ACOUSTIC ANALYSES OF MANDARIN TONES PRODUCED BY BILINGUAL ELEMENTARY STUDENTS IN CANADA

Youran Lin[1], Karen Pollock[1], Benjamin Tucker[1,2], Fangfang Li[3]

[1]University of Alberta, Canada, [2]Northern Arizona University, USA, [3]University of Lethbridge, Canada
youran1@ualberta.ca, karen.pollock@ualberta.ca, benjamin.tucker@nau.edu, fangfang.li@uleth.ca

ABSTRACT

In Canada, children can learn Mandarin in bilingual schools as a heritage language (HL) or second language (L2). Many Canadian students have greater difficulty acquiring Mandarin tones than other aspects of the language. The present study investigates bilingual students' productions of Mandarin citation tones. Pitch contours are modelled using generalized additive mixed models for HL and L2 learners across grades to examine the roles of speech input. Results suggest that early home language experiences and recent school learning experiences interact with the tone targets and impact students' tone productions differently. These results provide evidence on Mandarin phonetic learning in Canada among bilingual students from diverse backgrounds and expand the evidence for L2 phonetic learning theories in the suprasegmental domain among child learners of non-English languages.

Keywords: Mandarin, tone, acoustic, bilingual, children

1. INTRODUCTION

In Western Canada, English is the majority language, but two-way bilingual education is available in many international languages. The current study concerns a Chinese-English bilingual program in Edmonton, Alberta. This program delivers half of its academic content in Mandarin Chinese and the other half in English, thus attracting both students who learn and maintain Mandarin as a heritage language (HL) and those who learn Mandarin as a second language (L2).

It has been argued that a two-way design supports students with diverse backgrounds to develop functional bilingualism [1]. Empirical evidence from Indo-European languages suggested that English-speaking children can learn the pronunciation of a minority L2 similarly to their HL peers, despite their different home language input [2, 3]. However, little is known about how school-aged children in an English-dominant environment learn a minority language from another language family, for example, a tonal language like Mandarin.

Mandarin tones primarily use pitch contours to differentiate meanings, which can be measured acoustically by a sequence of fundamental frequency (*f0*) over time. There are four citation tones in monosyllabic productions: high-level ◌̄ (T1), mid-rising ◌́ (T2), low-dipping ◌̌ (T3), and high-falling ◌̀ (T4) tones. Pitch contours of Mandarin tones are often described in a 5-degree convention, i.e., T1 [55], T2 [35], T3 [214], and T4 [51], where a larger number represents a higher pitch. The developmental order is typically T1 > T4 > T2 > T3, where T2 and T3 are confusable due to their phonetic similarity [4, 5].

Not only the phonetic characteristics of the targets but also the amount and quality of speech input play a role in L2 phonetic learning, as suggested by the Speech Learning Model (SLM) [6]. Furthermore, early input and recent input may contribute differently, which could be attributed to differences in neurocognitive maturation [7] or differences in the development of phonetic categories [6]. The Native Language Magnet theory (NLM) claims that the perception of phonetic categories is biased by language-specific input early in life, which impacts perceptual learning of other languages later in life [8]. Therefore, we hypothesize that bilingual students' tone learning is influenced differently by early home language input and recent school input, as well as by the specific tone targets.

In this paper, we examine the four citation tones produced by students in a Chinese bilingual program in Canada. We compare pitch contours between HL and L2 groups and between grades 1 and 3 to determine the roles of early home input and recent school input: HL and L2 students differ in their home language input; Grade 1 and grade 3 students differ in their amount of Mandarin input at school. By examining the learning of Mandarin tones, this paper expands the evidence for L2 speech learning theories in child learners of a non-English language. To eliminate the impact of tone errors on pitch contours, this study only examines the correct productions, thus investigating the phonetic refinement of tones instead of the perceived accuracy.

2. METHOD

2.1. Participants

HL and L2 students were recruited from grade 1 (age (months) $\mu = 77.35$, $\sigma = 3.65$) and grade 3 ($\mu = 103.16$,

σ = 4.13). Parent questionnaires [9, 10] indicated that the HL group (N = 26) has strong home input in Mandarin and a wide range of English exposure before school, and the L2 group (N = 30) receives predominantly home input in English. Table 1 depicts the differences between groups with regard to the onset of exposure to both languages, the time their parents spend speaking both languages, and their parents' proficiencies in both languages.

	HL	L2
Onset age of regular exposure to Mandarin (months)	0 – 1	37 – 96
Onset age of regular exposure to English (months)	0 – 93	0 – 1
Parent % of time speaking Mandarin*	50 – 100	0 – 15
Parent % of time speaking English	0 – 100	80 – 100
Parent self-reported Mandarin proficiency on a scale of 0-5*	4 – 5	0 – 3
Parent self-reported English proficiency on a scale of 0-5	1 – 5	4 – 5
* The parent who is stronger at Mandarin.		

Table 1: A comparison of the home language environment of HL and L2.

In addition to the students described above, we recruited 12 Chinese teachers from the bilingual program to provide their speech samples as a reference. Among them, seven were native (L1) speakers of Mandarin, five were L1 speakers of another Chinese language and started learning Mandarin at school age, and two were born in Canada and graduated from the bilingual program.

2.2. Procedures

2.2.1 Speech sample collection

A picture-based elicitation task [11, 12] was used to elicit a list of 72 words. Three Mandarin-native speakers who were proficient in English administered the task. Spontaneous productions were encouraged but imitative models were provided when necessary. Two examples for each tone were selected based on the spontaneity of production (Table 2). Therefore, in total 544 productions of HL, L2, and teachers were collected (68 speakers × 8 words = 544 words).

Tone	Word	Pinyin	IPA	Meaning	Spontaneity
T1 ˉ	三	sān	[san^{55}]	three	98%
	八	bā	[pa^{55}]	eight	96%
T2 ˊ	蓝	lán	[lan^{35}]	blue	92%
	鱼	yú	[y^{35}]	fish	86%
T3 ˇ	手	shǒu	[ʂo͡ʊ^{214}]	hand	95%
	狗	gǒu	[ko͡ʊ^{214}]	dog	92%
T4 ˋ	二	èr	[ɐ˞^{51}]	two	98%
	绿	lǜ	[ly^{51}]	green	94%

Table 2: Selected target words for the four tones.

2.2.2 Transcription

Speech samples were transcribed by four Mandarin-speaking researchers [13]. Each tone production was transcribed as either one of the four citation tones, an inappropriate allophonic variation [5], or an uncategorizable production. Spontaneity was coded. In the whole wordlist, 23% of the samples were transcribed by a second transcriber and reached 90% inter-transcriber reliability. The first author re-examined the selected words and identified questionable transcriptions. Transcribers reached full consensus through blinded voting, revising 44 out of the 47 questionable transcriptions. Only spontaneous productions that were recognized as correct were included, which left 406 productions, 317 from students and 89 from teachers. The accuracy of each tone by each group is presented in Figure 1.

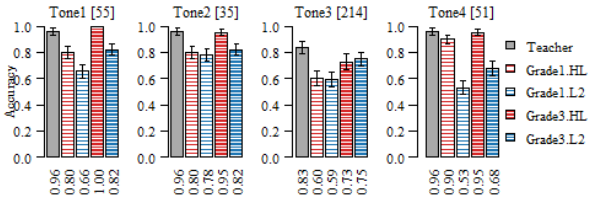

Figure 1: Accuracy of the four lexical tones among teachers and students with 95% confidence intervals.

2.2.3 f0 extraction

The onset and offset of pitch contours were labelled, including the nucleus and voiced coda. The first and last four cycles were excluded to avoid perturbations [4]. ProsodyPro was used to extract *f0* [14], where automatically recognized voicing pulses were examined and manually adjusted. Ten *f0* values with equal time intervals were extracted for each syllable. Trimming and smoothing functions were disabled to obtain raw *f0* values. The *f0* values were normalized into T values within each speaker according to Function (1) [15]. It logarithmically compressed a speaker's *f0* productions into a 5-degree scale [16].

(1) $T = 5 \times \frac{log_{10}^{x_i} - log_{10}^{x_{min}}}{log_{10}^{x_{max}} - log_{10}^{x_{min}}}$.

2.2.4 Statistical analysis

To model pitch contours, generalized additive mixed models (GAMMs) [17] were conducted using `mgcv` and `itsadug` in R [18, 19]. GAMMs were used due to the inclusion of parametric (linear), smooth (nonlinear), and random terms. The nonlinear terms were especially suitable to model the nonlinear pitch contours of Mandarin tones.

3. RESULTS

3.1 Model comparisons

Pitch contours were modelled with the fixed effects of group (HL and L2), grade (ordered, 1 and 3, and 0 for teachers), and tone (T1 to T4), as well as random smooths of speaker and word token. Interaction effects were implemented with contrast coding to allow interactions with time (the x-axis of contours). Models were compared using the `compareML` function [19]. The selected model with the lowest AIC included parametric and smooth terms of the 3-way interactions among tone × group × grade and the random smooths. This model suggested that pitch contours were impacted by the interactions among tones, groups, and grades. Therefore, a model was built for each tone to examine group and grade effects.

3.2 Group and grade effects

Each tone's model included group × grade interaction as parametric and smooth terms, as well as the random smooths. As an example, in Figure 2 for T1 model, the top row shows predicted contours of T1 by subgroups. The following rows include difference plots to examine the contrasts of interest: The middle row presents in grade 1, how HL and L2 differed from teachers and how L2 differed from HL. The bottom row presents the same for grade 3. In a difference plot, a section above zero suggests a higher pitch contour in the first group and *vice versa*. The percentage of sections of significant differences can be calculated to indicate the extent of differences.

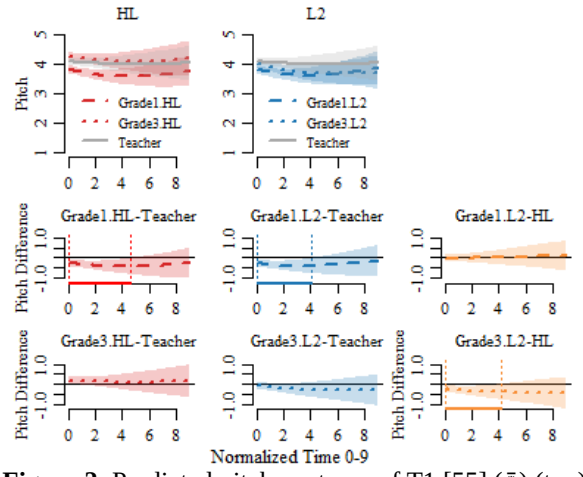

Figure 2: Predicted pitch contours of T1 [55] (◌̄) (top), and group differences in grade 1 (middle) and grade 3 (bottom). HL is coded in red, L2 is in blue, and L2-HL difference is in orange. Grade 1 is coded in dashed lines, grade 3 is in dotted lines, and teachers are in grey solid lines. Ribbons describe 95% confidence intervals. Vertical lines mark sections of significant differences.

For the high-level (◌̄) T1 (Figure 2), all subgroups produced high-pitch contours. In grade 1, HL's and L2's contours were different from teacher models (the percentage of sections with significant difference to the full range of normalized time for HL was 52%; L2: 45%), both with a lower pitch onset. This means both groups did not acquire the "level" feature in grade 1 and produced a slightly rising contour. In grade 3, both groups showed no significant difference from teachers, although L2's onset was significantly lower than HL (46%). This suggests both groups were able to acquire T1's features in grade 3, although the two groups' productions were slightly different phonetically.

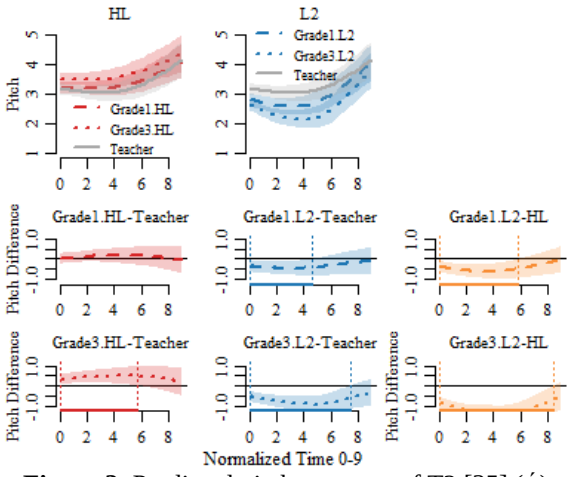

Figure 3: Predicted pitch contours of T2 [35] (◌́), differences between students and teachers, and group differences in both grades.

For the mid-rising (◌́) T2 (Figure 3), in grade 1, L2's pitch contour was significantly different from teacher models (52%) with a lower pitch onset. Unsurprisingly, L2's contour was also different from their HL peers in grade 1 (65%). In grade 3, the two groups produced opposite patterns: HL's contour was significantly different from teachers' (64%) with a higher onset. On the other hand, L2's contour was also different from teachers' (83%) but showed a lower onset and an even lower middle section of the pitch contour. Consequently, HL's and L2's difference was more distinct in grade 3 (94%). This suggests that while HL produced a mid- to high-rising contour in grade 3, L2 produced a mid to low onset with a dipping contour. It indicates that L2's T2 productions in grade 3 could be more subjected to being misrecognized as T3 by listeners, as a dipping contour with a falling portion is a characteristic feature of T3 compared to T2 [20, 21]

For the low-dipping (◌̌) T3, all subgroups produced a low-dipping pitch contour (Figure 4). Compared to the teacher models, HL in grade 1 produced a slightly higher offset (5%); HL in grade 3 produced a slightly higher onset (26%); L2's contours were not different from teacher models in both grades. Furthermore, L2's and HL's pitch contours were

significantly different in grade 3 (34%). However, given the complex phonetic specifications of T3, it remains a question whether these differences were perceptually significant.

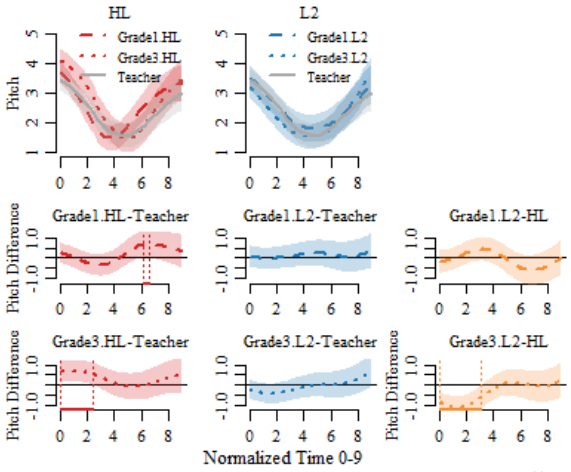

Figure 4: Predicted pitch contours of T3 [214] (˘), differences between students and teachers, and group differences in both grades.

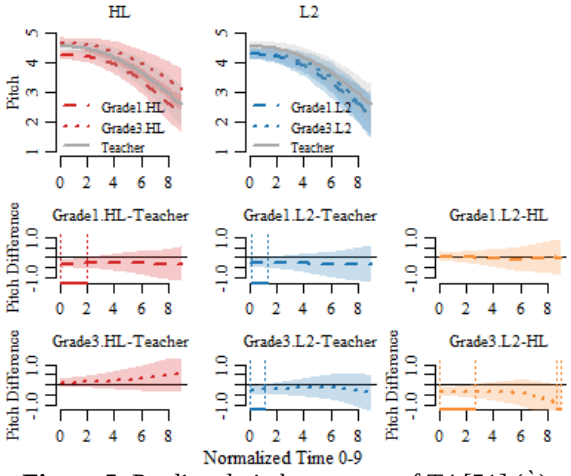

Figure 5: Predicted pitch contours of T4 [51] (`), differences between students and teachers, and group differences in both grades.

For the high-falling (`) T4, all subgroups produced a falling pitch contour (Figure 5). In grade 1, both HL's (22%) and L2's (13%) contours were significantly different from teacher models with a lower pitch onset, and the two groups showed no difference between each other. This suggests that both groups did not fully acquire the "high" feature of T4 in grade 1. However, in grade 3, HL's contour had no difference from teacher models, while L2's contour still had a lower onset than teachers' (12%), causing a significant difference between the HL and L2 groups (32%). This suggests that HL was able to acquire the high feature in grade 3, but L2 still had marginal difficulty producing a high-pitch onset. Similar to T3, the short sections of differences in T4 might be perceptually insignificant.

4. DISCUSSION

Both HL and L2 students produced some key features of citation tones as early as in grade 1, suggesting the efficiency of the two-way bilingual program [20]. However, their phonetic specifications sometimes differed from teacher models, suggesting a process of phonetic learning [4]. In this process, the impacts of speech input interacted with tone targets. Specifically, when the target was simple, bilingual students followed the same developmental trajectory despite their different backgrounds. However, when the target was complex, bilingual students' phonetic learning was impacted by their home input.

As a reminder, L1 children's developmental order of tones is T1 > T4 > T2 > T3 [4, 12]. For the earliest-developmental high-level T1, five-year-old L1 children's productions tend to be less level [4, 5]. Bilingual students demonstrated a similar pattern: Both HL and L2 produced the high feature in grade 1 and acquired the level feature later in grade 3. For the high-falling T4, young L1 children's productions tend to show a falling trend that is not steep enough [4, 5]. HL and L2 both produced the falling feature in grade 1, but only HL produced a high onset in grade 3, therefore a steeply falling contour. The mid-rising T2 is perceptually and productively confusable with the low-dipping T3, because both contours include a rising section [4, 21, 22]. The current study shows that HL produced the rising feature in grades 1 and 3, while L2 in grade 3 produced a dipping contour. It seems that the increasing input at school did not help L2 students refine their T2 productions. Instead, they were at an increased risk of T2-T3 convergence. A possible explanation is that L2 students were less sensitive to pitch due to limited early exposure and did not form the same phonological representations as HL students [8], which hindered their phonetic refinement of complex tone targets. No conclusive patterns were identified for the latest-developmental T3. A closer look at the tone errors is warranted.

These results differ from the evidence in Indo-European languages that home input difference can be levelled out by school learning [2, 3]. The current study indicates that the role of home input in L2 phonetic learning is impacted by the targets [6], with the learning of late-developmental targets more impacted by early home input. Moreover, it provides evidence in the suprasegmental domain for how speech input impacts phonetic learning [6] and how early input might tune learners' speech perception and impact future learning [8]. It needs to be noted that this study focused on phonetic specifications of productions judged as accurate, and future studies will further examine the relationships between acoustic measurements and perceived accuracy.

5. ACKNOWLEDGEMENTS

This paper draws on research supported by the Social Sciences and Humanities Research Council (SSHRC). The authors would like to thank the participating schools, students, parents, and teachers, as well as the research assistants who helped with transcription and tone labeling.

6. REFERENCES

[1] Cummins, J. 1979. Linguistic Interdependence and the Educational Development of Bilingual Children. *Rev. Educ. Res.* 49(2), 222–251.

[2] Menke, M. R. 2017. Phonological development in two-way bilingual immersion. *J. Second. Lang. Pronunciation*. 3(1), 80–108.

[3] Nance, C. 2020. Bilingual language exposure and the peer group: Acquiring phonetics and phonology in Gaelic Medium Education. *Int. J. Biling.* 24(2), 360–375.

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

[5] Rattanasone Xu, 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. *Front. Psychol.* 9.

[6] Flege, J. E., Bohn, O.-S. 2021. The revised Speech Learning Model. In: Chen, S. (ed), *Second language speech learning: Theoretical and empirical progress*. Cambridge University Press, 3–83.

[7] DeKeyser, R. M. 2000. The robustness of critical period effects in second language acquisition. *Stud. Second. Lang. Acquis.* 22(4), 499–533.

[8] Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., Nelson, T. 2008. Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Phil. Trans. R. Soc. B.* 363, 979–1000.

[9] Marian, V., Blumenfeld, H. K., Kaushanskaya, M. 2007. The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *J. Speech. Lang. Hear. Res.* 50, 940–967.

[10] Paradis, J. 2011. Individual differences in child English second language acquisition. *Linguist. Approaches. Biling.* 1(3), 213–237.

[11] Zhao, J., Bernhardt, B. M. 2012. Mandarin Single-Word Elicitation Tool for Phonology. http://phonodevelopment.sites.olt.ubc.ca/mandarin-picture-elicitation_2012/.

[12] Zhu, H. 2002. *Phonological Development in Specific Contexts: Studies of Chinese-speaking Children*. Multilingual Matters.

[13] Hedlund, G., Rose, Y. 2020. Phon 3.1. https://phon.ca.

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

[15] Shi, F., Wang, P. 2006. A statistic analysis of tone groups in Beijing Mandarin. *Contemporary Linguistics* 8(4), 379–380.

[16] Chao, Y.-R. 1968. *A Grammar of Spoken Chinese*. University of California Press.

[17] 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. *J. Phon.* 70, 86–116.

[18] Wood, S. N. 2017. *Generalized additive models: An introduction with R (2nd ed)*. CRC Press.

[19] van Rij, J., Wieling, M., Baayen, R., van Rijn, H. 2022. itsadug: Interpreting time series and autocorrelated data using GAMMs. R package version 2.4.1.

[20] Xu, Y. 1997. Contextual tonal variations in Mandarin. *J. Phon.* 25, 61–83.

[21] Wang Y., Li, M. 2010. The effect of tone pattern and register in perceptions of Tone 2 and Tone 3 in Mandarin. *Acta. Psychol. Sin.* 42(9), 899-908.

[22] Wang, Y., Jongman, A., Sereno, J. A. 2003. Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *J. Acoust. Soc. Am.* 113(2), 1033–1043.