

# Study on Stress Models of Chinese Disyllable

DENG Dan, ChEN Ming, LU Shinan\*

Infoquick Sinovoice Corporation

Institution of Acoustics\*

dengdan6@163.net

## Abstract

This paper investigates the stress distributions of Chinese disyllables in 1+1, 2+2 and 4+4. The result shows that: 1) most of Chinese disyllables do not have fixed stress models, moreover, two syllables have equal opportunities of being stressed; 2) disyllables in 2+2 and 4+4 have obviously initial-stressed inclination due to “neutralization” in final syllable; 3) not only pitch and duration, but also break level and the position in the prosodic phrase influence the distributions of word stress.

## 1. Introduction

There are two opinions about Chinese word stress to date. First, most of Chinese disyllables are final-stressed. From phonematic point of view, Chao (1979) divided Chinese stress into three kinds, i.e. normal, strong and weak. He also said that the stress degree of syllable is not equal to that of normal stress, furthermore, stress often locates on the last syllable. Lin *et al.* (1984) investigated 103 disyllables in 16 kinds of tone combinations, and concluded that Chinese disyllable has a tendency to be final-stressed. Second, the stress location is not fixed. Wang *et al.* (2003) figured that the stress position of a word in utterance was influenced by prosodic position of the word in a sentence, so there were not steady stress models.

The study on acoustic correlates of stress shows that both pitch and duration influence word stress (Lin *et al.* 1984; Xu *et al.* 2000). Lin (1983) proposed that duration is more important in stress, but Wang *et al.* (2003) said that the correlation coefficient between the stress degree and the pitch is larger than that between the stress degree and the duration.

The synthesized speech with high intelligibility has been realized by concatenation of non-uniform units. Word stress regulation is important in improving the naturalness of synthesized speech. In this paper we will investigate the stress location of Chinese disyllables in different prosodic conditions. Research method will be introduced in section 2. The results are analyzed in section 3. Final conclusions are given in section 4.

## 2. Research method

Chinese disyllables taken from TTS speech corpus of Infoquick Sinovoice Corporation are used in this study, which were read by a professional broadcast woman with neutral intonation. The corpus is composed of isolated sentences. Most sentences were selected from newspaper, and others from novels.

The type of tones born by the syllables of a word is one of the elements that influence the perception of word stress. Wang *et al.* (2003) investigated the perceptual stress of Chinese disyllables in 16 kinds of tone combinations, and proposed that the proportion of initial-stressed is high when initial syllable is tone-1 and tone-4, while the proportion of final-stressed is high when final syllable is tone-1. To reduce the diversity of pitch in different tones, we investigate Chinese disyllables composed of the same tones in this paper. Because 3+3 has become 2+3 in natural utterance, 3+3 will not be included in this study. 327 disyllables with 1+1, 303 disyllables with 2+2, 312 disyllables with 4+4 were selected randomly from 500 sentences in corpus.

### 2.1 Measurements of acoustic data

Firstly, we measure duration and pitch of syllables in spectrum and pitch plot by CSL3700 respectively. Each tone exists pre-onset and post-offset in natural utterance. Lin (1988) proposed that pre-onset and post-offset are not related to tone perception. The pitch values of syllables will not comprise pre-onset and post-offset in this study. Because tone-1 is high and level tone, its pitch value is only one, i.e. the average value of pitch. Tone-2 and tone-4 include high point and low point value. Syllable duration will not include section of closure.

Secondly, we calculate the high point difference( *FOH*), low point difference( *FOL*), and duration difference( *DUR*) in disyllables respectively by first syllable minus second syllable.

### 2.2 Prosody labeling

The subjects of this experiment are 4 staffers working in Infoquick Sinovoice Corporation. Two of them are born and grown up in Beijing; others have lived in Beijing for a long time and can speak standard mandarin. They all have natural audition. There is 10 minutes drill before experiment. The results of drill will not be used in calculation later.

#### 2.2.1 Stress labeling

Disyllables including some disturb words are played one by one randomly with 3 seconds break in between. Subjects are asked to do a forced choice in initial-stressed, equal-stressed and final-stressed. In listening process, subjects can also ask for repeated. The stress model of disyllable will be determined to be initial-stressed or final-stressed if above 3 subjects all choose initial-stressed or final-stressed. Otherwise, the stress model will be equal-stressed.

### 2.2.2 Break labeling

Sentences including disyllables are played one by one randomly with 3 seconds break in between. Subjects are asked to choose the break level behind disyllables from no-break (B0), small-break (B1) and big-break (B2). B2 also includes the break of sentence end. If less than 3 subjects make agreement on the break level of the same words, the corresponding disyllables will be played to the 5th subject. The break level chosen by most subjects are used in statistic calculation later.

### 2.2.3 Phase position labeling

According to Wang, B. *et al.*'s proposition (2003) that the acoustic correlates of prosodic phase boundaries are silence and pitch reset, we think a prosodic phase appears if the silence exceeds 50ms. Then we divide these disyllables into head-phase, middle-phase and end-phase.

## 3. Results and analyses

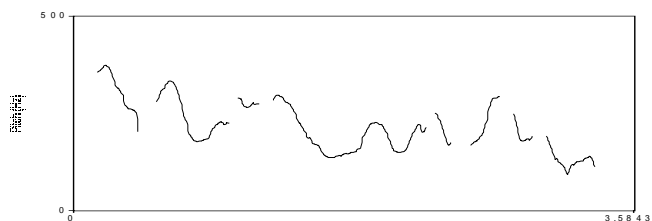
### 3.1 Distribution of word stress

Table 1 lists the numbers and percents of word stress in three kinds of tone combinations. It is easy to see in Table 1 that: 1) Most disyllables are judged equal-stressed in three kinds of tone combinations, namely, 91.8% for 1+1, 85.5% for 2+2, 68.6% for 4+4. 2) The stress distribution of tone combinations is different. As for disyllables of initial-stressed and final-stressed, the most lies in 4+4, next in 2+2, and the least in 1+1. 3) The number of final-stressed is a little higher than that of initial-stressed in 1+1, while the number of initial-stressed is higher than that of final-stressed distinctly in 2+2 and 4+4. These results show that the stress model of Chinese disyllables composed of the same tones is not fixed in natural utterance.

Moreover, most disyllables incline to be equal-stressed.

Our result of stress distribution is the same as Wang *et al.* (2003), but the percent of equal-stressed disyllables in our study (more than 65%) is larger than hers (close to 50%). The reason is that disyllables used by Wang included 16 kinds of tone combinations, while only 3 kinds of the same tone combinations are used in this paper. It implies that the percent judged to be equal-stressed would rise if the diversity of pitch in different tones were removed. Our result of stress distribution is different from Lin *et al.* (1984). Lin argued that the percent of final-stressed was close to 90%. We think these differences attribute to different study objects. Our disyllables come from natural utterance while his from laboratory. Because coarticulation and sentence prosody would influence the word pronunciation easily, the perceptual results of different objects are divergent.

Chinese tone-2 is "rise" normally, but some become "fall" and some become "flat" in our data, furthermore, these tones are similar to tone-5 that follows tone-2. Such as Figure 1, both 结 and 合 are tone-2 marked in dictionary, but the pitch of 合 become "fall". Since the final syllable of words is neutralized in utterance, the percent of being judged initial-stressed in 2+2 is higher than that in 1+1.



印地安人将他们的名字结合起来

Figure 1: Pitch plot of sentence-1

Table 1: Distribution of word stress in three kinds of tone combinations

Stress Model \ Tone Combination	1+1		2+2		4+4	
	Number	Percent (%)	Number	Percent (%)	Number	Percent (%)
Initial-stressed	8	2.4	32	10.5	82	26.3
Final-stressed	19	5.8	12	4	16	5.1
Equal-stressed	301	91.8	259	85.5	214	68.6

Table 2: Correlation coefficients of stress, pitch and duration

Tone combination	Stress and FOH	Stress and FOL	Stress and DUR
1 + 1	-0.080		0.098
2 + 2	0.005	0.170**	-0.147**
4 + 4	-0.339**	0.026	0.081

("\*\*") represent significant difference in 0.01 level)

Some words of 4+4 show initial-stressed inclination, such as

“变化、部队、但是、动物、干部、告诉、坏事、技术、烈

曰”。 This is due to three reasons. Firstly, frequency of using these words is very high. Their tones are marked tone-4 in dictionary, but their pronunciations are close to tone-5. Secondly, both tone-4 and tone-5 are falling tone. Their distinctive features are probably neutralized because of coarticulation in natural utterance. Thirdly, Deng *et al.* (2003) proposed that tone-4 of the second syllable in disyllable is easy to be neutralized. So we can conclude that the disyllables of 4+4 with high use frequency are inclined to initial-stressed, because of the second syllable neutralization.

### 3.2 Word stress, pitch and duration

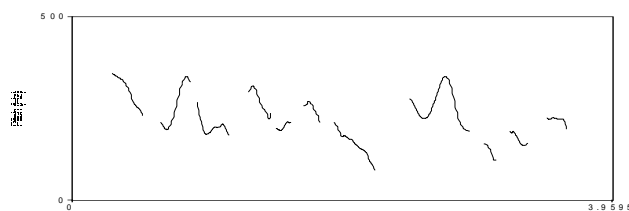
Table 2 lists the correlation coefficient of stress- *FOH*, stress- *FOL*, stress- *DUR*. From Table 2, we can see that: word stress correlates with both *FOH* and *FOL* in 1+1, but the correlation coefficient is not very high. The main reason is that only a few words are judged initial-stressed and final-stressed, while others are equal-stressed. We conclude that the stress model of most 1+1 disyllables is equal-stressed, and we can use only one model to synthesize 1+1 disyllables.

Word stress does not correlate with *FOH*, whereas correlates significantly with *FOL* and *DUR* in 2+2. Shen (1994) argued that the top point of syllable correlates to the stressed word while the low point correlates to the rhythm of the sentence. Our study shows that word stress does not correlate to top point while correlates to low point. This is

because the top feature of tone-2 is very unsteady for it is often neutralized in natural utterance (Chen, *et al.*2003; Deng, *et al.*2003). When locating in the second syllable of disyllable, tone-2 probably loses top feature, therefore, the stress information is diverted to the low point. In Figure 2, both 韩 and 国 are tone-2 marked in dictionary, but the pitch of 国 is lower than that of 韩 distinctly.

Word stress correlates significantly to *FOH*, weakly to *DUR*, not to *FOL* in 4+4. That shows the stress location of 4+4 is influenced by pitch and duration of disyllable, especially influenced by high point while independent of low point.

In a word, both pitch and duration influence the stress location. Stress is influenced mostly by high point. If high feature of syllable is neutralized in utterance, stress is influenced by low point.



但 韩国的爱国志士们却宁死不屈

Figure 2: Pitch plot of sentence-2

Table 3: Distribution of word stress in different break

Tone Break	1 + 1		2 + 2		4 + 4	
	Initial- Stressed	Final- Stressed	Initial- Stressed	Final- Stressed	Initial- Stressed	Final- Stressed
B0	6	4	28	1	59	2
B1	2	7	4	7	21	10
B2	0	7	0	4	2	4
Total	8	18	32	12	82	16

Table 4: Word stress distribution in different position of prosody phrase.

Tone Position	1 + 1		2 + 2		4 + 4	
	Initial- Stressed	Final- Stressed	Initial- Stressed	Final- Stressed	Initial- Stressed	Final- Stressed
Head	7	3	17	3	43	5
Middle	2	6	14	3	31	4
End	0	9	3	6	8	7
Total	8	18	32	12	82	16

### 3.3 Word stress and break

The correlating analysis to word stress and break level shows that correlation coefficients are 0.156\*\* in 1+1, 0.236\*\* in 2+2 and 0.368\*\* in 4+4. The results imply that type of stress correlates significantly to break level where word located.

Table 3 is the distributions of word stress that is judged initial-stressed and final-stressed in different break levels. It is easy to see that the words that are judged initial-stressed incline to exist in front of B0 while hardly in front of B2. Because there are a few numbers of initial-stressed and final-stressed in 1+1, we cannot deduce affirmative conclusion. According to

2+2 and 4+4, the neutralization of the second syllable in disyllable often appears in no-break utterance, whereas the neutralization barely appears in front of big-break or the end of sentence. Table 3 also shows that some disyllables that locate in front of B1 in 4+4 are judged initial-stressed. We find that the usage frequency of those disyllables is very high and the pronunciation of those disyllables is similar to tone-5.

Most disyllables that are judged final-stressed exist in front of B1, some in front of B2, and hardly in front of B0. That implies the disyllables in front of pause incline to final-stressed, especially in front of the small-break. These results coincide with the opinion that the stress often appears in the back. Deng *et al.* (2003) investigated the disyllables that include tone-4 and tone-5, and proposed that the pitch change is more distinct within the disyllables in front of small-break than that in front of big-break or no-break, moreover, the range of the second syllable is bigger, and the duration of the second syllable is longer. These results all reveal that the disyllables in front of small-break have final-stressed inclination.

### 3.4 Word stress and the location in prosody parse

The correlating analysis to word stress and location in phrase shows that correlation coefficients are 0.219\*\* in 1+1, 0.235\*\* in 2+2 and 0.346\*\* in 4+4. That shows the type of word stress correlates significantly to the location of word in phrase. Table 4 lists the distributions of word stress that are judged initial-stressed and final-stressed in different location of prosody phrase. Table 4 indicates that most disyllables that are judged initial-stressed are located in head-phrase while most disyllables that are judged final-stressed are in end-phrase. In other words, the disyllables located in head-phrase have initial-stressed inclination and the disyllables located in end-phrase have final-stressed inclination. That coincides with the conclusion of Chen *et al.* (2003). In addition, due to neutralization of the second syllable, some disyllables located in middle-phrase of 2+2 and 4+4 are judged initial-stressed.

## 4. Conclusion

The distributions of prosodic disyllables in 1+1, 2+2 and 4+4 are investigated in this paper. The results show that: 1) the common model of Chinese disyllable is equal-stressed. Because of neutralization of the second syllable, the disyllables are inclined to initial-stressed in 2+2 and 4+4. 2) Word stress correlates to pitch and duration. The type of word stress in 1+1 and 4+4 is influenced by duration and high point of the first and second syllable. Because the high point of tone-2 is often neutralized in natural utterance, the type of word stress in 2+2 is influenced by duration and low point. 3) The distribution of word stress correlates to break level where word locates. The disyllables that locate in front of no-break incline to be initial-stressed and those in front of small-break incline to be final-stressed. 4) The distribution of word stress correlates to the location of word in prosodic phrase. The disyllables located in head-phrase incline to be initial-stressed and those located in end-phrase incline to be final-stressed.

## 5. References

- [1] Chen, M.; Wang, A. H.; Xiao, N., *et al.*, 2003. The pitch analysis of mandarin continuous disyllables. *6th NCMP proceedings*. Tian Jin: China, 155-160.
- [2] Chao, Y. R., 1979. *A grammar of spoken Chinese*. Bei Jing: China. Commercial press.
- [3] Deng, D.; Chen, M.; Lu, S. N., 2003. The study of the prosodic features of Mandarin Tone-4 and Tone-5. *Language science*, in press.
- [4] Lin, T., 1983. The primary experiment on Chinese mandarin tone-5. *Linguistics series 10*, Bei Jing: China. Commercial press.
- [5] Lin, M. C.; Yan, J. Z.; Sun, G. H., 1984. The pilot study of normal stress in mandarin disyllables. *Dialect*, 335-339.
- [6] Lin, M. C., 1988. Acoustic feature and perceptual sign of mandarin tones, *Chinese 3*, 437-445.
- [7] Shen, J., 1994. Perceptual acoustic correlates of semantic stress in mandarin. *Chinese language and writing 3*, 10-15.
- [8] Wang, B.; Yang, Y. F.; Lu, S. N., 2003. Acoustic correlates of hierarchical prosodic boundary in mandarin. *Chinese Journal of acoustics*, in press.
- [9] Wang, Y. J.; Chu, M.; He, L. *et al.*, 2003. Stress perception of Chinese disyllable words in utterance, *Chinese journal of acoustics 22*(1), 86-96.
- [10] Xu, J. P.; Chu, M.; He, L. *et al.*, 2000. The influence of Chinese sentence stress on pitch and duration. *Chinese journal of acoustics 25*(4), 335-339.