INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Perception of tone in whispered Mandarin sentences: the case for Singapore Mandarin

Yuling Gu, Boon Pang Lim, Nancy F. Chen

Institute for Infocomm Research, Singapore
yuling.gu@outlook.com, bplim@i2r.a-star.edu.sg, nfychen@i2r.a-star.edu.sg

Abstract

Whispering is commonly used when one needs to speak softly (for instance, in a library). Whispered speech mainly differs from neutral speech in that voicing, and thus its acoustic correlate F0, is absent. It is well known that in tonal languages such as Mandarin, tone identity is primarily conveyed by the F0 contour. Previous works also suggest that secondary correlates are both consistent and sufficient to convey Mandarin tone in whisper. However, these results are focused on Standard Mandarin spoken in Mainland China and have only been obtained via small-scale experiments using citation-form speech. To investigate whether these results will carry over to continuous sentences in other variations of Mandarin, we present a study that is the first of its nature to explore native Singapore Mandarin. Unlike related works, our large-scale perceptual experiment thoroughly investigates lexical tones in whispered and neutral Mandarin by involving more diverse speech data, greater number of listeners and use syllables excised from continuous speech to better simulate natural speech conditions. Our findings differ significantly from earlier works in terms of the recognition patterns observed. We present further in-depth analysis on how various phonetic characteristics (vowel contexts, place and manner of articulation) affect whispered tone perception.
Index Terms: speech perception, whispered Mandarin, tone recognition, Singapore Mandarin, continuous speech

1. Introduction

Whispering is a natural mode of speech communication which in recent years has garnered more attention in the speech community [1, 2, 3, 4, 5, 6]. During whisper, vocal folds do not vibrate and consequently pitch is absent [7]. In spite of this, people can purportedly perceive tone distinctions during whisper for tonal languages such as Thai [8], or Mandarin [9]. Pitch in Mandarin is widely thought to be the main acoustic cue that conveys tone [10, 11], but secondary correlates are sufficient to convey tone in Standard Mandarin citation-form syllables, as demonstrated by several small-scale experiments in [9], [12] and [13]. In [9], the author investigated this using both stand-alone citation syllables and syllables in carrier sentences. Her acoustic analysis and listening tests suggest that whispered tones are better perceived when speakers lengthen the vocalic duration or emphasize amplitude contour, and showed that over 60% correct tone identification rate could be attained. Similarly, [12] and [13] also found tone identification rates above chance for Standard Mandarin citation-form syllables.

However, these previous experiments have been small-scale and limited due to the scarcity of whispered corpora. None of these experiments recruited more than 10 listeners and even the most comprehensive study amongst them did not have more than 4 speakers, reflecting a paucity of listener and speaker diversity. In [9], the stimuli used also had only limited phonetic contexts (two vowels, /a/, /i/ and three consonants /b/,/f/,/m/). Whether the results obtained from these experiments apply to syllables from continuous sentences in other variations of Mandarin (such as Singapore Mandarin) remains an important gap in our understanding of tone perception. A larger-scale perceptual experiment, using more speakers and listeners, more phonetic variations and source material that is more representative of daily use is thus desirable. In this paper we describe such an experiment using whispered Mandarin continuous speech, recruiting native Singapore Mandarin speakers and listeners as subjects.

2. Tone Perception in Mandarin Sentences

2.1. Mandarin Tone and Phonology

Mandarin is a tonal language. Its syllables may be written in *pinyin* orthography as a sequence of four phonemes and a tone marker such as `[C][G]V[N]T`. The optional consonant `[C]`, glide `[G]` and nasal or off-glide of diphthong `[N]` surround vowel `V` to form a base syllable [14]. The lexical tone marker `T`, denotes one of four citation tones – high level (Tone 1), evenly rising (Tone 2), falling rising (Tone 3) and falling (Tone 4). These describe the F0 contour as shown in Figure 1. For example the syllable verb"$xuan1$" would have consonant [x], glide [u], vowel [a], final [n] and Tone 1.

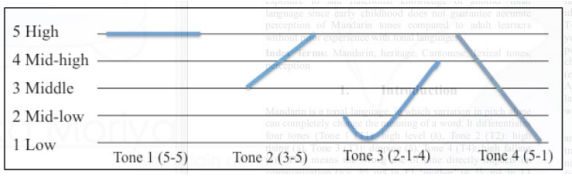

Figure 1: Four Tones in Standard Mandarin.

Although *pinyin* orthography specifies the exact tone, it is less commonly used compared to the Chinese character script. Individual characters may have multiple pronunciations, sometimes differing in the base syllable but more commonly in tone. Lexical items usually have fixed citation tones, but this might vary from dialectal varieties of region to region [15]. Mandarin also exhibits prevalent tone sandhi, in which citation tones may change in the context of other tones [16]. Some analysis of Mandarin include the neutral tone that arises out of sandhi rules. The realized F0 contour may also depend on the position of the syllable in the sentence [9]. These factors may further complicate analysis when we consider the entire speech communication channel, beginning from production from the commonly

http://dx.doi.org/10.21437/Interspeech.2016-297

used orthographic source to perception of tone in individual syllables.

2.2. Singapore Mandarin Tone System

The pronunciation of Singapore Mandarin differs from Standard (Beijing) Mandarin in Mainland China in terms of salient segmental features [17]. An acoustic analysis in [18] also showed that tone is differently realized. It is noted that in Singapore Mandarin, Tone 2 syllables showed an initial fall followed by a low level stretch before the final rise. This gives Singapore Mandarin Tone 2 a dipping contour, similar to that of Tone 3 in Standard Mandarin instead. In addition, Tones 3 and 4 were both realized as falling tones with Tone 4 starting at a higher initial F0 and both had variants that showed a small final rise. Thus, unlike Standard Mandarin, Tone 3 in Singapore Mandarin is not differentiated by a significant final rise. The same study also found that while Tone 4 was significantly shorter than the other tones, the other three tones were not significantly different in terms of duration. The phenomenon that Tone 3 has the longest duration as agreed upon by [9], [12] and [13] for citation-form syllables in Standard Mandarin is not shown in Singapore Mandarin. In this study, we contribute to this current understanding of the Singapore Mandarin tone system through further analyzing its impact on tone perception in Section 3.

2.3. Continuous speech versus citation-form syllables

Citation-form syllables differ significantly from those in continuous speech. In continuous speech, there exists various manipulations including assimilatory changes, vowel reduction and syllabic simplifications [19]. These manipulations affect the acoustic features observed. Vowel reduction, for instance, involves formant values shifting towards those of a central vowel, greater overlap between different vowel types in formant space and usually decreasing length of vowels [20]. In addition, F0 contours in continuous speech also vary while those in citation-form syllables have stable shapes [21, 22]. Since vowels are the "carrier" of tones [9] and tones are primarily conveyed by the F0 contour [10, 11, 21, 22], we will analyze these changes alongside with tone perception accuracies obtained in our experimental results.

2.4. Tone Perception Experiment

We designed and conducted perceptual experiments to quantify how well lexical tones can be identified by human listeners when whispered. The `iWhisper-Mandarin` corpus, a parallel whispered speech corpus [23] was used. It comprises 40 male and 40 female speakers from Singapore, each reading 100 sentences in both neutral and whispered speaking modes.

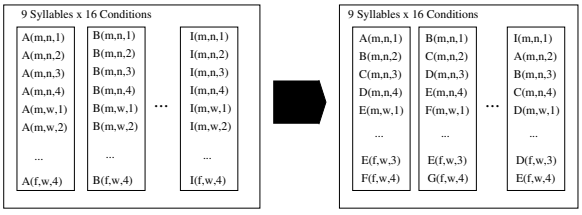

Figure 2: Rotation Scheme in Stimuli Generation.

We used the Kaldi speech recognition toolkit [24] to train two DNN-HMM acoustic models for each speaking mode, and

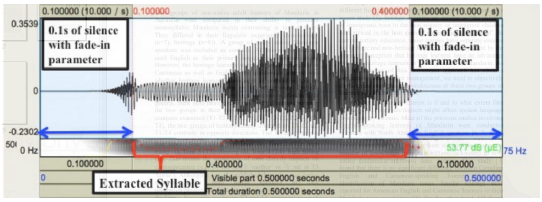

Figure 3: Extracted Syllable from iWhisper-Mandarin.

used them to force-align utterances, giving us syllable boundaries. A small portion of these alignments were manually verified and found to be accurate to within 50ms of the actual syllable boundary. A character to *pinyin* pronunciation lexicon was used to obtain canonical *pinyin* labels for words in each utterance. The transcripts were analyzed for counts of each syllable with tone – base syllables without sufficient counts in all four tones were removed, leaving 54 base syllables remaining. Examples from the corpus were randomly drawn for 16 distinct conditions of each base syllable – two genders, four tones, and two speaking modes (neutral and whispered) – resulting in a large balanced block of $54 * 16 = 864$ unique tonal syllables. This was rearranged into 9 column-sets, each column comprising of 6 unique syllables in 16 conditions. A rotating scheme, illustrated in Figure 2 distributes base syllables across the 9 columns. Each column eventually is assigned to a single listener. This ensures that for every such block, the priors for all conditions and base syllables are balanced, and no listener ever gets more than 2 of the same base syllable. We produced 4 such blocks, two assigned to female listeners and two to male listeners. This resulted in equal numbers of stimuli for all syllables and tones, and roughly balanced distributions presented to each listener. The time-marks from forced alignments were then used to extract audio waveforms according to these lists. Fade-in and fade-out from 100ms before and 100ms after each syllable was applied to produce stimuli that are more comfortable for listening. This is illustrated in Figure 3.

We recruited 18 male and 18 female listeners. These listeners are native Singapore Mandarin speakers who lived in Singapore for at least 10 years, speak fluent Mandarin on a regular basis, and were deemed to have no auditory impairments. They were screened to ensure that they could perceive lexical tones in Mandarin. The listening tests were conducted in a quiet office environment using a custom program running on a personal computer with USB headphones (Ars Technica ATH750-COM USB). We randomized each set of stimuli presented to ameliorate hysteretic effects. For each trial, the stimulus was played back through headphones, and the listener was asked to choose one of four options indicating the tone. They were allowed a fifth option (`''I don't Know''`) and encouraged to avoid random guessing. Listeners were also able correct responses, and replay stimuli up to three times. Familiarization with the software occurred prior to actual test. Tone identification rate is computed and expressed in percent correct (PC) in section 3.

3. Results and Analysis

3.1. Significance Testing for Independent Bernoulli Trials

Our setup can be treated as a sequence of independent Bernoulli trials, the result of each trial is either 1 or 0 corresponding to whether tone for the syllable was correctly identified. After ignoring the `''I don't know''` responses to eliminate ef-

fects of random guessing, the test becomes a four way forced choice response. We formulate a statistical test with the null hypothesis that the listeners are simply making random guesses for all the stimuli. The distribution of PC follows a binomial distribution [25], which can be approximated by a normal distribution when $np > 5$ and $n(1-p) > 5$, for n number of trials, and a chance level of $p = \frac{1}{4} = 0.25$. The corresponding two-sample test was used to check if the PC for one category is significantly above another.

3.2. Percent Correct Identification

In syllables excised from Singapore Mandarin continuous sentences, tone identification PC is 52.13% ($p < 0.001$) for normal speech, and 27.73% ($p = 0.004$) for whisper.

Contexts	Publication	Overall PC	
		Normal	Whispered
Mono-syllables produced in isolation	Tone in Whispered Chinese: Articulatory Features and Perceptual Cues. - M. Gao [9]	> 90	60.1
	Study on Tones in Whispered Chinese. - D. Q. Sha, X. L. Li and B. L. Xu [12]	NA	62.1
	The establishment of a Chinese Whisper Database and Perceptual Experiment. - L. L. Yang, Y. Li and B. L. Xu [13]	NA	60.8
Mono-syllables excised from continuous speech	Current work	52.1	27.8

Table 1: Comparison of overall PC.

The tone identification rate we have achieved using syllables excised from Singapore Mandarin continuous speech is lower compared to that reported for citation-form syllables in Standard Mandarin [9, 12, 13]. This suggests that the acoustic differences introduced when going from citation-form syllables to continuous speech [19, 20, 21, 22] are likely to play a part in affecting tone identification. For syllables excised from continuous speech, the lack of a stable F0 contour through which tone is primarily conveyed [10, 11, 21, 22] renders accurate tone identification to be more challenging even in normal speech. Furthermore, the introduction of various manipulations such as vowel reduction may alter the acoustic features (vocalic duration and formant values) [19, 20] that served as secondary correlates for whispered tone identification in [9] [12] and [13]. This result may also be partly attributable to the characteristics of the Singapore Mandarin tone system as described in Section 2.2. We further explore how these characteristics affect tone identification below in Section 3.3.

3.3. Tone Confusion Analysis

Table 2 shows the confusion patterns in tone identification. Bolded numbers in the table indicate the tone option chosen by the most number of listeners when each category of tone was presented. For normal speech, all tones have PCs significantly above chance level ($p < 0.001$). However, for whispered speech, only Tones 2 ($p = 0.051$) and 4 ($p < 0.001$) seem to be above chance level. Furthermore, some confusions occur at a rate above chance. The confusions for $1 \rightarrow 4$ ($p < 0.001$) and $3 \rightarrow 2$ ($p = 0.031$) appear more significant, while $3 \rightarrow 4$ ($p = 0.207$) and $4 \rightarrow 3$ ($p = 0.488$) less so. Interestingly,

(a) Normal Speech

Tone	Actual Listener choices (%)			
	1	2	3	4
1	**59.06**	14.82	13.41	12.71
2	19.42	**56.59**	16.31	7.67
3	22.75	22.04	**39.81**	15.40
4	25.71	7.38	13.81	**53.10**

(b) Whispered Speech

Tone	Actual Listener choices (%)			
	1	2	3	4
1	23.85	23.58	20.05	**32.52**
2	26.02	**28.57**	21.94	23.47
3	18.37	**29.08**	25.77	26.79
4	17.90	24.30	25.06	**32.74**

Table 2: Tone Confusions Patterns.

this pattern of confusions above chance also occurs for non-native learners of Mandarin [26], suggesting some similarities in tone perception patterns between Singapore Mandarin speakers and L2 learners. The above chance tone confusion patterns for whispered speech - that when presented with Tones 1 and 4, listeners perceived both as Tone 4, while Tones 2 and 3 are both perceived as Tone 2 - aligns with the observation of binary map perception and lack of distinct tonal categories [11] in L2 learners of Mandarin as well.

Specifically, Singapore Mandarin with citation-form syllables Tone 2 being realized with a dipping contour similar to that of Tone 3 in Standard Mandarin [18] could serve as an important reason for the confusion between the 2 tones in Singaporeans' understanding of the Mandarin tone system. In our experiment using syllables excised from continuous speech, we also observed lower PC for normal speech tone identification in Tone 3 as compared to the other tones and above chance confusion $3 \rightarrow 2$ in whisper when F0 is absent. Despite Tones 3 and 4 being both realized as falling tones in Singapore Mandarin citation-form syllables, Tone 4 syllables are significantly shorter [18], possibly resulting in less confusion. Interestingly, our results also show less confusion between these 2 tones.

Contexts	Publication	Ease of identification
Mono-syllables produced in isolation	Tone in Whispered Chinese: Articulatory Features and Perceptual Cues. - M. Gao [9]	$Tone3 > Tone4 > Tone1 > Tone2$
	Study on Tones in Whispered Chinese. - D. Q. Sha, X. L. Li and B. L. Xu [12]	$Tone3 > Tone4 > Tone1 > Tone2$
	The establishment of a Chinese Whisper Database and Perceptual Experiment. - L. L. Yang, Y. Li and B. L. Xu [13]	$Tone3 > Tone4 > Tone2 > Tone1$
Mono-syllables excised from continuous speech	Current work	$Tone4 > Tone2 > Tone3 > Tone1$

Table 3: Comparison of whispered tone identification patterns.

Comparing our results with whispered tone perception experiments done on Standard Mandarin citation-form syllables in Table 3, Tone 3 has always been the easiest to identify in these experiments yet it was the second most difficult tone to perceive from the responses given by listeners in our experiment. These previous studies have attributed the higher whispered tone identification rate for Tone 3 to its longer duration compared to other tones. However, in the case of Singapore Mandarin, the lack

(a) Normal speech

Manner \ Place	Labial	Dental	Palatal	Retroflex	Velar
Nasals	**M** /m/ 46.87 N=32 p=0.0021	**N** /n/ 38.70 N=31 p=0.0390			**NG**/ŋ/ (not in initial)
Plosives (Unaspirated)	**B** /p/ 46.87 N=32 p=0.0021	**D** /t/ 48.38 N=31 p=0.0013			**G** /k/ 38.09 N=63 p=0.0082
(Aspirated)	**P** /pʰ/ (absent)	**T** /tʰ/ 50 N=92 p<0.0001			**K** /kʰ/ 45.16 N=31 p=0.0048
Affricates (Unaspirated)		**Z** /ts/ 56.25 N=32 p<0.0001	**J** /tɕ/ 49.60 N=127 p<0.0001	**ZH** /tʂ/ 54.48 N=156 p<0.0001	
(Aspirated)		**C** /tsʰ/ 53.12 N=32 p=0.0001	**Q** /tɕʰ/ 52.38 N=63 p<0.0001	**CH** /tʂʰ/ 42.18 N=64 p=0.0007	
Fricatives	**F** /f/ 59.67 N=124 p<0.0001	**S** /s/ (absent)	**X** /ɕ/ 59.81 N=219 p<0.0001	**SH** /ʂ/ 59.67 N=128 p<0.0001	**H** /x/ 68.75 N=32 p<0.0001
Lateral		**L** /l/ 35.48 N=31 p=0.0888		**R** /ɻ/ (absent)	
Approximant (Glides)			**Y** /j/ 49.46 N=279 p<0.0001		
			W /w/ 49.43 N=89 p<0.0001		

(b) Whispered speech

Manner \ Place	Labial	Dental	Palatal	Retroflex	Velar
Nasals	**M** /m/ 22.22 N=27 p=0.3694	**N** /n/ 22.58 N=31 p=0.3779			**NG**/ŋ/ (not in initial)
Plosives (Unaspirated)	**B** /p/ 33.33 N=27 p=0.1587	**D** /t/ 31.81 N=22 p=0.2301			**G** /k/ 30.90 N=55 p=0.1558
(Aspirated)	**P** /pʰ/ (absent)	**T** /tʰ/ 29.06 N=86 p=0.1917			**K** /kʰ/ 37.03 N=27 p=0.0743
Affricates (Unaspirated)		**Z** /ts/ 24.13 N=29 p=0.4573	**J** /tɕ/ 26.27 N=118 p=0.3749	**ZH** /tʂ/ 22.72 N=154 p=0.2574	
(Aspirated)		**C** /tsʰ/ 28.12 N=32 p=0.3415	**Q** /tɕʰ/ 23.72 N=59 p=0.4108	**CH** /tʂʰ/ 25.80 N=62 p=0.4417	
Fricatives	**F** /f/ 23.57 N=123 p=0.3578	**S** /s/ (absent)	**X** /ɕ/ 29.57 N=213 p=0.0614	**SH** /ʂ/ 31.93 N=119 p=0.0404	**H** /x/ 31.03 N=29 p=0.2265
Lateral		**L** /l/ 34.78 N=23 p=0.1393		**R** /ɻ/ (absent)	
Approximant (Glides)			**Y** /j/ 30.25 N=238 p=0.0307		
			W /w/ 24.28 N=70 p=0.4451		

Table 4: PC for Different Front Consonants.

of temporal differences amongst syllables bearing Tones 1 to 3 [18] would imply that this acoustic feature most likely can no longer provide the basis for distinguishing these tones, leading to lower identification PCs. Instead, Tone 4, which has shown to be the second easiest to perceive in Standard Mandarin citation-form syllables, was the easiest to recognize in our experiment. Tone 4 being significantly shorter than the other tones in the Singapore Mandarin tone system [18] could possibly provide a key perceptual cue for its identification. This suggests that secondary correlates which help to convey tone in whispered citation-form syllables for Standard Mandarin may not directly carry over to whispered Singapore Mandarin continuous speech.

3.4. Phonetic Characteristics

We analyzed in detail several new aspects that are not explored by previous works – identification rate for various subcategories of varying phonetic contexts such as the preceding consonants, vowel contexts and final consonants. Table 4 shows PC under different preceding consonant contexts. Each box shows the pinyin orthography, corresponding IPA symbol, along with percent correct (PC) identification, the number of syllables in the experiment, and corresponding p-value under the Z-test for PCs above chance. Tone identification is significantly above chance for all initial contexts in normal speech, but for whispered speech, only syllables with initial consonants /ʂ/ and /j/ are identified significantly above chance.

	none	dental [N]	velar [NG]
Normal	**49.9** $\mathbf{N = 1059}$ $(\mathbf{p < 0.001})$	**51.8** $\mathbf{N = 220}$ $(\mathbf{p < 0.001})$	**59.5** $\mathbf{N = 375}$ $(\mathbf{p < 0.001})$
Whispered	**28.1** $\mathbf{N = 964}$ $(\mathbf{p = 0.0128})$	27.2 $\mathbf{N = 202}$ $(p = 0.2323)$	28.3 $\mathbf{N = 353}$ $(p = 0.0743)$

Table 5: PC for Different Ending Consonants. Here, N is the number of trials, p is the p-value for PC above chance levels (calculated using a one-sample Z-test).

Vowels can be classified according to vertical position of the tongue – high, mid and low [27]. However, separate two-sampled Z-tests (not shown) did not find correlation of higher vowel height to increasing PC as significant. Similar analysis for nasal endings arranged by place of articulation is shown in Table 5. Tone identification PC for syllables with nasal endings are higher than those with vowel endings in normal speech but this pattern is not observed for whispered speech. Amongst the various conditions, whispered syllables with /e̯i̯/ as their endings attained the lowest tone identification PC of 12.0% ($p = 0.0667$) which is below chance level while syllables with /ən/ as their endings had the highest tone identification PC of 39.3% ($p = 0.0404$). These analysis thus surface phonetic characteristics that are likely to help in tone identification. Finally, two-sample Z-tests showed that tone identification in neutral speech is better than for whispered speech for all categories, reaffirming the importance of F0 in conveying tone.

4. Discussion and Conclusion

We have presented a perceptual experiment that shows tone identification results using citation-form syllables for Standard Mandarin do not carry over directly to continuous speech in other varieties of Mandarin. Our analysis is useful in bringing the current understanding of tone perception to a new level - from results obtained using citation-form syllables to syllables from continuous speech that better simulate natural speech conditions. We also bring into attention how differences between Singapore Mandarin and Standard Mandarin may affect tone identification patterns. We present in-depth analysis on the case for Singapore Mandarin and explore the impact of various phonetic characteristics on tone identification. Future works can expand upon this study to further explore neutral and whispered tone identification in syllables from continuous sentences in other varieties of Mandarin (such as Standard Mandarin and Taiwan Mandarin). These will contribute towards a better fundamental understanding of how tone is conveyed and serve as an important basis for building better Mandarin speech recognizers - ones that are able to maintain robust performance when automatic speech recognition is performed on long utterances in different speaking modes and in different varieties of Mandarin.

5. References

[1] Z. Chi and L. Hansen, "Whisper-island detection based on unsupervised segmentation with entropy-based speech feature processing," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 4, pp. 883–893, May 2011.

[2] H. R. Sharifzadeh, I. V. McLoughlin, and M. J. Russell, "A comprehensive vowel space for whispered speech," *Journal of Voice*, vol. 26, no. 2, March 2012.

[3] X. Fan, C. Busso, and J. H. L. Hansen, "Audio-visual isolated digit recognition for whispered speech," *19th European Signal Processing Conference*, 2011.

[4] T. Ito, K. Takeda, and F. Itakura, "Analysis and recognition of whispered speech," *Speech Communication*, vol. 45, pp. 139–152, October 2003.

[5] S. T. Jovicic and Z. Saric, "Acoustic analysis of consonants in whispered speech," *Journal of Voice*, vol. 22, no. 3, pp. 263–274, 2008.

[6] D. T. Grozdic, B. Markovic, J. Galic, and S. T. Jovocic, "Application of neural networks in whispered speech recognition," *Telfor Journal*, vol. 5, pp. 103–105, 2013.

[7] Y. Zhao, L. Zhao, and C. Zou, "A survey of whisper analysis and its processing," *Technical Acoustics*, vol. 27, pp. 562–569, 2008.

[8] A. S. Abramson, "Tonal experiments with whispered Thai," in *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre*, A. Valdman, Ed. The Hague, Netherlands: Mouton, 1972, pp. 31–44.

[9] M. Gao, "Tones in whispered Chinese: Articulatory features and perceptual cues," Master's thesis, University of Victoria, British Columbia, Canada, 2002.

[10] K. Yu, R. Wang, L. Li, and L. Ping, "Processing of acoustic and phonological information of lexical tones in Mandarin Chinese revealed by mismatch negativity," *Frontiers in Human Neuroscience*, 2014.

[11] B. Yang, "A model of Mandarin tone categories – a study of perception and production," Master's thesis, University of Iowa, 2010.

[12] D. Q. Sha, X. L. Li, and B. L. Xu, "Study on tones in whispered Chinese," *Audio Engineering*, vol. 11, no. 4, pp. 4–6, 2003.

[13] L. L. Yang, Y. Li, and B. L. Xu, "The establishment of a Chinese whisper database and perceptual experiment," *Journal of Nanjing University*, vol. 41, no. 3, pp. 311–317, 2005.

[14] D. San, "Phonology of Chinese (Mandarin)," in *Encyclopedia of Language and Linguistics, 2nd edition*. University of Michigan, MI USA: Elsevier Publishing, 2006.

[15] Z. Jie and J. Liu, "Tone sandhi and tonal coarticulation in Tianjin Chinese," *Phonetica*, vol. 68, no. 3, pp. 161–191, 2011.

[16] M. Y. Chen, *Tone sandhi: patterns across Chinese dialects*. Cambridge: Cambridge University Press, 2000.

[17] C. Chung-yu, *Salient Segmental Features in the Mandarin Spoken in Singapore*. Chinese Language Research Centre National University of Singapore, 1982.

[18] L. Lee, "The tonal system of Singapore Mandarin," in *Proceedings of the 22nd North American Conference on Chinese Linguistics (NACCL-22) and 18th International Conference on Chinese Linguistics (IACL-18)*, vol. 1, 2010, pp. 345–362.

[19] J. Laver, *Principles of Phonetics*. Cambridge University Press, 1994.

[20] J. Harrington and S. Cassidy, *Techniques in Speech Acoustics*. Kluwer Academic Publishers, 1999.

[21] K. Hirose and J.-S. Zhang, "Tone recognition of Chinese continuous speech using tone critical segments," in *Sixth European Conference on Speech Communication and Technology, EUROSPEECH 1999, Budapest, Hungary*, vol. 2, 1999, pp. 879–882.

[22] Y.-R. Wang and I.-B. Liao, "An overview of Mandarin-speech tone recognition," *Journal of The Chinese Institute of Electrical Engineering*, vol. 7, no. 2, pp. 145–155, 2000.

[23] P. X. Lee, D. Wee, B. P. Lim, N. F.-Y. Chen, and B. Ma, "A whispered Mandarin corpus for speech technology applications," in *INTERSPEECH*, 2014.

[24] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, and K. Vasely, "The Kaldi speech recognition toolkit," in *IEEE Workshop on Automatic Speech Recognition and Understanding*, 2011.

[25] R. C. Sprinthall, *Basic Statistical Analysis*, 9th ed. Allyn & Bacon, Inc, 2011.

[26] N. F. Chen, V. Shivakumar, M. Harikumar, B. Ma, and H. Li, "Large-scale characterization of Mandarin pronunciation errors made by native speakers of European languages," in *INTERSPEECH*, 2013.

[27] P. Lieberman, "Some effects of semantic and grammatical context on the production and perception of speech," *Language and Speech*, vol. 6, no. 3, pp. 172–187, 1963.