A CONTRASTIVE INVESTIGATION OF STANDARD MANDARIN AND ACCENTED MANDARIN[*]

*Aijun Li   * Xia Wang*

Institute of Linguistics, CASS   * Nokia Research Center, Beijing
Liaj@linguisitics.cass.net.cn   *Xia.S.wang@nokia.com

ABSTRACT

Segmental and supra-segmental acoustic features between standard and Shanghai-accented Mandarin were analyzed in the paper. The Shanghai Accented Mandarin was first classified into three categories as light, middle and heavy, by statistical method and dialectologist with subjective criteria. Investigation to initials, finals and tones were then carried out. The results show that Shanghainese always mispronounce or modify some sorts of phonemes of initials and finials. The heavier the accent is, the more frequently the mispronunciation occurs. Initials present more modifications than finals. Nine vowels are also compared phonetically for 10 Standard Chinese speakers and 10 Shanghai speakers with middle-class accent. Additionally, retroflexed finals occur more than 10 times in Standard Chinese. No significant difference exists on durations of initials and finals for these 20 speakers. And no phonological difference is found on four lexical tones. It seems that the prosodic difference is mainly on rhythmic or stress pattern.

1. INTRODUCTION

Spoken Chinese comprises many regional varieties, called dialects. There are 9 dialectic areas in China: Guan, Jin, Wu, Hui, Xiang, Gan, Kejia, Yue and Min. Guan (Mandarin) was referred to as a common language which covers a very large regional area from north east to south west of China, with over 800 million speakers. Most Chinese speak one of the Guan (Mandarin) dialects, which are largely mutually intelligible. The dialect spoken in Beijing constitutes the base for Standard Chinese. It forms the basis both of the modern written vernacular, Baihua, which supplanted classical Chinese in the schools after 1917, and of the official spoken language, Putonghua, prescribed in 1956 for nationwide use in schools. Nowadays Standard Chinese (Putonghua, hereafter referred to as SC or Mandarin) is widely used all over China on almost every activity from broadcast news to commercial trades. [8]

People from different dialectal areas might not be able to communicate with each other simply because the differences among the dialects are so significant. Mandarin, or Putonghua, would be a good choice as a sharing basis. Most people in China are bilingual Chinese speakers, i.e. native dialect and Mandarin. Although lots of people CAN speak Mandarin, they speak it with different accents, depending on how well they grasp the language. The Mandarin they speak is always affected by their native dialects phonetically, lexically and syntactically.

Wu dialect is a group of dialects spoken in ShangHai, ZheJiang, southern JiangSu, and part of FuJian and AnHui. Wu dialect has about 70 million speakers, which makes it the second biggest dialect running after Mandarin. The dialect of interest in this paper is Shanghainese, the native dialect spoken in Shanghai covering more than 11,850,000 populations. Although it is rather young in Wu dialect family, Shanghainese becomes more and more interesting to researchers because of its economical and political importance.

In this paper we will mainly focus our contrastive study on segmental and supra-segmental acoustic features between standard and Shanghai-accented Mandarin (ASH). Accent of Shanghai Mandarin is classified into three categories as light, middle and heavy by subjective criteria from dialectologist and objective criteria from statistical results obtained from the annotation. Then the phonetic level analysis on Shanghai accent is made for the three accent categories. A phonetic articulation contrastive table for Standard and Shanghai accent Mandarin will be generated for initials and finals with occurrence frequency for each substitutional pronunciation in Shanghai accented Mandarin. Occurrence times of retroflexed finals and neutral tone syllable are also analyzed for SC and ASH. Formant values F1-F2 for each vowel, durations of initials and finals, and pitch patterns for four lexical tones are measured and investigated statistically to study both micro-segmental features on phonemic level and supra-segmental prosodic features on lexical level.

Dialectal differences are widely investigated for dialect identification, language (L2) learning and pronunciation modeling for Automatic Speech Recognition (ASR) [1-7]. Especially in Chinese ASR systems, how to deal with and tackle the accent issue is a big challenge due to the variability of the language. We hope the contrastive study from phonetic point of view on regional accented Mandarin will shed a light on the Chinese ASR framework.

2. MATERIAL

The database used in this study is the SpeeCon Mandarin Chinese speech database collected by Nokia in the framework of SpeeCon [10]. It covers spoken Chinese in four regional accents. The speech of 50 Beijing speakers and 51 Shanghai speakers recorded in office environment is selected for contrastive study. Each speaker has 321 utterances including phonetically rich words, phonetically rich sentences, application commands, proper names, numbers, time expressions and spontaneous speech. All the recorded utterances were phonetically annotated on orthographic and pronunciation tiers. Pronunciation variables or phonemic changes caused by dialects were annotated dedicatedly.

3. ACCENT CATEGORY OF ASH

3.1. The phonology of SC and SHD

There are 54 finals, 34 initials and 5 lexical tones in Shanghainese [12], 21 initials, 38 finals and 4 lexical tones (excluding neutral tone) in SC [13]. Compared with SC, Shanghainese has checked tone, velar nasal [Ð], voiced stop [b d g], voiced affricate [d¯], voiced fricative [v, ¯,z,ä] and two more nasal initials [Ð,Â] besides [m,n]. But Shanghainese does not have retroflex initials [t©, t©H,©] and voiced fricative [¸] as in SC.

3.2. Accent category

Accent of ASH is categorized into three levels and each level can then be further divided into two classes [15]. We first calculated the frequencies of pronunciation variability caused by dialect for all 51 Shanghai speakers. Then clustered them into three categories as light, middle and heavy, and send them to a dialectologist to make judgments of the accents objectively. Finally we got 25, 19 and 7 speakers as light, middle and heavy accent respectively.

We also made correlative analysis on accent and age, accent and education background for 51 speakers and found that no significant correlative relation exits between accent and age ($r^2=0.23$), but high correlative relation exits between accent and education background ($r^2=0.86$). The lower the education is, the heavier the accent is.

3.3. Extended phonetic annotation

Extended phonetic annotation was made for initials and finals of each syllable with time alignment. Prosodic and stress structure was annotated by C-ToBI [14]. Those extended annotations were only made for 10 SC speakers (5 female and 5 male speakers selected from 50 SC speakers) and 10 ASH speakers (2 male and 8 female speakers selected from 19 middle-class accent speakers). The annotation software used is praat [11].

4. NEUTRAL TONE AND RETROFLEXED FINALS

Neutral tone and retroflexed finals are two characteristics differentiating SC from many other dialects. Some function words must be neutralized in running speech in SC. But many syllables are unnecessarily lightly read or unstressed with the same phonetic features as neutral tone syllables. Some light tone syllables can distinguish lexical meanings, Such as, "地道 [ti4 tao4]" means "tunnel" with normal stress pattern, and means "purely" with the second syllable neutralized.

The SC speakers also would like to make some finals retroflexed in many words, such as "油饼[iou2 piÐ3]" with [iou2 piÈ‹ !‹ r3]. But ASH speakers seldom do in this way.

Table 1 shows the distribution of neutral tone and retroflexed finals for ASH and SC. The average number of retroflexed finals is about 11 for each SC speaker, while it is only 1.4 for each light accent ASH speaker. The neutral tone syllables also occur less in ASH speakers. But the number becomes even less for heavier accent group. It can also be concluded from the table that ASH speakers seem to produce neutral tone syllables better than retroflexed syllables.

Table 1: Distribution of neutral tone and retroflexed finals

	accent	Neutral tone	Ave./spk	Retroflexed Finals	Ave./spk
SC	SC	3917	78.34	543	10.86
AS H	light	1623	61.92	38	0.94
	middle	1087		3	
	heavy	386		6	

5. CONTRASTIVE STUDY ON SEGMENTS

5.1. Initials and finals

Table 2 is the examples of the correct pronunciation rate of initials and finals and the first alternative pronunciations of 51 ASH speakers for three accent groups. Symbol like [t©H+o] stands for an aspirated affricate initial [t©H] followed by a final beginning with vowel [o]. The statistic results were analyzed and the following observations were obtained:

A) ASH speakers can not distinguish retroflex initial [t©, t©H, ©] with plain initial [ts,tsH,s] for they do not have retroflex initial [t©, t©H, ©] in Shanghainese. But this kind of mix-pronunciation is unsymmetrical. For the heavy accent group, most retroflex initials are uttered as plain ones. For an instant, 80.5% of [©+È] is pronounced as [s+È], while only 10% [s+È] is pronounced as [©+È]. In middle and light accent groups, the figure is 37% and 1.5% for the former case and 30% and 3% for the later case. So the mix-pronunciation of initials of these two groups reduces with accent reduced. A great decrease can be found from the statistics of the heavy accent group to that of the middle accent group.

B) About 50% of [Ä] is pronounced as [QÄ] for heavy accent and 30% for middle accent group. Monophthong [y] or onset with [y] is often replaced with corresponding [i] vowel.

C) For light accent group, the mispronunciation exists mainly for finals: some finals with [Ð] and [n] codas are always mispronounced. For example, 30-40 % of [ÈÐ] and [iÐ] is pronounced as [Èn] and [in]. Additionally, the speakers with light accent can distinguish retroflex and plain initials very well. The correct rate is more than 90%. So with the accent decreases, the speakers can improve the pronunciation of retroflex initials better than that of finals with [Ð] or [n] coda.

Therefore, the mispronunciation for ASH comes from two sources, firstly from initials, then from finals.

5.2. Acoustic features of vowels

There are 9 vowels in SC: [a,o,È,i.u,y, ¡,Ÿ,Ä]. The acoustic features of them have been widely studied [9,13]. Here we will compare these vowels' spectral features for 10 SC and 10 ASH speakers. For the limited material of 20 speakers, we cannot get isolated vowels from isolated monophthongs. But we can guarantee that the vowels are from the same contexts as shown in table 3. Fig. 1 is the vowel chart drawn by the average values of F1 and F2 for SC and ASH speakers. The ellipses are not drawn with the radii of standard deviations along the two principal components of each vowel clusters. They are plotted to represent the contrastive vowels of two groups by the author.

Table 2: Correct pronunciation rate of initials and finals in 3 level accents

Initial /Final(SC)	First alternation	Heavy	Mid	Light
t©H+o	tsH+o	0.0000	0.288	0.945
©+a	s+a	0.0349	0.468	0.947
t©H+a	tsH+a	0.0357	0.406	0.914
t©+a	ts+a	0.0385	0.275	0.867
t©H+È	tsH+È	0.0641	0.452	0.934
t©+o	ts+o	0.0761	0.254	0.916
t©+u	ts+u	0.0784	0.489	0.965
t©+ Ÿ	ts+¡	0.0811	0.446	0.981
t©+È	ts+È	0.1083	0.470	0.942
t©H+ Ÿ	tsH+¡	0.1111	0.525	0.100
©+o	s+o	0.1163	0.495	0.983
©+u	s+u	0.1183	0.485	0.926
t©H+u	tsH+u	0.1636	0.363	0.927
¸+a	l+a	0.1739	0.50	0.921
©+È	s+È	0.2018	0.632	0.985
©+ Ÿ	s+¡	0.2049	0.687	0.992
¸+u	l+u	0.2800	0.59	0.927
¸+È	l+È	0.4151	0.656	0.995
uÈn	uÈÐ	0.4848	0.841	0.984
in	iÐ	0.5207	0.553	0.701
tsH+u	tsH +È	0.5429	0.610	0.864
in	iÐ	0.5449	0.678	0.706
Ä	QÄ	0.5736	0.763	0.991
¸+o	l+o	0.5833	0.784	0.98
-Èn	Èng	0.7098	0.851	0.967
n+u	l+u	0.7500	0.818	0.1
¸+ Ÿ	l+È	0.7838	0.967	0.129
uÈn	uÈÐ	0.7857	0.782	0.983
n+È	l+È	0.8125	0.950	0.100
m+o	m+u	0.8214	0.891	0.100
iE	i	0.8267	0.952	0.998
m+u	m+o	0.8333	0.818	0.983
¯iÐ	in	0.8613	0.946	0.979
iÐ	in	0.8667	0.966	0.966
n+o	l+o	0.8667	1.00	0.100
¯uaÐ	aÐ	0.8667	0.982	0.998
l+y	l+i	0.8710	0.972	0.989
yn	iÐ, in	0.8750	1.00	0.991
pH+u	f+u	0.8750	0.889	0.1
f+u	f+o	0.8772	0.943	0.1
s+u	s+È	0.8861	0.776	0.978
s+È	©+È	0.8966	0.705	0.968
¯y	yÈ	0.8977	0.959	0.997
-ÈÐ	Èn	0.9000	0.677	0.665

Table 3: Context for vowels

a	o	e	i	u	y	¡		Ä
pa 八	po 播	t©È 者	i 一	u 五	y 于	s¡ 四	© 十	Ä 二

What we can conclude from the figure 1 is as follows:

A) [a,u,y,¡] has no difference between SC and ASH.

B) [i] is more front and close in ASH.

C) [o] is approaching to [u] in ASH. (it is not a genuine monophtong in syllable [po], it should be a diphthong as [uo])

D) Retroflex [Ä] is treated as a diphthong [ÈÄ] by some phoneticians [9]. And it is characterized with a lower F3 and a rapid downdrift of F3. However, [Ä] in ASH is a more front vowel approaching [QÄ].

E) [Ÿ] is more back in ASH closing to [}].

6. PROSODIC ASPECTS

6.1. Duration of initials and finals

Initials are classified into 8 classes according to their method of articulation and finals into 3 classes as monophthong, diphthong and triphthong. Table 4 and 5 show the duration of initials and finals for 10 SC and 10 ASH speakers based on all words with one to four syllables. Statistic analysis was made for duration contrast between SC and ASH:

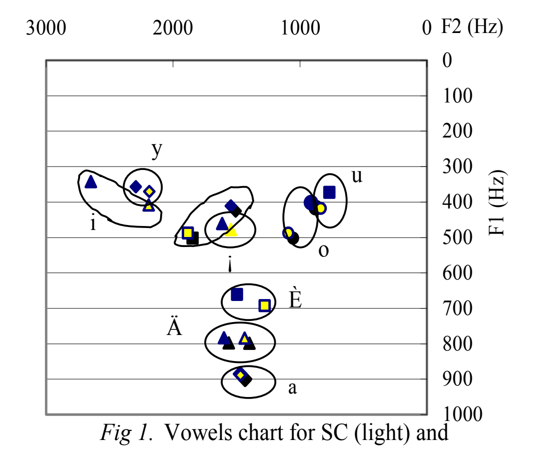

Fig 1. Vowels chart for SC (light) and ASH(Dark)

A) Pearson correlation analysis was made for duration of each speaker within two groups respectively. No significant difference exists between any two speakers in each group for their initials and finals durations (0.01 level, 2-tailed), i.e. speakers agree very well with their initials and finals durations.

B) T-Test analysis shows that for SC speakers, durations of class 3 and 5, class 6, 7 and 8 have no significant difference (p>0.01). Only 6 and 7 has no significant difference on 0.05 levels. For ASH speakers, initial durations of class 3 and 5, class 4, 6, 7 and 8 have no significant difference (p>0.01).

C) T-Test analysis for initial duration of SC and ASH indicates that initial durations of SC and ASH are two independent variants (p>0.05). All initials have no significant difference except [¸](p>0.05). [¸] and [l] has significant difference on 0.2 level, i.e. p=80%. But stops, aspirated stops and aspirated affricates of SC are longer than those of ASH; others are shorter than those of ASH.

D) T-Test analysis also shows that durations of finals between SH and ASH have no significant difference for all 3 classes (P>0.05). But the finals of ASH are longer than the corresponding ones of SC.

Table 4: Duration of initials (second)

Initial class	SC		ASH	
	Ave	Stdev	Ave	Stdev
1-ptk	0.0210	0.0034	0.0170	0.0045
2-pHtHkH	0.0915	0.0109	0.0831	0.0176
3-fs©»x	0.1283	0.0149	0.1376	0.0269
4-ts s t»	0.0670	0.0096	0.0680	0.0159
5-tsH t©Ht»H	0.1381	0.0159	0.1354	0.0204
6-mn	0.0506	0.0098	0.0611	0.0148
7-l	0.0473	0.0106	0.0605	0.0139
8-¸	0.0392	0.0091	0.0712	0.0206

Table 5: Duration of finals (second)

finals	SC		ASH	
	Ave	Stdev	Ave	Stdev
monophthong	0.1568	0.0190	0.1800	0.0235
diphthong	0.1760	0.0183	0.2033	0.0255
triphthong	0.1988	0.0177	0.2242	0.0264

6.2. Lexical tones

Four lexical tones of SC are H-H, L-H, L-L and H-L. Isolated syllables of 10 SC speakers and 10 ASH speakers are used to get the F0 contours of four lexical tones. All F0 values (in SemiTone) are normalized according to each speaker's pitch range, and the duration of each syllable is also normalized.

Fig 2 and 3 are the F0 curves of four lexical tones for SC and ASH speakers respectively. It shows that there is no phonological difference between lexical tones of SC and ASH. The tonal range is about 0.4 for two groups. But the tonal register of ASH is lower about 0.1. The F0 values are disturbed by different initials. Zero-initial and lateral initial has lower F0 onset values, (values are limited for nasal initial). Aspirated affricates have higher F0 values than other initials.

6.3. Speech rate and prosodic units

No significant difference has been found for speech rate, durations of prosodic word and minor prosodic phrase, and syllable number of prosodic units, except major prosodic phrase (T-test, P>0.05).

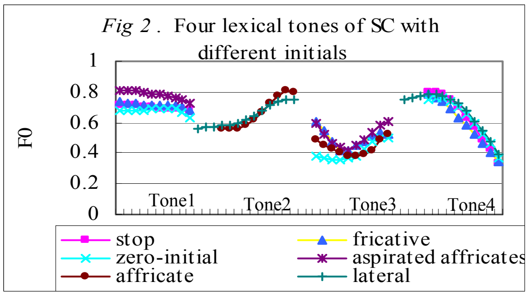
Fig 2 . Four lexical tones of SC with different initials

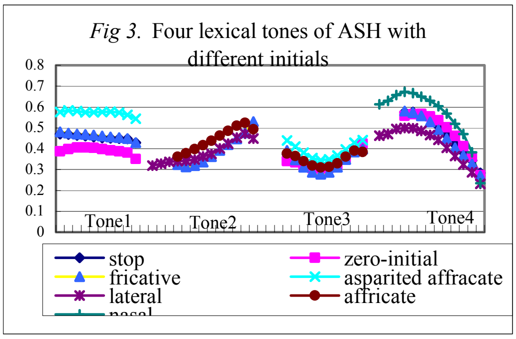
Fig 3. Four lexical tones of ASH with different initials

7. SUMMARY AND DISCUSSION

In this paper, we showed the preliminary results on the contrastive study of standard and Shanghai accented Mandarin. Mix-pronunciation of retroflex initials, and alveolar and velar coda finals was observed in ASH. Fewer retroflexed finals were detected as well in running speech. But no phonological difference was observed for lexical tones and durations of initials and finals. It seems that the stress pattern or stress structure for ASH is significantly different from SC. One of the evidence is that the frequency of neutral tone or light tone occurrences for SC is much greater than that for ASH. Other evidence comes from the analysis of the tonal pattern and stress structure of prosodic words, which was not presented in this paper due to the space limit.

In spontaneous speech of ASH, the speech rate is faster than in read speech. And differences are also found in lexical, syntactic and intonational aspects for Shanghainese. In addition to the sound variability, such as deletion, insertion and voicing, phonemic changes or modifications are also detected and they are correlated with the lexical frequency Further researches will be carried out, based on carefully designed materials, with more considerations on spontaneous speech.

ACKNOWLEDGEMENTS

Many thanks to Fang Wu for his suggestion and discussion, and Zhigang Yin and Qiang Fang for their help in data processing, and to Jue Yu and Juanwen Chen for their help in annotation checking.

REFERENCES

[1] Tsukada, K., “An Acoustics Comparison Between American English and Australian English Vowels”, *ICSLP’2002*, 2257-2260.

[2] Pellegrino, F. and Barkat M., “Investigating Dialectal Differences via Vowel System Modeling: Application to Arabic”, *ICPhS99,* 145-148.

[3] Goreman, C.G., “Diallect Identification From Prosodic Cues”, *ICPhS99*, 1237-1240.

[4] Peters, J., “The Timing of Nuclear High Accent in German Dialects”, *ICPhS99*, 1877-1880.

[5] Burger, S. and Oppernann D., “Regional Variants of German: Categories of Pronunciation Deviation From Standard German”, *ICPhS99*, 1589-1592.

[6] Petek, B., “Identification of Regional Variants in the Standard Slovenian Speech”, *ICPhS99*, 1681-1684.

[7] Gronnum, N., “Rhythm in Regional Variants of Standard Danish”, *Working Papers of Lund UNIV.* Dept. of Ling. 1993, 41, 20-23.

[8] Hou, J. Y., “ *The Outline of Modern Chinese Dialects*”, Shanghai Education Publishing House, 2002.

[9] Zee, Z., “The Phonetic Value of the Vowels, Diphthongs, and Triphthongs in Beijing Mandarin”, in *the proceeding of 5th national Conference on Modern Phonetics*, p54-60, 2001,Beijing.

[10] http://www.speecon.org

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

[12] Xu, B. H. and Tao, H., “*Dictionary of Shanghai Dialect*”, Jiangsu Education Publishing House.

[13] Wu, Z.J., Lin. M.C., et al. “*Outline of Experimental Phonetics*”, Beijing high education publishing house.

[14] Li, A.J., Chinese Prosody and Prosodic Labeling of Spontaneous Speech, *Prosody Speech 2002*, AIX-EN-PROVENCE France, 2002.

[15] National education ministry, “*Putonghua Shuipingceshi Dagang*”.

[*] EUROSPEECH2003, Sept. 2003 Geneva