Contrastive study on prosodic aspects for Standard and regional accented Chinese

*Juanwen Chen, Aijun Li‡, Xia Wang**

School of foreign languages, Zhejiang University
ˋInstitute of Linguistics, Chinese Academy of Social Sciences
* Nokia Research Center, Beijing
juanwenchen@163.com, ‡Liaj@cass.org.cn, *Xia.S.wang@nokia.com

Abstract

It is commonly held that there exists some prosodic differences between standard Chinese and accented Mandarin. In this paper, a perceptual test was first made by filtering the signal above 500Hz to conceal the segmental information and make utterances meaningless. It is proven that people could identify Standard Chinese (SC) and Shanghai-accented Mandarin (ASH) only by the prosodic features. Then, we studied the word stress of disyllables in citation form and found the word stress distribution and its acoustic features are different between SC and ASH.

1. Introduction

In China there are 9 dialectic areas and these regional variants are different from each other to some extent. To make them mutually intelligible, most Chinese speak one of the Guan dialect, which is widely used all over China to make communication possible. This Guan dialect is Standard Chinese (Putonghua, hereafter referred to as SC or Mandarin). However, when one in the dialectic areas speaks Mandarin, the segmental and prosodic features in his dialect are inevitably brought in subconsciously. It may cause some difficulties in second language learning and pronunciation modeling for Automatic Speech Recognition (ASR). [1] How to deal with and tackle the accent issue promotes us to investigate the different aspects lied in SC and regional accented Mandarin. Shanghai-accented Mandarin (hereafter referred to as ASH) becomes more and more interesting to researchers because of the economical and political importance of Shanghai.

Because Shanghainese may bring the dialectical features into Mandarin they speak to a different degree, Shanghai Mandarin is classified into three categories as light, middle and heavy according to the accent. The criteria include subjective one from dialectologist and objective one from statistical results obtained from the annotation. [1] For the significance of study, ASH with middle-degree accent is the most frequently appeared and so is the one mostly worthy of study. Therefore, Twenty Shanghainese joining this study belong to this category. At the same time, ten Beijingese also read the same material as Shanghainese did to make a contrastive study.

There are many studies accounting for dialectal prosody such as SweDia 2000 project aiming at capture the tonal identity of a dialect (or a dialect type) (http://www.swedia.nu). Our previous study has shown that the rising tone in ASH is like the low tone in SC according to the perception though it is still a rising tone phonologically, the low tone has a more obvious falling-rising part in SC than in ASH, and the falling tone in ASH cannot falls as low as those in SC. In this paper, we first used a perception experiment to prove that only by the prosodic features could people identify SC and ASH. Then we studied on the distribution and acoustic features of disyllabic word stress in citation form. The results reveal that there exists significant difference caused by regional factor.

2. Perception Experiment

When studying the prosodic factors in speech reorganization, people always employed the perception experiments to confirm the importance of the prosody. These experiments showed that the variety of languages and dialects could be distinguished only according to the information revealed by pitch. [4] The linguistic background of the listeners was also found to affect performance in the experiments. [5]

In our study, we employed the perception experiment to confirm the prosodic features indeed play an important role in identifying SC and ASH. The results may, on one hand, tell the prosodic features can successfully identify SC and ASH no matter it is read speech or spontaneous speech. On the other hand, the error rate can tell which one of SC and ASH may depend on more prosodic features in identification.

2.1. Material and Method

In the perception experiment, 30 sentences with different length were selected from SPEECON speech corpus (http://www.speecon.com), including 10 read utterances and 20 spontaneous utterances. These utterances were filtered above 500Hz in the Praat software [8]. In this way, the segmental information was concealed and so only the prosodic information was remained. The subjects should judge whether the speaker was a Shanghainese or a Beijingese according to the filtered sound they heard. In this experiment, we presented the subjects with the text of the 30 sentences to avoid the misguidance caused by tone. For example, the rising tone in ASH is somewhat like the low tone in SC as we have studied before. Therefore, if without the help of the text, the subjects might take the rising tone in ASH as the low tone in SC and could destroy the validity of this experiment. The speech samples were randomized to give objective judgment.

2.2. Result

Eight subjects participated in this experiment. They are young naïve listeners who speak Standard Chinese fluently and have good hearing. They were exposed both to SC and to ASH so they obtained the needed linguistic background. Each of the randomized speech samples was judged by force as those of SC and ASH. Listeners themselves decided how many times each sample was listened. The result from this listening test is shown as follows:

Table 1: *the results of perception experiment*

Perception	Result
Correct rate of identifying SC	73.96%
Correct rate of identifying ASH	73.81%
Error rate of taking SC as ASH	25.00%
Error rate of taking ASH as SC	27.38%
Error rate of perception in spontaneous speech	20.83%
Error rate of perception in read speech	36.67%
Error rate of perception in short sentences	36.11%
Error rate of perception in long sentences	14.10%

The result shows that only by the prosodic features could people identify SC and ASH because the correct rates of identifying SC and ASH were 73.96% and 73.81% respectively. Error rate of taking SC as ASH and taking ASH as SC are 25.00% and 27.38% respectively, which shows the prosodic features play similar roles in identification. Error rate in spontaneous speech and in read speech are 20.83% and 36.67% respectively, which shows the prosodic features might have a relative important role in spontaneous speech than in read speech. Error rate in short sentences and in long sentences are 36.11% and 14.10% respectively, which shows the prosodic features are much more important in the long sentences than short sentences. In conclusion, the prosodic features play somewhat different role in identifying SC and ASH. It is the extrinsic difference and we tended to find the intrinsic difference laying in the acoustic features, such as the duration, the pitch contour in citation form and in sentences, in our contrastive study.

3. Stress distribution and its acoustic features

Chinese is a stress-free language, i.e. its stress is variable except for words necessarily with neutral tone. The term "word-stress" will be used here to refer to those syllables that would be perceived as stressed in word of citation form or word in sentences. As for the phonetic features, the major acoustic correlates of the relative stress of a syllable are its pitch, its duration and its intensity. When syllables are stressed, they are generally louder, the F0 peak is higher in the case of Tone 1, 2 and 4 or the F0 trough is lower in the case of Tone 3, and their duration greater, especially in case of Tone 3, in relation to their non-prominent correlates. [3]

The study on the word stress [6] argued that the second syllables are stressed more heavily than the first syllables in the isolated disyllabic words. In the prosodic unit, the final syllable is primarily stressed, the initial one is secondly stressed and others are weaker than the initial and the final ones. [2] Concerned with the sentential stress, it is concluded that the high tone makes a syllable to have the largest chance to obtain a sentence stress, and the low tone makes the possibility least, and the effect of the rising tone and the falling tone are in between. [7]

3.1. **Material**

As a tone language, different tones, with different pitch patterns and duration, play a very important role in the realization of stress in Chinese. Moreover, the word constituent is also a factor in influencing stress. Therefore, we studied the different representations of stress in SC and ASH with different word constituents according to different tones.

In this study, we used 2221 disyllabic words in ASH and 1203 in SC. These disyllables cover different tonal combinations, intersyllabic junctures as well as word constituent, each of which includes a word stress perceived by force. Four young people annotating for a long time annotated the stress. The consistence is 73%, which shows a high reliability. Our study was done according to the information getting from the annotation.

3.2. Word stress distribution

When analyzing, we used initial-stressed/final-stressed syllables to stand for the first/second syllables are stressed. The percentages of initial-stressed syllables and final-stressed syllables in SC and ASH are computed according to the different tones.

The result shows that the descending sequence of occurrence frequency for the stressed initial syllables with the four tones in SC are the falling tone, the high tone, the rising tone and the low tone. ASH has the same distribution in this case, but the stressed falling tones are more than the stressed high tone compared with SC. As for the stressed final syllables,the descending sequence is the high tone, the falling, the rising tone and the low tone in SC, but the falling tone, the high tone, the rising tone and the low tone in ASH. Moreover, there is also some difference in the position of the stressed syllables. All the ratios of the pre-/final-stressed syllables in four tones in SC are higher than those in ASH. It shows

Table 2: *Distribution of word stress with different tones in SC*

Distribution Tone	Occurrence frequency of initial-stressed syllables	Occurrence frequency of final-stressed syllables	Ratio of pre-/final-stressed
High	68.00%	48.77%	1.39
Rising	61.35%	38.76%	1.58
Low	40.59%	16.67%	2.44
Falling	70.41%	45.82%	1.54
Total	61.10%	38.90%	1.57

Table 3: *Distribution of word stress with different tones in ASH*

Distribution Tone	Occurrence frequency of initial-stressed syllables	Occurrence frequency of final-stressed syllables	Ratio of pre-/final-stressed
High	59.55%	49.71%	1.20
Rising	56.35%	42.70%	1.32
Low	32.13%	17.99%	1.79
Falling	73.63%	57.06%	1.29
Total	56.51%	43.49%	1.30

Shanghainese have a tendency to put stress on the latter syllables compared to Beijingese. Shanghainese need to prepare to pronounce ASH consciously or subconsciously. Since their consciousness on the latter syllable is stronger than Beijingnese, Shanghainese shows a tendency to put stress on the final syllables compared to Beijingese.

3.3. Acoustic features

For the acoustic parameters on stress, we studied the pitch and duration of the stressed syllables. Figure 1 and 2 are the pitch contrast of word stress for females and males in SC and ASH, which is made based on the maximum, minimum and mean of average pitch values (in semi tone). The referent frequency is 64.44 Hz. As shown from the result, the pitch registers of the stressed syllables with four tones in ASH are higher than those in SC on the whole. To analyze the pitch range, we chose the maximum pitch value in the falling tone and the minimum pitch value in the low tone because this range can represent the pitch range of a speaker. In this way, the pitch range for females and males are 14.039St and 12.087St respectively in SC; and 11.098St and 10.932St in ASH. It shows the pitch range in SC is wider than that in ASH both for females and males although there is larger difference in females.

ANOVA test was used to find the factors that caused the difference. The result tells us that the acoustic parameter of F0max, F0min and F0mean is significantly different by the regional factor (P=0.0).

For the duration, we categorized the duration and compared them. The distribution of duration for stressed syllables was analyzed based on the different tones as shown in Figure 3-6. The areas with centralized distribution of duration in ASH precede to that in SC in the high tone, the rising tone and the falling tone. That is to say, the duration of syllables with these tones in ASH is shorter than that in SC. However, an opposite representation occurs in the low tone.It tells the factor of duration plays a more important role in realizing stress in the low tone in ASH. Since there is no falling-rising part in the high tone, rising tone and falling tone, Shanghainese tend to change the pitch to realize stress when speaking Mandarin and are not likely to lengthen the duration. For the low tone, for Shanghainese cannot fulfill the falling-rising pitch pattern as Beijingese can, they tend to lengthen the duration of low tone to realize stress. Due to the influence of the Shanghai dialect, the realization of stress is different between SC and ASH. ANOVA test also shows the regional factor influences the duration of stressed syllable significantly (P=0.0).

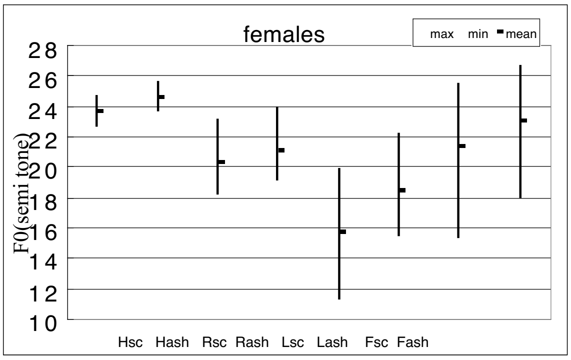

Figure 1: *Pitch of word stress for females in SC and ASH*

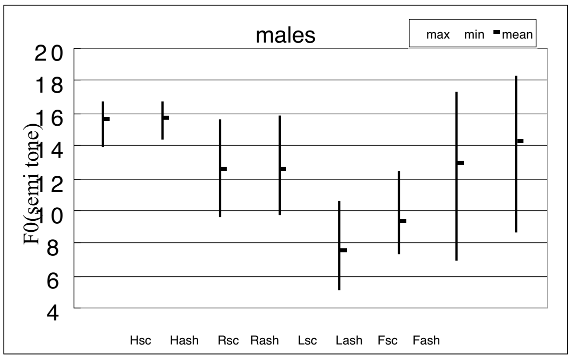

Figure 2: *Pitch of word stress for males in SC and ASH*

Note: H stands for the high tone, R for the rising tone, L for the low tone and F for the falling tone. Top, middle and bottom points of each bar stand for the maximum, mean and minimum of average pitch values.

Concluded from the acoustic representation of pitch and duration, the factor of duration plays a more important role in realizing stress in the low tone in ASH. For the low tone, the Shanghainese cannot fulfill the falling-rising pitch pattern as Beijingese can, they tend to lengthen the duration of low tone to realize stress. Influenced by the Shanghai dialect, the realization of stress in pitch and duration is different between ASH and SC.

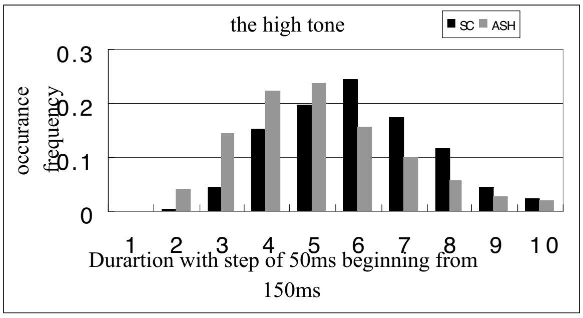

Figure 3:*Distribution of duration for the stressed high tones in SC and ASH*

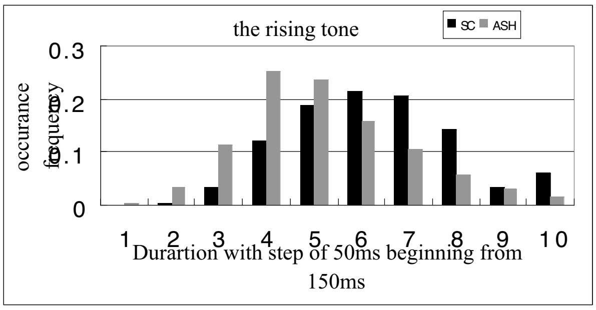

Figure 4:*Distribution of duration for the stressed rising tones in SC and ASH*

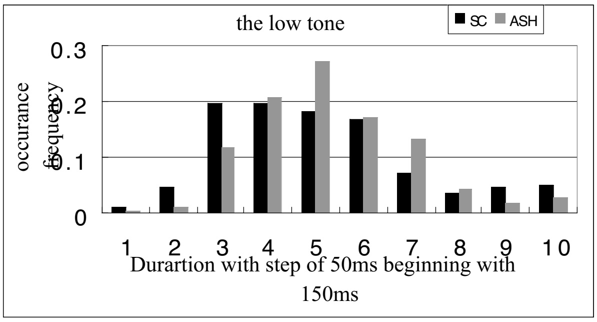

Figure 5: *Distribution of duration for the stressed low tones in SC and ASH*

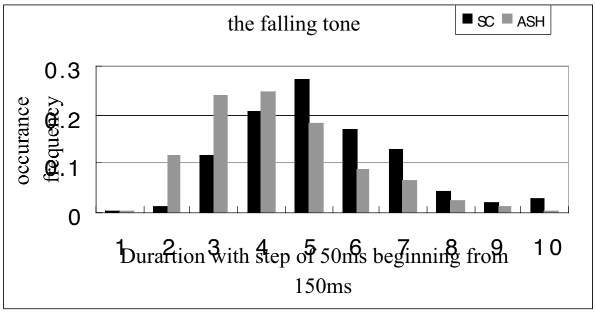

Figure 6: *Distribution of duration for the stressed falling tones in SC and ASH*

4. Conclusion

The perception experiment proves that SC and ASH can be identified based on the prosodic information, i.e. there are prosodic difference between SC and ASH. It seems that the stress pattern or stress/rythmic structure for ASH is significantly different from SC.

One of the evidence comes from the contrastive analysis of the word stress with different tones. As for the word stress, the descending sequence of occurrence frequency for the stressed initial syllables with the four tones in SC are the falling tone, the high tone, the rising tone and the low tone. ASH has the same distribution in this case, but the stressed falling tones are more than the stressed high tone compared with SC. As for the stressed final syllables, the descending sequence is the high tone, the falling, the rising tone and the low tone in SC, but the falling tone, the high tone, the rising tone and the low tone in ASH. There is also some difference in the position of the stressed syllables. Shanghainese have a tendency to put stress on the latter syllables compared to Beijingese.

As for the acoustic cues on word stress, the pitch registers of the stressed syllables with four tones in ASH are higher than those in SC on the whole and the pitch range in SC is wider than that in ASH both for females and males although there is small difference in females. The duration of stressed syllables with the high tone, the rising tone and the falling tone in ASH is shorter than that in SC, but an opposite representation occurs in the low tone.

5. References

[1] Li, A., 2003. A Constastive Investigation of Standard Mandarin and Accented Mandarin. *Eurospeech* 2003.

[2] Chao, Y.R., 1997, *A Grammar of Spoken Chinese.* being translated by S. Lu, Beijing: Commercial Press, 1979.

[3] Hirst, D.J., Di Cristo, A., 1998. *Intonation Systems*. Cambridge University Press, 1998.

[4] Peters, J., Gilles, P., Auer, P., Selting, M., 2003. Identifying regional varieties by pitch information: A comparison of two approaches. *ICPhS* 2003.

[5] van Leyden, K., van Heuven, V.J., 2003. Prosody versus segments in the identification of Orkney and Shetland dialects. *ICPhS* 2003.

[6] Lin, M.C., Yan, J.Z., Sun, G.H., 1984. A Primary Experiment on the Stress Pattern of Normal Disyllabic Words in Mandarin. *Dialect*, 1984.

[7] Wang, Y., Chu, M., He, L., 2003. Location of Sentence Stresses within Disyllabic Words in Mandarin. *ICPhS 2003*.

[8] http://www.praat.org