PRP—IL(CALL)/1990,65—79

A TONAL MODEL FOR SYNTHESIZING POLYSYLLABIC WORDS AND PHRASES IN STANDARD CHINESE

Yang, Shun—an

普通话词语合成中的声调模型

杨顺安

[详细提要] 汉语普通话是一种声调语言。所有语音学研究都认为，在普通话的单音节语素的音高模式上，能看到四种有区别性的声调。但是，在连续的语流中又存在着大量的包括连读变调在内的复杂的协同调音效应，使得语流中的每一音节的声学表现与单念时不一样。语流中这种由音高、音长和强度等方面的变化所表现出来的韵律特征(即超音段特征)，对于合成普通话语句的自然度和连贯性关系极大。

另一方面，普通话中的音节又是形音义的结合体，对语流中某一个音节来说，音变大多发生在语音平面，而不在音位平面上，就语义而言，单音节在语流有相对的独立性。正因为如此，许多合成系统，在合成出普通话全部约 1300 个单音节后，就采用音节串接的方案，合成出词、短语和句子，其合成音质，在清晰度或可懂度方面尚可接受，但自然度和连贯性就很差，听起来就有一股子"机器味"或"洋腔洋调"。为了改善合成汉语普通话的音质，就必须建立一套能模拟音节间协同调音效应的协同调音规则和一套能模拟连读变调等的韵律规则。

为了处理普通话多音节词语和语句合成中的声调问题，人们提出了许多方案。(李子殷,1985; Hsieh, et al. ,1989; 张家录,1990; Fujisaki et al. ,1990)。

我们在韵律特性方面进行一系列研究，初步归纳出包括声调协调规则、时长协调规则和幅度协调规则在内的韵律规则。本文主要讨论用于合成普通话多音节词语的声调模型，它能模拟自然语流中因协同发声而产生的变调现象，以提高合成词语的自然度和连贯性。

在研究普通话全部单音节的合成时，我们提出了一种"普通话的归一化字调模型"(杨顺安，1986)，此字调模型(图 1)可用下式表出：

$$Foi(\tau)=\log^{-1}[Fc+Fd * foi(\tau)] \qquad i=1,2,3,4 \qquad (1)$$

这个字调模型共有 3 个输入参数：(1)调基值 Fc，它能体现姓别不同的人的嗓音高低；(2)调域 Fd，它能反映出音节重音不同时所有声调的音高变化范围；(3) 调型码 i。调形函数 foi(τ) 是作为中间参量的，它表征了某一特定声调的音高变化模式，根据实测数据而得。采用这个字调模型合成了普通话的全部单音节，经试听，合成音节的声调很自然。(Yang & Xu,1988)

如果能找出在特定的语境中某个音节的声调特性的变化规律，例如，调型有无改变、调基值和调域如何变化等，那么，我们就可以把描述单音节声调特性的字调模型，扩充成描述多音节词和短语以至语句的声调模型。

根据这种设想，我们在合成双音节词语时，采用过一个这样的声调模型，取得了较满意的合成效果。(杨顺安，1990) 经过修改和完善，一种用于合成多音节词和短语的声调模型的框图如图 2 所示，按此模型，词语中某一音节的基频随时间的变化 Foi(τ) 可用下式来计算：

$$Foi(\tau)=\log^{-1}[Fc+Fd*foi(\tau)+Fb*Lb(\tau)+Fe*Le(\tau)] \qquad (2)$$

采用合成分析法，我们可以得到该模型的输入参数。见图 ‹ 的例子。通过对大量多音节词语的窄带语谱图分析和反复的合成试验，我们认为，普通话中的重音对音节的调域、时长和浊音强度等都有很大的相关性。下面我们将详细讨论重音与调域的关联性，并归纳出一套合成多音节词语的声调规则。

普遍认为，普通话中的词语，其轻重格式不尽相同，重音是影响韵律特征的一重要因素。传统语音学的研究指出，“汉语的重音首先扩大音域和持续时间，其次才是增加强度”(赵元任，1968,p. 23)。换言之，普通话中的重音跟音节的音高、时长和响度密切相关。最近，一些声学语音学的研究也证实了这种相关性（如：林茂灿等,1984；陆致极,1984；颜景助等,1988)。因此，在研究多音节词语的合成中，我们将把语流中各音节的重音等级，当做控制韵律特性的主要参量，根据每一音节的轻重等级，调节这个音节的调域、声韵母时长和浊声源幅度。

一个词语的轻重格式，是在长期使用过程中“约定俗成”的，经过语音学家的归纳，从音系学的角度加以标定。一个词语的轻重格式虽然被定下来了，但在语流中还会变化。我们认为，所有语流中音变，可以分做两个层次来考察：一是音系学层次，二是语音学层次。参考一些语音学的研究和根据我们的实验，音节轻重程度，即重度 Sd 从音位上分为四级是必要的，即：4(重)、3(中)、2(次轻)、1(轻声)。一般而言，重度的数值越大，音节的时长、调域和幅度的数值也越大，而且，重度为 1 的轻声音节在调域和音色上还会有特殊改变。

经实验研究，重度 Sd 和 调域 Fd 的关系可用下式近似描述：

$$Fd = 0.1 + 0.05 * Sd \qquad (6)$$

现在，以合成短语“提琴师的示范表演”为例，来说明如何应用此声调协调规则和声调模型最终生成此短语的基频曲线的全过程 (图 5)：

(a) 单音节处理。给定短语中的每一单音节的调型码和重度。图中每一小框对应于一个音节，框长代表该音节的时长，框高代表该音节的调域，框中横虚线代表音节的调基值框中的纵虚线代表清声母和韵母的分界线(如果有的话)，这里的单音节的重度都被标定 4，在此阶段，因为都作单音节

处理，所以，各音节不相连，各小框互相分开。框中的实线表示该音节的的基频曲线；

(b) 词处理阶段：根据音系学上的约定，为短语中的每一音节分配一个调型码和重度，并进行音系上变调处理。

(c) 短语处理：在由词组成短语的过程中，短语中的各音节的重度会发生语音学层次上的变化。我们认为，这种变化主要来自所谓"位置效应"和"音节数效应"。按式(3－6)重新计算短语中各音节的 Sd 和 Fd.

(d) 优化处理阶段：对每一音节的声调曲线作自然化和光滑化的处理。所谓"自然化"是给每一音节的声调曲线，加上适当的"弯头"和"降尾"，使得合成音质更加自然；而"光滑化"是对那些浊声母或零声母音节而言的，如此例中的"演"，它的基频曲线应与前一音节的光滑相连。

在我们的普通话语音合成系统上，已合成过大量的多音节词和短语，合成音质在自然度和连贯性上，都比用单音节拼接的有较明显提高，有些合成词语几乎听不出什么"机器味儿"，有些几乎达到了"以假乱真"的程度。图 6 是合成的一个多音节词语的语图。

值得一提的是，用这种声调模型可以较合理地解释许多语音学中讨论的连读变调现象。例如：两去声音节的词，按"中重"格式读出时，前音节之所以会变成半去，是因为该音节的调域较小的缘故；同样，三音节词的中间音节常被听成阴平，也是因其调域较小的缘故，它们都属语音学上的变调。

当然，这个用于合成普通话的声调模型及规则还是相当粗糙的，还需要不断地修正和完善。

ABSTRACT

Standard Chinese (SC) is a tonal language. Differences in the tone patterns of a syllable may lead to lexically semantic distinction, while a lot of complicated tone-sandhi arise in running speech. Some synthesis systems have synthesized out all monosyllables in SC, and have made up synthesis of words, phrases or sentences and even a text by directly concatenating the syllables. Intelligibility of this kind of synthesized speech may be accepted, but naturalness and fluency are very poor. In order to improve the quality of synthesized Chinese, it is important to lay down a coarticulation rules which should be simulated coarticulation effects across syllables and a set of prosodic rules which should be simulated tone-sandhi effects, etc.

On the basis of the tone model for synthesizing monosyllables in SC, a new tonal model for synthesizing polysyllabic words and phrases is built up. In this model, there are three model parameters: (1) the tonal base value Fc which controls intonation, (2) the tonal range Fd which is related to the stress degree of the syllables and (3) the tone pattern codes. In synthesizing a polysyllabic word or phrase, it is necessary only to assign the tone pattern codes and the stress degrees of all syllables, the model can produce the Fo-contour as a whole. Listening test shows that the synthesized speech are very close to natural ones in both naturalness and fluency.

Using this tonal model, many complicated tone-sandhi in running speech may also be interpreted reasonably.

INTRODUCTION

Standard Chinese (SC) is a tonal language. All phonological descriptions of SC have agreed on the existence of four distinctive tones which can be observed in the picth contour of monosyllabic morphemes, but, on the other band, it is well known that several types of coarticulation effects in syllable sequences arise in running speech. The coarticulatoin effects involves segmental and prosodic characters of each syllable, so that the acoustic manifestatatin of a syllable in running speech differs more and less from that of the syllable which is pronounced while it is a monosyllable. Prosodic features in running speech determine the naturalness and fluency of synthetic words, phrases and sentences for SC. Thus, it is important to handle well tones and tone-sandhi for developing any speech synthesis system in SC.

A syllable in SC is a unit combining of graphemic, phonemic and semantic features together, and sound changes of a syllable in running speech basically take place at phonetic level rather at phonemic level, monosyllables are relatively stationary in running speech. On account of this, some synthesis systems have synthesized out all monosyllables in SC, and have made up synthesis of words, phrases or sentences and even a text by directly concatenating the monosyllables concerned. Intelligibility of such synthesized speech may be accepted, but naturalness and fluency are very poor. In order to improve the quality of synthesized Chinese, it is important to lay down a coarticulation rule which simulated coarticulation effects across syllables and a set of prosodic rules which simulates tone-sandhi effects, etc.

In order to handle tone matters for synthesizing polysyllabic words andsentences in SC, several strategies been have suggested. Li(1985) used 15 types of Fo patterns for synthesizing disyllabic sequences. On phonemic level, the disyllabic sequences in SC have only 15 ($=4*4-1$) types of Fo pattern indeed, but the Fo patterns of speech sequences may augment with the stress pattern of the sequences, and there are many stress patterns for disyllabic words or sequences, so that only 15 types of Fo pattern are not enough for synthesizing all disyllabic words in SC. If Li's strategy was adopted for synthesizing trisyllabic, quadro-syllabic and more syllabic words, phrases and sentences, innumerable types of Fo patterns must be used. It is evident that thus strategy will get absolutely nowhere. Hsieh et al. (1989) after analyzing pitch contour shapes of speech in their sensentes database, 14 types of representative tone patterns, including 2 types of 1st tone, 2 of 2nd tone, 3 of 3rd tone, 4 of 4th tone and 2 of neutral tone, and 9 tone concatenation rules were summed up. In a text-to-speech system developed by Zhang(1990), Lund's generative intonation model (Gading et al., 1983) was adopted, and the intonation marks are added manually to the model parameters, and the pitch contour of a sentence can be generated by the model. Recently, based on an extension of a model that has already been proved to be applicable to Japanese, Fujisaki el at. (1990) presented a model which could be used in analyzing and synthesizing the fundamental frequecy contour of words and phrases in SC.

To improve quality of synthetic speech, especially words and phrasesin SC, we stutied as a whole the prosodic characters of polysyllabic sequeces, and preliminarily summed up the prosodic rules, including the tonal rules, duration rules and amplitude rules. This paper mainly discusses a tonal model and its tonal rules for synthesizing polysyllabic words and phrases in SC. This model can embody the tone-sandhi caused by coarticulation effects in running speech and can improve the naturalness and fluency of synthetic speech.

THE TONAL MODEL

When developing the synthesis system for all monosyllable in SC, we have presented a normalized tone model (Yang,1986), which is similar in form to Fujisaki's model for Japanese (Fujisaki et al. ,1971). The block gram of the normalized tone model is shown in Fig. 1 and the formula for generating Fo-contour of a syllable is as follows:

$$Foi(\tau)=\log^{-1}[Fc+Fd * foi(\tau)] \qquad i=1,2,3,4 \qquad (1)$$

Where τis the normalized time, $\tau=t/T$, t is the actual time and T is the duration of a syllable. The index i , i=1,2,3,4 , is called as the tone type codes, corresponding to four types of tone respectively. Foi(τ) is Fo-contour for i'th tone. Three input parameters are needed for this model: (1) Fc, the cardinal pitch (or tonal base) represets the neutral picth of the voice; (2) Fd, the tonal range represets the sphere of picth variation for four types of tone; (3) the tone type code i, the tone contour function foi(τ) is a data array. Using this model, we have synthesized all the 1268 monosyllables in SC. Listening test showed that quality of the synthetic syllables were close to that of natural syllables with respect to both intelligibility and naturalness. (Yang & Xu,1988)

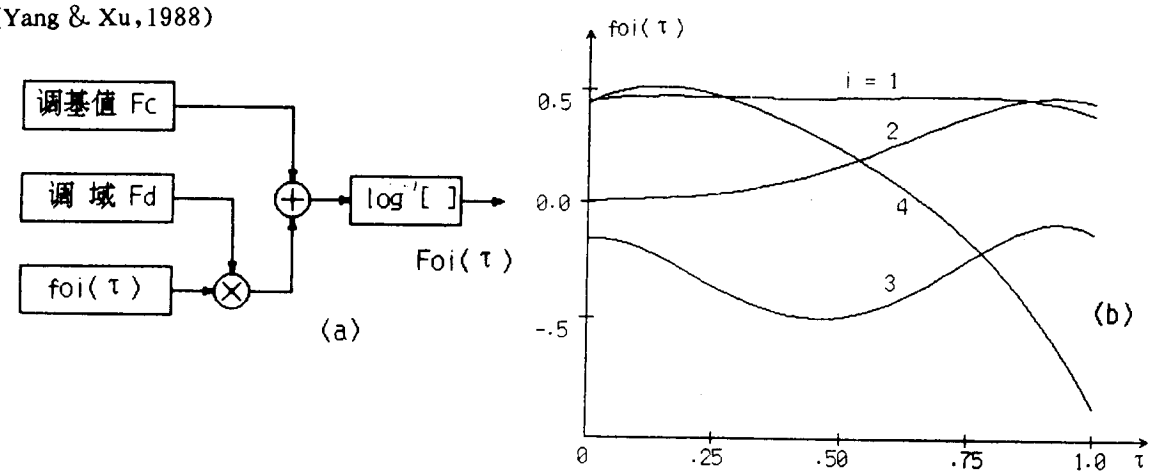


Fig. 1 Block diagram of the tone model for synthesizing monosyllables

图 1 合成单音节用的字调模型框图

If variation laws governing Fo contour for a given syllable in a syllabic sequence, e. g. how do the tone types, Fc and Fd vary, can be found, it is possible to turn the tone model for monosyllable into a tonal model for polysyllabic sequences.

On the bases of this tentative idea, we presented a tonal model by which the batter rusults for synthesizing disyllebic words would be attained (Yang, 1990a). After modifying and perfecting, now the tonal model (Fig. 2) for synthesizing polysyllabic words and phrases in SC is presented. The Fo-contour of a certain syllable in the syllabic sequence is genetated by the following formula :

$$Foi(\tau)=\log^{-1}[Fc+Fd * foi(\tau)+Fb * Lb(\tau)+Fe * Le(\tau)] \qquad (2)$$

where τ is still the normalized time, and the defination of Fc, Fd and Foi(τ) are as same as the ones for the model of monosyllable. The tone type codes i , i=1,2,3,4,5. When i=5, it is the half-3rd tone. The Fb and Fe are so-called former- and latter-linking coefficents respectively, and Lb(τ) and Le(τ) are the former- and latter-linking curves respectively. When syllable begins with a vowel or a voiced consonant, Fb will be not zero, while when a succesive syllable begins with a vowel or a voiced consonant, the Fe will be not zero. The Lb(τ) and Le(τ) are two data arraies.

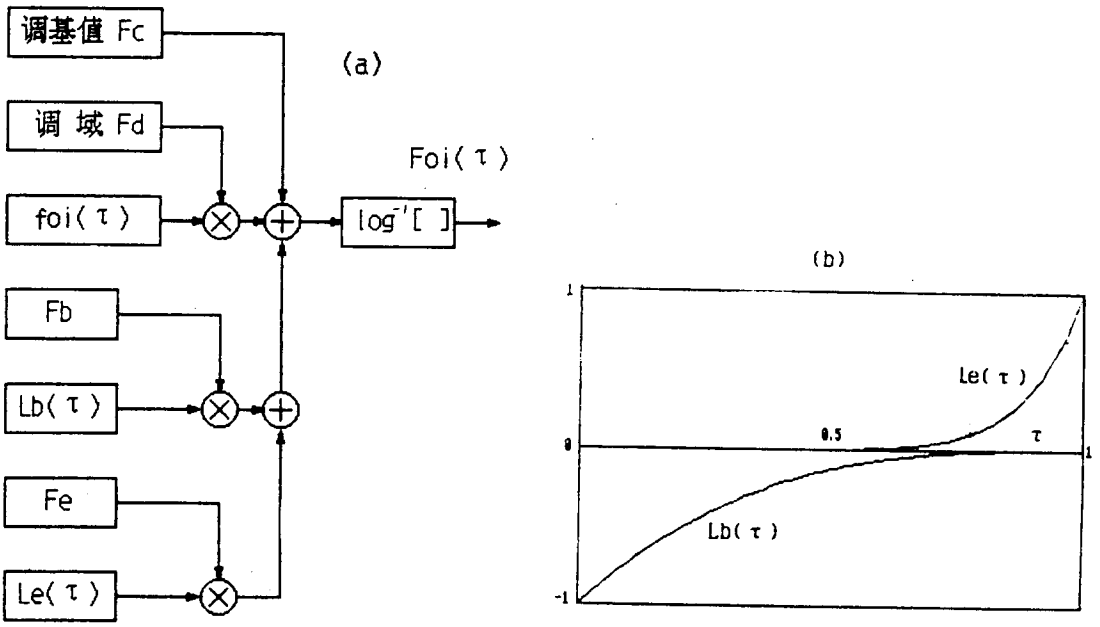


Fig. 2 Block diagram of the tone model for synthesizing polysyllables

图 2 合成多音节词语用的声调模型框图

Using the method of analysis-by-synthesis, the input parameters of the model for a syllabic sequence can be extracted. As an example, Fig. 3a shows the narrow-band spectrogram of the quadrosyllabic word "枪林弹雨" (a hail of bullets). First, from the spectrogram we measured a number of Fo values which were shown by small circles in the figure, then the optimal input paramters for the model would be attained by trail and error. Fig. 3(b) shows a calculated rusults macthed with the measured value shown by more thick lines in the figure. It is seen that the calculated curve of first syllable lies over the measured points, and that implys that Fc is too small. While, the calculated curve at front half part of third syllable closes to the measured points, and they at back half under the measured points. That implys that the Fd of the syllable is less. Fig. 3(c) shows a better macthed rusult, and it is seen that the calculated curve passes through or close to the measured points.

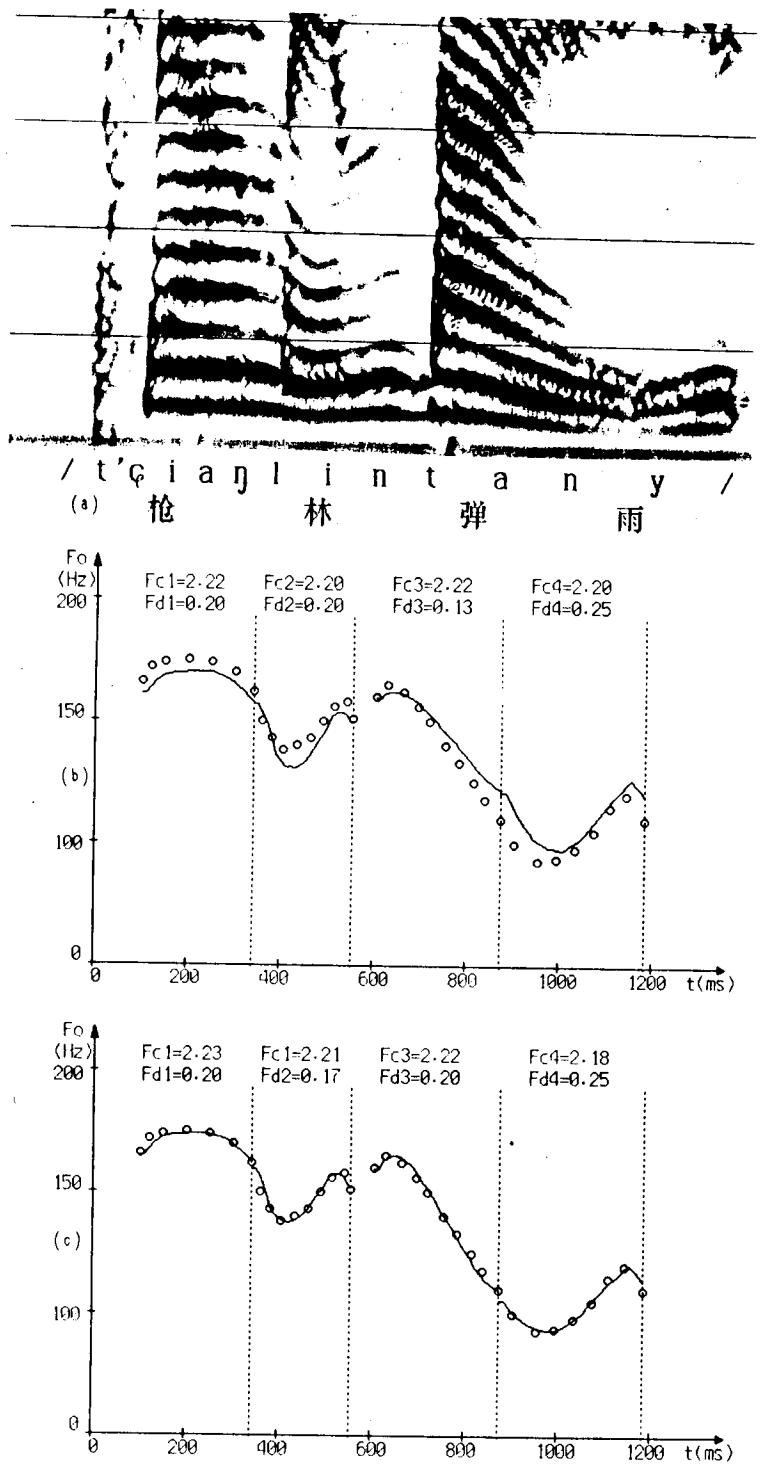


Fig. 3 The example for extracting the model parameter

图 3 抽取模型参数的例子

After applying the method of analysis-by-synthesis to a lot of polysyllabic words and phrases, we find that Fd, one of the most important model parameters, greatly correlates with the stress degree Sd of the syllable, and Fc correlates to the position at which a given syllable lie in a syllabic senquence. Below, we will discuss in grealer detail the correlation between Fd and Sd, and sumup a tonal rules for synthesizing polysyllabic words and phrases.

THE CORRELATION BETWEEN Sd AND Fd

It is commonly considered that the stress patterns by which each syllable in words are pronounced, are different, and it is an important factor influencing prosodic feacture. Zhao pointed out that "Stress in Chinese is primarily an enlargement in pitch range and time duration and only secondarily in loudness." (Zhao, 1968, p. 35) In other words, the stress degree of a syllable is closely correlated with pitch, duration and loudness. Recently, some acoutic-phonetic studies, e. g. Lin, et al. (1984), Lu (1984) and Yan & Lin (1988) also demonstrated such correlation. So that, at studying synthesis of polysyllabic words and phrases, the stress degree of a syllable is considered as one parameter controlling prosodic feacture, and according to the stress degrees of each syllable, their tonal ranges, durations and source amplitudes can be assinged. In this paper, the influence of stress on tonal range is discussed emphatically.

In fact, the stress patterns of words in SC are established afterlong-term usage, and can be marked phonemically through phoneticians' sum-up. For example, there are two utterances: "他分不清东西" (he can not distinguish east from west) and "他去卖东西" (he goes and bay things). The disyllabic word "东西" of the former should be pronouced by middle-heavy stresed pattern, while that of the latter by heavy stressed-neutral tone pattern and confusion in meanning may arise or listened unnatural, if they are pronounced in other ways.

The stress pattern of a certain word pronounced individually may be appointed, but their stress degrees of each syllable in this word would change with speech context. We consider that sound change with speech context can be observed from two level ; phonemic level and phonetic one. As an example, there are two sentences here: "他不辨南北和东西"and "他不辨东西和南北" . The disyllabic words "东西" in the two sentences above are appointed as middle-heave streess pattern on phonemic level, but in the actural utterances, the stress of the word "东西" may sound different, a result of changes caused from phonetic level. In the first sentense, the "东西" lies at sentence's end, and it is basically still pronounced with middle-heavy stressd pattern, but the "东" is pronounced more unstressed. While "东西" in the second sentence lies at sentence's middle, and "西" is pronounced more stressed then "东" .

Till now. No final conclusion has yet been reached on stress patternsof all words in SC. For disyllbic words, most of phonetic studies considered that there were only two stress patterns: middle-heavy and heavy stress-neutral tone. While some other students pointed out that there were still other two stress patterns: heavy-weak and heavy-midlle. Taking some phonetic studies (e. g. Yin, 1982; Xu, 1982; Lu, 1984; Yu, 1988) into consideration and basing on our exepriments, we consider it necessary to classify the stress degrees of syllables as four grades: heavy stress, middle stress, weak stress and neutral tone, and the syllabic stress grade may be marked numerically as 4, 3, 2, 1 respectively. In general, the larger the numeric value of stress degree of a

syllable is, the larger the numeric value of its tonal range, duration and amplitude. As for neutral tone, specfic changes in tonal base and timbre may occured in syllables whose stress degree are marked as1 .

Linguistic studies consider commonly that most of monosyllabic wordsand disyllabic words in SC are basic units using which all words of more than three syllables can be made. For example, "语音学" = "语音" + "学" and "语音合成" = "语音" + "合成" . In our study, the stress degrees of monosyllabic and disyllabic words are appointed as the following:

monosyllabic word : 4, 3, 2, 1 ;

disyllabic word: 3—4, 4—1, 4—2, 4—3 .

The stress pattern of more than three syllabic words can be appointed phonemically, or establshed according to their word-formations. For most of tri- and quadro-syllabic words, their stress patterns can be appointed as 3—2—4 and 3—2—3—4 respectively. For many compound words with affix, their stress patterns should be establshed separately.

In running speech, a sentence ,especially a longer one, is pronounced by one phrases after another, and a phrase is composed of two or more than two words. In the course of composing words into phrase, the stress degrees of each syllable will change more or less on phonetic level. It is considered that such a change is caused from so-called "position effect" and "syllabic number effect" . In view of the two effects, the stress degree Sd' of the syllable in a phrase should be computed through the following formula:

$$Sd' = Sd - Dp - Dn \tag{3}$$

Where Sd is the stress degree of the sylable in a word; Dp and Dn are so-called "position decrement" and " syllabic number decrement" respectively, and may be obtained with experimental formulae.

The so-caled "position effect" is that the stress degree of a syllable which lies at the word boundary in a phrase may decrease. Accoding to our measurements, the Dp at differant positions are as follows:

$$\left.\begin{array}{ll} Dp=0 & \text{at phrase beginning (PB)} \\ Dp=(Sd-1)/3 & \text{at word beginning after word boundary (WB)} \\ Dp=0 & \text{at word middle (WM)} \\ Dp=2(Sd-1)/3 & \text{at word end before word boundary (WE)} \\ Dp=0 & \text{at phrase end (PE)} \end{array}\right\} \tag{4}$$

The so—called "syllabic number effect" is that the stress degrees of syllables may decrease with increase of syllabic number in a phrase. By try and error, Dn can be calcutelated with the following formula:

$$Dn=0.1+0.4\{1-\exp[-0.23(N-1)]\} \tag{5}$$

Where N is the syllabic number in the phrase.

As an example, in phrase "提琴师的示范表演" (a violin player's example domonstration) composed of eight (N=8) sylables, the Dp,Dn and Sd' of each syllable have been calculated by using the formulea above, and are listed in Table 1 .

Table 1 The Sd and Fd of an example

syllable	提	琴	师	的	示	范	表	演
position	PB	WM	WE	WM	WB	WE	WB	PE
Sd	3	2	4	1	3	4	3	4
Dp	0	0	2	0	0.67	2	0.67 0	
Dn	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42
Sd'	2.58	1.58	1.58	0.58	1.91	1.58	1.91	3.58
Fd	0.229	0.179	0.179	0.129	0.196	0.179	0.196	0.279

Now, we can more reasonably appoint the stress degrees for a word, and recalculate the stress degr each syllable in the phrase. The acoustical relation of stress degree have been studied preliminarily, and t lationship between the tonal range Fd and the stress degree Sd is described roughly by the following form

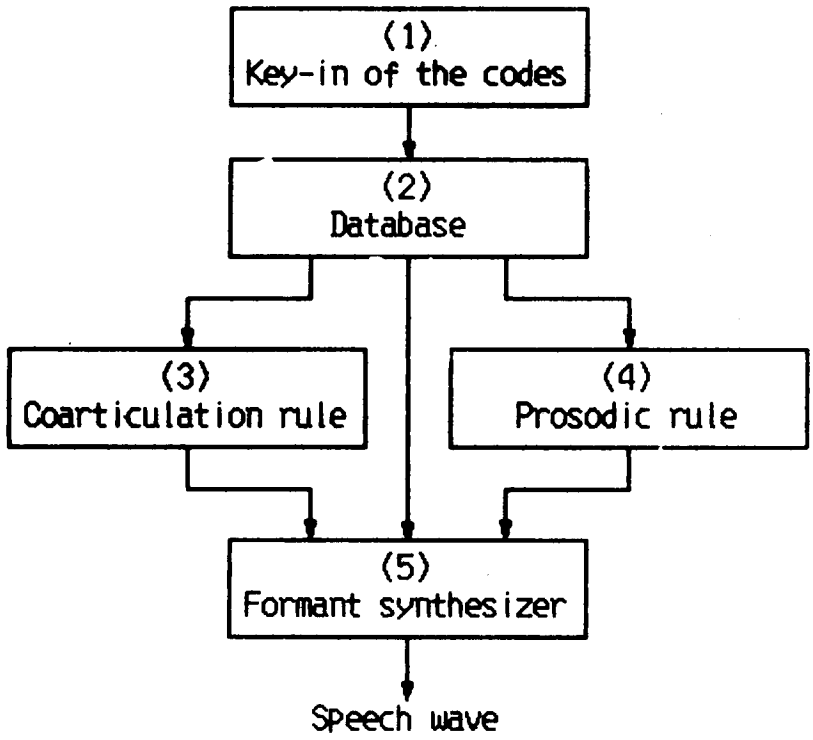
$$Fd = 0.1 + 0.05 * Sd \quad (6)$$

As an example, the Fd of each syllable for the phrase "提琴师的示范表演" are listed in the last r Table 1.

APPLICATION OF THE TONAL MODEL AND ITS TONAL RULES

In our speech synthesis system, the synthetic parameters data base with units of the initials (includi loinitials) and the finals has been established. To synthesize a polysyllabic word or a phrase, it is only r sary to input a series of the initial codes, the final codes, the tonal codes and the stress degrees which are r to the syllables in the word or phrase. Fig. 4 shows a block diagram for synthesizing polysyllabic wor phrase. The coarticulation rules shown by block (3) are used to modificate and calculate the segment pa ters (e.g. the formant frequecies of syllabic end), so that coarticulation effect in timbres between the ad sylables can be simulated (Yang, 1990b). The prosodic rules shown by the block (4) include the tonal r duration rules and amplitude rules.

Below, we will illustrate with an actual example how to generate ultimately a fundamental freq curve of a gaven phrase by means of the tonal model and rules mentioned above. A phrase to be synth still the "提琴师的示范表演". Fig. 5 shows the whole procedure generating the Fo-curve of the phras



'ig. 4 Block diagram for synthesezing polysyllbic words and phrases

图 4 合成多音节词语的流程图

(a) Monosyllable processing stage: to appoint a certain tone type and the stress degree for each syllable in he phrase. In the figure, each small block corresponds to a syllable, and the width and height of the block repesent the duration and the tonal range of the syllable, respectivily. The horizontal dotted line and vertical one, f any, represent the tonal bese Fc and the boundary between un-voiced initial and voiced final of the syllable espectively. In this stage, the stress degree of each syllable was appointed as 4, and the every block was to be eparated. The real lines at the blocks represent the syllabic Fo-pattern.

(b) Word processing stage: to appoint the stress degree Sd of each word in accordance with its phonological appointment and to process the phonological tone-sandhi.

In the example just cited, one trisyllabic word "提琴师", one monosyllabic auxiliary "的", and two disyllabic words "示范" and "表演" were included in the phrase. The stress degree Sd of the each syllable in the our words were listed in the Table 1 now.

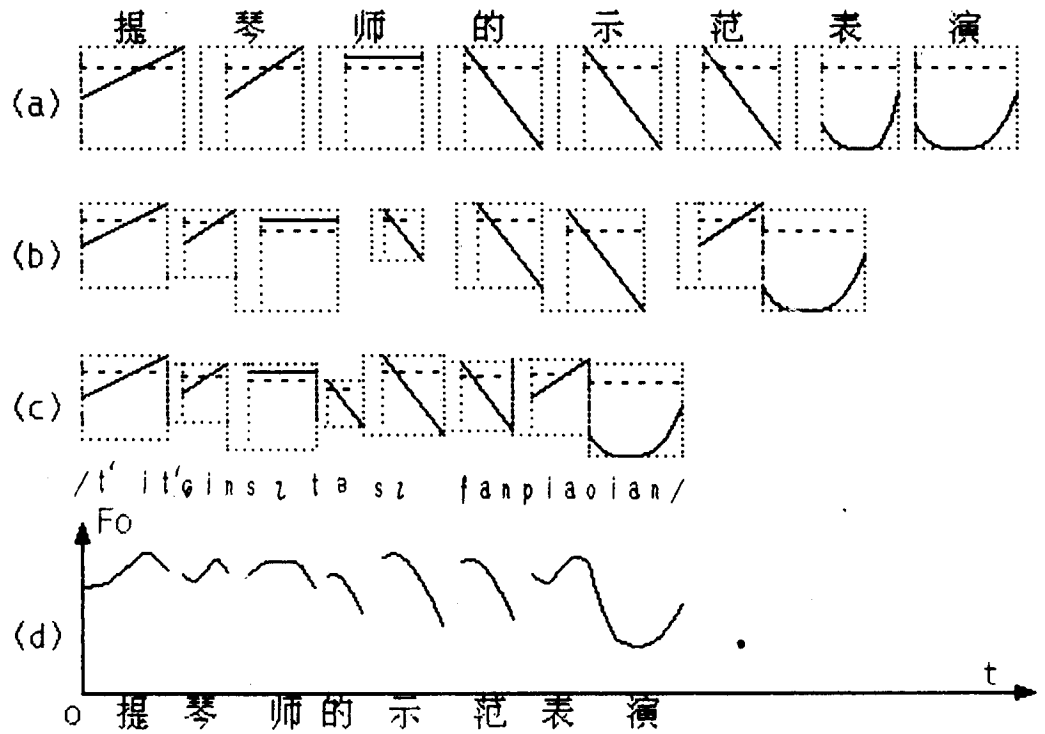


Fig. 5 An example generating the Fo-curve of the phrase

图 5 用声调模型最终生成一个短语的基频曲线的过程示意图

In this stage, there are the phonological tone-sandhi rules as thefollowing :

(1) 3rd tone + 3rd tone ⟶ 2nd tone + 3rd tone;

(2) 3rd tone + non−3rd tone ⟶ half−3rd tone + non-3rd tone;

(3) "一, 七, 八, 不" tone-sandi rules (see,Xu,1980,pp. 124−127);

(4) the parameters of the tonal model for a neutral tone syllableshould be procesed by'the neutral tone rules' (Yang,1991).

In this example, "表演" is composed of "表" and "演" with 3nd tone, the tone type of "表" should be modified as 2nd tone. In Fig. 5(b), the blocks are different in width and height (i. e. the duration and tonal range Fd), because the stress degree of each syllable has been modificated by formule (6). In addition , Fc of the syllables in the same words should also fall successively.

(c) Phrase processing stage: to recalculate Sd and Fd of each syllable in the phrase by formulea (3) - (6).

In this example, the phrase is composed of words: "提琴师", "的" , "示范" and "表演" , and the resulting Sd' and Fd are already listed in the Table 1. It is seen from Fig. 5(c) that the Fd of the syllables: "师", "范" and "表" which lie at word boundaries in the phrase, decrease obviously.

(d) Optimization precessing stage: to do the so-called imitating and the so-called smoothing to the pitch contour of each syllable.

The so-called imitating means that the onset and offset curving sections should be added in the pitch contour of each syllable. While the so-called smoothing means that the Fo curve of a syllable whose initial was voiced should link with the Fo curve of the preceding syllable. Through imitating and smoothing, naturalness and fluency of the synthetic sequences can be improved.

Using our speech synthesis system, quite a number polysyllabic words and phrases have been synthesized, and their quality in naturalness and fluency are better than that of speech made up of concatenation of monosyllables. As an example, Fig. 6 show the spectragram of the synthesized phrase "提琴师的示范表演".

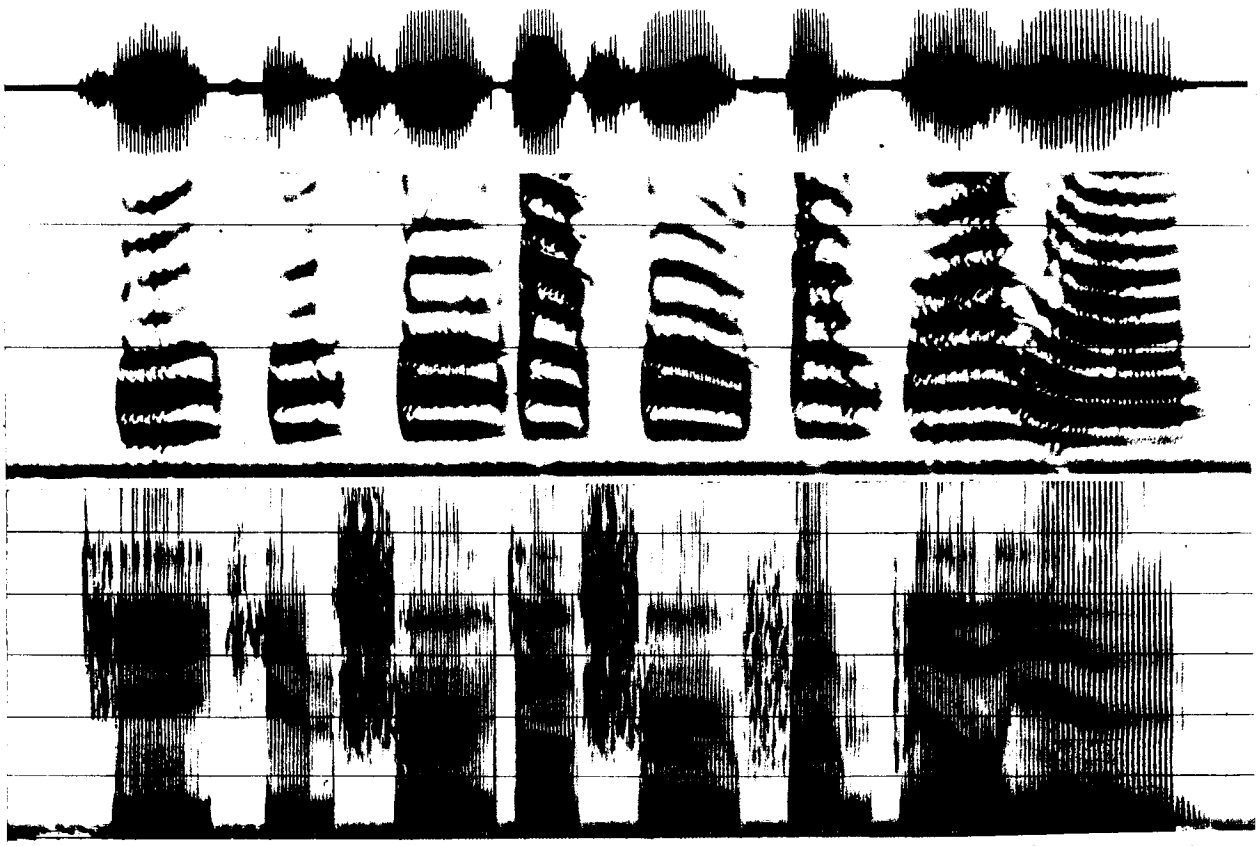


Fig. 6 the spectragram of the synthesized phrase

图 6 一个合成的短语的语图

It is worth mentioning that many tone-sandhi which were described over and over again by phonetics can be more reasonably explained by our tonal model and its tonal rules. For example, phonetics has pointed out

that " A 4th Tone followed by another 4th does not fall to bottom; that is , 51: + 51: ⟶ 53: 51:" (Zhao,1968,p. 28). According to our view, such tone-sandhi is one which takes place on phonetic level. As such disyllabic words are pronounced with 'middle-heavy stress pattern', the stress degree Sd of the former syllable is less than that of the latter syllable and the tonal range Fd of the former syllable is less than that of the latter syllable, while the tonal base value Fc of the former syllbale is larger then that of the latter syllable, so that the tone value of the former syllable is 53 instead of 51 (see the word "示范" in the Fig. 5(b)). As another example, phonetics has pointed out that, " If in a three-syllable word or phrase ABC, A is in the 1st or 2nd Tone, B in the 2nd Tone, and C in any except the neutral tone, then changes into 1st Tone for speech at conversational speed, but does not change at a deliberate speed, " (Zhao,1968,p. 27) According to our view, such trisyllabic words are pronouced with'midlle-weak-heavy stress pattern', and the stress degree Sd and the tonal range Fd of the 2nd syllable is smallst among the three syllables, so a rising of its Fo-contour is not obvious , and it may be heard as 1st Tone. In our example shown in Fig. 5(b), "提琴师" belongs to such tone-sandhi on phonetic level.

Of course, the tonal model and its tonal rules for synthesizing speech in SC are still tentative, and it will be necessary to modify and consummate them further.

REFERENCES

Garding,E , Zhang,J. & Svantesson,J. (1983), A generative model for toneand intonation in Standard Chinese, Working Papers 25, Linguistics-Phonetics, Lund University, pp. 53—65.

Fujisaki,H. & Sudo,H. (1971), A model for the generation of fundamental frequncy contours of Japanese word accent, J. Acoust. Soc. Japan, 27, 445—453. (in Japanese)

Fujisaki,H. et al. (1990), Analysis and modeling of tonal features in polysyllabic words and sentences of the Standard Chinese, Proc.
ICSLP'90, Vol. 2, pp. 841—844, Nov. 18—22,1990, Kobe, Japan.

Hsieh,C. et al. (1989), An improved Chinese text-to-speech system based on formant synthesis techniques, Computer Porc. Chinese & Oriental Lang. , Vol. 4,No. 2,153—170.

Lin , Maocan et al. (1984), The stress pattern and its acoustic corelates in Mandarin, Fanyan, No. 3, pp. 57—63. (in Chinese)

Wu, Z. (1982), Tone-sandhis in sentences in Standard Chinese, Zhongguo Yuwen, No. 6, pp. 439—450. (in Chinese)

Xu, Shirong(1980), Elements of Phonetics for Stardard Chinese, (in Chinese).

Xu, Shirong(1982), Volume analysis of disyllabic word, Language Teaching and Linguistic Studies, No. 2, pp. 4—19. (in Chinese)

Yan , J. & Lin, M. (1988), The acoustic manifestation for stress in trisyllable groups in Standard Chinese, Fanyan, No. 4, pp. 227—237. (in Chinese)

Yang, S. (1986), The effect of the dynamic charateristics of voiced source upon the quality of synthesized speech, Zhongguo Yuwen,pp. 173—187. (in Chinese)

Yang,S. & Xu,Y. (1988), An Acoustic-phonetic oriented system for synthesizing Chinese, Speech Communication, Vol. 7, pp. 317—325.

Yang, Shun-an (1990a), On tone-sandhi in synthesizing disyllabic words in Standard Chinese, Yuyan Yanjiu, No. 2, (in Chinese).

Yang, Shun-an (1990b), Articulation rules for synthesis of polysyllabic words and phrases in Standard Chinese, Proc. ICSP'90, Vol. 1, pp. 397—400, Oct. 22—26, Beijing, China.

Yang, Shun-an (1991), Synthesis by rule for the neutral-tone syllables in Standard Chinese, Applied Acoustics, Vol. 10, No. 1, pp. 12—18(in Chinese).

Yin , Zuoyan (1982), A preliminary study of accents and atonics in disyllabic words in common use, Zhongguo Yuwen, No. 3, pp. 168—173. (in chinese)

Yu, Min (1988), Words of the "看不见、找不着" in the Beijing dialect, Fangyan (Dialect), No. 4, pp. 247—256. (in Chinese)

Zhang, Jialu (1990), Acoustic parameters and phonological rules of a text-to-speech system for Chinese, Acta Acustica, Vol. 15,No. 2, pp. 113—120. (in Chinese)

Zhao, Yuanren(1968), A Grammar of Spoken Chinese, University of California, Press Ltd.