Acoustic characteristics of bilingual children's Mandarin tone productions that are important for tone identification: Effects of home language and schooling experiences

Youran Lin[1], Karen Pollock[1], and Fangfang Li[2]

[1]University of Alberta (Canada), [2]University of Lethbridge (Canada)

Tone categories are regarded as early established among Mandarin-native children [1]. However, tone productions involve complex acoustic characteristics. Children take more than five years to refine their tonal characteristics in the time domain (i.e., pitch and duration [2,3]) and more than ten years in the spectral domain (i.e., voice quality [4]). These nuanced characteristics in tone productions contribute to listeners' tone identification [4-6]. However, these have barely been researched among second language (L2) learners [7], especially child learners. Research among child learners allows us to observe L2 phonetic learning in progress, which is influenced by the L2 speech input across contexts [8].

The current study addresses two research questions: (1) How do the multidimensional acoustic characteristics of child Mandarin learners' tone productions influence listeners' tone identification? (2) How is child learners' production of tonal characteristics related to language experiences at home and at school? The study concerns a unique population of students in Chinese bilingual education programs in Canada. We hypothesize that (1) children utilize both time- and spectral-domain characteristics in tone productions, and (2) children with more input at home and school utilize acoustic characteristics more effectively, which facilitates listeners' tone identification.

Participants included 82 students in grades 1, 3, and 5 at Chinese bilingual schools in Canada and 12 Chinese teachers at the same schools. Students were divided into two groups based on home language histories, 38 heritage language (HL) speakers with early Mandarin input (N_{Grade1} = 15, N_{Grade3} = 11, N_{Grade5} = 12) and 44 L2 learners from English-speaking households (N_{Grade1} = 16, N_{Grade3} = 14, N_{Grade5} = 14). Productions of 32 monosyllabic words that involved the four Mandarin tones were elicited [1,9]. Speech samples were transcribed by four native speakers, and tone boundaries were labeled [2]. More than 50 acoustic measures were extracted using ProsodyPro [10], including time-domain measures such as mean fundamental frequency (*f0*), *f0* excursion (range), and duration and spectral-domain measures such as cepstral peak prominence (cpp) and harmonicity. *F0* values were modeled with a second-order polynomial function using numpy [11], the first- and second-order coefficients representing the slope and curvature of the *f0* contour, respectively [3]. Acoustic characteristics of bilingual children's tone productions were used to predict listeners' tone identification using randomForest [12]. If acoustic measures explained a large amount of variance in tone identification, it would be interpreted that children utilized these acoustic characteristics effectively, which facilitated listeners to take these acoustic cues for their tone identification.

Variances explained differed across domains and speaker groups (Table 1). Time-domain characteristics explained more variances than spectral-domain. More variances were explained for HL's productions than L2 and for students in higher grades. However, the variances explained in students' productions never approached the teachers' productions. The top three characteristics that served as the most important cues for listeners' tone identification were identified for each tone across speaker groups (Figure 1). Listeners relied on category-specific cues to identify tones produced by bilingual children. For example, the identification of Tone 1 (high-level tone) relied on excursion, minimal f0, and harmonicity (i.e., small pitch range, high pitch, and modal phonation). The cues for each tone were similar across speaker groups, although listeners relied on different cues to identify Tone 2 produced by L2 (i.e., location of maximum instead of minimum *f0*).

This study is among the first to address the linkages between child learners' tone productions and listeners' perception of such productions, as well as their association with learners' language experiences. Results suggest that acoustic characteristics of tones take years to be refined even for school-aged learners, which is in line with the evidence among younger monolingual Mandarin speakers [2-4]. When identifying Mandarin tones produced by bilingual children, listeners rely more on time-domain cues while integrating spectral cues, which indicates that the acquisition and production of time-domain tonal characteristics are more important for bilingual learners. Results also demonstrate the positive effects of both home and school input on students' production of tonal characteristics, which suggests a life-long development of bilingual speech [6].

Table 1. Variances in listeners' tone identification explained by time-domain and spectral-domain tonal characteristics produced by speakers from different language backgrounds and grade levels.

% Variance explained	HL			L2		
	Time-domain cues	Spectral-domain cues	Combined cues	Time-domain cues	Spectral-domain cues	Combined cues
Teacher	74.37	50.27	73.52			
Grade 1	35.86	18.41	33.17	15.47	9.62	18.14
Grade 3	51.46	29.43	52.45	18.18	9.57	20.40
Grade 5	55.20	35.19	55.27	24.42	16.15	26.77

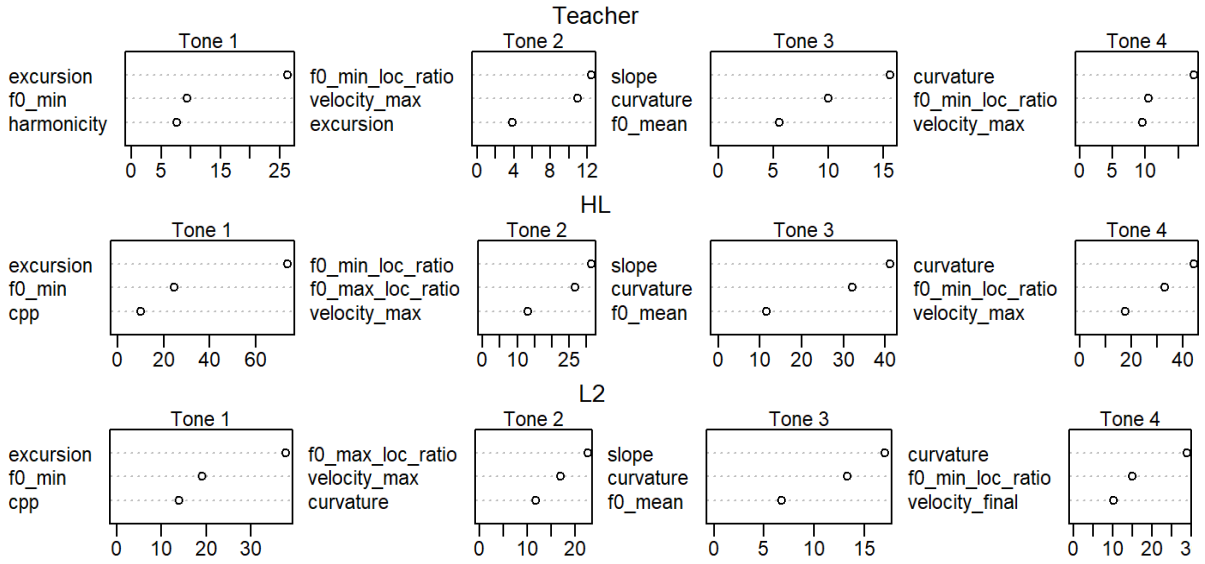

Fig. 1. The three most important acoustic characteristics for listeners' identification of each tone produced by teachers, HL students, and L2 students.

References

[1] Zhu, H. (2002). *Phonological development in specific contexts: Studies of Chinese-speaking children*. Multilingual Matters.

[2] Wong, P. (2012). Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic Mandarin lexical tone productions. *Journal of Phonetics, 40*(1), 141-151.

[3] Xu Rattanasone, N., Tang, P., Yuen, I., Gao, L. & Demuth, K. (2018). Five-year-olds' acoustic realization of Mandarin tone sandhi and lexical tones in context are not yet fully adult-like. *Frontiers in Psychology*, 9, 817.

[4] Rhee, N., Chen, A. & Kuang, J. (2019). Integration of spectral cues in the development of Mandarin tone production. In *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 3135-3138). Australasian Speech Science and Technology Association Inc.

[5] Blicher, D. L., Diehl, R. L., & Cohen, L. B. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement. *Journal of Phonetics, 18*(1), 37-49.

[6] Massaro, D. W., Cohen, M. M., & Tseng, C. Y. (1985). The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese. *Journal of Chinese Linguistics,* 267-289.

[7] Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America, 113*(2), 1033-1043.

[8] Flege, J. E., & Bohn, O.-S. (2021). The Revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3-83). Cambridge University Press.

[9] Zhao, J., & Bernhardt, M. B. (2012). *Mandarin single-word elicitation tool for phonology* ©, SSHRC.

[10] Xu, Y. (2013). ProsodyPro-A Tool for Large-scale Systematic Prosody Analysis. *TRASP*, Aix-en-Provence, 7-10.

[11] Harris, C. R., Millman, K.J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., …, Oliphant, T. E. (2020). Array programming with NumPy. *Nature, 585*, 357-362.

[12] Liaw, A. & Wiener, M. (2002). Classification and Regression by randomForest. *R News 2*(3), 18-22.