ON THE SYLLABLE STRUCTURES OF CHINESE RELATING TO SPEECH RECOGNITION*

Jialu Zhang

Institute of Acoustics, Academia Sinica
Beijing P.O.Box 2712, 100080, China

ABSTRACT

It is well known that Chinese is a tone language with multi-tone system, but the distinctive syllable structures relating to speech recognition have not brought to phoneticians' attention yet. The syllable structures, the phonotactic rules were discussed and the joint probability of the initials and the finals were given in this paper. A comparative study of the relative information transmitted by the place channel between Chinese and English shows that the syllable structrues of Chinese are advantage to increasing the place recognition rate and the syllable intelligibility. It was shown that perceiving a phoneme is based on a syllable in which it exits.

1. INTRODUCTION

It is well known that Chinese is a tone language with multi-tone system and the Chinese character is a kind of ideogram. Every character in written Chinese is a syllable and a morpheme too. But a syllable in spoken Chinese can be corresponding to two characters, such as "*mianr*"[miar], although it occurs not so frequently. And the amount of syllables used in real speech is only about 1200 syllables with different lexical tones. So that the syllables take a very important place in designing both speech dictation systems and text-to-speech systems for Chinese.

Two thousands years ago, during Han dynasty, traditional Chinese phonologists had understood that a syllable can be divided into two parts - the initial and the final, and they used two commonly used characters (syllables) to denote the pronunciation of a new character using the initial consonant of the first syllable and the final part of the second syllable. This kind of sound notation is called "*Qieyun*". In Qing dynasty(1616-1911) the phoneticians divided a syllable into four parts called the head, the neck, the belly and the tail. After doing some experimentally phonetic studies on tonal patterns another part called spirit--lexical tone was added [1]. That is really the same with morden phonetic system.

In fact the traditional phonological system of Chinese--the initials, the finals and the tones is still being used in Chinese phonetic transcription and language teaching, because it is easy to learn and more suitable to Chinese syllable structures.

Based on some psychophysical experiments and analysis methods of information theory, the Chinese syllable structures relating to speech recognition were studied in this paper.

2. SYLLABLE STRUCTURES

From point of view of phonetics, there are four basic syllable structures: v, v-c, c-v , and c-v-c in Chinese as in other stress languages. However, some significant differences in syllable structures exist in between Chinese and European languages. Some special rules are described as follows.

A. There is no consonant clusters in spoken Chinese. Only single consonant appearing at the initial or/and final position of a syllable.

B. In Standard Chinese (*putonghua*) just nasals /n/** and /ng/ can appear at the final position of syllables, but in addition to the nasals the plosives [p], [t], [k] are possible in Guangdong and Fuzhou dealects.

C. Generally no more than four speech sounds exist in a Chinese syllable, and no more than three phones in the final, so that only single vowel and diphthong can be with nasal coda.

D. The back nasal /ng/ is never used as initial.

E. Each syllable has certain lexical tones, in standard Chinese they are five: level, rising, dipping, falling and atonic(light, it occurs only at word level).

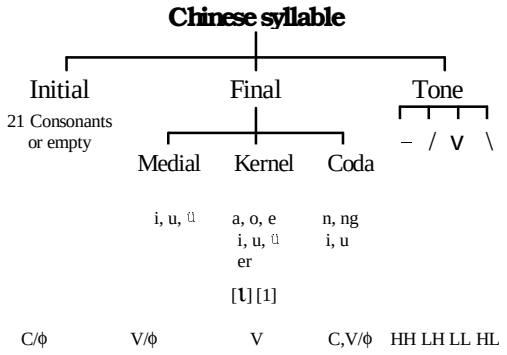

Fig.1. The syllable structures of standard Chinese.

* This work was supported by the key project of NSF of China.

** Note: The symbol in brackets is IPA, and that in slants the Chinese phonetic transcription.

The syllable structures of standard Chinese are illustrated in Fig.1.

Some other distinctions of spoken Chinese are deduced from the syllable structures. Firstly the syllable structure is concise and the boundary of the syllable is clear. The lexical tone as a ribbon ties the syllable into a boundle, while the consonant or the semi-vowel is a symbol of the beginning of a syllable. Second the vowels have the advantage of the consonants in spoken Chinese, the occurrence frequency of vowels is 54.2%, and that of English 38%, Russian 43.3%. All of these are beneficial to speech recognition and speech synthesis for Chinese at syllable level.

3. PHONOTACTIC RULES

Some strict phonotactic rules which control a initial in combination with a final to form a syllable exist in Chinese. In general these phonotactic rules mainly manifest the relation between the place of articulation of the initial and the quality of the medial vowel in a syllable. According to the traditional phonology the finals of Chinese can be divided into four classes: 1. The finals with no medial vowel called *Kaikou;* including /a, o, e, er, ai, ei, ao, ou, an, en, ang, eng/and vowels [ɿ] and[1]; 2. The finals with medial [i] called *Qichi* including /i, ia, ie, iao, iou, ian, in, iang, ing/; 3. The finals with medial [u] called *Hekou* including /u, ua, uo, uai, uei, uan, uen, uang, ong/; 4. The finals with medial [y] called *Cuokou* including /ü, üe, üan, ün, iong/.

Table 1. The occurrence frequency of the initials and the joint probability of the initials in combination with the finals, %.

Initials	Finals			
	Kaikou	*Qichi*	*Hekou*	*Cuokou*
b, p, m 5.15, 0.98, 3,74	47.98	33.33	18.68	0
f 2.45	84.62	0	15.38	0
d, t 12, 3.53	59.04	20.87	20.09	0
n, l 2.53, 5.69	46.38	41.58	10.17	2.03
z, c, s 3.01, 1.15, 1.08	54.81	0	45.19	0
zh, ch, sh, r 7.18,2.75,7.66,1.94	75.13	0	24.87	0
j, q, x 6.98, 3.11, 4.86	0	78.73	0	21.27
g, k, h 5.50, 1.83, 4.42	58.81	0	41.19	0
ϕ 12.45	5.91	55.18	26.14	13.95

Note: ϕ stands for zero initial.

A statistical study based on a corpus of one million syllables was carried out in 1960s, in order to get the occurrence frequency of the initials and the finals and the joint probability of a initial in combination with a final to form a syllable as while[2]. The average joint probability of the finals in each of the four classes and the occurrence frequency of the initials in percentage are listed in Table 1.

Table 1. shows a clear picture of the quantitative relation of the phonotactic rules. It can been seen that the distribution of the joint probability is not so even and there are 12 impossible combinations of which the entry are zero. For examples, the dorso-palatal /j, q, x/ can only be combined with the finals of *Qichi* and *Cukou* and never with the *Kaikou* and *Hekou*. Inversely the apico-palatal /zh, ch, sh, r/, of which the place of articulation is near the dorso-palatal, have different syllable structures, and then the perceptual distance between them is enlarged.

So that 22 initials including the “zero initial” and 37 finals in Chinese can make 814 possible combinations, but only about 410 syllables without lexical tones exist in real speech. That means much more redundancy in phoneme coding.

It can be also seen that the amount of the finals of *Kaikou* is higher than 50% of syllables with initial consonants, while the occurrence frequency of *Qichi* is higher that 50% of the syllables without initial consonants or with “zero initial”.

The significantly uneven distribution of the joint probability resulting from the strict phonotactic rules increase the redundancy of the information source(phoneme)coding and then decrease the decoding errors at the receiving end. In other words, the phonotactic rules which should be learned by the speakers and listeners during the language acquisition process can be treated as the internal information of the syllable structures. And the role playing by these rules - internal information in speech communication is discussed as follows.

4. MANNER OF ARTICULATION IS SUPERIOR IN SPEECH PERCEPTION

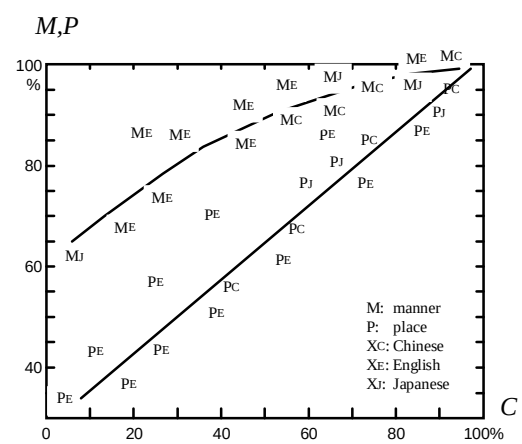

Fig.2. The manner recognition rate M and the place recognition rate P versus the consonant recognition rate C for Chinese, English, and Japanese.

In order to explor the perceptiual importance of different features of speech, the recognition rate of manner and of place versus consonant intelligibility for three different languages were investigated. The English data including four manner and three place channels were cited from Miller and Nicely [3] and the Japanese data including five manner and three place channels from Nagai, Sato, and Sato [4]. The manner recognition rate M and the place recognition rate P versus consonant intelligibility C were shown in Fig. 2. It can be seen that the manner recognition rate is always higher than the place for all languages compared under the same transmisson conditions.

5. RELATIVE INFORMATION TRANSMITTED BY PLACE CHANNEL

It is well known that the manner of articulation is superion in speech perception and more perceptual confusions of consonants are in between different places. In order to quantitatively prove that the phonotactic rules play a role of internal information in speech perception, the relative information transmitted by the place channel and the voiced channel of Chinese was calculated and compared with that of English given by Miller and Nicely[3].

The relative information Tr(x, y) transmitted by different perceptual feature channels is defined as

$$T_r(x,y) = T(x,y) / H(x) \qquad (1)$$

$$T(x,y) = -\sum_i \sum_j P(i,j) \log_2 \frac{P(i)P(j)}{P(i,j)} \qquad (2)$$

where T(x, y) is the information transmitted by a feature channel; H(x) the entropy of the information source; P(i) the occurrence probability of speech sound at the transmitting end; P(j) the occurrence probability of speech sound at the receiving end ; P(i, j) the confusion probability between sound i and j.

The differences of Tr(x,y) transmitted by the voiced channel and the place channel for different low pass filtering speeches of Chinese and English are shown in Fig. 3. It is worth to notice that the relative information transmitted by place channel of Chinese is much higher than that of English, while the relative information transmitted by the voiced channel is quite close each other.

In addition, it is worth to point out that six places were used to calculate Tr(x,y) for Chinese, but only three places for English, so that the entropy of information source for Chinese is 0.938 bits higher than that for English. Actually the relative information transmitted by place channel of Chinese is even higher than that shown on Fig. 3.

We must be clearly aware that some side effects of the phonotactic rules which detract the perceptual features of manner of articulation exist in spoken Chinese. For example, the nasal[n] and lateral[l] have different manners, and larger distance in perceptual space for English. They have, however, same syllable structures in Chinese and then they are easily confused each other for Chinese listeners. Similarly three dorso-palatal consonants /j/, /q/, /x/ belong to three different manners but in the same cluster in perceptual space [5], for they have same syllable structures. But this is not so serious for polysyllabic word perception.

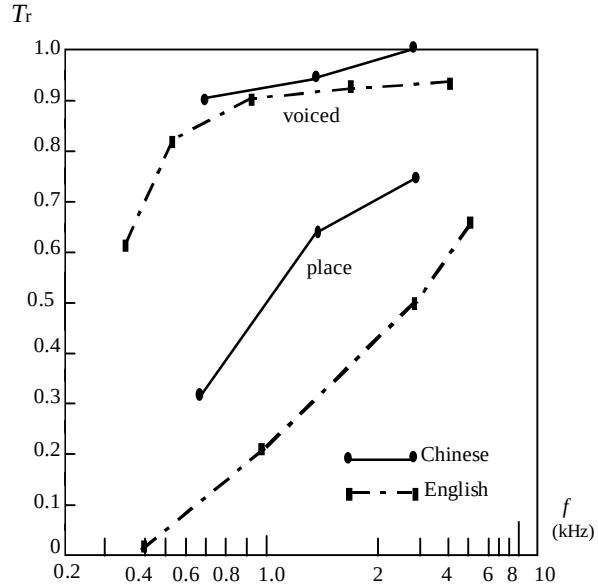

Fig.3. The relative information Tr transmitted by voiced channel and place channel for low-pass filtering speech.

6. R-COLOURED SYLLABLES

It should be mentioned that the r-coloured syllables which exist in some dialets (Beijing and Sichuan) are produced by some specific sound modification, when a syllable is followed by an /er/ syllable in a word, such as *zher*, *Harbin*. The r-coloured syllables are used to give some one (thing) a pity name, but sometimes they make big differences in meaning, for example, *Baimian* means flour, but *Baimianr* heroin. The phonetic rules and the acoustic parameters of r-coloured syllables were disscussed elsewhere in detail [6].

7. CONCLUSIONS

The phonological system , the initials, the finals, and the tones as the structural elements of syllables, is suitable to Chinese, and there are some strict phonotactic rules based on it in syllable structures. In speech perception the manner of articulation has advantage of the place. And the phonotactic rules of Chinese are dependent on the place of articulation of the initials and the quality of medial vowels. These rules can be considered as the internal information of spoken Chinese and are beneficial to increase syllable intelligibility. The fact that the speech perception is influenced by the syllable structures has proved that perceiving a phoneme is based on a syllable in which it exists. So that a intelligent speech recognition system should be

based not only on the individual features of speech sound but also on the feature combinations in the context.

A Chinese character is a syllable and a morpheme or a word also, therefore the syllable intelligibility is very important to speech recognition. May be that is why the specific syllable structures of Chinese were developed during a long evolution process.

8. REMARKS

We have to gain an insight into the word structures and semantic and syntactic rules in next step. Although Chinese syllable structures have advantage of English in speech communication, the sentence intelligibility of both languages is almost the same under a certain Articulation Index(AI).

9. REFERENCES

1. Luo, C.P. and Wang, J., "Essentials of General Phonetics", *Science Publishing House, Beijing (1957) (in Chinese).*

2. Zhang, J.L., in "Handbook of Acoustics", *edt. by Maa, D.Y. and Shen, H., Science Publishing House, Beijing(1983)(in Chinese).*

3. Miller, G.A. and Nicely, P.E., "An analysis of perceptual confusions among some English consonants", *J. Acoust. Soo. Am., 27(2), (1955), 338-352.*

4. Nagai, K., Sato, R. and Sato, T., "An analysis of phonemes by articulation tests", *J. Acoust. Soc. Japan, 12 (1956), P148.*

5. Zhang, J.L., Lu, S.N., and Qi, S.Q., "A cluster analysis of the perceptual features of Chinese speech sounds", *J. Chinese Linnguistics, 10, (1982), 190-206.*

6. Zhang, J.L., "Acoustic parameters and phonological rules of a text-to-speech system for Chinese", *ICASSP'86, Tokyo, 38.7.1-4.*