The perception of lexical tones in emotional speech by Dutch learners of Mandarin

Yachan Liang, Aoju Chen

Utrecht University
y.liang@uu.nl, aoju.chen@uu.nl

ABSTRACT

Past work on tonal perception in L2 Mandarin has focused on the citation forms of the tones produced in neutral emotion. However, an additional tone can be added to the citation tone or the citation tone can end at higher pitch level to express emotion in Mandarin. The current study examined tonal perception in different emotions (i.e., *neutral, happy, angry, sad*) and tonal contexts in Mandarin by Dutch learners of Mandarin, compared to native controls. It was found that both the L2 learners and native controls were significantly less accurate in the anger condition, compared to the neutral condition. Furthermore, although the intermediate-advanced learners performed better than the beginners, neither learner group differed significantly from the native controls. These results suggest that the L2 learners have acquired the phonological forms of the tones and their tonal perception was only affected when the surface forms underwent substantial changes in emotional speech.

Keywords: Mandarin, Dutch, lexical tones, emotional intonation, perception

1. INTRODUCTION

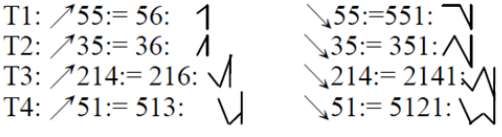
As a tone language, Mandarin uses pitch to convey lexical meanings. For example, the same syllable *ma* has different meanings when it is pronounced in different tones, namely *ma1* (mother), *ma2* (hemp), *ma3* (horse), *ma4* (scold). T1 is a high tone, T2 is a rising tone, T3 is a falling-rising tone, T4 is a falling tone. In contrast, pitch variation in non-tone languages like English and Dutch does not differentiate lexical meanings, but functions at sentence and discourse levels.

Previous studies have examined tonal perception typically in isolated syllables in Mandarin by both tone and non-tone language speakers [13, 6, 18]. It has been suggested that pitch height and direction of pitch movement are the main parameters to characterize Mandarin tones. But non-tone language speakers process pitch information in Mandarin tones differently than native speakers of Mandarin. For example, Gandour [5] compared the perception of Mandarin tones by speakers of English, Thai, Taiwanese, Cantonese and Mandarin, and found that Mandarin and Cantonese speakers paid more attention to the height dimension than speakers of the other three languages, while Mandarin and Cantonese speakers focused more on the direction dimension than English speakers. According to Ladd [10, 11], non-tone language speakers might process tones as intonation in their L1 because of the mismatch between their native intonational language and L2 tone categories. Therefore, the source of difficulty in learning Mandarin tones by non-tone language speakers has been attributed to the interference from their native language's suprasegmental features [17, 1, 9, 3].

Proficiency of Mandarin also matters to the perception of Mandarin tones by non-native speakers. For example, Yang [20] examined the perception of Mandarin tones and intonation by American learners at different proficiency levels, compared to native controls. It was found that both native controls and advanced learners performed much better than the first- and second year learners in tone identification. Hao [7] investigated the discrimination of Mandarin vowels and tones by native English speakers with different Mandarin proficiency levels via an AXB discrimination task. It was found that non-tone language speakers without any Mandarin training were less sensitive to tonal distinction than to vowels, whereas the learner groups were highly accurate in perceiving all contrasts except for the T2-T3 pair.

However, Mandarin also uses pitch to express emotion; the pitch characteristics of tones under certain changes in emotional speech, compared to neutrality [12]. Cross-linguistically, anger and happiness are characterized by longer duration, higher pitch accompanied by wider ranges, whereas sadness is characterized by a decrease in mean pitch, a slightly narrow pitch range, and a slower speaking rate, compared to neutral speech [19]. According to [2], Mandarin uses the successive additional tone (right panel, and T4 in the left panel, Figure 1) as well as modification to the pitch offset of the tone (T1, T2, T3 in the left panel, Figure 1) to realize the emotion.

Figure 1: The successive additional tones in Mandarin [2]



Recently, Li et al. [12] investigated the expression of seven emotions (*neutrality*, *sadness*, *happiness*, *fear*, *surprise*, *anger*, *disgust*) in Mandarin by analysing monosyllabic utterances produced by a male speaker. They found that *disgust* and *anger* were expressed via an additional falling tone, while *happiness* and *surprise* were expressed via an additional rising tone. Furthermore, *happiness* and *surprise* were characterized by a higher pitch range and pitch register (mean pitch), compared to *sadness* and *disgust*. *Surprise* was realized with a slightly narrower range than *happiness*.

The co-existence of both tone and emotional intonation in Mandarin raises the question as to whether changes in tones in emotional speech affect the perception of lexical tones by learners of Mandarin from a non-tone language background. To address this question, we have investigated tonal perception in emotional speech by Dutch learners of Mandarin, compared to tonal perception in neutral speech. Four commonly used emotions were chosen, i.e., *neutrality*, *happiness*, *anger*, *sadness*. Considering that emotions change the surface forms of the tones to various degrees in Mandarin, we predicted that L2 Mandarin learners may be the least accurate in ANGER, compared to NEUTRALITY, similar to native controls. We also expected that intermediate-advanced L2 Mandarin learners would be more accurate in tonal perception in different emotions than beginning L2 Mandarin learners but both would be less accurate than native controls.

2. METHOD

A tonal perception experiment was carried out, following the ethical guidelines of the institution.

2.1. Participants

Twenty-four adult Dutch learners of Mandarin (12 males, 12 females, age: M = 25.5 years, SD = 5.3) participated in the experiment. Eight of them were from the Chinese Education Center in the Netherlands, and the others were students of Chinese Studies at Leiden University. The learners started to learn Mandarin Chinese at the age of 18 in the Netherlands. None of them spoke a tonal dialect of Dutch (e.g., the Limgurgian dialect) or had studied tone languages other than Mandarin prior to this study. The participants were divided into two groups of twelve participants (beginners and intermediate-advanced learners) on the basis of their HSK score. HSK is a standardised Chinese proficiency test, standing for *Han Yu Shui Ping Kao Shi*. The participants who passed HSK level 1 & 2 were grouped as beginners; the participants who passed level 3 or higher were grouped as intermediate-advanced learners. In addition, 12 native speakers of Mandarin (mean age: 28 years) participated in the experiment as the controls. They were graduate students from University and did not major in Linguistics. None of the participants reported having speaking, reading or hearing defects.

2.2. Experimental materials

The stimuli were composed of 256 short sentences, containing 64 tokens of each of the four target words (i.e., *mong*, *ging*, *ra*, *bü*). To minimize lexical familiarity effects in tonal perception, we used pseudo words as target words. The target words were pronounced in each of the four tones and embedded in four carrier sentences, adopted from the carrier sentences used in [4], as shown in example (1). The carrier sentences were semantically neutral and thus could not provide any semantic cues to the tonal perception. Considering tonal coarticulation, we varied the tone preceding the target word in the carrier sentences such that each target word was preceded by each of the four tones, namely T1 (*chu1*), T2 (*du2*), T3 (*xie3*), T4 (*lian4*). Because tones at the initial position have the highest pitch level, and tones at the final position have "sentence-final tone lowering effects" [15], we placed the target word in sentence medial position.

All stimuli were recorded at a sampling frequency of 44.1 kHz with 16 bits resolution by a female native speaker of Mandarin with training in acting at the Linguistics Laboratory at University. The stimuli were subsequently checked by another native Mandarin speaker in a perception test. The listener reached 100% accuracy in perceived emotion, suggesting that the stimuli were produced in the intended emotions.

(1) a.	指出__这个字。
Pinyin (with tone)	zhi3 chu1 __ zhe4 ge4 zi4.
English translation	Please point the word __ out.
b.	我会读__这个字。
Pinyin (with tone)	wo3 hui4 du2__zhe4 ge4 zi4.
English translation	I can read the word __.
c.	我会写__这个字。
Pinyin (with tone)	wo3 hui4 xie3__zhe4 ge4 zi4.
English translation	I can write the word __.
d.	我想练__这个字。
Pinyin (with tone)	wo3 xiang3 lian4__zhe4 ge4 zi4.
English translation	I can practice the word __.

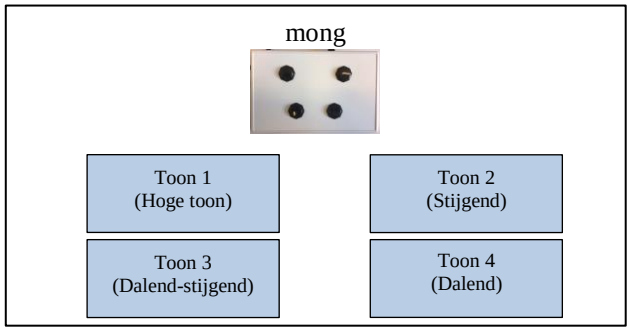
In order to keep the experiment within a feasible length and minimize boredom during the experiment, we divided the 256 stimuli into four lists according to a Latin square design. Each list contained 64 stimuli, including 16 representations of each target word in each tone in each emotion. Each list contains four different emotions, four target words, four tonal contexts, and four different lexical tones. Each stimulus list was presented in a randomized order in ZEP [16] according to the following constraints: the same tone/ carrier sentence/ syllable/ emotion should not appear more than two times in a row.

2.3. Procedure

The participants were semi-randomly assigned to a stimulus list so that there was an approximately equal number of participants tested using each list. The participants were tested individually in a quiet place in their university. Before the experiment, each participant was presented with written instructions in their native language (Dutch or Chinese), in which the task and procedure of the experiment were explained.

The experiment was administered using the software ZEP [16]. Each trial started with the presentation of a fixation cross ("+") on the computer screen, which lasted for 1000 ms. Then the fixation cross disappeared and the stimulus sentence in Chinese characters and Pinxin, its Dutch translation appeared on the screen, which lasted for 2400 ms. Then a sound icon (🔊) popped out below the target word (i.e., *mong*) and stayed on the screen for 400 ms. After that the sound of the stimulus sentence was played to participants via headphones. At the end of the sound file, the target word, the image of the button box, and the four response boxes appeared on the screen (Figure 2), representing the four lexical tones. The appearance of the response boxes acted as a signal to the participant that he or she could indicate the tone of the target word by pressing the corresponding button of the button box. The inter-trial interval was set at 500 ms. The experiment lasted on average about 10 minutes.

Figure 2: The layout of the perception experiment



3. STATISTICAL ANALYSIS AND RESULTS

The participants' responses were coded as either "correct" or "incorrect" for each trial depending on whether the selected tone was the intended one. A mixed-effect binary logistic regression in SPSS was used to assess the effects of the predictors on the perception of tone. The predictors included the main effects of four factors, i.e., EMOTION (*neutrality*, *happiness*, *anger*, *sadness*), TONAL CONTEXT (T1, T2, T3, T4), TONE of the target word (T1, T2, T3, T4), and PROFICIENCY (1-beginners, 2-intermediate-advanced; 3-native) and four two-way interactions (i.e., EMOTION x PROFICIENCY, TONAL CONTEXT x PROFICIENCY, TONE x PROFICIENCY, and EMOTION x TONE). Participants and words were included as random factors. The outcome variable was the perception of tone (i.e., *correct* vs. *incorrect*). Our modelling yielded three significant main effects, (i.e., EMOTION, TONE and PROFICIENCY), and three significant interactions, (i.e., EMOTION x TONE, TONE x PROFICIENCY, and TONAL CONTEXT x PROFICIENCY). The summary of the statistically significant results can be seen in Table 1.

Regarding the main effect of emotion with NEUTRALITY as the reference category, in general the participants were less accurate in tonal perception in the three non-neutral emotions, compared to the neutral condition (happiness: *coefficient* = -0.062, sadness: *coefficient* = -1.130, anger: *coefficient* = -1.815). However, only the difference between ANGER and NEUTRALITY reached statistical significance (*coefficient* = -1.815, *SE* = 0.558, t = -3.254, $p < .05$). Regarding the main effect of TONE of the target word with T1 as the reference category, the coefficients suggested that the participants performed less well in T2 (*coefficient* = -1.947, *SE* = 0.628, t = -3.100, $p < .01$), T3 (*coefficient* = -1.243, *SE* = 0.633, t = -1.965, $p < .05$), T4 (*coefficient* = -2.038, *SE* = 0.635, t = - 3.208, $p < .01$), compared to T1. The participants did worst in T4. However, the interaction of EMOTION x TONE revealed that the participants were more accurate in the perception of T4 than the perception of T1 in ANGER, suggesting that tonal perception can differ between tones in the same non-neutral emotion.

With respect to the main effect of proficiency with PROFICIENCY 3 as the reference category, the participants from PROFICIENCY 1 (beginners) and PROFICIENCY 2 (intermediate-advanced) did not differ from the native controls from PROFICIENCY 3. This is the case across tones and tonal contexts. However, when we ran the analysis again with PROFICIENCY 1 as the reference category, we found that the participants from PROFICIENCY 2 (*coefficient* = 2.670, *SE* = 0.996, t = 2.681, $p < .01$)

were significantly more accurate in tonal perception than the participants from PROFICIENCY 1. This advantage of a higher level of L2 proficiency was primarily driven by more accurate perception of T3 (*coefficient* = -1.728, *SE* = 0.823, *t* = -2.099， $p < .05$), compared to the perception of T1 and more accurate perception of tones when preceded by T4 in the carrier sentence (*coefficient* = -1.130, *SE* = 0.509, *t* = -2.222， $p < .05$), as shown by the interaction of TONE x PROFICIENCY and TONAL CONTEXT x PROFICIENCY.

Table 1: Summary of statistically significant results of the mixed model linear regression analysis. The reference category was INCORRECT for the outcome variable, NEUTRALITY for the fixed factor emotion, T1 for the fixed factor tone, PROFICIENCY 3 (native) for the fixed factor proficiency, TONAL CONTEXT 1 (i.e., the target word was preceded by a T1 word in the carrier sentence) for the fixed factor tonal context.

	Coefficient	Std. Error	t	Sig.	95% Confidence Interval	
					Lower	Upper
Intercept	2.810	0.583	4.819	.000	1.667	3.954
Anger	-1.815	0.558	-3.254	.001	-2.909	-0.721
Tone 4	-2.038	0.635	-3.208	.001	-3.284	-0.792
Tone 3	-1.243	0.633	-1.965	.050	-2.483	-0.002
Tone 2	-1.947	0.628	-3.100	.002	-3.179	-0.716
Proficiency 2	2.670	0.996	2.681	.007	0.717	4.624
Emotion*Tone						
Anger*Tone 4	2.751	0.717	3.839	.000	1.346	4.156
Tone*Proficiency						
Tone 3* Proficiency 2	-1.728	0.823	-2.099	.036	-3.343	-0.113
Tone 2*Proficiency 3	2.454	0.919	2.671	.008	0.652	4.256
Context*Proficiency						
Context 4*Proficiency 2	-1.130	0.509	-2.222	.026	-2.128	-0.133

4. CONCLUSIONS AND FUTURE RESEARCH

Different from previous studies which merely focused on the perception of Mandarin tones in isolated syllables produced in neutral emotion, the current study examined tonal perception in different tonal contexts in emotional speech by Dutch learners of Mandarin at different levels of proficiency. We have found that both the L2 Mandarin learners and native controls were significantly more accurate in tonal perception in the neutral emotion than in the angry emotion. These results indicate that emotions can influence tonal perception, especially those that would cause great changes to the citation form of the tone, as predicted. However, not all tones were equally difficult to perceive in the angry emotion. Tone 4 was more accurately perceived than Tone 1 in the neutral emotion. Furthermore, we have found that the intermediate-advanced L2 Mandarin learners were more accurate than the beginning L2 Mandarin learners. This difference appears to be driven by a difference in the perception of Tones 2, 3 and 4 when preceded by Tone 4 by the intermediate-advanced learners. Unexpectedly, we have found no evidence for more accurate tonal perception in the native controls than in the L2 Mandarin learners.

Our inspection of the percentage of correct responses showed that the native controls made a higher percentage of correct responses in all the fours emotions (Native controls: 24,35% in “anger”, 24,35% in “happy”, 24,74% in “sad”, 24,22% in “neutral”; intermediate-advanced learners: 21,74% in “anger”, 21,61% in “happy”, 22,27% in “sad”, 22,79% in “neutral”; beginning learners: 17,32% in “anger”, 17,19% in “happy”, 16,28% in “sad”, 19,01% in “neutral”). Nevertheless, these differences did not show up in our mixed-effect binary logistic regression. Future research involving more L2 learners and more stimuli is needed to validate whether the similarity in accuracy of tonal perception in our study can be replicated.

Another direction for future research is examining the effect of speaker gender. In this study, one female speaker recorded all the stimuli. However, there has been reports on gender-related differences in prosodic expression of emotion [8]. For example, the realization of anger was different in terms of amplitude between the male and female speakers. In addition, previous research suggests that female and male voices are processed differently [21]. Female voices might impede language processing which is attributed to the high acoustic salience and complexity of female voices [14]. In future research, it might be interesting to examine the relationship between the gender and the perception of tones in emotional speech in L2 learners and native speakers of Mandarin.

5. ACKNOWLEDGEMENTS

The authors wish to thank all participants from Chinese Education Center in the Netherlands, Leiden University, and Utrecht University. Many thanks go to Shuangshuang Hu, who helped to record the stimuli for the experiment.

6. REFERENCES

[1] Broselow, E., Hurtig, R. R., Ringen, C. 1987. The perception of second language prosody. In G. Ioup & S.H. Weinberger (eds.), *Inter-language Phonology, The Acquisition of Second Language Sound System* (pp. 350-361). Cambridge: Newbury House Publishers.

[2] Chao, Y. R. 1933. Tone and intonation in Chinese. *Bulletin of the Institute of History and Philology*, 4, 121-134.

[3] Chen, Q-H. 1997. Toward a sequential approach for tonal error analysis. *Journal of the Chinese Language Teachers' Association*, 32(1), 21-39.

[4] Francis, A. L., Ciocca, V., Ma, L., Fenn, K. 2008. Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, 36(2), 268-294.

[5] Gandour, J. T. 1983. Tone perception in far Eastern languages. *Journal of Phonetics*, 11, 149–175.

[6] Gandour, J., Xu, Y., Wong, D., Dzemidzic, M., Lowe, M., Li, X., & Tong, Y. 2003. Neural correlates of segmental and tonal information in speech perception. *Human Brain Mapping*, 20, 185–200.

[7] Hao, Y. C. 2018. Second language perception of Mandarin vowels and tones. Language and Speech, 61(1), 135-152.

[8] Klatt, D., & Klatt, L. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, 87(2), 820–857.

[9] Leather, J. 1987. F0 pattern inference in the perceptual acquisition of second language tone. In A. James & J. Leather (eds.), *Sound Patterns in Second Language Acquisition* (pp. 59-81). Dordrecht: Foris Publications.

[10] Ladd, D. R. 1992. An introduction to intonational phonology. In G. J. Docherty & D. R. Ladd (eds.), *Papers in laboratory phonology II: Gesture, segment prosody* (pp. 321–334). Cambridge: Cambridge University Press.

[11] Ladd, D. R. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.

[12] Li, A., Fang, Q., Dang, J. 2011. Emotional tone in a tone language: experimental evidence from Chinese. Icphs XVII, 17-21.

[13] Moore, C. B., Jongman, A. 1997. Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America*, 102, 1864 –1877.

[14] Magnuson, J. S., Nusbaum, H. C. 2007. Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance*, 33, 391–409.

[15] Vance, T. J. 1976. An Experimental investigation of tone and intonation in Cantonese. *Phonetica*, 33, 368-392.

[16] Veenker, T. J. G. 2017. The Zep Experiment Control Application (Version 1.10). Beexy Behavioral Experiment Software. Retrieved from http://www.beexy.org/zep/

[17] White, C. M. 1981. Tonal perception errors and interference from English intonation. *Journal of Chinese Language Teachers Association*,16, 27–56.

[18] Xu, Y., Gandour, J. T, Francis, A. 2006. Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *Journal of the Acoustical Society of America*, 120(2), 1063–1074.

[19] Yildirim, S., Bulut, M., Lee, C. M., Kazemzadeh, A., Busso, C., Deng, Z., et al. 2004. An acoustic study of emotions expressed in speech. *Interspeech*, (2004).

[20] Yang, C., Chan, M. K. (2010). The perception of Mandarin Chinese tones and intonation by American learners. *Journal of Chinese Language Teachers Association*, 45(1), 7-36.

[21] Yang, H., Yang, S., Park, G. 2013. Her voice lingers on and her memory is strategic: effects of gender on directed forgetting. *Plos One*, 8(5), e64030.