CHINESE TONES AND THEIR DURATION

M.K. RUMYANTSEV

COLLEGE OF ASIAN AND AFRICAN STUDIES, MOSCOW UNIVERSITY,
USSR

ABSTRACT

In strictly phonological sense the parameter of duration can be regarded as redundant in so far as the system of Chinese tones is concerned. However, when their constitutive function is taken into account duration as well as intensity prove to be indispensable for producing "natural" Chinese speech.

In isolating languages tones play an important role: their primary - phonological or distinctive - function is to distinguish one morpheme from another. But it is not their only function. Various combinations and modifications of tones constitute the prosodic system of an isolating language as a whole. As a result of their interaction different rhythmic structures of words are created, different emotional and evaluative overtones are superimposed on information proper.

The main physical correlate of a tone is its fundamental frequency. Tones differ as to their register, pitch direction (contour - level, rising, falling, falling-rising) and the respective intervals. For a considerable period of time these parameters of tones have been in the focus of attention of experimental phonetics. The other physical correlates of tones - intensity and duration - have not yet received all the attention they deserve.

A considerable amount of data obtained during speech synthesis experiments [1] has convinced us that the parameter of duration cannot be ignored if our aim is "natural" Chinese speech. This feature is redundant only on the distinctive level: morphemes can be distinguished even if their duration is equal, provided, of course, that their registers and contours are different. Thus the simplest opposition is observed in the case of the first and the fourth Chinese tones: a level tone on a high note is opposed to a falling tone which starts quite high and then falls as low as possible. From the point of view of their constitutive function duration and intensity prove to be absolutely indispensable for making tones sound 'natural', really Chinese.

When we stress the importance of duration we do not mean to say that it is only the total duration of each tone that really matters. The parameter in question is highly relevant for determining the inner structure of complex tones. Thus the third (circumflex) tone is not only longer than any other basic tone, but is also characterized by a certain time relationship between its components: its rising part cannot exceed a certain limit, otherwise the listener may mistake it for the second (rising) tone.

The duration of tones as integral prosodic units of syllables has a functional load in words and through the latter in all the other prosodic layers of the language.

Our synthesis of two-syllable words has shown that the role of duration for producing natural (proper) Chinese sound cauls cannot be overestimated. There are 19 models of Chinese words and each model has its own time relationship of the constituting tones. If this or that time relationship is not observed the sound caul becomes unacceptable from the point of view of the language norms.

Under the influence of the higher prosodic levels tones may vary in length as well as in range, pitch direction and intervals. Thus different degrees of prominence - sentence stress, logical stress etc - can affect the duration of tones, but the time relationship typical of the respective words should not be distorted. Otherwise the listener may fail to identify the word as such.

Experimenting with syllables of different duration in speech synthesis enables us to solve at least two problems: 1) finding the optimal rhythmic models in each of the above mentioned 19 groups of words, 2) determining tolerance zones for each of these models.

The rhythmic function of duration is best seen in those models which are represented by combinations of tones of the same type. Thus, for instance, in the model constituted by a sequence of two first tones the second syllable is either longer than the first or at least is of the same length. The normative time relationship is equal to 1.3. The equal length is apparently the threshold because when the second syllable is shorter than the first the model is rejected as false, unnatural. Synthesis has so far not revealed the predominance threshold, beyond which the realisations of words are perceived as exaggerated or unacceptable. In the model of two second tones the second syllable is also longer than the first, but in contrast with the above mentioned model the equal length is not tolerated. At the same time the above limit of exceeding the length of the first syllable can be established somewhere in the region of 1.76. It should be noted, however, that although this realization of the model appears to be fairly acceptable, some auditors characterized the second syllable as unusually long.

The model of two fourth tones gives a completely different picture: the first syllable here is longer than the second. The variation zone is rather wide, ranging from 1.04 to 2.11.

If we try to correlate the so far discussed time relationships with the parameter of interval we shall come to the conclusion that the situation in different models is different. Thus in the model of two second tones we observe a direct proportion (the wider the interval the longer the syllable), whereas in the model of two fourth tones the reverse proportion is true: a wider interval is correlated with a shorter syllable.

As a rule, the duration relationship of tones in isolation is preserved when different tones are used in the modelled words. For instance, the third tone being originally the longest will remain to be so when it is part of this or that word. To what extent it will be longer than the other tones depends, however, on the concrete rhythmic model used. In the model of I+3 the duration of the second syllable (the third tone) exceeds the duration of the first syllable considerably (the proportion is I.67). There can be no question in this case of making the syllables equally long or reversing the proportion. That would be absolutely unacceptable. In the model 2+3 the original duration relationship of the tones is also preserved: the second syllable (the third tone) is longer than the first. To reverse the duration relationship would be out of the question, but the variation zone is rather wide (I.I - I.7). It remains to be seen whether making the syllables equal would be rejected by auditors or not.

If the third tone is used before the first, the second or the fourth it will be longer than the fourth but shorter than the first or the second. In the normative synthesized realization of the word kǎoyā the proportion is I.2 or it can be even increased. The first tone cannot be, however, shorter than the third tone. The synthesized word kǒuyǐn is the case in point, the proportion between the second syllable and the first is I.03.

When the third tone is combined with the fourth, the former should be longer than the latter and the proportion varies from I.4I to I.66.

The fourth tone before the first, the second or the third tone is always shorter than any of them. In the model 4+I the first tone is longer than the fourth and the duration proportion is I.4I. It can be slightly increased, but when in the synthesized word - like unit xìàyī the proportion turned out to be I.93, this realization was rejected by the auditors. The longer duration of the first tone there went beyond the accepted norm. It should be pointed out that as far as the parameters of register and interval are concerned the constituent tones were within the norm and the reaction of the auditors cannot be accounted for by these parameters. The lessening of the duration of the first tone to the point when it becomes equal to the fourth tone or when the duration relationship shifts undermines the rhythmic characteristics of words. For example, in one of the programmes of the word àixī the first tone in the final of the syllable xī was shorter than the fourth tone in the syllable ài, which immediately caused a negative reaction from the auditor.

No less important is the time relationship in various models of the type "basic tone + neutral tone". Even the best programmes in so far as the parameters of register and interval were concerned were often rejected by the auditors because of some wrong duration proportions. The optimal proportion in the model "the first tone + the neutral tone" requires that the neutral tone should be twice as short or even shorter than the first tone. The proportions of I.36-I.38 formed the threshold. The time proportion was markedly improved in the realizations with the duration relationship equal to I.45.

The tolerance zone of the relationship between the second tone and the neutral one is about 2.II. The I.56 proportion was rejected. As far as the relationship between the fourth and the neutral tones is concerned, 2.51 appears to be within the norm. The I.75 proportion was rejected by the auditor, who insisted on the longer fourth tone.

Wrong time proportions interfere with the production of normative rhythmic characteristics in words even if the register and interval relations are correct. Nor can the correct time proportions alone,

without the normative register and interval proportions, ensure good rhythmic characteristics, which are produced by the sum total of features. The close interconnection of duration and interval proportions and their combined effect are borne out by their combined interpretation by the auditors and the undifferentiated perception of the duration, register and interval parameters of words. Some realizations of the words tóufa and xífu were interpreted in precisely this way by the auditors. The rhythmic parameters of the synthesized word tóufa were deemed unsatisfactory. The auditor described the neutral tone in the beginning of the syllable fa as too high (tóu, 139-166, fa, 146-134) and suggested that the tone of the syllable tóu be prolonged, even though the time relationship in this case was quite normative: 2.24 (tóu - 415, fa - 185). The desired prolongation of the second tone in the first syllable was probably automatically associated in the auditor's mind with an increase of the interval, which would really improve the interval proportion between the end of the second tone and the beginning of the neutral tone. Merely to prolong the second tone without increasing its rising interval is not enough to improve the rhythmic parameters of the word. In one of the realizations of the word xífu the rhythmic parameters were also deemed unsatisfactory. The frequency interval between the end of the second tone in the syllable xí and the beginning of the neutral tone proved too small (I.08 against the normative I.23). The modelled time proportions (I.54 against the normative 2.II) were also unsatisfactory. The auditor insisted on lowering the neutral tone and on increasing the interval in the first syllable. The auditor failed to notice certain duration disproportions in the given word sample and sought to improve the rhythmic parameters by correcting only the register and interval proportions: both the lower register of the neutral tone and the increased rising interval of the tone in the syllable xi aim at one and the same thing, i.e., at increasing the interval between the end of the second tone and the beginning of the neutral tone.

Tones act in words as prosodic factors forming morphemes and words. Duration, as one of their constituents, is functionally important in words: time proportions in a word cannot be broken without distorting its prosodic make-up.

The most intimate time mechanisms of tone are manifest in the fine spectra of speech signals, responsible for their different quality. In different tones the finales of Chinese syllables are known to be perceived by ear as slightly differing in quality. In different tones and finales these distinctions are not the same but they are indisputably functionally important to the Chinese ear in the sense of the national specificity of sounds. In any case it is important to determine what spectral parameters account for this specificity. Analysis of natural tones and their synthesis elucidate primarily the role of the frequency and amplitude parameters of the spectrum. For instance, in the natural realizations of syllables in our material a rise of fundamental frequency of the rising tone causes a progressive shift of the first formant. With a male speaker the shift proceeded as follows: at the beginning of the finale í of Tone 2 the first formant was 250 Hz, in the middle it became 300 Hz and at the end it was 350 Hz. With a female speaker the shift was even more pronounced: 350 Hz, 400 Hz and 500 Hz.

Analysis of the synthesized syllables with the finale í recognized by the auditors as "natural", that is undistinguishable from the natural sounding shows that their naturalness is accounted for precisely by this fine correction (correlation) between the fundamental frequency and amplitude values and different formants and by the coordinated function of all the spectral components. The measure and concrete proportions of that coordination are not universal and depend on the linguistic system, their main purpose being to ensure the normative quality of sounding. It is not by chnace that the attainment of this goal also has to do with making the synthesized signal natural or close to it. The impression of naturalness is produced by the absence of monotony (machine-like quality) in the spectrum of the synthesized vowel which, just like it is in natural speech, is not uniform, as far as its quality is concerned, at different segments of the sounding and evolves from the beginning to the middle and the end. For example, F_1, which is coordinated in frequency and amplitude with F_0 in keeping with the rules of the system, is represented by a set of coordinated values within the formant itself and among them rather than by one and the same value throughout the signal. In coordinating the values at every given segment the synthesized signal is in fact ascribed frequency and amplitude "micro-variations". These variations are not universal or determined by the human organs of speech but systemic and linguistic, that is characteristic of a given phonetic norm. In our case we get the syllabic tone with all its inherent systemic characteristics. The latter are determined in the spectral structure not only by the corresponding frequency and amplitude coordination but also by time coordination: frequency and amplitude dynamics of the spectrum unfolds in its portions of time, which correspond to different segments of the sounding tone from its beginning to end.

The role of the coordination of frequency, amplitude and time in the spectrum of Chinese finales of different tones is well illustrated by the synthesized tones which were rejected as unsatisfactory. Tone, as a phonological unit which distinguishes syllabic morpheme, in its model variant is a set of features functioning in unison: if within a certain period fundamental frequency values form an even contour, the envelope of amplitude values forms the same contour. The rise (fall) of fundamental frequency is accompanied by corresponding changes in amplitude values. Inadequate coordination of parameters often results in the inadequate synthesis of syllabic tone. However, the proportion of this correlation may differ, depending on the linguistic system. For example, the auditors rejected the realizations of the sharply falling Chinese (fourth) tone, whose programmes envisaged falling frequency intervals equal to 1.62 and 1.51. The intervals were not only within the norm but the optimal ones in fact. The amplitude values in principle changed in the same direction and, nevertheless, the tones were characterised by the auditors as "passive, inert" and the interval seemed to be inadequate (!). Consequently, the synthesized signals failed to reproduce the amplitude, frequency and time relationships that were worked out by the given linguistic system. The fall in the amplitude values at every given segment of the tone failed to fit the norm prescribed by the fall of the fundamental frequency values or to be synchronized with the time segments of the realization, during which these spectral changes took place.

When separate tone in which the fundamental frequency and amplitude parameters changed in different directions were given to the auditors for identification they were often confused or totally rejected as unacceptable within the given orthoepic norm. This is not to say that lack of coordination is always a defect. On the contrary, it is in many cases a normal phenomenon in connected speech. It is explained by the fact that the parameters coordinated in the units of speech pronounced separately or in the strong position are assigned different roles in connected speech. Thus at the level of syllable fundamental frequency always differentiates lexical meanings in the Chinese language system, i.e., acts as different tones, whereas the amplitude and times values of formants provide for other prosodic distinctions, such as the word's rhythm and intonation contrasts. It is necessary to learn to model this lack of coordination in simulated speech in order to get the needed sounding at every given point of speech continuum. This calls for great efforts on the part of linguists because the measure of this uncoordination, too, is being worked out within the language systems.

1. Chinese tones were synthesized in the laboratory of experimental phonetics at the College of Asian and African Studies (Moscow University) with the help of a formant synthesizer (SPPI-75).