THE PERCEPTUAL CUES OF TONES IN STANDARD CHINESE

MAO-CAN LIN

Institute of Linguistics
Chinese Academy of Social Sciences, Beijing, China

ABSTRACT

The synthesized speech of /shi/ /tuo/ and /ai/ were utilized to investigate the perceptual cues for tones .

The result of this experiment indicated that the four tones can be generated alone by Fo pattern with the possibility of about 95% , whereas the four tones can not be distinguished by amplitude contours alone . It also showed that the effect of duration on the naturalness of tone-3 and tone-4 is greater than that on the rate of identification of tone-3 and tone-2

INTRODUCTION

In 1924 , Liu Fu discovered the important role of Fo in Chinese tone (1). It was found that the Fo curve in syllable not only has a " tone-section " , but also generally has a " onset-curving section " and " end-falling section " (2) . Chuang et al. made the Fo analysis and identification test for colloquial Standard Chinese (3) .

Howie demonstrated the primacy of Fo pattern in the identification of the four tones (4) . Wang talked about the role of Fo and amplitude in the four tones (5) . Lin and Wang discovered that the judgement of tone category of the first syllable in bisyllabic word is often influenced by pitch of the second syllable and duration of the first one (6) .

This experiment tried to investigate the role of Fo , amplitude and duration in the four tones by varying these parameters in the synthesized speech .

THE PHYSICAL MANIFESTATION OF TONES

We made an acoustical analysis of 138 monosyllables consisted of 38 different Initial and Final Combinations with tones spoken by two speakers (m and f) .

Fig. 1(m) and 1(f) were the Fo pattern of the two speakers . It can be seen from fig. 1 that each tone generally has its own peculiar Fo pattern .

Although the durations of the four tones did not show a regular relative relation , comparatively speaking , the duration of tone-3 were in most of the cases the longest .

Four different types of amplitude contour could roughly be drawn from the amplitude curves in 276 monosyllables, namely: mid-hump , back-hump , two-hump and front-hump . It can be seen that the amplitude contours in tone-3 spoken by m were all two-hump , but those spoken by f were two-hump only in 60% of the cases .

The peak of intensity in tone-3 showed in most of the cases the lowest .

PERCEPTUAL EXPERIMENT OF TONES

The syllables of /shi/ , /tuo/ and /ai/ were synthesized by a synthetic system (7) under five conditions shown in the left column of tables given below . All the speech sounds were randomized to make it impossible for the 14 subjects (as listeners) to predicate under which condition the speech sound were synthesized while he or she heard it . The average rate of identification of tone by subjects (14) was displayed in percentage in the right column of each table . The figures in parentheses in the tables represented the percentage of the speech sounds in good timbre judged by the subjects .

The data of the parameters in condition one roughly corresponded to the physical manifeslation of tones . A sonagram of /tuo/ synthesized by condition one was displayed in fig. 2 . Table 1 showed that the rate of correct identification of tone was 98.8% , and the speech sounds in good timbre amounted to 70.7% .

In condition two , the amplitude contours were only varied , e.g. , the amplitude contour of low-falling-rising of Fo varied to mid-hump from two-hump , but other parameters were the same as those in condition one . A sonagram of /tuo/ in this condition was displayed in fig. 3 . Table 2 showed that the rate of correct identification of tone was 97.6% , and the speech sounds in good timbre amounted to 67.1% .

In condition three , Fo patterns were all mid-level , and durations were the same as those in condition one , but the amplitude contours had the four different types of mid-hump , back-hump , two-hump and front-hump . A sonagram of /tuo/ in this condition was displayedin fig. 4. Table 3 showed that the subjects (14) identified the speech sounds as tone-1 about 90% . No one identified them as other tones , namely , no one identified the speech sounds with amplitude contours of two-hump or front-hump and with mid-level of Fo as tone-3 or tone-4 . This result indicated that the four tones can not be distinguished by amplitude contours alone .

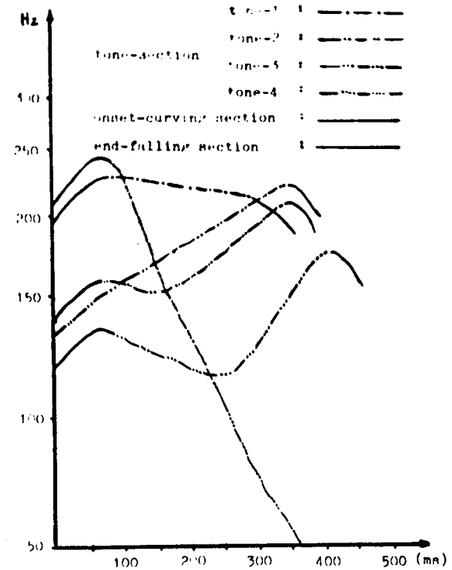

Fig. 1(m) average fundamental frequency curves in mono-syllables for Beijing speaker m

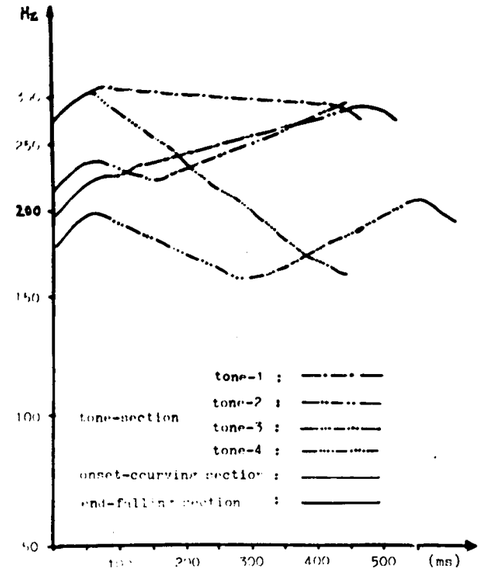

Fig. 1(f) average fundamental frequency curves in mono-sullables for Beijing speaker f

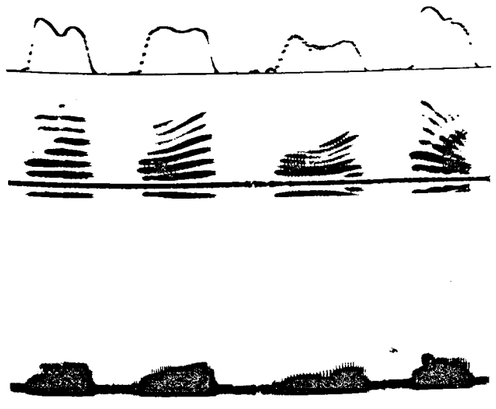

Fig. 2 sonagrams of /tuo/ synthesized in accordance with condition 1

TABLE 1

Condition one			The rate of identificationof tones			
			tone-1	tone-2	tone-3	tone-4
1.1	F_0:	high-level				
	Amp.:	front-humo back-hump	97.6 (75.8)			
	T:	348ms				
1.2	F_0:	mid-rising				
	Amp.:	back-humo front-humo		100 (76.2)		
	T:	390ms				
1.3	F_0:	low-falling-rising				
	Amp.:	two-humo	2.4 (2.4)		97.6 (61.9)	
	T:	470ms				
1.4	F_0:	high-falling				
	Amp.:	front-hump mid-hump				100 (69.0)
	T:	307ms				

While the Fo patterns and the amplitude contours in condition four and five remained the same as those in condition one , the durations in condition four and five were different from those in condition one . In condition four , the durations of four different sounds were regulated as the same as those of tone-4 in condition one ; In condition five , the durations of four different sounds were done as same as those of tone-3 in condition one . The speech sounds synthesized by condition four were correctly identified as tone-1 , tone-2 and tone-4 95.8% in average , but as tone--3 90.5% , namely , the rate of correct identification of tone-3 decreased about 7% compared with that in condition one . This time , the number of speech sounds with tone-3 judged to be in good timbre decreased 22% from those in condition one.

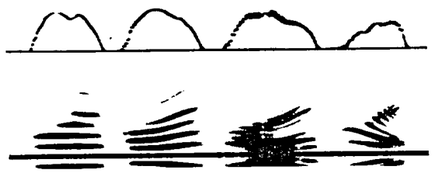

Fig. 3 sonagrams of /tuo/ synthesized in accordance with condition 2

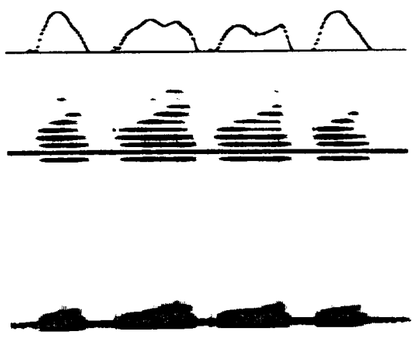

Fig. 4 sonagrams of /tuo/ synthesized in accordance with condition 3

And the speech sounds in condition five were correctly identified as tone-1 , tone--3 and tone-4 95.8% , but as tone-2 88.1% , namely , the rate of correct identification of tone-2 decreased 12% compared with those in condition one . This time , the number of speech sounds with tone-4 judged to be in good timbre decreased 19% from those in condition one . These two results indicated that the effect of duration on the naturalness of tone-3 and tone-4 was greater than that on the rate of identification of tone--3 and tone-2 .

CONCLUSION

We may conclude that the four tones can be generated by Fo pattern alone with the possibility of about 95% ; The effect of duration on the naturalness of tone-3 and

TABLE 2

Condition two			The rate of identification of tones (%)			
			tone-1	tone-2	tone-3	tone-4
2.1	F_0:	high-level	100 (92.7)			
	Amp.:	mid-hump two-hump				
	T:	382ms				
2.2	F_0:	mid-rising		97.6 (64.3)		
	Amp.:	front-hump				
	T:	412ms				
2.3	F_0:	low-falling-rising			97.6 (57.1)	
	Amp.:	mid-hump				
	T:	468ms				
2.4	F_0:	high-falling				
	Amp.:	back-hump				95.2 (54.7)
	T:	370ms				

TABLE 3

Condition three			The rate of identification of tones (%)			
			tone-1	tone-2	tone-3	tone-4
3.1	F_0:	mid-level	92.2 (38.1)			
	Amo.:	mid-hump				
	T:	332ms				
3.2	F_0:	mid-level		85.7 (42.7)		
	Amo.:	back-hump two-hump				
	T:	413ms				
3.3	F_0:	mid-level			85.7 (26.2)	
	Amo.:	two-hump				
	T:	443ms				
3.4	F_0:	mid-level				92.9 (57.1)
	Amp.:	front-hump				
	T:	310ms				

TABLE 4

Condition four			The rate of identification of tones (%)			
			tone-1	tone-2	tone-3	tone-4
4.1	F_0:	high-level	100 (78.5)			
	Amp.:	front-hump back-hump two-hump				
	T:	310ms				
4.2	F_0:	mid-rising		95.2 (66.6)		
	Amo.:	back-humn front-hump				
	T:	310ms				
4.3	F_0:	low-falling-rising		4.7 (2.4)	90.5 (39.7)	
	Amp.:	two-hump				
	T:	310ms				
4.4	F_0:	high-falling				97.6 (73.8)
	Amp.:	front-hump mid-hump				
	T:	310ms				

TABLE 5

Conditin five			The rate of identification of tones (%)			
			tone-1	tone-2	tone-3	tone-4
5.1	F_0:	high-level	100 (83.3)			
	Amp.:	front-hump back-hump two-hump				
	T:	407ms				
5.2	F_0:	mid-rising	7.1 (7.1)	88.1 (66.7)		
	Amp.:	front-hump back-hump				
	T:	467ms				
5.3	F_0:	low-falling-rising			95.2 (69.0)	
	Amp.:	two-hump				
	T:	467ms				
5.4	F_0:	high-falling				100. (50.0)
	Amp.:	front-hump mid-hump				
	T:	467ms				

tone-4 is greater than that on the rate of identification of tone-3 and tone-2 ; The four tones can not be distinguished by amplitude contour alone .

REFERENCES

(1) Liu Fu , An experimental record of Chinese tone , 1924 , Shanghai Qunyi book--store .

(2) Lin Mao-can , The pitch indicator and the pitch characteristics of tones in Standard Chinese , Acta Acustica , Vol. 2 No. 1 , 1965 .

(3) Chuang , c.k. , Hiki , s. , Sone , T. & Nimura , T. , The acoustical features and perceptual cues of the four tones of colloquial Standard Chinese , In Proceedings of 7th Inter . Cong. of Acoustics , Vol. 3 1972 .

(4) Howie , j.m. , Some experiment on the perception of Mandarin tones , In proceedings of the 7th Inter. Cong. of Phonetic Sciences , eds. by A. Rigault & R. Charbommeau , 1972 . The Hague : Mouton . PP. 900-904 .

(5) Wang , William S-Y. , Lectures on experimental Phonetics , in Collective Work of Linguistics , No. 14 , ed. by Beijing University , 1983 . Commercial Press .

(6) Lin Tao and Wang , William S-Y., The Problem of tonal perception , The Journal of Linguistic Society of China , No. 2 , 1984 .

(7) Yang , Shuen-an , The effect of the dynamic characterictis of voiced source upon the quality of systhesized speech , ZHŌNGGÓ YŬWÉN , No. 3 , 1986 .