Can we sing the tones of a tonal language?
The duration of Mandarin Tone 3 and Tone 4 in the music context

Qianyutong Zhang[1], Lei Zhu[1], Xiaoming Jiang[1]

[1]*Institute of Linguistics, Shanghai International Studies University (China)*

Backgrounds: While pitch is the primary cue for tone realization and recognition [1], secondary cues like duration can play indispensable roles when the primary cue is insufficient or unavailable, e.g. in whispering speech [2-5]. Singing, like whispering, is a context where pitch realization is suppressed for tones, but it can be more complex in the sense that the suppression comes from a competition between tone and melody in which the latter, however incongruent with the former in height and contour, apparently wins. The actual realization of tones in singing being still largely unknown, this study endeavors to explore whether speakers maintain intrinsic duration differences of Mandarin tones in singing where pitch is regulated.

In the normal speech of Mandarin Chinese, Tone 3 has the longest duration and Tone 4 the shortest [6-7]. However, this pattern may vary as the register changes. [8] compared the durations of Mandarin tones in monosyllable reading and casual conversational speech, and found that in the latter environment, the four lexical tones become shorter and tend to be similar in length. Therefore, while focusing on Tone 3 and Tone 4, the two Mandarin tones furthest apart in length, this study sets up a monosyllable-singing and a sentence-singing session as two contrastive environments for tone realization, taking the possible influence of register into consideration.

Methods: Six native Mandarin speakers were recruited to record the singing tones (3 female; mean age = 24.83). Twelve words comprised of three syllables (/ta//ti//tu/) with four Mandarin tones were set as materials. Tone 3 and Tone 4 were target tones, and Tone 1 and Tone 2 were irrelevant stimuli to prevent fatigue in tone production. The stimuli were presented in the form of Chinese characters (see Table 1). The carrier sentence, 这是 X 字 ("This is the word X"), was designed to place isolated syllables in a continuous speech. The regulated musical notes for target tones were G3 (≈197Hz), A3 (≈221Hz), and B3 (≈247Hz), considering the common vocal range of both male and female speakers. For other words in the carrier sentences, the note A3 is assigned. There were 36 stimuli per part (3 syllables × 4 tones × 3 notes). For each stimulus, the participants first read it and then listened to the notes (piano timbre), each note lasting 500ms. The speaker then sang the word/sentence three times according to the notes played. The tone of each syllable in all recordings were labeled in Praat [9]. The average duration of each labeled tone was extracted in Praat, and normalized among speakers. Linear mixed effects were modeled in R using the "lme4" package for tonal duration [10].

Results: Table 2 shows the average durations of Tone 3 and Tone 4 in the two sessions. In the monosyllable-singing session, the duration of Tone 3 was significantly longer than that of the Tone 4 (p = 0.008, see Table 3 and Fig. 1). Musical note has no significant main effect and no significant interaction effect with Tone. In the sentence-singing session, no significant difference in duration was found between Tones 3 and 4, nor was there any significant interaction between Musical note and Tone (see Table 3 and Fig. 2).

Conclusion: In general, this study found that in singing contexts where single musical notes dominate pitch realization, the tones that are sung still follow their internal duration patterns to a certain extent, regardless of melody accompaniment. In monosyllabic words, the average duration of Tone 3 was significantly higher than that of Tone 4, as in normal speech. The result was not affected by the change of musical notes, indicating that tone duration does not covary with pitch, and that duration patterns of tones may have been internalized. However, when positioned in a singing sentence, the tones showed no significant differences in duration, a finding consistent with previous research on continuous speech. The duration of each syllable was shortened, as were the differences between variant tones, both in normal speech and in singing. Future studies are needed to explore the dynamic F0 curves with a focus on their variation during tone production in different singing contexts.

Table 1. Word list for recording.

	Tone 1	Tone 2	Tone 3	Tone 4
/ta/	搭	答	打	大
/ti/	低	笛	底	地
/tu/	督	读	堵	度

Table 2. The mean durations of Tone 3 and Tone 4 in two sessions.

Session	Mean Duration（ms）	
	Tone 3	Tone 4
Monosyllable singing	448.95	414.45
Sentence singing	360.70	349.77

Table 3. Results of linear mixed model of Tone 3 and Tone 4 duration in two sessions.

Session	Sum Sq	Mean Sq	Num DF	Den DF	F value	Pr(>F)
Monosyllable singing	0.056	0.221	1	16.178	9.242	0.008**
Sentence singing	0.019	0.019	1	15.997	0.945	0.346

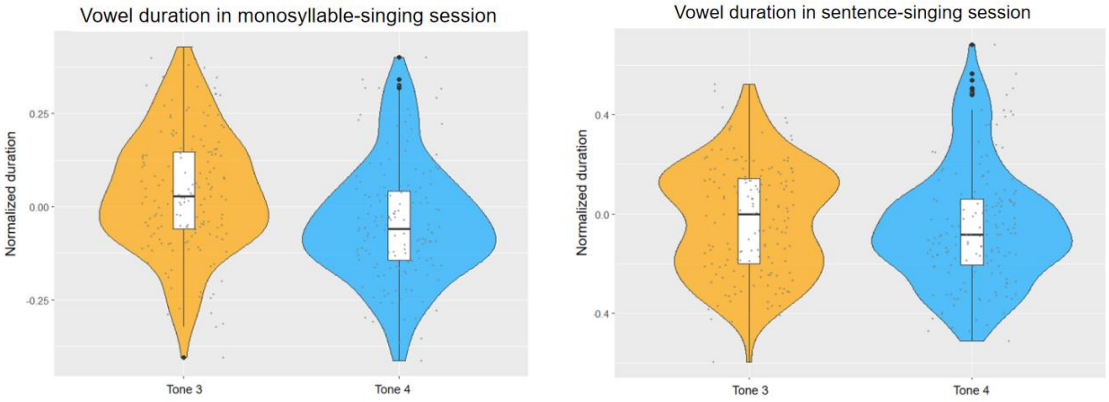

Fig. 1. Durations of Tone 3 and Tone 4 in monosyllable-singing and sentence-singing sessions.

References

[1] Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of phonetics*, 11(2), 149-175.

[2] Heeren, W. F., & Lorenzi, C. (2014). Perception of prosody in normal and whispered French. *The Journal of the Acoustical Society of America,* 135(4), 2026-2040.

[3] Higashikawa, M., & Minifie, F. D. (1999). Acoustical-perceptual correlates of" whisper pitch" in synthetically generated vowels. *Journal of Speech, Language, and Hearing Research,* 42(3), 583-591.

[4] Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America,* 119(5), 3059-3071.

[5] Llanos, F., Dmitrieva, O., Shultz, A., & Francis, A. L. (2013). Auditory enhancement and second language experience in Spanish and English weighting of secondary voicing cues. *The Journal of the Acoustical Society of America*, 134(3), 2213-2224.

[6] Ho, A. T. (1976). The acoustic variation of Mandarin tones. *Phonetica*, 33(5), 353-367.

[7] Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, 49(1), 25-47.

[8] Yang, J., Zhang, Y., Li, A., & Xu, L. (2017). On the Duration of Mandarin Tones. *In INTERSPEECH* (pp. 1407-1411).

[9] Boersma, P. (2011). Praat: doing phonetics by computer *[Computer program]*. http://www. praat. org/.

[10] Bates, D., Mächler, M., Bolker, B., Walker, S., Christensen, R. H., Singmann, H., & Dai, B. (2015). lme4: Linear mixed-effects models using Eigen and S4. *R package version* 1.1–7. 2014.