TONE PRODUCTION IN STANDARD CHINESE: EMG DATA AND COMMAND-RESPONSE MODELLING

P.A. Hallé

Laboratoire de Psychologie Expérimentale, Paris, France.

ABSTRACT

A model of tone production in Standard Chinese is presented and confronted to phonetic and EMG data. The model is of the command-response type: Fo is viewed as the response of the laryngeal structure to excitation commands. The same speech material was used to obtain EMG data and to model Fo contours so that tone production can be viewed from two different perspectives. EMG data reveal stable patterns of laryngeal muscle activity attached to each tone. Similar patterns obtain for the model commands.

1. INTRODUCTION

Second order linear systems are widely used for approximating the behaviour of various physical or biological sytems, with respect to some given dimension. Indeed, one reason for this is that linear systems and their mathematics are well understood. For many natural systems however, biomechanical properties make the approximation by a linear system reasonable. For the laryngeal structure, Öhman [8] proposed the first to model Fo contours by the response of a linear system to excitation commands. Fujisaki [2] took up the idea for speech and singing voices, and proposed a biomechanical interpretation of his model [3]. In short, commands can be understood as controlling the length of the vocal folds, hence their longitudinal tension, hence Fo (for a more accurate account, see Boë [1]). However close to the physics these models may come, they would not be of much use if linguistically relevant patterns of commands could not be identified among the stream of the many commands required for modelled contours to closely follow actual Fo data. Öhman and Fujisaki both identified simple patterns attached to the production of pitch accent. We applied the same ideas to model Fo contours in Standard Chinese, proposing that qualitatively stable patterns of commands be attached to each tone type. Starting with the simplest patterns [9], we gradually came to the patterns presented in section 4. Our model not only provides an economical account of tone production, but also explain tone contour changes due to the tonal coarticulation that occurs in running speech.

EMG studies of laryngeal muscles have evidenced stable patterns of activity attached to each tone type [5]. It is tempting then, to bring together Fo modelling and EMG data obtained with the same speech material.

2. MATERIAL

The speech material was designed for EMG experiments, where Cricothyroid (CT) and the Sternohyoid (SH) were examined. We used target syllables embedded in a frame sentence: /yi2ge X zi4/ (a character X). Target syllables X belonged to minimal series sharing the same segmentals in the four tones: [ʝi], [pi], [mi], and [xu] (in Pinyin transcription, /yi/, /bi/, /mi/, and /hu/). Those segmentals were chosen in order to minimize SH contribution to supralaryngeal articulation (some SH activity related to tongue backing was expected for /hu/). The target syllable X does not occur in prepausal position, is stressed and surrounded by unstressed syllables to avoid strong tonal context effects, as well as intonation downdrift on the last syllable of breath groups.

Hooked wire EMG electrodes were inserted in the CT, Vocalis, and SH. Correct insertion was checked with various non-speech manoeuvres before and after the experiment, and periodically during its course. Subjects pronounced the 16 sentences (4 segmentals x 4 tones) at a normal speech rate, in 10 separate blocks.

Correct insertion of the electrodes in the CT and SH could be achieved for 2 subjects, both male native speakers of Standard Chinese, born and raised in Beijing, aged 26 and 38, with no known speech pathology. Similar data were obtained for both. We use here the data from the first subject.

3. EMG PATTERNS

For each sentence, all repetitions were lined-up and time-normalized, using 2 reference events. this technique allow for averaging utterances on a wide domain, and for coping with speech rate fluctuations. One utterance per sentence, the closest to the mean with respect to the duration between line-up events, served for time scale reference.

Patterns of CT/SH activity related to tone production are found to be stable across segmental variations. The time relationships of the patterns are found to be stable and consistent with respect to the rime --not to the entire voiced part of the syllable. This confirms that the rime is the domain of tone [6].

Patterns can be described as follows:

- tone 1: CT activity begins to increase at about 200 ms before rime onset, reaches a peak of moderate intensity at 75-80 ms before rime onset, and finally decreases to a steady level that is maintained until the end of the rime.
- tone 2: SH activity reaches a peak value 70-80 ms before rime onset. CT activity starts much later in the syllable than for tone 1, and is more concentrated. It parallels the Fo contour, but precedes it by 75-80 ms.
- tone 3: SH activity is extremely intense for this tone. It begins to increase at about 100 ms before rime onset, and drops down a little before rime offset. There is no CT activity for tone 3 (the CT activity at the end of a target syllable must be related to the next syllable /zi/, in tone 4).
- tone 4: CT activity is very intense and parallels the Fo contour with a lead of 70-80 ms. CT peak activity occurs at about 45 ms before rime onset. A moderate concentration of SH activity consistently appears, centered a little before the mid point of the rime.

Note that what one may call "secondary activities" of the SH in tones 2 and 4 are found for both subjects. In order to show that these activities are tone-related, we have compared tone 2 or 4 to tone 1, where the smallest SH activity, presumably segment-related, is observed. Comparisons were made at each point of time between sets of utterances (see [5] for details). The region where tones 1 and 2 significantly differ with respect to SH activity is the region where "secondary" SH activity is found before rime onset. Similar results obtain for tone 4 versus tone 1: Fo fall in tone 4 is assisted by SH activity. Interestingly, these EMG patterns explain puzzling phonetic data on running speech: the longer a tone 2 syllable, the lower its tone contour onset, and the longer a tone 4 syllable, the lower its tone contour offset [7]. This can only be the result of an active Fo lowering device for tones 2 and 4. Indeed, SH activity is such a device.

Let us see now how much Fo modelling comes close to these data.

4. MODEL COMMAND PATTERNS

The model we propose here is adapted from Fujisaki's model for Japanese [2]: we use impulse commands to produce the "phrase component", which is assumed to represent the overall intonation, and step commands to produce local variations of Fo in the syllable domain. For Japanese, step commands are paired to form "accent commands": one onset step command followed by one offset step command of opposite amplitude. For Chinese, we call such pairs of commands "tone commands". We use both "positive" and "negative" tone commands: positive ones have an onset step command of positive amplitude and raise Fo, while negative ones have the opposite pattern and lower Fo. Time constants and damping coefficients characterize the responding

system. They are kept constant within a given utterance. However, the system is allowed to respond differently to onset versus offset step commands, and to positive versus negative commands. Critical damping is assumed for phrase commands but not for tone commands. Amplitudes and time locations of the commands characterize the excitation to the responding system. Practically, for a given utterance, the input to the model comprises the actual Fo data, and the initial estimates of excitation and system parameters. The latter are then optimized to minimize the discrepancy between the response of the system and the actual Fo data. Indeed, the optimization process does not lead to an unique solution. However, qualitative patterns of commands for each tone have emerged from our previous studies [4]. We use them as initial estimates, in order to reduce the search space of the optimization process. They may be summarized as follows: one positive tone command for tone 1, and, likewise, one negative command for tone 3, roughly spanning the whole rime; one main positive command followed by a weaker negative one for tone 4, and the opposite pattern for tone 2.
These patterns are qualitatively similar to the observed EMG patterns.

5. COMPARISON

We examined further the analogy by applying the model to the speech material described earlier. For each sentence, we analysed the utterance that had served for time scale reference in the processing of EMG data. Care was taken to standardize analysis conditions for all utterances. In particular, parameters for the optimization process were the same for all utterances, and initial estimates were similar across segmentals. Fig. 1 shows CT and SH activities, together with tone commands obtained for the segmentals/mi4/. Similar results obtain for other segmentals. Tone commands and CT/SH activities related to target syllables are compared with respect to their amplitude and their timing relative to the target syllable rime. Results are summarized in Table I.

For timing, there is a good agreement between CT/SH activities and tone commands. Positive tone commands parallel CT activity, while negative ones parallel SH activity. However, amplitudes are poorly correlated.

6. DISCUSSION

EMG activity reflects an internal force developped within a muscle, whereas commands just indicate target Fo values. Contraction of the CT for example, produces a motive force f_c which tends to lengthen the vocal folds. The linear system approximation entails that f_c counteracts mechanical resistances to motion: inertia, frictions, and elasticities. As simple mathematics can show, in order to raise Fo from a rest level to a high level, as in tone 1, rapidly enough to keep pace with the speech flow, f_c must overshoot the target value corresponding to the high level static equilibrium. When this level is reached, f_c drops down to the target value, and eventually faints away when the high level is given up. Hence, the typical profile of CT activity in tone 1.
That similar timing are observed for EMG activity and commands indicates that target values of Fo are programmed as target values of muscle tensions. Amplitudes of commands and EMG activities may correlate where Fo adjustments are stabilized, as after the onset of tone 1. Elsewhere, EMG amplitudes reflect dynamic aspects of Fo control, while commands reflect static equilibrium, that is, target Fo values.

REFERENCES

[1] BOE,L.J., COTTET,O. & PARADIS,L. (1983), "Modélisation des évolutions de Fo", *Bull.Inst.Phon.Grenoble 12*, 48-65.
[2] FUJISAKI,H. & SUDO,H. (1971), "A model for the generation of fundamental frequency contours of Japanese word-accent", *J.A.S.A. 27*, 445-453.
[3] FUJISAKI,H., TATSUMI,M. & HIGUCHI,N. (1981), "Analysis of pitch control in singing", in K.N.STEVENS & M.HIRANO (eds.) *Vocal Fold Physiology*, Univ. of Tokyo Press, 347-362.

[4] FUJISAKI,H., HALLÉ,P. & LEI,H. (1987), "Application of Fo contour command-response model to Chinese tones", *Proc. ASJ fall meeting*, 197-198.
[5] HALLÉ,P.A., NIIMI,S., IMAIZUMI,S. & HIROSE,H. (1990), "Modern Standard Chinese 4 tones: EMG and Acoustic patterns revisited", *Ann.Bull.RILP 24*, 41-58.
[6] HOWIE,J.M. (1974), "On the domain of tone in Mandarin", *Phonetica 30*, 129-148.
[7] KRATOCHVIL,P. (1985), "Variable norms of tones in Beijing prosody", *CLAO 13 (2)*, 135-174.
[8] ÖHMAN,S. (1967), "Word and sentence intonation: A quantitative model", *KTH STL-QSPR 2-3*, .20-54.
[9] SAGART,L., HALLÉ,P. & de BOYSSON-BARDIES,B. (1987), "Modélisation des évolutions de Fo en Pékinois", *Rap.ATP 955316*, 20 pages.

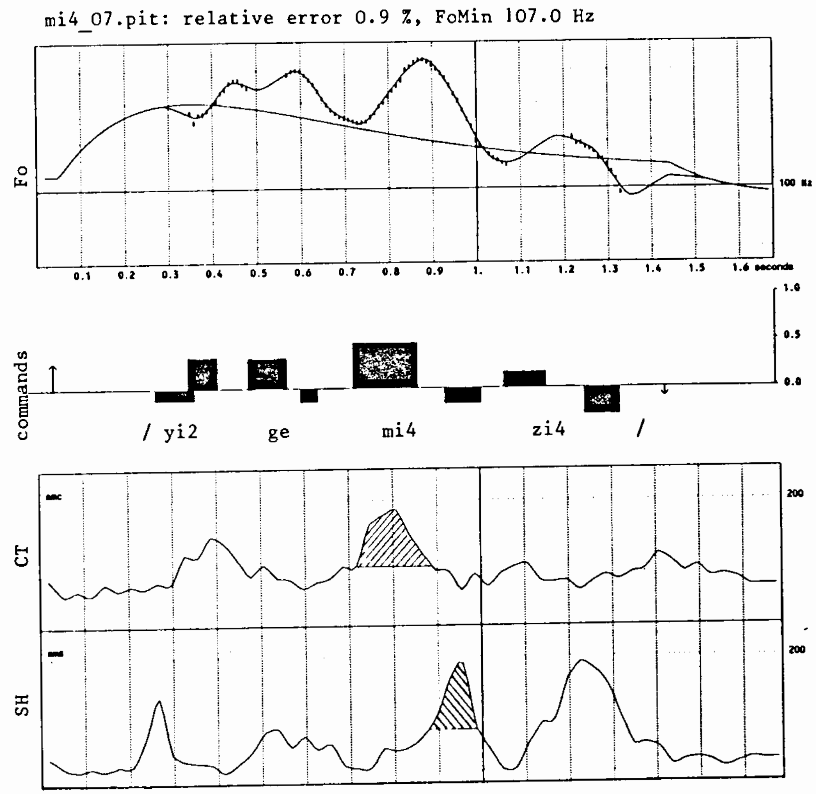

Figure 1. /mi4/: Actual and modelled F_0, model commands, CT and SH activities.

Table 1. EMG versus Commands: differences of timing (EMG-command, ms), ratios of amplitudes (EMG/command, arbitrary unit).

	Fo-raising			Fo-lowering		
	Δonset	Δoffset	ampl. ratio	Δonset	Δoffset	ampl. ratio
tone 1	+16 ms	+70 ms	3.78			
tone 2	+6 ms	+45 ms	3.63	-70 ms	-37 ms	11.2
tone 3				0 ms	-17 ms	12.6
tone 4	+25 ms	+52 ms	4.15	-21 ms	+18 ms	13.4