INTERSPEECH 2020
October 25–29, 2020, Shanghai, China

Tone variations in regionally accented Mandarin

Yanping Li, Catherine T. Best, Michael D. Tyler, Denis Burnham

The MARCS Institute, Western Sydney University, Australia
{yanping.li,c.best,m.tyler,d.burnham}@westernsydney.edu.au

Abstract

The present study investigated tone variations in regionally accented Mandarin (i.e., Standard Mandarin [SM] spoken by dialectal Chinese speakers) as influenced by the varying tone systems of their native dialects. 12 female speakers, four each from Guangzhou, Shanghai and Yantai, were recruited to produce monosyllabic words in SM that included minimal contrasts among the four Mandarin lexical tones. Since SM developed from the Beijing dialect, their pronunciations were compared to the same Mandarin words produced by four Beijing female speakers. Regional Mandarin speakers successfully produced the four Mandarin lexical tones, but their productions varied from SM. Two crucial acoustic measures for Mandarin lexical tones, $F0$ (fundamental frequency) and duration values, were fitted into linear mixed-effects models on differences between regional and Beijing accents. Regional speakers had longer word duration and different $F0$ height when producing SM, resulting in variations in Mandarin lexical tones across the regional accents. These findings shed light on regional accent variations in Mandarin lexical tones and lay a foundation for deeper understanding of their impact on perception of accented Mandarin lexical tones by native (Beijing) Mandarin listeners.

Index Terms: lexical tones, regional accent variation, tone production, Mandarin, Chinese dialects

1. Introduction

There are between 7 and 10 dialects in Chinese [1-2]. Speakers of different dialects experience difficulties in understanding each other's dialects, to the extent that dialects may be mutually incomprehensible. Chinese dialects are therefore widely accepted to be different languages [3]. To ensure that people in China to understand each other across various regions, Standard (Beijing) Mandarin (SM or simply Mandarin) was taken as the official language as developed from Beijing dialect [4]. Because of phonological differences between regional dialects and Beijing dialect, the pronunciation of SM varies among regional speakers, resulting in regionally accented Mandarin.

Studies documenting regionally accented Mandarin show contrastive findings in terms of segments and lexical tones. On the one hand, there are phonological differences in regionally accented Mandarin segments. For example, Shanghai Mandarin speakers replace the Mandarin velar coda with alveolar coda because there is no [ŋ]/[n] contrast in the Shanghai dialect [5]. Mandarin lexical tones in regional accents, on the other hand, are reported to display almost no phonological differences. An investigation of Mandarin lexical tones in the Shanghai accent showed that the tone contours were not *phonologically* distinct from SM [6]. However, this claim seems at odds with the predictions of well-established second language learning theories, such as the Speech Learning Model (SLM) [7] and the Perceptual Assimilation Model (PAM) [8-9]. Both models claim that second language (L2) learners are unable to discern certain phonetic differences between segments in their first language (L1) and L2 because of their "equivalence classification" (SLM) or "perceptual assimilation" (PAM) of L2 phones to their native phonemes, resulting in inaccurate L2 pronunciation. From SLM and PAM viewpoints, we would expect Shanghai speakers' production of Mandarin tones to differ from those by native Beijing Mandarin speakers given that SM has four lexical tones (i.e., T1 has high-level pitch, T2 high-rising pitch, T3 low-dipping pitch, and T4 high-falling pitch [10]), but the five-tone inventory of the Shanghai dialect lacks a dipping tone. Since Tone 3 in SM has a *low* falling-rising contour, its pronunciation in Shanghai Mandarin is expected to be a distorted dipping tone that combines the Shanghai dialect's *high* falling and high rising tones (i.e., an overall higher dipping contour than in SM). Previous studies only subjectively (rather than statistically) compared the pitch contours of regionally accented Mandarin lexical tones with SM, simply categorizing the pitch contours as same or different. Any category-goodness differences between SM and accented tones would therefore have been ignored. But it is exactly those category-goodness details that are likely to arise in accented tone variations, and that presumably allow native Beijing listeners to detect regional accents. The failure to consider category-goodness differences between SM and regional accented tones may explain why the findings reported in [6] are inconsistent with SLM and PAM predictions.

Mandarin tone variations resulting from regional accents are expected to influence native Beijing listeners' perception because they use two complementary principles to detect speech information [11]: *phonological distinctiveness*, by which critical differences between contrasting phonetic segments (or tones) can distinguish a word from similar-sounding words or non-words (e.g., c<u>a</u>ke from c<u>o</u>ke) and *phonological constancy*, which keeps word identity intact across lexically irrelevant variations, such as regional accents. Clearly detailing the acoustic properties of tone variations in regionally accented Mandarin lexical tones will thus lay a foundation for a better understanding of how native Beijing Mandarin listeners perceive regionally accented tones, which will be examined in a separate study. The current study explored tone variations in regionally accented Mandarin tone productions.

2. Experiment

2.1. Method

2.1.1. Regional dialects

Table 1 shows the tone systems of Beijing, Guangzhou, Shanghai and Yantai dialects, characterized using Chao's tone number system [12] in which tone heights are labelled from 1

http://dx.doi.org/10.21437/Interspeech.2020-1235

to 5, with 5 as the highest pitch and 1 as the lowest. T1, T2, T3, and T4 in (Beijing) Mandarin are labelled as 55, 35, 214, and 51, respectively. When producing Mandarin lexical tones, we predicted that regional speakers would produce tone contours that are similar to those in their native tone systems, yielding systematic tone variations from SM. For example, Yantai Mandarin speakers are predicted to produce the SM falling tone using the Yantai falling tone, which starts with a lower pitch. Moreover, tone variations in regionally accented Mandarin should deviate from Beijing Mandarin to different extents. When producing the dipping tone, Yantai Mandarin speakers are likely to produce similar curves to SM speakers, because the dipping tone in Yantai dialect shares the same pitch height as SM. However, the dipping tone in Guangzhou and Shanghai accents are more likely to be different from SM. As mentioned above, SM dipping tone incorporates a falling and a rising pitch contour. Guangzhou and Shanghai Mandarin speakers are predicted to combine their own dialect's falling and rising tones, as neither dialect has a dipping tone. The pitch heights of the falling and/or the most similar rising tones in these two dialects differ from the dipping tone in Beijing Mandarin. Its curve in Guangzhou and Shanghai Mandarin is thus predicted to differ from Beijing Mandarin, resulting in larger differences from SM dipping tone than Yantai accent.

Table 1: *Tone systems of Beijing, Guangzhou, Shanghai and Yantai dialects*[a]

	Beijing	Guangzhou	Shanghai	Yantai
Level	55	55	55	55
		33		
		22		
Rising	35	25	34	
		23	23	
			12	
Dipping	214			214
Falling	51		53	31
		21		

[a]Note: This study followed the widely accepted descriptions of the tone systems of Beijing [12], Guangzhou [13], Shanghai [14] and Yantai [15].

2.1.2. *Participants*

16 female speakers, four each from Beijing (M_{age} = 24 years, *SD* = 1), Guangzhou (M_{age} = 20 years, *SD* = 1), Shanghai (M_{age} = 24 years, *SD* = 2) and Yantai (M_{age} = 21 years, *SD* = 2) participated in this research, when they were studying in Beijing at various stages of their higher education. The speakers from the Guangzhou, Shanghai, and Yantai regions acquired their regional dialects as an L1 and had been speaking Mandarin as L2 since primary school. They did not leave their hometown for more than one month at a time before studying at Beijing. All were non-musicians, defined as no more than 3 years of private lessons in any combination of instruments [16], because musical training can facilitate tone production [17].

2.1.3. *Stimulus materials and apparatus*

Four consonant-vowel syllables (i.e., *ba, di, du, gu*) were selected, which cover the most universal type of consonants (stop) and vowels (*a, i, u*) in Mandarin. Imposing the four Mandarin tones on each syllable created 16 monosyllabic Mandarin words. Daily used Chinese characters were selected.

Chinese characters were presented on an external monitor via E-prime Professional 2 running in a Dell Latitude 7280 laptop. Their productions were captured by a Shure SM57 dynamic microphone and recorded by Audacity Windows 2.3.0 running in a Dell desktop with 44.1 kHz sampling rate (32-bit) via an M-audio M-track II external sound card. The productions were high-pass filtered at 70 Hz and then manually segmented into individual utterances using Praat. These were presented in a word similarity test using E-prime Professional 2 on a Dell OptiPlex 3030 AIO desktop via AKG K272 headphones.

2.1.4. *Procedure*

Participants were tested individually in a soundproof booth at the Speech Acquisition and Intelligent Technology Lab, Beijing Language and Culture University (BLCU), Beijing, China. They completed both language and music background questionnaires before the experiment.

Participants were required to do a single-word reading task. They were first familiarized with the target Mandarin words to minimize frequency effects on production. During recording, they were instructed to read the word in Mandarin immediately after seeing its Chinese character on the screen, and to produce it as faithfully as possible. Repetition and correction were encouraged whenever they felt a production had been non-ideal. There were 10 blocks, each including the 16 Mandarin words, which were displayed in random order. A total of 2560 (16 speakers × 10 tokens per word × 16 words) Mandarin words was collected. Four female listeners each from Beijing (M_{age} = 21 years, *SD* = 2), Guangzhou (M_{age} = 20 years, *SD* = 1), Shanghai (M_{age} = 24 years, *SD* = 2) and Yantai (M_{age} = 21 years, *SD* = 2) were recruited to judge 640 (4 speakers × 10 tokens per word × 16 words) Mandarin words, respectively, which were produced by their native dialect speakers as introduced above. On a given trial, they saw a single Chinese character and listened to one of the productions. They were required to decide to what extent the presented pronunciation was similar to that in their own Mandarin community on a scale of 1-7, from not similar to similar. Tokens with average rating scores of 4 or above were retained for acoustic analysis (*n* = 2450).

2.1.5. *Data analyses*

Mandarin lexical tones are characterized by different fundamental frequency (*F*0) values, with *F*0 height and *F*0 contour as the primary acoustic parameters [18]. Robust Epoch And Pitch EstimatoR (REAPER) [19] was used to extract *F*0 values, which has been demonstrated to return effective *F*0 measurements at low pitch ranges [20], such as creaky voice in the low-dipping tones. To compare each Mandarin lexical tone produced by Guangzhou, Shanghai, and Yantai speakers with Beijing speakers' pronunciations on the same scale, both time and pitch values were normalized. The *F*0 values of each word were measured at 11 equidistant points (P_0, $P_1 \ldots P_{10}$), generating a set of time-normalized *F*0 values. Only the *F*0 values of P_1 to P_9 (the most stable part) were used for further analyses. To preserve accent phonetic variation, *F*0 values were converted to semitones using the equation provided in [21]:

$$F_0^{ST-ref} = \frac{12}{\log_{10} 2} \times \log_{10} \frac{x_i}{ref} \quad (1)$$

where x_i refers to raw values of *F*0 in Hz with *i* taking the value 1 to 9 for measuring points P_1 to P_9 and *ref* is relative to each speaker's average pitch. *F*0 mean ($F0_{mean}$) captures the steady pitch height for level tones, *F*0 maximum ($F0_{max}$) and *F*0 minimum ($F0_{min}$) capture pitch height of the other three tones in Mandarin. The extreme *F*0 values, along with their corresponding locations, determine the pitch contour. In addition, $F0_{min\ location}$ indicates the turning point of the low dipping tone. Duration values for each word, one of the crucial acoustic parameters for differentiating among Mandarin lexical tones [22], were also measured with a Praat script developed by [23].

2.2. Results

2.2.1. Acoustic analysis of Beijing Mandarin lexical tones

In order to determine the contributions of pitch (i.e., $F0_{mean}$, $F0_{max}$, $F0_{max\ location}$, $F0_{min}$, $F0_{min\ location}$) and duration to distinguishing Mandarin lexical tones, six linear mixed-effects models were built with the six measures as dependent variables, tone types as the fixed-effects factor, and participants and vowels as random-effects factors. The Kenward-Roger approximation to the degrees of freedom was used to calculate the *p* values for the fixed effects and *Anova* function from *car* package in R was adopted to calculate *F*. Significant main effects of tone type were found for all six measures: $F0_{mean}$, $F(3, 8) = 31$, $p < .001$; $F0_{max}$, $F(3, 7) = 23$, $p < .001$; $F0_{max\ location}$, $F(3, 7) = 162$, $p < .001$; $F0_{min}$, $F(3, 7) = 49$, $p < .001$; $F0_{min\ location}$, $F(3, 7) = 57$, $p < .001$; duration, $F(3, 7) = 63$, $p < .001$. These results suggest that Mandarin lexical tones are reliably distinguished by the proposed six measures.

Multiple comparisons were conducted with the R-package *lsmeans* to determine how the acoustic measures distinguish the four Mandarin lexical tones. $F0_{mean}$ and $F0_{max}$ for T3 were significantly lower than the other three tones, which did not differ from each other. $F0_{min}$ for the four Mandarin lexical tones all differed significantly except T1 vs. T2. Both $F0_{min\ location}$ and $F0_{max\ location}$ differed significantly among the Mandarin lexical tones except $F0_{min\ location}$ for T1 vs. T3. $F0_{max}$, $F0_{min}$, $F0_{min\ location}$ and $F0_{max\ location}$ together indicate that T3 was a low tone with a similar falling part to T4. When it comes to tone durations, there were significant differences among four Mandarin lexical tones. T3 had the longest duration, followed by T2, T1, and T4.

2.2.2. Tone F0 measures across the four regional accents

Mandarin lexical tones in four regional accents are plotted in Figure 1. The pitch contours of the four Mandarin lexical tones in each of the three regional accents are level, rising, dipping, and falling, respectively, as in Beijing Mandarin. There are, however, some deviations between regional Mandarin lexical tones and SM, especially for T3 (the dipping tone). Guangzhou, Shanghai, and Yantai T3 are the most visibly distinct from the Beijing accent. Most notably, the *F*0 of the turning point in all three regional accents is higher than in Beijing Mandarin. It is highest in the Guangzhou accent, followed by the Shanghai and Yantai accents. Although T3 begins at a similar *F*0 in all four accents, it ends at the highest value in Guangzhou accent, followed by Beijing, Shanghai, and Yantai accents. Beijing Mandarin has the deepest curve, followed by Yantai, then Shanghai. Guangzhou's curve is shallowest. In addition, Yantai T3 reaches its turning and final points ahead of those in Beijing T3. T4 is similar in Guangzhou and Beijing accents. Yantai T4 has a similar contour but a lower offset *F*0 than Beijing T4. Shanghai T4 is quite different from Beijing T4, with a higher onset and lower offset. The *F*0 contour of Yantai T2 is similar to Beijing T2, but is higher overall. Shanghai and Guangzhou T2 both start at a lower *F*0 than Beijing Mandarin. Guangzhou T2 rises gradually to end at the same *F*0 as Beijing T2, whereas Shanghai T2 is higher than both Beijing and Yantai T2 by the middle and rises even higher near the end. T1 in Guangzhou, Shanghai, and Yantai accents are virtually identical to Beijing accent except that Beijing T1 is higher overall.

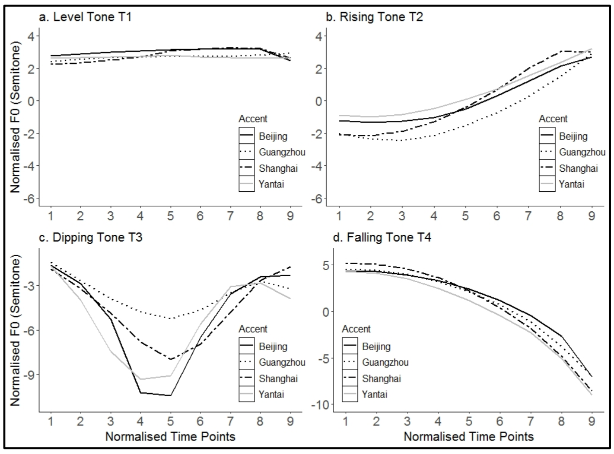

Figure 1: *Lexical tone F0 trajectories for each regional accent. Lines indicate the mean of the normalized F0 contour across speakers in terms of normalized time.*

$F0_{mean}$, $F0_{max}$, $F0_{max\ location}$, $F0_{min}$, and $F0_{min\ location}$ were each fitted with a linear mixed-effects model to investigate tone variation between the three regional accents and Beijing accent. For each model, a given acoustic measure was the dependent variable, with tone types (the four Mandarin tones) and accent types (Beijing, Guangzhou, Shanghai, Yantai) as the fixed-effects factors, and participants and vowels (*a*, *i*, *u*) as the random-effects factors. Five models were built to test all possible main effects and interactions. Results showed that there were significant main effects for tone types with regard to each of the proposed acoustic measures (see Table 2). However, there were no significant main effects for accent types, suggesting that all four Mandarin lexical tones are distinguished according to the proposed acoustic measures across accents. The acoustic details of the tones were significantly modulated by accent, as indicated by significant interactions between tone types and accent types for all *F*0 measures. Only two significant accent differences involved tones other than T3: $F0_{min\ location}$ occurred significantly earlier in Guangzhou than Beijing T1, although they did not differ in $F0_{min}$; and $F0_{mean}$ was significantly lower in Yantai than Beijing T4. As for the more numerous dipping tone accent differences, $F0_{max\ location}$ occurred significantly later in Shanghai than Beijing, but they did not differ in $F0_{max}$. The turning point value ($F0_{min}$) was significantly lower in Beijing than in both Shanghai and Guangzhou accents, but when it occurred, $F0_{min\ location}$ did not differ among accents. Finally, $F0_{mean}$ was significantly higher Guangzhou than Beijing T3.

2.2.3. Tone durations across the four regional accents

Figure 2 illustrates tone durations across Beijing, Guangzhou, Shanghai, and Yantai accents. Durations of the four Mandarin lexical tones are longer in all three regional accents than in the Beijing accent. Meanwhile, T3 is the longest tone while T4 is the shortest in each accent, with T1 and T2 of intermediate durations. A linear mixed-effects model was built with duration as the dependent variable, tone types and accent types as the fixed-effects factors, and participants and vowels as the random-effects factors. Significant main effects were found for tone types, $F(3, 2468) = 981$, $p < .001$, as well as a significant interaction between accent and tone types, $F(9, 2467) = 27$, $p <$

Table 2: *Statistical results for detailed F0 measurements across regional accents*

	Tones	Interactions	Pairwise differences between tone types and accent types
$F0_{mean}$	$F(3, 8) = 40, p < .001$	$F(9, 2414) = 16, p < .001$	T4: Beijing vs. Yantai, $\beta = 1.176$, $SE = 0.303$, $t(27.77) = 3.884$, $p = 0.037$; T3: Beijing vs. Guangzhou, $\beta = -1.613$, $SE = 0.310$, $t(30.56) = -5.199$, $p < .01$
$F0_{max}$	$F(3, 8) = 28, p < .001$	$F(9, 2414) = 2, p = 0.009$	
$F0_{max\ location}$	$F(3, 7) = 605, p < .001$	$F(9, 2415) = 14, p < .001$	T3: Beijing vs. Shanghai, $\beta = -27.023$, $SE = 6.76$, $t(15.9) = -3.995$, $p = 0.051$
$F0_{min}$	$F(3, 7) = 57, p < .001$	$F(9, 2414) = 27, p < .001$	T3: Beijing vs. Shanghai, $\beta = -4.4911$, $SE = 0.836$, $t(22.5) = -5.371$, $p = 0.0016$; T3: Beijing vs. Guangzhou, $\beta = -6.611$, $SE = 0.834$, $t(22.3) = -7.928$, $p < .0001$
$F0_{min\ location}$	$F(3, 7) = 215, p < .001$	$F(9, 2414) = 18, p < .001$	T1: Beijing vs. Guangzhou, $\beta = 22.982$, $SE = 4.38$, $t(18.2) = 5.250$, $p = 0.0036$

.001. Multiple comparisons were conducted using R *lsmeans* to test the pairwise differences in tone types and the interaction. All four Mandarin tones were significantly different from each other, but there were no significant differences among the four regional accents for each Mandarin tone category, indicating that Mandarin tones differed based on duration across accents.

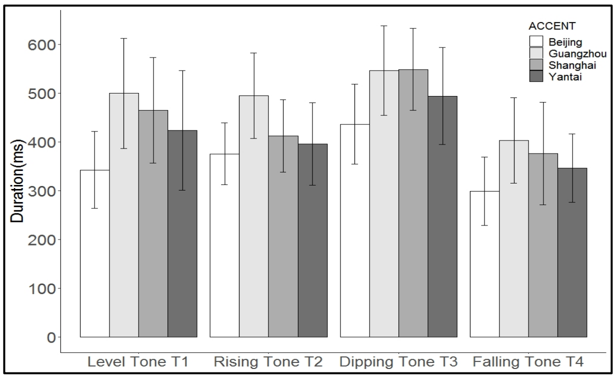

Figure 2: *Tone durations across regional accents*

3. Discussion and Conclusions

Mandarin lexical tones produced as an L2 by regional dialect speakers from Guangzhou, Shanghai, and Yantai were documented in this study. Consistent with existing findings [6], regional speakers successfully produced all four Mandarin lexical tones.

Interestingly, as expected, even though speakers of Guangzhou, Shanghai and Yantai produced the tones of the Mandarin words in the correct categories, the phonetic details of their pronunciations varied from Beijing Mandarin. Regionally accented Mandarin tones were generally longer than native Beijing Mandarin speakers'. In fact, SM is an L2 for regional Mandarin speakers. Influenced by "equivalence classification" (SLM) or "perceptual assimilation" (PAM) of their native tone systems, regional Mandarin speakers may have devoted more effort and/or attended more closely to their pronunciations [24] to ensure that they were standard when acquiring similar tone categories. Even though they had spoken Mandarin as a daily L2 from childhood, this may have resulted in lengthened Mandarin tones. A similar result was found for L1-English L2-Mandarin speakers [25], who may have perceptually assimilated the tones to their L1 intonation categories [26]. Although tones in Guangzhou, Shanghai, and Yantai accents did not differ from the Beijing accent by overall $F0$ values, they did differ significantly in more detailed $F0$ measures. T3 in Guangzhou, Shanghai, and Yantai accents had low $F0_{mean}$ values as native Beijing Mandarin, whereas the dipping curves were shallower, and the $F0_{mean}$ in Guangzhou accent as well as the offset in Shanghai accent were higher. These results demonstrate category-goodness variations in regionally accented Mandarin tones, which contrasts with the category-shifting segmental variability (i.e., deviant vowels and consonants that are assimilated to a different, contrasting native phoneme, which is a novel extension of PAM's Two Category assimilation type [27]) seen in regionally accented Mandarin consonants (i.e., Shanghai speakers produce the alveolar coda in Mandarin as a velar coda [5]).

Given that Mandarin T1 in all four regional accents was a high-level tone, but T3 varied from the Beijing accent to varying extents, it is safe to conclude that tone deviations from Beijing T3 were larger than T1, suggesting differences in the degree of variability across the four Mandarin tones. Similarly, tone variations varied among the three regional accents. As a tone language, Chinese is phonologically classified into seven dialects (i.e., Guanhua, Wu, Gan, Xiang, Min, Hakka, and Yue) based on tone features [28]. Yantai and Beijing dialects belong to Guanhua, while Guangzhou and Shanghai dialects come from the Yue and Wu dialect families, respectively. The Yantai dialect is supposed to be much more similar to the Beijing dialect than Guangzhou and Shanghai dialects, resulting in smaller regional accents than Guangzhou and Shanghai accents. Indeed, our results show that the Yantai accent was less distinct from the Beijing accent than Guangzhou and Shanghai, especially for the dipping tone T3. However, we did not observe differences between the Guangzhou and Shanghai accents in their degree of deviation from the Beijing accent. More detailed statistical analyses such as difference scores [29] between regional and Beijing accents based on $F0$ measures and duration may be required to capture the phonetic differences that characterise the differences between them.

Guangzhou and Shanghai Mandarin speakers successfully produced a dipping contour despite the lack of any dipping tones in their native regional dialects. This contrasts with [30], who found that female speakers of the Xiamen dialect, which also lacks a dipping tone, produced a falling tone for Mandarin T3. This difference may be due to uncontrolled factors that are known to affect influence regional accents. For example, female speech is characterised as clearer [31] and more variable [32] than male speech in terms of speech production, the degree of perceived regional accent is positively correlated with age [7], and it is negatively correlated with level of education [6]. It is possible that the speakers in [30] came from different age and/or education groups from those in this study, but no participant information was introduced in [30]. To reconcile these findings, future research on accent variability in Mandarin tones should control for the speakers' gender, age, and education.

4. Acknowledgements

Supported by CSC scholarship (No. 201707110014) and ALS research grant 2018 to the 1st author. We are grateful for tech support from the Speech Acquisition and Intelligent Technology Lab (BLCU, China), to Juqiang Chen for assistance with analyses, and above all to the participants.

5. References

[1] H. Chappell, "Sinitic Grammar: Synchronic and Diachronic Perspectives", UK, Oxford: Oxford University Press, Chap. 1, pp. 3-28, 2001.

[2] J. Yuan, "An introduction to Chinese dialects", 2nd Ed., Beijing: Wenzi Gaige Chubanshe, Chap. 3-4, pp. 16-55, 1989 (in Chinese).

[3] D. C. S. Li, "Chinese as a lingua franca in greater China", *Annual Review of Applied Linguistics*, vol. 26, pp. 149-176, 2006.

[4] J. J. Wrenn, "Popularization of Putonghua", *Journal of Chinese Linguistics*, vol. 3, no. 2-3, pp. 221-227, 1975.

[5] Y. Guan, "Nasal coda realization in speech production of Shanghai Mandarin" in *Proceedings of the 19th International Congress of Phonetic Science*, 2019.

[6] A. Li, & X. Wang, "A Contrastive Investigation of Standard Mandarin and Accented Mandarin" in *Proceedings of the 8th European Conference on Speech Communication and Technology (EUROSPEECH 2003)*, pp. 2345-2348, 2003.

[7] J. E. Flege, "Second-language speech learning: Theory, findings, and problems" in *Speech perception and linguistic experience: Issues in cross-language research*, W. Strange, Ed., pp. 229-273, 1995.

[8] C. T. Best, "A direct realist view of cross-language speech perception," in *Speech perception and linguistic experience: Issues in cross-language research*, W. Strange, Ed., Timonium, MD: York Press, pp. 171-204, 1995.

[9] C. T. Best, & M. D. Tyler, "Nonnative and second-language speech perception: Commonalities and complementarities" in *Second language speech learning: The role of language experience in speech perception and production*, M. J. Munro & O.-S. Bohn, Eds., Amsterdam: John Benjamins, pp. 13-34, 2007.

[10] Y. R. Chao, "Mandarin primer: An intensive course in spoken Chinese", Cambridge, MA: Harvard University Press, 1948.

[11] C. T. Best, "Devil or angel in the details? Perceiving phonetic variation as information about phonological structure," in J. Romero and M. Riera, Eds., *Phonetics-Phonology Interface: Representations and Methodologies*, pp. 3-31, 2015.

[12] Y. R. Chao, "A system of tone-letters", *Le Maitre Phonetique*, vol. *45*, pp. 24-27, 1930.

[13] J. G. Barry, & P. J. Blamey, "The acoustic analysis of tone differentiation as a means for assessing tone production in speakers of Cantonese", *The Journal of the Acoustical Society of America*, vol. 116, no. 3, pp. 1739-1748, 2004.

[14] B. Xu, & Zh. Tang, "Shanghai Shiqu FangyanZhi" [A description of Shanghainese spoken in the urban districts of the Shanghai City], Shanghai: Shanghai Education Press, 1988.

[15] Zh. Bao, "The Structure of Tone", New York: Oxford University Press, 1999.

[16] P. C. M. Wong, & T. K. Perrachione, "Learning pitch patterns in lexical identification by native English-speaking adults", *Applied Psycholinguistics*, vol. 28, no. 4, pp. 565-585, 2007.

[17] T. L. Gottfried, A. M. Staby, & C. J. Ziemer, "Musical experience and Mandarin tone discrimination and imitation", *The Journal of the Acoustical Society of America*, vol. 115, no. 5, pp. 2545-2545, Apr. 2004.

[18] J. M. Howie, "Acoustical Studies of Mandarin Vowels and Tones", Cambridge, MA: Cambridge University Press, 1976.

[19] D. Talkin, https://github.com/google/REAPER, 2015.

[20] K. Dallaston, & G. Docherty, "Sociophonetic variation in the prevalence of creaky phonation in the speech of young adults from two capital cities in Australia", *Language Variation and Change (LVC-A3) Workshop*, 2017.

[21] J. Zhang, "A Comparison of Tone Normalization Methods for Language Variation" in *Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation*, pp. 823-831, 2018.

[22] D. H. Whalen, & Y. Xu, "Information for Mandarin tones in the amplitude contour and in brief segments", *Phonetica*, vol. 49, pp.25-47, 1992.

[23] M. Antoniou, https://markantoniou.blogspot.com, 2011.

[24] W. Strange, "Automatic selective perception (ASP) of first and second language speech: A working model", *Journal of Phonetics*, vol. 39, no. 4, pp. 456-466, 2011.

[25] M., Peabody, & S. Seneff, "Annotation and features of non-native Mandarin tone quality", *INTERSPEECH 2009 - 10th Annual Conference of the International Speech Communication Association*, September 6-10, Brighton, UK, Proceedings, pp. 460-463, 2009.

[26] C. K. So, & C. T. Best, "Categorizing mandarin tones into listeners' native prosodic categories: The role of phonetic properties", *Poznan Studies in Contemporary Linguistics*, vol. 47, no. 1, pp. 133–145, 2011.

[27] C. T. Best, J. A. Shaw, & E. Clancy, "Recognizing words across regional accents: The role of perceptual assimilation in lexical competition", *INTERSPEECH 2013 - 14th Annual Conference of the International Speech Communication Association*, August 25-29, Lyon, France, Proceedings, pp. 2128-2132, 2013.

[28] P. Ting, "Hanyu Fangyan Qufen de Diaojian" [Phonological features for classification of the Chinese dialects], *Tsing Hua Journal of Chinese Studies*, no. 14, pp. 257-273, 1982.

[29] J. Chen, C. T. Best, & M. Antoniou, "Cognitive factors in Thai-naïve Mandarin speakers' imitation of Thai lexical tones", *INTERSPEECH 2019 - 20th Annual Conference of the International Speech Communication Association*, September 15-19, Graz, Austria, Proceedings, pp. 2653-2657, 2019.

[30] A. Li, Q. Fang, R. Xu, X. Wang, & Y. Tang, "A contrastive study between Minnan-accented Chinese and Standard Chinese" in *Proceedings of the 8th International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques*, pp. 86-92, 2005.

[31] A. Suppes, "Gender, Power and the Human Voice" in *The Exercise of Power in Communication*, R. Schulze and H. Pishwa, Ed, UK: Palgrave Macmillan, pp. 187-198, 2015.

[32] H. Caroline, "Cross-language variation in the vowels of female and male speakers" in *Proceedings of the 13th International Congress of Phonetic Sciences*, pp. 420- 423, 1995.