

THE STABILITY ANALYSIS OF DISYLLABIC STRESS IN MANDARIN SPEECH

Xiaoying Xu^{a,b}, Ya Li^b, Jianhua Tao^b & Yingchao Lu^{a,b}

^aPSC Center, CLLD, Beijing Normal University, 100875, Beijing, China;

^bNLPR, Institute of Automation, Chinese Academy of Sciences, 100190, Beijing, China

xuxiaoying2000@bnu.edu.cn; yli@nlpr.ia.ac.cn;

jhtao@nlpr.ia.ac.cn; d_r_036@yahoo.com.cn

ABSTRACT

The paper focuses on the analysis of the stability of stress perception at the word level by using a large amount of corpus data, and tries to expand research on disyllabic stress patterns in continuous speech. The results show that: a) Only a small part of words have the stable stress patterns. These words appear more than three times in the corpus and get the full agreement among all annotators. b) There is no significant correlation between the stability of stress patterns and some frequently used phonetic features, e.g., the tone combination types, the types of the vowel, and the prosodic boundary types. Furthermore, the paper compares our results with two other stress perception works which are based on isolated words. Results show that there are obvious differences between disyllabic stress patterns in isolated words and continuous speech.

Keywords: stress stability, stress pattern, Mandarin speech

1. INTRODUCTION

For a typical stress language, e.g. English, most of the words have relatively stable stress patterns despite that the word stress patterns may change under the impact of prosody rhythm or other context information [1, 2] and a small amount of compounds may have variable stress modes.

However, for a typical tone language, e.g. Mandarin, it is still hard to get some stress patterns due to the complicated confusion between tones and stresses. Many researchers [2, 5, 7] argued that it is difficult for human to identify the stress pattern of the full-tone disyllabic in Mandarin. On the contrary, lots of researchers [10, 12] insisted that the stress does exist in Mandarin despite the difficulties in identifying them and they categorized the Mandarin disyllabic stress pattern into three types, left-stressed, right-stressed and neutral-tone respectively, based on their personal perceptual judgment and some other native

speaker's perceptual judgment [4, 6, 8, 9, 10, 12]. Furthermore, although it has been a long time for the research on stress of Mandarin speech, there is no previous works focusing on the stability of the disyllables in continuous natural speech.

The stability analysis of the word stress plays a very important role in the field of language teaching and speech synthesis. In this paper, we focus on the analysis of the stability of disyllabic stress patterns by using a large amount of corpus which contains 6,000 sentences and 223,339 disyllables.

The rest of the paper is organized as follows. Section 2 introduces the corpus and stress labeling method in our research. Section 3 provides the stability analysis of disyllables in our corpus. Section 4 makes the comparison of our work with Song's work [10] and Feng & Wang's work [6] separately. Section 5 makes the conclusion and the discussion of the future research.

2. THE CORPUS AND STRESS LABELING

The speech corpus used in our work contains 6,000 sentences, which is built for speech synthesis with rich context information and read by a professional female speaker. All of the sentences had been labeled with Pinyin script, Syllable segmentation

Three professional research assistants are asked to do the stress annotation by listening to the prosodic words and labeled the stress level in the above mentioned text. In order to avoid the perceptual disturbing of the adjacent syllable, the sequence of the prosodic words has been disarranged, and the annotators only listen to one prosodic word per time.

Stress annotation in our works focus on both the contrast relation within the prosodic word and the stress level of each syllable. According to annotators personal auditory perceptual felling, the annotators were asked to rate each syllable into 3, 2, 1 which is correspond to the three stress level, i.e. SS (Stressed Syllable), RS (Regular Syllable)

and US (Unstressed Syllable). There is no restriction for the stress type of the prosodic word. If there is no perceptible prosodic prominence, all syllables in the prosodic word can be marked with the same score.

3. THE STABILITY ANALYSIS OF DISYLLABIC STRESSES

In our work, three annotators labeled the stress pattern of the 41,038 disyllabic respectively. The results show only 16.8% (6894 words) can get the total agreement of the three annotators, 83.2% of them have different annotations among three annotators.

The annotation of the corpus showed that the vast majority of prosodic disyllabic can't achieve the full agreement of different annotators. In previous studies, the inconsistencies of stress labeling among different annotators are found in both the isolated word level and the natural speech level. This phenomenon can be explained in two different directions. One is taking the different annotations as the reflection of no significant salient syllable within the prosodic word. Based on this kind of view, the word without full agreement in stress annotation can be considered as the equal-stressed word. The other is taking the different annotations as the difference of the personal perceptual feeling. Based on this kind of view, the majority annotation will be taken into consideration, and the minority annotation will be ignored.

In our work, we focus on the issue of stability of word stress pattern, we think that this stress labeling inconsistencies mean that most of the word stresses cannot be easily fixed with specific patterns. To avoid the bias, only the words with a full agreement of annotation will be discussed in the following of this paper.

3.1. The stability of disyllabic stress patterns

The total amount of the words which got the full agreement among annotators and appear more than 3 times in our corpus is 492. The stability of their stress patterns is as follows.

Table 1: Stability of disyllabic stress pattern.

Stress pattern		Numbers of words (%)
Stable stress pattern	Left-stressed stable	246 (50.0%)
	Right-stressed stable	131 (26.6%)
Unstable stress pattern	Left-stressed prominence	59 (12.0%)
	Right-stressed prominence	56 (11.4%)

The words with stable stress patterns account for 76.6%, while the words with unstable stress patterns account for 23.4%. Among all stable stress patterns, 50.0% are left-stressed (named as left-stressed stable), and 26.6% are right-stressed (named as right-stressed stable). It shows that most of the stress patterns keep stable and the left-stress stable have numerical superiority in our corpus. In unstable stress patterns, if one disyllable was labeled as left-stressed more than right-stressed, we name this pattern as left-stressed prominence. With the same idea we get right-stressed prominence. We nearly get the equal rate of left-stressed prominence and right-stressed prominence.

3.2. Impact factor of the stability of disyllabic stress pattern

A large number of researches proved that the impact factors of the distribution and perception of the disyllabic stress pattern are the tone combination types, the contrast degree of aperture of the vowel, and the prosodic boundary types [3, 11, 13]. The following is the discussion of the effects of these three factors.

3.2.1. Tone combination types

The tone combinations of the disyllabic in Chinese used in our work are $4 \times 5 = 20$ types, only 19 occurred in our corpus. Table II demonstrates the statistic percentage of left-stressed among all disyllables. R refers to the type of the right tone in the disyllable, and L refers to the type of the left tone.

Table 2: Tone-combination type and the stability of disyllabic stress pattern.

I \ R	1	2	3	4	5
1	0.41	0.82	0.91	0.75	1.00
2	0.49	0.47	0.81	0.24	0.74
3	0.24	0.31		0.15	0.87
4	0.56	0.84	0.95	0.86	0.99

The results indicate that only the tone combination type "15" is left-stress stable, all rest tone combination types don't have stable stress pattern. Different tendency of stress distribution existed, such as 13, 14, 23, 25, 35, 42, 43, 44, 45 tone combination types tend to be left-stressed prominence 24, 31, 32, 34 tone combination types tend to be right-stressed prominence and 11, 21, 22, 41 have equal rate in left-stressed and right-stressed, but the disyllabic with neutral-tone can

also have right-stressed pattern and the third tone syllable which is normally unstressed in Chinese can be addressed the stress regardless of its head position and final position in the disyllabic.

3.2.2. The contrast degree of aperture of the vowel

We classify the aperture degree of the vowel “a” in Chinese into 4, “ē” into 3, “o, e, er” into 2, and “i, u, ü, -i, -i” into 1. Then, according to the contrast degree of aperture of the vowel, we divide the prosodic disyllabic into three types, i.e. equal type (the degree of the right syllable equal to that of the left syllable), right type (the degree of the right syllable larger than that of the left syllable) and left type (the degree of the left syllable larger than that of the right syllable). The results are as follows.

Table 3: The contrast degree of aperture of the vowel and the stability of disyllabic stress pattern.

The contrast degree of aperture of the vowel	Numbers of the words	Left-stressed stable	Right-stressed stable	Unstable stress pattern
Equal type	141	72	29	40
Right type	175	33	96	46
Left type	176	140	7	29

For the disyllabic with a contrast degree of aperture of the vowel which is the right type, the ratio of left-stressed stable and right-stressed stable is 33:96, for that which is left type, the ratio of left-stressed stable and right-stressed stable is 7:140. The results indicate that the contrast degree of aperture of the vowel do impact the disyllabic stress pattern. But there are also no significant correlation between the contrast degree of aperture of the vowel and the stability of the disyllabic stress pattern. There are many unstable stress patterns among all types of the contrast degree.

3.2.3. Prosodic boundary type

In this part, we use b1, b2, b3 to represent the three prosodic boundary types in our corpus, i.e. the prosodic word, the prosodic phrase and the intonational phrase.

Table 4: Prosodic boundary type and the stability of disyllabic stress pattern.

Prosodic boundary type	Left-stressed stable and left-stressed prominence	Right-stressed stable and right-stressed prominence
b1	0.757774	0.241571
b2	0.529478	0.470522
b3	0.463435	0.534014

We can find that the majority of the disyllabic in b1 boundary are left-stressed and the in b2, b3

boundary the ratio of left-stressed and right-stressed is relatively closer, but stress patterns of the disyllabic in different prosodic boundary are also not stable.

3.3. The comparison with other stress labelling work

3.3.1. Stress patterns in isolated words and the stability of disyllabic stress patterns

Lots of previous works categorized the disyllabic stress pattern base on the personal perceptual judgment of the isolated words or the designed sentences embedded with the relative words [10, 12]. In order to examine whether the stress pattern displayed in the isolated word can maintain the same tendency in the real speech or not, we compared our results with the data showed in Song [10] which consists of 6,595 disyllabic. The numbers of the disyllabic co-occurred in both works are 360, the results are as follows.

In Song’s work, among 314 right-stressed disyllables in [10], 119 maintain the right-stressed pattern while 195 change to left-stressed. Among 46 left-stressed disyllabic in [10], 39 maintain the left-stressed pattern while 7 change to left-stressed. It is common for a disyllabic to have one kind of stress pattern in isolated word and change to another stress pattern in the real speech. Song [10] emphasis that “we will feel stiff and consider it as not up to standard if the light-stressed word to be pronounced as the right-stressed word in the real speech no matter in reading or talking”. The analysis results in our corpus don’t support this conclusion.

3.3.2. Sentential focus stress pattern and the stability of disyllabic stress pattern

Feng and Wang [6] argued that tonal trochees do exist in Beijing Mandarin, especially in colloquial speech, and that these tonal trochees should join the non-tonal trochees to form the language’s lexically stressed category. Feng and Wang [6] displayed 268 disyllabic’s stress pattern totally and emphasis that the stress pattern do exist when the word are located in the sentential focus position. In order to examine whether the stress pattern displayed in the sentential focus position can maintain the same tendency in the real speech context or not, we compared our results briefly with the data showed in Feng and Wang [6]. The numbers of the disyllabic co-occurred in both works are 51, the results are as follows.

Feng and Wang [6] displayed 29 left-stressed words, in our corpus, 22 maintain the left-stressed stable stress pattern, 4 are right-stressed stable and 3 have both left-stressed and right stressed. Feng and Wang [6] displayed 22 right-stressed words, in our corpus, only 7 maintain the right-stressed stable stress pattern, but 12 are left-stressed stable and 3 have both left-stressed and right stressed. Results show that the stress pattern displayed in the sentential focus position can't maintain the same tendency in the read speech context, the words with right-stressed in the sentential focus position are more likely to change into left-stressed pattern.

4. DISCUSSION AND CONCLUSIONS

In our work, the following results have been concluded:

a) In the natural speech, among the 41,038 prosodic disyllabic in our corpus, only 16.8% can get the full agreement of the three annotators, the result is much lower than the result of the previous studies. We will do more in-depth analysis on the annotation disagreement in the future research.

b) Among words which got the full agreement of the three annotators and the frequency is more than three times in our corpus, 76.6% have the stable stress mode and 33.4% don't. The results indicate that most of the disyllabic with a salient stress pattern can maintain the stable stress pattern in the real speech. According to this conclusion, the stress pattern for this part of word can be labeled in the lexicon, and the information can be used in Language teaching and speech synthesis directly.

c) There are no significant correlations between the stability of disyllabic stress patterns and the tone combination types, the contrast degree of aperture of the vowel and the prosodic boundary types, which indicate that the disyllabic stress patterns in Chinese exist in the lexical level, it can't be derived according to some linguistics hints.

d) Comparing our results with the Song's work [10], it shows that there are obvious differences between the disyllabic stress pattern of the isolated word and that of the words in the real stream. The conclusion displayed in [10] that "we(the tester) will feel stiff and consider it as not up to standard if the light-stressed word to be pronounced as the right-stressed word in the real speech no matter in reading or talking" are problematic, the previous practice of language teaching and testing based on this kind of lexicon must be improved.

e) Comparing our work with Feng and Wang's work [6], it shows that the disyllabic stress pattern of the sentential focus is also different from the data of our corpus. According to this conclusion, we suggest establishing two separating lexicon for the speech synthesis system, one lexicon labeling the relatively stable sentential focus stress pattern of the words, the other lexicon labeling the common context stress pattern. For the second lexicon, little words have stable stress pattern, little words have no stable stress pattern, majority of the words don't have stress salience pattern and can't get the full agreement of the different annotators. This work can provide some primarily results for this lexicon, more reliable data depend on the analysis of the much larger scale audio corpus.

5. ACKNOWLEDGMENTS

The work was supported by the National Science Foundation of China (No. 60873160, 90820303), 863 Programs (No.2009AA01Z320), China Singapore Institute of Digital Media (CSIDM), Beijing Normal University(004-127028) and the Fundamental Research Funds for the Central University.

6. REFERENCES

- [1] Bolinger, D.L. 1961. Contrastive accent and contrastive stress. *Language*. 37, 83-96.
- [2] Couper-Kuhlen. E. 1986. *An Introduction to English Prosody*. Baltimore: Edwards Arnold.
- [3] Deng, D. 2007. *Experimental Study of Chinese Prosodic Word*. Beijing: Beijing University Press.
- [4] Duanmu, San. 2000. *The Phonology of Standard Chinese*. New York, Oxford: Oxford University Press.
- [5] Duanmu, San. 2000. The rhythm of Chinese. *Contemporary Linguistics* 4, 203-209.
- [6] Feng, S.L, Wang, Z.J. 2006. Tonal contrast and disyllabic stress patterns in Beijing Mandarin. *Language Science*. Vol. 1.
- [7] Gao, M.M.K., Shi, A.S. 1963. *A General Introduction to Linguistics*. Beijing: Zhonghua book Company.
- [8] Lin, M., Yan, J.Z, Sun, G.H. 1984. A primarily study on normal stress of bi-syllable word in Beijing Mandarin. *Dialect* 1, 57-73.
- [9] Lin, T., Wang, L.J. 1992. *A Course on Phonetic*. Beijing: Beijing University Press.
- [10] Song, X.Q. 2004. *A Handbook for PSC Tester*. Beijing: Commercial Press.
- [11] Wang, Y.J, Chu, M., He, L. 2003. Labeling stress in continuous mandarin speech perceptually. *Proceeding of the 15th International Congress of Phonetic Science* Barcelona, 2095-2098.
- [12] Xu, S.R. 1980. *Phonetics Commonsense*. Beijing: Press of Chinese character reformation.
- [13] Ya, L., Lu Y.C., Xu, X.Y., Tao J.H. 2010. The influence of prosody levels and tone patters for stress percetion in contineuos speech. *The Nineth Phonetic Conference of China*, Tianjin.