The Research of the Relationship between Perceived Stress Level and Times of Vibration of Vocal Folds

Yin Zhigang
Phonetics Laboratory, Institute of Linguistics,
Chinese Academy of Social Sciences
Beijing, China

***Abstract*—Whether a syllable is perceived as stressed or not and whether the stress is strong or weak are hot issues in speech prosody research and speech recognition. A focus of the stress study is on the investigation of the acoustic factors which contribute to the perception of stress level. This study examined all possible acoustic/physiological cues to stress based on data from Annotated Chinese Speech Corpus and proposed that times of vibration of vocal folds (TVVF) reflects stress level best.**

It is traditionally held that pitch and duration are the most important acoustic parameters to stress. But for Chinese which is a tone language and features special strong-weak pattern in prosody, these two parameters might not be the best ones to represent stress degree. This paper proposed that TVVF, reflected as the number of wave pulses of the vocalic part of a syllable, is the ideal parameter to stress level. Since number of pulses is the integral of pitch and duration (Pulse=∫f(pitch)dt), TVVF can embody the effect of stress on both pitch and duration. The analyses revealed that TVVF is most correlated with the grades of stress. Therefore, it can be a more effective parameter indicating stress level.

Keywords- Times of vibration of vocal folds (TVVF); stress level; parameter.

I. INTRODUCTION

An important aspect of the study of natural language processing is the research of speech stress—for natural language generation, the stress manipulation of synthetic speech units will affect the result of speech synthesis; for natural language understanding, stress-location can affect the comprehensions of sentences or some words. For example, in English, *[re'cord]* is a verb but *['record]* denotes a noun. In some of the other languages, such as Chinese, stress plays a more important role in speech understanding.

But what is stress? What factors contribute to the perception of stress levels? The past research shows that stress is the psychological reflection of the prominence of prosodic units and this reflection is influenced by the objective acoustic factors [1]. These factors include pitch (f0), duration, intensity, lexical tones, and prosodic location. Of these factors, some factors are more correlated with stress than others. [2]

In order to estimate stress levels of syllables automatically, the relationship between perceived stress level and attainable acoustic factors should be analyzed. In this study, we investigated all possible acoustic/physiological parameters to stress levels of syllables in order to find out the one that is most correlated. To a certain degree, it can be an effective indicator to stress grade.

II. POSSIBLE PARAMETERS TO THE STRESS LEVEL

A. Duration and Pitch

According to traditional theory, pitch and duration are most closely related to stress level.

Generally, the prosodic unit which has stress will have raised upper limit of pitch range (PitchMax), enlarged pitch range [3] and lengthened duration [4]. More specifically, the phonetic changes caused by stress can be explained as the intensification of articulation. As shown in figure1, this effect will lengthen the duration and make prominent the pitch features by raising pitch when it has [+high] feature and lowering pitch when it has [+low] feature.

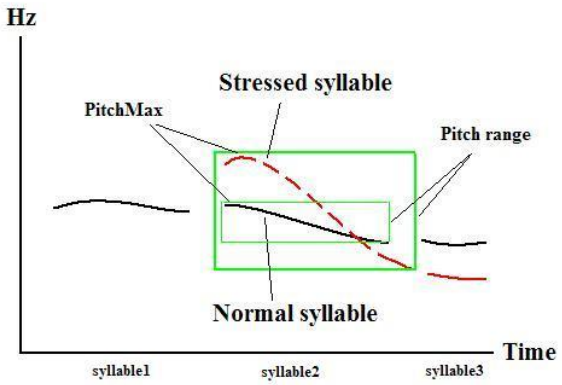

Figure 1. The pitch contours of a normal syllable and a stressed syllable

The above theory suggests that the pattern of pitch change of a stressed syllable will be influenced by its tone type in a tone language. It could be illuminated by the example of Mandarin Chinese. Mandarin is a tone language and the figure2 shows its 4 basic tones—Tone 1(HH), Tone 2(LH), Tone 3(LL) and Tone 4(HL). Of these tones, Tone 1(HH), Tone 2(LH) and Tone 4(HL) all have [+high] feature. When the syllables with these tones are stressed, the [+high] feature will be amplified, with the raised upper limit of pitch range and the enlarged pitch range. The only tone in Mandarin with all [+low] feature is Tone 3(LL). So when the syllables with Tone 3 are stressed, the lower limit of pitch range will be lowered to enlarge pitch range.

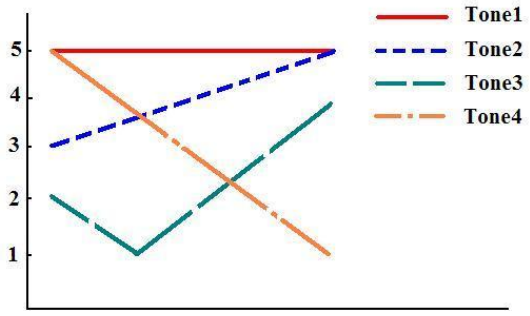

Figure 2. Diagrammatic sketch of Mandarin tones

As for pitch and duration, which one influences the perception of stress most in Mandarin? The issue is controversial. One school of researchers holds that duration contributes most to the perception of stress [5]. The other school thinks that pitch is the acoustic cue to stress. One study [6] found that pitch is most correlated with stress in disyllabic words. Cao [7] studied the phonetic realization of focus stress. His results showed that the most important acoustic cue is the change of pitch register. The change of duration is the secondary cue to stress. Cai [8] proposed an equation to simulate the perception of stress:

$$L=1.5F(\text{pitch})+0.95(\text{duration})+0.65R(\text{pitchRange}) \quad (1)$$

The coefficients implied that pitch is the most influential factor. Other studies demonstrated complimentary relationship and interaction effect between them [9]. Even with complimentary relationship, pitch is more important than duration.

B. *Limitation of regarding pitch and duration as acoustic parameters to stress level*

Although pitch-related acoustic parameters (like pitch range and upper limit of pitch range) and duration are important parameters to stress level, they still have many limitations:

1） The parameter like upper limit of pitch range (pitchMax) can not reflect the perception of stress of Tone 3 in Mandarin. The feature of Tone 3(LL) is [+low]. The stress will lower the lower limit of pitch range. However, the effect of stress for the other tones in Mandarin is to raise the upper limit of pitch range.

2） Another limitation of taking pitchMax as the acoustic cue to stress is that the declination of pitch range happening in the large prosodic unit will make the pitch of the syllables in latter position lower than that of the previous syllables. The declination phenomena of pitch curve could be found in figure3.This in turn will weaken the effect of stress. That is, the higher pitch of the previous syllables doesn't mean these syllables are more stressed (local relative pitch differences are more important).

3） The limitation of pitch range is that it cannot represent the range change of Tone 1(HH) under stress, for the pitch range of Tone 1 is very narrow by nature. The change of pitch range is hard to be detected.

4） The limitation of duration as the cue to stress is that the lengthening of syllables can be due not only to the effect of stress but also the effect of prosodic boundaries [3]. If a syllable is at a prosodic boundary, it will be lengthened, but this lengthening doesn't mean the syllable is stressed. The lengthening phenomena before a boundary could also be found in figure3 (the last syllable).

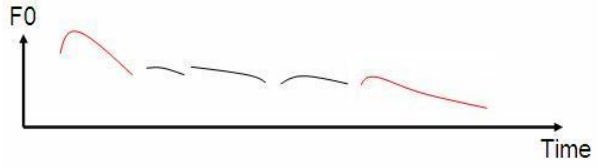

Figure 3. The pitch curve of a sentence

In general, pitch and duration alone cannot effectively cue stress in Mandarin. A new cue which can fully embody stress is needed.

C. *A new parameter to stress level–Times of vibration of vocal folds (TVVF)*

In this research, *Times of vibration of vocal folds* (TVVF) is regarded as an effective cue to stress level. TVVF is a physiological parameter. The times of vibration of vocal folds of a syllable is equal to the number of wave pulses of the voiced part of a syllable (usually, this part is vowel).

The figure below (based on one syllable) shows the relationship among TVVF, pitch and duration on syllable final.

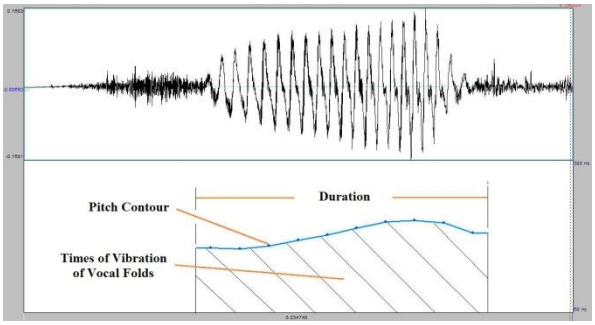

Figure 4. The relationship among TVVF, pitch and duration

In the figure, the blue curve means the pitch contour of a syllable. It could be regarded as a function curve of independent variable (*time*). Because TVVF means the total number of pitch periods in the *duration* of the syllable, it could be represented as the integral of *pitch* and *time* (*duration)* value of the syllable in mathematics. In figure4, the area (signified by diagonal lines) under the blue curve of f0 means TVVF. The relationship among TVVF, *pitch* and *time* could be expressed as the formula:

$$P = \int f(pitch)dt \quad (2)$$

In this formula, *P* means TVVF（*the times of vibration of vocal folds*）. *t* means *time* and its maximum is equal to the *duration* value of the syllable. Therefore, the times of vibration of vocal folds (TVVF) is the integral of pitch and duration value of the syllable.

The times of vibration of vocal folds (TVVF) is related to both pitch and duration.

1) The higher or lower of pitch will increase or decrease the times of vibration of vocal folds.

2) The lengthening or shortening of duration will also increase or decrease the times of vibration of vocal folds.

3) Although the TVVF (times of vibration of vocal folds) is a physiological cue, it has an acoustic counterpart, i.e. number of wave pulses. It is convenient to measure.

4) The limitation of using TVVF as a parameter to stress is that like pitch, it refers to the voiced part of the syllable. So TVVF can be used for the vocalic part. It can't be used to analyze the consonant part. But some research [10] demonstrated that the lengthening of the syllable is mainly due to the lengthening of the vowel. Therefore, the effect of stress can be embodied with the lengthening of vowel alone.

III. Analyses

According to the above discussion, theoretically, TVVF should be more correlated with perceived stress level. The following analyses will test the assumption by comparing the correlation coefficients between stress levels and all acoustic parameters.

A. Corpus

The data analyzed in this study are from ASCCD Annotated Speech Corpus of Chinese Discourse, made by the phonetic laboratory of the institute of linguistics of CASS (Chinese Academy of Social Sciences). The corpus is made of the recordings of 18 passages of different styles (about 10,000 syllables). These passages were read by 10 Mandarin speakers. In this paper, we only analyzed the data of 2 female speakers (F001, F002) and 2 male speakers (M001, M002). The styles of reading between the 4 speakers are very different.

The recordings were made in stereo, with a sampling frequency of 16 kHz and 16 bits per sample resolution. The corpus is about 1.5GB. All the recordings were annotated by manual and the consistent rate of 6 annotators was about 87.5%. The annotations include orthography, syllable initials, syllable finals, prosodic boundaries and grades of stress. The stress is indexed as 0, 1, 2, 3. 0 means weak, 1 means normal, 2 means strong and 3 means very strong. All the labels of stress levels were annotated based on perception of stress without consulting any acoustic parameters.

The following analyses were made to examine the correlation between the acoustic parameters and stress.

The data analyzed in this part were not normalized. In the following tables showing results, the figures marked by ** mean the correlation is significant at the level of 0.01 (two-tailed).

B. Correlation between TVVF and stress

Correlation analysis was conducted between TVVF and the grades of stress for F001, F002, M001 and M002. The results showed that the correlation coefficients were 0.50, 0.56, 0.55 and 0.58 (significant at the level of 0.01). The mean of 4 correlation coefficients was 0.55.

TABLE I. Correlation between TVVF and stress level

Correlation between TVVF and stress level					
		F001	**F002**	**M001**	**M002**
stress level	Pearson Correlation	0.50**	0.56**	0.55**	0.58**
	Sig. (2-tailed)	0.00	0.00	0.00	0.00
	N	8762	8762	8759	8760

C. Correlation between upper limit of pitch range (pitchMax) and stress

The upper limit of pitch range (pitchMax) was the maximum pitch of each syllable.

The following table showed that with all data included (including all the lexical tones, Tone 1, Tone 2, Tone 3, Tone 4 and neutral tone), the 4 speakers' correlation coefficients between pitchMax and stress level are 0.37, 0.45, 0.48 and 0.51 (significant at the level of 0.01), which were both lower than those between TVVF and stress. The mean of this kind of correlation coefficients was about 0.45.

TABLE II. Correlation between upper limit of pitch range and stress level (including all tones)

Correlation between pitchMax and stress level (including all tones)					
		F001	**F002**	**M001**	**M002**
stress level	Pearson Correlation	0.37**	0.45**	0.48**	0.51**
	Sig. (2-tailed)	0.00	0.00	0.00	0.00
	N	8762	8762	8759	8760

Since the effect of stress on the syllable with Tone 3 is different from the syllables with other tones (for syllables with

Tone 3, being stressed will enlarge the pitch range by lowering the lower limit of pitch range), the following correlation is conducted when syllables with Tone 3 were excluded.

TABLE III. Correlation between pitchMax and stress level (excluding syllables with Tone 3)

Correlation between pitchMax and stress level (excluding Tone 3)					
		F001	**F002**	**M001**	**M002**
stress level	Pearson Correlation	0.43**	0.52**	0.55**	0.56**
	Sig. (2-tailed)	0.00	0.00	0.00	0.00
	N	7273	7274	7268	7270

The results showed that the coefficients of all speakers were higher in different degree and the mean of them was about 0.52. This means that the effect of stress imposed on syllables with Tone 3 is different from those with other tones. Excluding syllables with Tone 3, the correlation between pitchMax and stress is stronger and approaches to the level between TVVF and stress.

D. Correlation between pitch range and stress

The pitch range is the difference between pitch maximum and minimum.

TABLE IV. CORRELATION BETWEEN PITCH RANGE AND STRESS LEVEL (INCLUDING ALL TONES)

Correlation between pitch range and stress level (including all tones)					
		F001	**F002**	**M001**	**M002**
stress level	Pearson Correlation	0.09**	0.25**	0.39**	0.41**
	Sig. (2-tailed)	0.00	0.00	0.00	0.00
	N	8762	8762	8759	8760

The results showed that with all tones included, the correlation coefficients for 4 speakers are 0.09, 0.25, 0.39, 0.41 (significant at the level of 0.01) and the mean of them is 0.29.

The results showed that the correlation between pitch range and stress is not as strong as indicated by the traditional sentential stress theory. To further examine the data, another analysis is conducted between stress and pitch range based on ST (Semitone) scale. f0 is changed into ST through the formula St=12*log2(F0 / F(reference)). The correlation based on St revealed stronger relation, 0.15 for F001, 0.30 for F002, 0.32 for M001, and 0.43 for M002 (significant at the level of 0.01). But still it is not strong enough.

The reason for the weak correlation between pitch range and stress might be due to Tone 1. Tone 1 is with very narrow pitch range by nature. So even the syllable with Tone 1 is stressed, its pitch range will not be enlarged, which in turn weakens the correlation between pitch range and stress. With Tone 1 excluded, the correlation coefficients are 0.12 for F001, 0.35 for F002, 0.50 for M001, and 0.55 for M002. (table V)

TABLE V. CORRELATION BETWEEN PITCH RANGE AND STRESS LEVEL (EXCLUDING TONE 1)

Correlation between pitch range and stress level (excluding Tone 1)					
		F001	**F002**	**M001**	**M002**
stress level	Pearson Correlation	0.12**	0.35**	0.50**	0.55**
	Sig. (2-tailed)	0.00	0.00	0.00	0.00
	N	7094	7096	7088	7092

The results of the above two correlation analyses (based on Hz and ST) are not consistent with the strong correlation between pitch range and stress proposed by the traditional sentential stress theory. The reasons might be:

1. The natural speech has the feature of pitch declination (at the same time with narrowed pitch range). The past research usually made normalization on this effect. All the data analyzed in this study were raw data without normalization.

2. The claim of the strong correlation between pitch range and stress is largely based on the local comparison in some adjoining prosodic units. However, in this study, all the data were compared in the overall instead of focusing on particular type of prosodic units.

3. According to traditional stress theory, the widening effect of stress on pitch range is with reference to larger prosodic units rather than the small unit like syllable.

The reasons mentioned above can explain why the correlation in this study is weaker than that proposed by the traditional theory. At the same time, these reasons also imply that using pitch range as cue to stress is not effective.

E. Correlation between duration and stress

The results are shown in the following table.

As shown in the table, the coefficients for F001, F002, M001, M002 were 0.46, 0.50, 0.54 and 0.57 (significant at the level of 0.01), which were good results but still lower than those between TVVF and stress.

TABLE VI. CORRELATION BETWEEN DURATION AND STRESS

Correlation between duration and stress level					
		F001	**F002**	**M001**	**M002**
stress level	Pearson Correlation	0.46**	0.53**	0.54**	0.57**
	Sig. (2-tailed)	0.00	0.00	0.00	0.00
	N	8762	8762	8759	8760

F. Correlation between intensity and stress

For speech, the intensity isn't generally acknowledged a dominant parameter to stress, compared to pitch and duration.

TABLE VII. CORRELATION BETWEEN INTENSITY AND STRESS

Correlation between intensity(Max) and stress level					
		F001	**F002**	**M001**	**M002**
stress level	Pearson Correlation	0.06**	0.05**	0.05**	0.08**
	Sig. (2-tailed)	0.00	0.00	0.00	0.00
	N	8762	8762	8759	8760

The correlation between maximum of intensity and stress were analyzed. In general, intensity were not strongly correlated with stress. For F001, F002, M001and M002, the correlation coefficients between stress grades and intensities were 0.06, 0.05, 0.05 and 0.08. The reason why intensity was only weakly correlated with stress was that it was hard to measure intensity under the standard condition and intensity will be influenced easily by factors other than acoustic ones. For example, the distance from the mouth to the microphone and the physical and emotionally status of the speaker will cause the change of intensity.

G. Correlation between energy and stress

TABLE VIII. CORRELATION BETWEEN ENERGY AND STRESS LEVEL

Correlation between energy and stress level					
		F001	**F002**	**M001**	**M002**
stress level	Pearson Correlation	0.05**	0.04**	0.02**	0.06**
	Sig. (2-tailed)	0.00	0.00	0.00	0.00
	N	8762	8762	8759	8760

As shown in the table VIII, the correlation coefficients for 4 speakers were 0.005, 0.04, 0.02 and 0.06 (significant at the level of 0.01). These coefficients meant that energy was only weakly correlated with stress. The reason for the weak correlation was that like intensity, energy will be affected by other factors (recording volume, distance, etc.) rather than acoustic ones. Therefore, it is not reliable to take energy as the parameter to stress.

IV. Conclusion And Discussion

Taken all the analyses together, based on the correlation coefficients derived, the acoustic parameters mentioned above are ranked as follows:

TABLE IX. THE RANKS OF THE ACOUSTIC PARAMETERS IN ACCORDANCE WITH THE CORRELATION COEFFICIENTS

Correlation ranks (stress level)	**F001**	**F002**	**M001**	**M002**	**Mean**
TVVF	0.50	0.56	0.55	0.58	**0.55**
Duartion	0.46	0.53	0.54	0.57	**0.53**
pitchMax (excluding Tone3)	0.43	0.52	0.55	0.56	**0.52**
pitchMax(all tones)	0.37	0.45	0.48	0.51	**0.45**
pitch range (excluding Tone 1)	0.12	0.35	0.50	0.55	**0.38**
pitch range (ST scale, all tones)	0.15	0.30	0.30	0.43	**0.30**
pitch range (all tones)	0.09	0.25	0.39	0.41	**0.29**
intensity	0.06	0.05	0.05	0.08	**0.06**
energy	0.05	0.04	0.02	0.06	**0.04**

As shown in the table, TVVF is most correlated with stress level while duration, upper limit of pitch range and pitch range are less correlated. The weakest correlation is between intensity and energy, and stress.

Though TVVF is not that strongly correlated with stress as indicated by the coefficient (less than 0.6), considering all data analyzed are not normalized, TVVF can be taken as a good parameter to stress.

Normalizing data can make the correlation stronger. Generally, the effects of types of syllable initials, syllable finals and prosodic location should be normalized. For a tone language, the type of syllable tones is another influencing factor and should be normalized too. Table X gives the correlation coefficients between stress level and 4 normalized global acoustic parameters-*TVVF, duration, pitchMax(all tones) and pitch range(all tones).*

TABLE X. THE CORRELATION COEFFICIENTS OF NORMALIZED DATA

Correlation ranks (stress level)	**Mean (original data)**	**Mean (normalized data)**
TVVF	0.55	0.62
Duartion	0.53	0.61
pitchMax(all tones)	0.45	0.54
pitch range (all tones)	0.29	0.37

The new result shows that the normalization processing could raise the correlation coefficients between stress level and the acoustic data, and TVVF was still the best parameter to stress.

All in all, according to the analyses of this study, TVVF is most correlated with perceived stress level. That means the stronger the stress is, the more of TVVF. It can be concluded that times of vibration of vocal folds (TVVF) is a sensitive parameter to stress in Mandarin Chinese.

It should be pointed out that the study is based on a Mandarin corpus but the same conclusion applies for other languages in theory. In the future, more work will be carried out with other language corpus to check the validity of the conclusion.

Acknowledgment

This work has been supported by Phonetic Laboratory, Institute of Linguistics, Chinese Academy of Social Sciences.

References

[1] Ye jun, "The Research of Rhythm of Modern Chinese", Shanghai Century Press, 40-49, 2008

[2] Yin Zhigang, "The Research of Rhythm of Standard Chinese", 61-71, PhD Academic Dissertation of CASS,2011

[3] Cao Jianfen, The relationship of Tone and Intonation, China Language, (3) 195-202, 2002

[4] Lin Maocan, Yan Jingzhu, Sun Guohua, The Stress Pattern of Foot in Beijing Dialect, Dialect (1), 1984

[5] Zhong Xiaobo, et al., The Stress Perception of Standard Chinese, Psychological Sinica, (6)vol31, 2001

[6] Wang Yunjia, Chu Min, He Lin, A preliminary study of Focus Stress and semantic Stress in Chinese, Chinese Courses for Foreigners, (2)86-98,2006

[7] Cao Wen, The Prosody of Chinese Focus, 5-10, Beijing Language Press, 2010

[8] Cai Lianhong, Wu Zongji, Tao Jianhua, The Computability Research of Chinese Prosody Characters, The Modern Phonetics of New Century, Tsinghua University Press, 2001

[9] Xu Jieping, Chuming, He Lin, Lv Shinan, The Infection of Utterance Stress to Pitch and Duration, Acoustics Journal, Vol25(4),2000

[10] Mei Xiao, Analysis of Duration of Mandarin Prosodic Structures, Journal of Chinese Information Processing, Vol.24, No.4, Jul 2010

Authors Profile

PhD Yin Zhigang is an assistant professor of Phonetic Laboratory, Institute of Linguistics, Chinese Academy of Social Sciences. His research interests include Phonetics, Phonology, Speech Database, Multi-modal, EEG/ERP, Eye-tracking research and Semantics.