A LARYNGEAL ULTRASOUND STUDY OF SINGAPOREAN MANDARIN TONES

Dawn Poh Zhi Yun[1], Scott Reid Moisik[1]

[1]*Linguistics and Multilingual Studies, Nanyang Technological University, Singapore*
zhiyun001@e.ntu.edu.sg, scott.moisik@ntu.edu.sg

ABSTRACT

Previous acoustic studies on Mandarin tones have focused on mainland (Beijing or Taiwan) dialects. With the aid of laryngeal ultrasound, this exploratory study provides an acoustic and articulatory examination of Singapore Mandarin and corroborates the findings of previous studies into the f_o of its tones. The relationship between vertical and horizontal movements of the larynx is then examined in relation to f_o. Generally larynx height patterns with f_o, but there was an inverse relationship between larynx height and f_o for the production for some tones. This could be attributed to laryngeal medialisation, which may be a mechanism Singapore Mandarin speakers employ in the lowering of f_o.

Keywords: Tone, laryngeal articulation, laryngeal ultrasound, Mandarin, Singapore.

1. INTRODUCTION

While the acoustic aspects of Mandarin tones (T1 = high level, T2 = mid rising, T3 = falling (rising), T4 = high falling) are well-documented [8, 11, 24], most of the work has only examined the varieties of Mandarin spoken in Beijing or Taiwan. In comparison, studies on Singaporean Mandarin (SgM) tones remain relatively scant [1]. Acoustic research has revealed non-trivial differences between SgM and other varieties of Mandarin. Lee's [13] study on SgM tones found variations in syllable duration and f_o contour compared to Beijing and Taiwan Mandarin. In addition, while T2 in these varieties remains level or falls slightly before rising towards the end, SgM T2 exhibits "a mid-low level stretch not found in the other varieties" [13]. The realisation of T3 in SgM is similar to that of Taiwan Mandarin: starting at a mid-high level, f_o falls steadily to a low level without a distinct final rise characteristic of Beijing Mandarin. In addition to f_o differences between male and female speakers (as is expected on physiological grounds), Lee's analysis also revealed systematic patterns of alignment related to gender, with male speakers producing a longer mid-low plateau in T2. While this could be a signal of gender identity, it could also be attributed to the biomechanics of pitch production.

As [9] asserts, "the characteristic patterns of vocal fundamental frequency in speech are derived from the morphology of the human speech organs". The production of tone is a physiologically complex process, involving the use of various laryngeal and extra-laryngeal mechanisms. Vertical larynx movement during phonation has previously been observed using X-ray [3, 2], electromyographic (EMG) [10, 16], and magnetic resonance imaging [7]. Larynx height in lexical tone production has been based on inferences made with EMG [18, 4, 6]. Direct observation of citation tone production in Mandarin has been made using simultaneous laryngoscopy and laryngeal ultrasound [15], which describes two strategies for low tone production: (i) larynx lowering and (ii) raising with (epi)laryngeal constriction. It was also found that the larynx rises continuously during T1 production. In [19], electromagnetic articulography evidence shows that tongue height for Mandarin /i/ and /a/ varies in relation to T2 and T3.

The present study aims to provide an acoustic and articulatory characterisation of SgM citation tones. Given the differences in f_o contour between SgM and other varieties of Mandarin [13], it may be that SgM also exhibits

articulatory differences. Not only would articulatory data contribute towards the growing corpus in SgM tone data, such data could also generate further empirical and theoretical questions concerning the physiology of tone production more generally.

2. METHODOLOGY

2.1 Speakers, speech material, and equipment

Eleven native speakers of SgM (7 females) participated in this study. The age range of the participants was 21 to 30 years old (*M*=24.4, *SD*=2.5). A language background questionnaire was used to ensure that the participants spoke SgM regularly and were proficient in reading Mandarin characters.

Participants were asked to produce three iterations of three maximally distal vowels (/i/-/a/-/u/) in bilabial- and alveolar-initial CV monosyllables, each repeated across the four lexical tones. We omitted combinations not associated with a meaningful word. In total, 53 target words were elicited, generating (53 tokens × 3 iterations × 11 speakers =) 1749 trials.

Audio signals were captured using an AT3035 Audio-Technica cardioid condenser microphone (powered by a Focusrite Scarlett Solo audio interface using 16-bit sampling at 44.1 kHz). Laryngeal ultrasound video data was collected using a L12-5L40S 64 (40 mm, 5–12 MHz) linear ultrasound transducer following [15] and the SonoSpeech micro ultrasound system operated via AAA (Articulate Instruments), which automatically synchronised all signals. The transducer was placed sagittally adjacent to the laryngeal prominence of the speakers.

2.2 Post-processing

Vertical and horizontal changes in larynx position were obtained using optical flow analysis, which tracks movement in a series of video frames as a time-varying vector field [15]. Here we used the `imregdemons()` algorithm in MATLAB 2018b on frame pairs (reduced to 25% original size) to obtain the displacement fields. The vertical and horizontal velocity signals were then obtained using the `trimmean()` function set to exclude the lowest and highest 25% values of the vertical and horizontal components of the fields. Only the superficial-most upper half of the video was used so as to exclude vocal fold flutter and other noise. Displacement over time was then computed via numerical integration using the `cumtrapz()` function. Using Praat, the voiced portion of each audio token was automatically segmented for analysis, and the f_o contour was obtained. Based on the segmentation, the relevant portion of the larynx displacement signals were then extracted. All signals were time-normalised. Manual checking was performed on 10% of the optical flow analysis to ensure that the segmentations were accurate.

2.3 Generalised Additive Mixed Modelling

Since the data consists entirely of time-series observations (f_o, larynx height, larynx medialisation) for each token, we employ Generalised Additive Mixed Modeling (GAMM) [20, 22] using the `mgcv` package [23] in R. We ran three separate models with (i) f_o, (ii) vertical larynx displacement, and (iii) larynx medialisation as DVs. For IVs, each model included difference smooth terms for tone, sex, and vowel quality (along with the associated parametric terms). No interactions were tested. To account for non-linear variation among the tones produced by participants, we used by-participant and by-tone random smooths. With the help of preliminary models, we then decided to incorporate an AR1 model to account for autocorrelation of residuals and fit using the scaled-t family to address non-normality of the residuals. Variable selection was made using shrinkage smoothers [14]. Model diagnostics and residual analysis (via `gam.check()` in `itsadug` [17]) indicated that final models were generally nearly normal in residual distribution and sufficient basis dimensions were used for the smooths. Deviance explained was (i) 72%, (ii) 37%, and (iii) 55%, suggesting reasonably good fits. Because of space limitations, we do not

report full results from GAMMs, instead relying on smooth plots using the confidence intervals.

3. RESULTS

Fig. 1 shows smooths for f_o. The f_o contours observed in this study are consistent with previous findings for SgM [13]. All smooth terms are significant except that contour shape does not significantly differ between T1 and T3. We also find the typical intrinsic f_o of vowels [21], with [i] and [u] significantly higher (by 11 and 15 Hz respectively) than the low vowel [a]. f_o across all tones appears to lower sharply initially, which may be an anticipatory effect associated with citation form. This effect was more evident for female speakers. The tones can be seen to cluster into two regions at onset and offset.

Figure 1: GAM smooth plots (nonlinear estimate plus confidence intervals) for f_o by sex and vowel (T1 = solid, T2 = dashed, T3 = dotted, T4 = dashed-dotted).

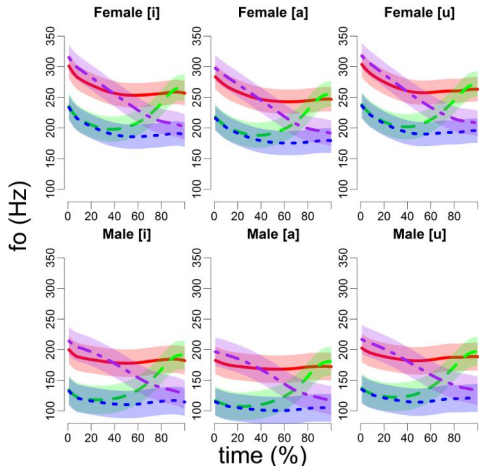


Fig. 2 shows smooths (all significant) for larynx height (vertical component), ranging around 6 mm for females and 7 mm for males. The [u] context is (significantly) lower by about 2 mm; [a] and [i] are virtually identical. The patterns are generally as expected except that male T1, T3, and T4 exhibit unanticipated raising towards tone offset (most noticeable in [u] context). Fig. 3 shows smooths (again, all significant) for larynx medialisation. The estimated extent of displacement is approximately 1 mm for both sexes, but the wide confidence intervals suggest large individual variation. T3 shows the greatest medialisation and T1 the least (with T2 and T4 being intermediate). This suggests that medialisation is serving low tone production (possibly in substitution for lack of falling larynx height during voicing).

Figure 2: Larynx height. (See Fig. 1)

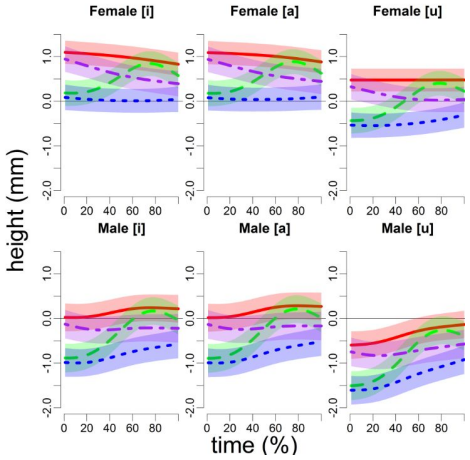
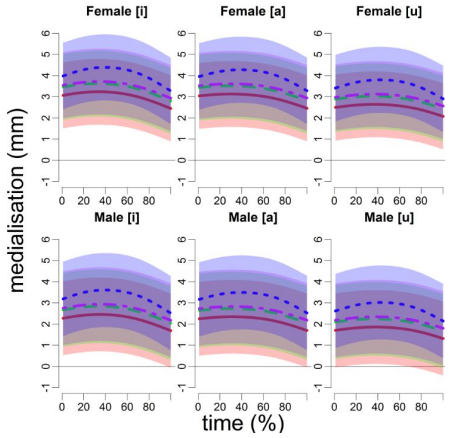


Figure 3: Larynx medialisation. (See Fig. 1)



4. DISCUSSION

4.1 f_o and larynx height

The organisation of pitch levels into two distinct onset and offset zones is consistent with Lee's findings [13]: similar to Beijing or Taiwan Mandarin, f_o contours for T1 and T4 begin with a high onset, with T4 lowering into a low offset. T2 however exhibited a slight lowering before

rising to a high offset, while T3 seems to follow a lowering pattern without the 'dipping' characteristic of Beijing Mandarin. When examined in relation to larynx height however, some notable differences emerge: while the larynx height contours generally reflect the pattern of the corresponding f_o contours for T1 and T2, larynx height for T4 appears to have a lower onset than T1 and a higher offset than T3. Additionally, laryngeal movement for T3 appears to diverge from its corresponding f_o contour—rather than ending on a level pitch as indicated by the f_o measurements, laryngeal height contour for T3 instead shows a slight rise towards the end of phonation (especially for males). While T3 in standard Mandarin speakers is associated with "larynx lowering, which reaches its nadir roughly halfway through the tone where the f_o is also low; the larynx then ascends in correspondence with the rising f_o at this point" [15], the laryngeal movement of SgM differs in that it does not correspond to the continual fall in f_o. Such an effect indicates the presence of other mechanisms other than larynx height involved in f_o lowering of T3.

4.2 Vowel effects on larynx height

In [19], it was found that tongue height is low for [a] but high for [i] for low onset tones. Likewise, given the coarticulatory relationship between lingual and laryngeal gestures [12], it is expected that larynx height should show a similar conditioning by physiological factors. The current study did not, however, find any effect of [i] (in distinction to [a]) on larynx height. A general lowering effect was found for [u]. The grouping of [i] together with [a] as higher in larynx height than [u] is consistent with the findings of older articulatory studies (e.g., [5]). This goes against the high-vowel grouping by intrinsic f_o, suggesting that hyoid pull (associated with [i] and [u]) is more probably the cause of this effect.

4.3 Laryngeal medialisation and tone

To our knowledge, the medial movement of the larynx is a novel finding of the present study. From the above data, medialisation appears to partially correlate with low f_o—hence, the relatively low medialisation in the production of T1 is unsurprising as the degree of vertical laryngeal movement and the f_o associated with the tone are comparatively high and level. The degree of medialisation for T2, T3, and T4 generally inversely reflect the corresponding f_o contours: the most medialisation is found in T3, which also exhibits the lowest mean f_o. It should be noted that medialisation does not reflect adductory action of the vocal folds for two reasons. First, the region of interest used for optical flow analysis was quite superficial, going only as far as about the depth of the thyroid cartilage. Second, the vocal folds should be adducted by phonation onset, yet medialisation can increase during tone production. The medialisation appears as drift and even expansion of the superficial laryngeal structures and infrahyoid musculature. While the medialisation may be related epilaryngeal constriction [15], which should be accompanied by changes to phonatory quality (yet to be assessed in our data), the impression is that the thyroid laminae become approximated.

5. CONCLUSION

Our acoustic results correspond to the findings of previous research on SgM, indicating that the citation tones of this variety differ from those of standard Mandarin. To fully understand the f_o contours of SgM tones, it will be beneficial to take into account contextual variation by examining sentential utterances (data for which has been collected) in addition to the citation forms. A comparison between f_o contours and laryngeal movements revealed some articulatory effects: the syllable-final raising of the larynx in the production of T3 does not correspond to the f_o contours, suggesting that some other articulatory mechanisms might be involved in pitch lowering. One such mechanism may possibly be larynx medialisation. More refined techniques looking specifically at medialisation should be considered for future studies.

6. REFERENCES

[1] Chen, N.F., Wee, D., Tong, R., Ma, B. and Li, H. 2016. Large-scale characterization of non-native Mandarin Chinese spoken by speakers of European origin: Analysis on iCALL. *Speech Communication* 84, 46–56.

[2] Chiba, T. and Kajiyama, M. 1941. *The Vowel: Its Nature and Structure,*. Tokyo-Kaiseikan Pub. Co., Ltd.

[3] Curry, R. 1937. The mechanism of pitch change in the voice. *Journal of Physiology* 91, 254–258.

[4] Erickson, D. 2011. Thai tones revisited. *Journal of the Phonetic Society of Japan* 15, 2, 74–82.

[5] Ewan, W.G. 1979. Laryngeal behaviour in speech. *Report of the Phonology Laboratory*.

[6] Hallé, P.A. 1994. Evidence for Tone-Specific Activity of the Sternohyoid Muscle in Modern Standard Chinese. *Language and Speech* 37, 2, 103–123. DOI:https://doi.org/10.1177/002383099403700201.

[7] Hirai, H., Honda, K., Fujimoto, I. and Shimada, Y. 1994. Analysis of magnetic resonance images on the physiological mechanisms of fundamental frequency control. *Journal of the Acoustical Society of Japan* 50, 296–304.

[8] Ho, A.T. 1976. The acoustic variation of Mandarin tones. *Phonetica* 33, 5, 353–367 DOI:https://doi.org/10.1159/000259792.

[9] Honda, K. 1995. Laryngeal and extra-laryngeal mechanisms of F0 control. *Producing speech: Contemporary issues: For Katherine Safford Harris*. F. Bell-Berti and L.J. Raphael, eds. AIP Press.

[10] Honda, K., Masaki, S. and Shimada, Y. 2000. Observation of laryngeal control for voicing and pitch change by magnetic resonance imaging technique. (Beijing, China).

[11] Howie, J.M. 1976. *Acoustical studies of Mandarin vowels and tones*. Cambridge University Press.

[12] Kühnert, B. and Nolan, F. 1999. The origins of coarticulation. *Coarticulation: Theory, Data and Techniques*. W. Hardcastle and N. Hewlett, eds. Cambridge University Press. 8–30.

[13] Lee, L. 2010. The tonal system of Singapore Mandarin. *Proceedings of the 22nd North American Conference on Chinese Linguistics (NACCL-22) & the 18th International Conference on Chinese Linguistics (IACL-18)*, (Cambridge, MA), 345–362.

[14] Marra, G. and Wood, S.N. 2011. Practical variable selection for generalized additive models. *Computational Statistics & Data Analysis* 55, 7, 2372–2387. DOI:https://doi.org/10.1016/j.csda.2011.02.004

[15] Moisik, S.R., Lin, H. and Esling, J.H. 2014. A study of laryngeal gestures in Mandarin citation tones using simultaneous laryngoscopy and laryngeal ultrasound (SLLUS). *Journal of the International Phonetic Association* 44, 01, 21–58. DOI:https://doi.org/10.1017/S002510031300003 27.

[16] Ohala, J. and Hirose, H. 1969. The function of the sternohyoid muscle in speech. *Reports of the 1969 Autumn Meeting of the Acoustical Society of Japan*, 359–360.

[17] van Rij, J., Wieling, M., Baayen, R.H. and van Rijn, H. 2016. *itsadug: Interpreting Time Series and Autocorrelated Data using GAMMs*.

[18] Sagart, L., Hallé, P., Boysson-Bardies, B. de and Arabia-Guidet, C. 1986. Tone production in Modern Standard Chinese: An electromyographic investigation. *Cahiers de Linguistique Asie Orientale* 15, 2, 205–221.

[19] Shaw, J.A., Chen, W.R., Proctor, M.I. and Derrick, D. 2016. Influences of tone on vowel articulation in Mandarin Chinese. *Journal of Speech, Language, and Hearing Research* 59, 6, 1566–1574.

[20] Sóskuthy, M. 2017. Generalised additive mixed models for dynamic analysis in linguistics: a practical introduction. *arXiv:1703.05339 [stat]*.

[21] Whalen, D.H. and Levitt, A.G. 1995. The universality of intrinsic F0 of vowels. *Journal of Phonetics* 23, 349–366.

[22] Wieling, M. 2018. Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics* 70, 86–116. DOI:https://doi.org/10.1016/j.wocn.2018.03.002.

[23] Wood, S.N. 2017. *mgcv: mixed GAM computation vehicle with automatic smoothness R package*.

[24] Xu, Y. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25, 1, 61–83. DOI:https://doi.org/10.1006/jpho.1996.0034