /bi/: Fo, Audio, CT, Vocalis, SH

CT activity starts much later in the syllable than for tone T1, and is more intense and concentrated. Its evolution parallels the Fo contour, but precedes it by 75-80 ms (inter-peak distance).

- Tone T4: CT activity is very intense and parallels Fo curve in a striking manner, with a lead of 70-80 ms (inter-peak distance). CT peak activity occurs at about 45 ms before rhyme onset. A moderate concentration of SH activity consistently appears, centered a little before the mid point of the rhyme. This SH activity does not appear at all at the other tones. At the beginning of the target syllable, the same weak SH activity as for T1 also appears.
- Tone T3: SH activity is extremely intense for this tone. It generally appears as consisting of 2 overlapping bulges, the 1st centered on a rather fuzzy peak 30-35 ms before rhyme onset, the second 55-75 ms after rhyme onset. As for T2, SH activity for T3 seems to add onto T1 SH activity. There is no CT activity for tone T3 (the CT activity at the end of the target syllable must be related to the following syllable /zi/ at tone T4).

The new finding in this study is the consistent SH activity for tones T2 and T4. Since SH is also involved in supralaryngeal articulation, the question arises as to whether SH activity found for T2 and T4 is related to Fo control or not. A simple examination of the curves shows that SH activity is minimum - in the domain of the target syllable - for tone T1. It seems significantly higher for other tones, for T3 indeed, as one can easily persuade oneself, but also for T2 and T4. In order to assess this difference, we have performed paired comparisons between SH activity values corresponding to 2 different tones (with the same segmentals), at each point of time. The result is plotted as a set of Students' t values along time. Fig. 8 illustrates the comparison of /bi1/ and /bi2/: the area where the Students' t values are highly significant lies precisely where Fo-oriented SH activity was suspected. The same results are obtained for all the T1-T2 and also T1-T4 comparisons, whatever the segmentals be.

Another means for quantifying differences consists in computing overall distances between EMG curves. We propose the following L1 normalized distance:

$$D(x,y) = SUM_{[t0,tf]}\{|x(t)-y(t)|\} \ / \ SUM_{[t0,tf]}\{x(t)+y(t)\}$$

where x() and y() represent 2 EMG curves that have been first lined-up and time normalized together in the range [t0,tf]. The second sum stands for amplitude/duration normalization (note that this normalization cannot compensate for EMG signals with a low signal to noise ratio, like Vocalis EMG signal in our data). Table 3 indicates that distances between tones within the same segmentals are always greater than distances between segmentals within the same tone category. It means that the effect of "tone" factor overrides the effect of "segment" factor when tone is varied from T1 to T4 and segmentals from /yi/ to /hu/. In other words, the EMG patterns that have been found are related to tone production. This result also holds for Vocalis activity, although the difference may seem tiny. However, more informative paired

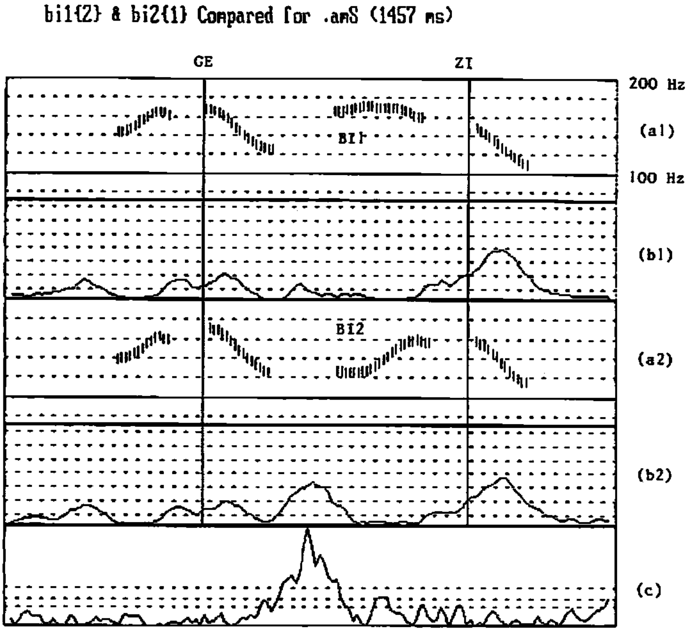


Fig. 8. Paired comparison along time of /bi1/ and /bi2/ for SH activity: SH activity at tone T2 syllables' onset is tone-related. Fo (a) and SH (b) for /bi1/ and /bi2/, Students' t along time (c). The 3 levels correspond to p=0.05, p=0.01, and p=0.001.

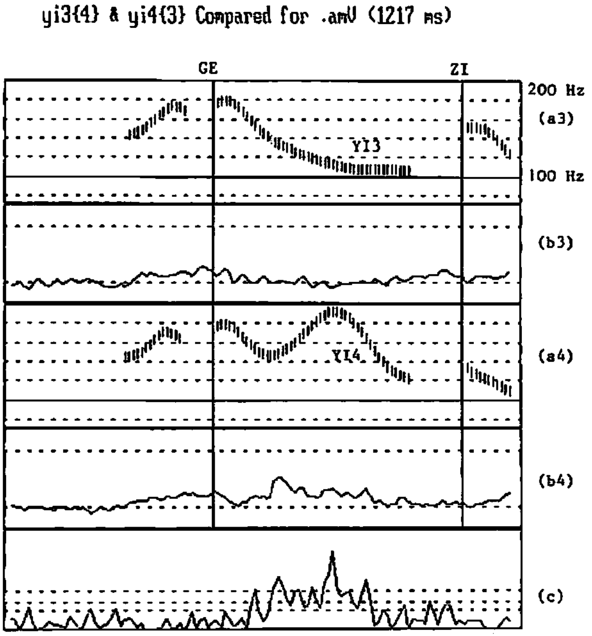


Fig. 9. Paired comparison along time of /yi3/ and /yi4/ for VOC activity: Vocalis activity at tone T4 syllables' onset is tone-related. Fo (a) and Vocalis (b) for /yi3/ and /yi4/. (c) same as in Fig. 8.

Table 3. Overall distances between EMG activities. Inter-tone within the same segmentals, and inter-segment within the same tones.

*Inter-tone, same segmentals*

| pair of tones | CT | SH | Vocalis |
|---|---|---|---|
| T1 T2 | 0.094 (.006) | 0.135 (.010) | 0.034 (.001) |
| T1 T3 | 0.124 (.012) | 0.260 (.014) | 0.041 (.002) |
| T1 T4 | 0.099 (.013) | 0.166 (.008) | 0.038 (.004) |
| T2 T3 | 0.092 (.006) | 0.192 (.020) | 0.035 (.003) |
| T2 T4 | 0.150 (.011) | 0.194 (.030) | 0.044 (.004) |
| T3 T4 | 0.146 (.012) | 0.220 (.019) | 0.039 (.004) |
| *mean* | 0.118 | 0.195 | 0.038 |

*Inter-segment, same tones*

| pair of segmentals | CT | SH | Vocalis |
|---|---|---|---|
| yi bi | 0.051 (.003) | 0.098 (.022) | 0.034 (.002) |
| yi mi | 0.042 (.003) | 0.090 (.012) | 0.034 (.003) |
| yi hu | 0.061 (.015) | 0.124 (.017) | 0.048 (.006) |
| bi mi | 0.044 (.004) | 0.080 (.016) | 0.029 (.001) |
| bi hu | 0.050 (.009) | 0.095 (.025) | 0.032 (.005) |
| mi hu | 0.052 (.012) | 0.098 (.021) | 0.035 (.002) |
| *mean* | 0.050 | 0.097 | 0.035 |

comparisons along time described above can be used to show that Vocalis activity is different between tones. In the target syllable domain, Vocalis has the lowest activity for tone T3 and is active otherwise in about the same way and the same areas as CT. This can be seen through T3-T1, T3-T2, or T3-T4 comparisons (Fig. 9). Therefore, in spite of the low amplitude and the low signal to noise ratio of the Vocalis EMG signal, we are entitled to conclude that Vocalis and CT have synergic activities, agreeing here with the findings of Erickson (1976). A closer inspection of Table 3 also reveals that distances between tones for CT and Vocalis are correlated (.844, df=4), whereas they are not for CT and SH (.551, df=4) or SH and Vocalis (.578, df=4). Again, this tends to show that CT and Vocalis work together.

Patterns relevant to breath group

All utterances end in the same manner. First, a very intense SH burst of activity, that cannot be ascribed solely to the production of tone T4 in /zi/, since it is almost twice as intense as tone T4 SH activity in the preceding stressed syllable. We suggest that this SH intense activity is responsible for the intonation downdrift that terminates breath groups. Second, there is a rather intense activity of both CT and Vocalis, centered on /zi/ syllable offset, that faints immediately away. This maneuver must increase vocal fold tension and thickness, and result in vocal fold adduction. Bearing in mind that it is performed 1) in a low pitch register, 2) at the end of the expiratory process, when transglottal pressure drop tends to vanish, we may think of it in terms of a voice termination maneuver that allows for only a few more glottal pulsations in a creaky mode (creaky voice was actually observed throughout at the end of utterances).

Other syllables

We have described sofar the EMG patterns attached to the target syllables. However, these patterns also apply rather well to other syllables. We have already pointed out that the CT activity that occurs just after the target syllable at tone T3 was relevant to the production of /zi/ syllable's tone T4. Let us now review each syllable of the utterances, starting with /zi/:

- /zi/: The EMG pattern for /zi/ falls in the tone T4 category: CT activity before rhyme onset and SH activity within the rhyme. However, the SH activity attached to that syllable participates primarily in the final intonation downdrift that terminates the breath group. CT activity seems to be scaled to the moderate degree of stress of syllable /zi/. After a tone T1, T2, or T4 syllable, it has similar low intensities (it overlaps with the CT activity attached to a preceding T1 or T2 syllable). It follows that /zi/ maintains a high-mid falling contour after T1 or T2, but has a low-mid falling contour after T4 whose Fo endpoint is very low. This would lead us to conclude that motor control is rather invariant, and tonal context effects merely mechanical. However, after T3, the SH activity attached to /zi/ is quite substantial, and /zi/ preserves a high-mid falling contour although T3 Fo endpoint is even lower than T4 Fo endpoint. This simple fact precludes the - tempting - idea of invariance at the motor control level.
- /yi/ and /ge/: The EMG pattern for utterance initial /yi/ fits perfectly with tone T2 pattern. This /yi/ is a rather special morpheme. It should be normally pronounced at tone T1 before a pause, at tone T2 before tone T4, and at tone T4 before other tones (Chao 1968). As a matter of fact, the following /ge/ has clearly a tone T4 contour. So that the sequence /yige/ has been pronounced T2-T4. The pattern (SH, CT) is clear for /yi/. As for /ge/, the initial CT activity is rather weak, apparently overlapping with the activity attached to /yi/. Otherwise, there is a noticeable SH activity at mid rhyme, consistent with tone T4 pattern. This probably Fo-oriented SH activity is preceded by another concentration that is segment-oriented. Note that Vocalis activity seems to extend the action of CT in order to maintain a proper vocal fold tension at /ge/ onset. Returning to /yi/, SH and CT intensities are scaled to /yi/'s duration, i.e. degree of stress. Its rhyme duration, considering all utterances, is about 120 ms, against 180 ms for the target syllable /yi2/. Interestingly, the time latency for CT (inter-peak distance) is 50-55 ms, against 75-80 ms for the target syllable /yi2/. Therefore, the latency time too appears to be scaled to the degree of stress. This unexpected result needs more data to be confirmed, and deserves, by itself, further experiments.

## General discussion

In this paper, we have presented a picture of EMG patterns for the production of MSC's four tones in stressed syllables within a quite neutral context, and at a rather invariant articulation rate - depicted as "comfortable". We call the patterns "canonical", suggesting that further research is needed to know how they would change with different 1) stress, 2) tonal context and 3) articulation rate conditions.

The finding of SH activity at the onset of tone T2 and in the middle of tone T4 explains Kratochvil's account of how tone contours vary according to duration. Kratochvil finds that the longer is a tone T2 syllable, the lower its tone contour onset, the longer is a tone T4 syllable, the lower its tone contour offset. This can only be explained if there exist an active Fo lowering device for these 2 tones: SH activity is such a device. Finally, all these data put together are coherent as soon as we assume that stress entails lengthening and generally more intense muscular activities, including laryngeal activities.

We believe that some aspects that are found consistently in our data deserve tentative interpretations. In particular, it can be asked why a non-stressed, prepausal syllable at tone T4 needs a more intense CT activity after tone T3 than after other tones ? This situation, we believe, is quite similar to the well known "third tone sandhi". If no sandhi occurred when two tones T3 appear in succession, Fo control would be achieved by two intense SH activity bulges in succession, and nothing more. In our data, for utterances terminated with a stressed tone T3 followed by an unstressed syllable /zi/ at tone T4 where a sharp intonation downdrift occurs, Fo control should similarly be achieved by two intense SH activity bulges in succession, and a moderate CT activity in between, scaled to the low degree of stress of /zi/, like in utterances with tone T1, T2, or T4 target syllables. We suggest that in both cases, phonation would not be possible without an active - and substantial - Fo resetting between the two SH activity bulges. This would explain both the larger than expected CT activity for /zi/ after tone T3, and the transformation of the first tone T3 of two consecutive tone T3 syllables. Since SH activity occurs later in the syllable /zi/ than it would in a tone T3 syllable, the Fo resetting caused by CT can also occur later. In other words, the basic timing pattern specific to tone T4 does not have to be changed, and a tone T4 contour is preserved for /zi/. As for two consecutive T3 syllables, Fo resetting has to intervene earlier, that is, at least slightly, before the second SH bulge, and this means within the first syllable rhyme. So, the EMG pattern of the first tone T3 becomes similar to tone T2 pattern. However, there should be some discrepancies such as, in particular, a larger SH intensity that might explain the lower Fo in the transformed T3 than in T2, observed by some authors (Kratochvil 1987).

## References

1) Chao, Y.R.: A Grammar of Spoken Chinese. Berkeley, Univ. of California Press, 1968.

2) Kratochvil, P.: The case of the third tone. In Wang Li Memorial Volume. Hong Kong, Joint Publishing Co. 1987.

3) Kratochvil, P.: Tone and stress discrimination in normal Peking dialect speech. Proceedings of the Second Conference on Chinese Language Use. Canberra. 1981.

4) Ohala, J.: Production of tone. In V. Fromkin (ed.), Tone: A Linguistic Survey. New York, Academic Press. 1978.

5) Hirose, H., Z. Simada, & O. Fujimura: An Electromyographic study of the activity of the laryngeal muscles during speech utterances. Ann. Bull. RILP, 4, 9-25, 1970.

6) Ohala, J. & H. Hirose: The function of the sternohyoid muscle in speech. Ann. Bull. RILP, 4, 41-44, 1970.

7) Gårding, E., O. Fujimura & H. Hirose: Laryngeal control of Swedish word tones. Ann. Bull. RILP, 4, 45-54, 1970.

8) Hirose, H. & T. Gay: The activity of the intrinsic laryngeal muscles in voicing control: an electromyographic study. Phonetica, 25, 140-164, 1972.

9) Collier, R.: Physiological correlates of intonation patterns. JASA, 58 (1), 249-255, 1975.

10) Erickson, D.: A Physiological Analysis of the Tones of Thai. Unpublished Ph.D. Dissertation, Univ. of Connecticut, 1976.

11) Sagart, L. & P.A. Hallé: Tone production in modern standard Chinese: an electromyographic investigation. CLAO, 14 (2), 153-174, 1986.

12) Atkinson, J.E.: Correlation analysis of the physiological factors controlling voice frequency. JASA, 63 (1), 211-222, 1978.

13) Honda, K.: Action of the cricopharyngeus muscle in lowering voice fundamental frequency. Proceedings of ASJ Spring Meeting, 169-170, 1988 (in Japanese).

14) Ohala, J.: How is pitch lowered ? Abstract in JASA, 52, 1.124 (A), 1972.

15) Yip, M.: The Tonal Phonology of Chinese Tones. (Ph.D.) Bloomington, Indiana Univ. Linguistic Club, 1980.

16) Kratochvil, P.: Variable norms of tones in Beijing prosody. CLAO, 13 (2), 135-174, 1985.

17) Hallé, P.A., S. Niimi & S. Imaizumi: Tone-specific patterns of laryngeal control in Chinese. Proc. of Speech '89 Congress, Budapest, 1989.

18) Howie, J.M.: Acoustical Studies of Mandarin Vowels and Tones. Cambridge University Press, 1976.

19) Howie, J.M.: On the domain of tone in Mandarin. Phonetica, 30, 129-148, 1974.