Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

Bai tone perception and production by Naxi speakers in Jiuhe: a preliminary study

Meihao Wan, Peggy Mok

The Chinese University of Hong Kong
wanmeihao@link.cuhk.edu.hk, peggymok@cuhk.edu.hk

Abstract

This study investigated the perception and production of Bai tones by native Naxi speakers in Jiuhe. Jiuhe Bai features six lexical tones distinguished by pitch and phonation, while Jiuhe Naxi has three tones differentiated only by pitch. We explored whether Naxi speakers could accurately perceive and produce the Bai tones, particularly those with similar pitch contours. 10 native Naxi speakers participated in the perception and production experiments. The perception experiment involved a discrimination task with all possible pairs of Bai tones. In the production experiment, participants were instructed to produce all the Bai tones with a wordlist of minimal pairs. Both acoustic and EGG signals were recorded and analyzed for pitch and phonation patterns. The discrimination results revealed that native Naxi speakers had difficulty distinguishing certain Bai tones, which were also merged in their own production. Interestingly, not only the tones with similar pitch contour, but also those with similar phonation patterns, are prone to confusion among the Naxi speakers. The findings can shed new light on the acquisition of L2 tone categories in nonnative speakers, specifically in languages that employ multiple cues for tone distinction.

Index Terms: L2 production, L2 perception, Bai tones, Naxi

1. Introduction

Both Bai and Naxi are spoken in Jiuhe township, Yunnan Province in China. The two languages belong to Tibeto-Burman languages, some of which have phonation-based register contrasts in tones. Tones with non-modal phonation are called tense tones, while tones with modal phonation are described as lax tones in these register languages, like Yi and Bai. Both Bai and Naxi are tone languages. Jiuhe Bai has a complicated tone system: it has six lexical tones and has the tense vs. lax contrast (see Table 1). T1, T3 and T5 are tense tones, T2, T4 and T6 are lax tones. Previous studies found that the tense tones in Bai are accompanied by non-modal phonation [1], but not with a specific phonation type. The non-modal phonation types used for the tense tones vary across speakers and dialects [1], [2]. Such variations can also be found in Jiuhe Bai. For instance, the tense tone T1 can be cued by a creaky, breathy, or harsh voice among different speakers.

The F0 contours of Jiuhe Bai tones are shown in Figure 1. The six lexical tones in Jiuhe Bai can be well distinguished by pitch, in which T1 and T2 are low-falling tones, T3 is a mid-falling tone, T4 and T5 are mid-level tones, and T6 is a high-level tone. Some Bai tones have similar pitch contours, such as T1/T2, T4/T5. For native Bai speakers, these different tones can be easily distinguished.

Table 1. *Tones of Jiuhe Bai*

Tone category	Pitch contour	Pitch value	Register
1	Low-falling	21	Tense
2	Low-falling	21	Lax
3	Mid-falling	31	Tense
4	Mid-level	33	Lax
5	Mid-level	44	Tense
6	High-level	55	Lax

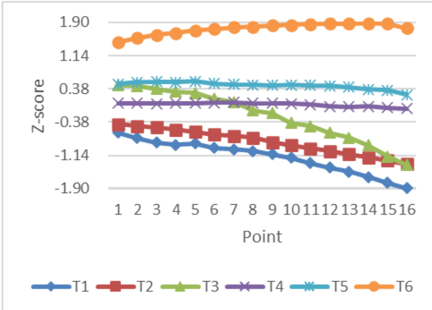

Figure 1. *Mean F0 contours of Jiuhe Bai tones*

On the other hand, Jiuhe Naxi has a relatively simple tone system: it only has three level tones distinguished by pitch, with no tense vs. lax contrast. Due to the fact that Jiuhe Naxi having fewer tones compared to Jiuhe Bai and the fact that the cues for Bai tone contrasts are more intricate, it is probable that native Naxi speakers will encounter much difficulty when acquiring Bai tones. How native Naxi speakers acquire the Bai tone system remains to be addressed.

Previous second language (L2) speech acquisition models propose that the similarity between two speech sounds can predict the discrimination accuracy for L2 learners. The Perceptual Assimilation Model (PAM) claims that the discrimination will be inaccurate if the two contrastive sounds are assimilated into one single category in the first language (L1) [3]. Another model, the Speech Learning Model (SLM), argues that L2 sounds that are similar to L1 sounds are prone to be pronounced inaccurately, while L2 sounds with no L1 equivalents will be produced more accurately [4]. However, these models mainly focus on the learning of L2 consonants and vowels instead of tones [5]. PAM was later extented to the suprasegmental features and the PAM for Suprasegmental (PAM-S) was proposed. It assumes that L2 tone acquisition is also constrained by the L1 tone system. L2 tones may be assimilated into one category if they are similar in the L1 tone system [6]. The previous L2 models indicate that the hardest parts for L2 learners are those that have similar characteristics, not those that are very different from their L1, which can be

10.21437/SpeechProsody.2024-53

demonstrated by earlier research on Cantonese tone. The tones with similar pitch contours in Cantonese, like T2 [25] and T5 [23], T3 [33] and T6 [22], are difficult to distinguish for both native and nonnative speakers [7], [8]. Thus, we postulate that tones with similar characteristics will be more difficult to discern and cause learning difficulties for L2 listeners.

Based on the above discussion, we hope to answer these research questions with the data of these two under-resourced languages: (1) Can native Naxi speakers distinguish the Bai tones with similar pitch contours? (2) If so, can they discriminate the tones in their perception? (3) Can native Naxi speakers acquire the contrastive phonation cue in Bai, which is new to them?

In this case, we predict that the Bai tones with similar pitch contours, such as T1 and T2, as well as T4 and T5, would be challenging for native Naxi speakers to distinguish both in their production and perception. These tones share similarities in both pitch heights and contours, which may lead to difficulties in accurately perceiving and reproducing them. Furthermore, we anticipate that the contrastive phonation cue present in Bai tones, which is novel to native Naxi speakers, may pose additional challenges in acquiring and differentiating these tones.

To investigate these questions, our study included both production and perception experiments. In the production experiment, native Naxi speakers were required to produce Bai tones, while in the perception experiment, they were tasked with listening to different pairs of Bai tones and determining whether they are the same or different.

2. Method

2.1. Participants

Ten native Naxi speakers (5 males, 5 females, age range: 21-61, M_{age} = 46.4) participated in the study. All participants took part in both the production and the perception experiments. They were born and raised in Jiuhe and had no long-term out-of-town experience. Before going to school, they only spoke Naxi. They all acquired Bai after school age. All of them can speak Naxi and Bai proficiently. No participants reported hearing and speaking problems. They were paid a nominal fee for participating in the experiments.

2.2. Materials

In the production experiment, a word list of 11 monosyllable minimal pairs were made for the recording. All the words were relatively balanced in frequency and were familiar to the participants. Each monosyllable was produced twice in a row. 132 tokens (11 syllables × 6 tones × 2 repetitions) were recorded for each participant.

The perception task included an AX discrimination task. Specifically, five distinct syllables ([kɯ], [pæ], [tɑ], [tɕi], [tsʊ]) were selected for the task, each of which can carry the six tones in Jiuhe Bai. The AA pairs consisted of monosyllables sharing identical segments and tones, but the two stimuli came from different recordings. To balance the number of AB pairs, each AA pair appeared five times. The AB pairs consisted of monosyllables with matching segments but different tones, e.g., T1/T2, T3/T5. The order of the AB pairs was counterbalanced. In total, there were 150 AA pairs (6 tones × 5 syllables × 5 repetitions) and 150 AB pairs (15 tone combinations × 2 orders × 5 syllables). All pairs were randomized within the discrimination task. A middle-aged female native Bai speaker who cannot speak Naxi produced all the stimuli.

2.3. Procedure

The experiments were conducted in a quiet room. The participants were asked to do the production experiment first. After the familiarization, they were instructed to produce the monosyllabic words at a normal speech rate. Both acoustic and EGG signals were recorded with Glottal Enterprises EG2-PCX EGG. The sampling frequency was 22050 Hz.

The perception experiment was conducted after the production tasks. The participants were asked to judge if the tones of the two monosyllables were the same or different after listening to the stimuli. The stimuli were presented on a laptop computer using PsychoPy via headphones. The task was divided into six blocks. The participants could have short breaks between blocks.

2.4. Analysis

Both F0 and EGG parameters were measured in this study. F0 was measured from the acoustic signals with VoiceSauce using the STRAIGHT algorithm [9]. The results were averaged on ten normalized-time points within each speaker.

EGG measurements were made using a MATLAB program, which was also used in [10]. EGG parameters included the Closed Quotient (CQ), the Speed Quotient (SQ), and the Peak Increase in Contact (PIC). CQ and SQ refer to the ratio of the duration of the closed phase to the whole period or the opening phase in a glottal cycle. PIC refers to the positive peak amplitude in the first derivative of EGG. The results were sampled at twelve points within the vowel. After removing the first and the last points, there were ten points left for analysis.

Addtionally, within-speaker normalization was done for each parameters. All the values, including F0 and EGG parameters, were transformed into z-scores for the comparison among different speakers [11]. Soothing-Spline ANOVA was conducted to compare the parameter differences between two tones.

3. Results

3.1. F0

F0 contours of Bai tones produced by native Naxi speakers are shown in Figure 3. Unlike native Bai speakers, whose tones can be distinguished by pitch like in Figure 1, some Bai tones are merged in the production of native Naxi speakers.

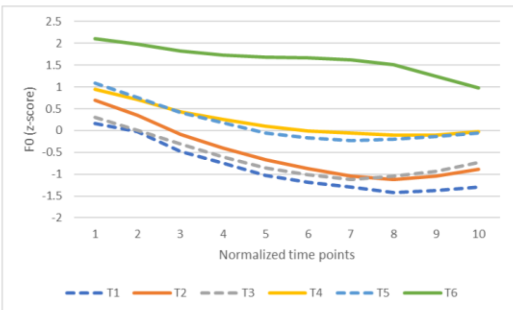

Figure 2. *Mean F0 contours of Bai tones in native Naxi speakers.*

For native Naxi speakers, they had difficulty in distinguishing between T1, T2 and T3. In addition to this, T4 and T5 are not well distinguished as well. The high-level tone T6 is distinct from other tones and is the least affected.

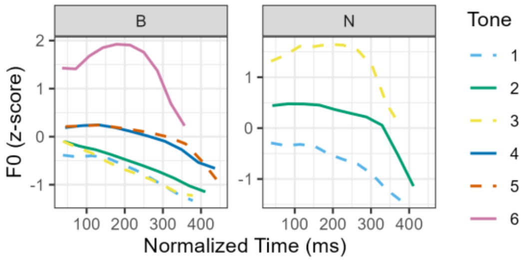

Figure 3. *Mean F0 contours of both Bai and Naxi tones of a female native Naxi speaker.*

Due to space constraints, Figure 3 shows the F0 contours of both Bai and Naxi tones of a female native Naxi speaker. The T4 and T5 of Bai are assimilated to Naxi's T2. And T1, T2 and T3 of Bai are assimilated to the T1 of Naxi. The general pattern conforms to the prediction of PAM.

3.2. Perception experiment

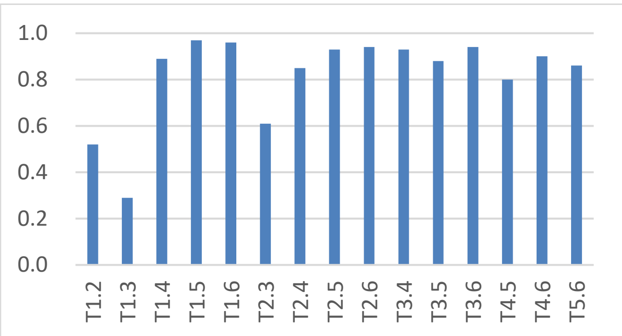

Figure 4. *Mean discrimination accuracy rate of Bai tone by native Naxi speakers.*

The participants' mean accuracy rates for discriminating the different tone pairs are presented in Figure 4. Native Naxi speakers demonstrated high discrimination performance (⩾ 80%) for most of the tone pairs.

However, certain tone pairs, specifically T1/T2, T1/T3, and T2/T3, were poorly discriminated by the participants. Among these tone pairs, T1/T3 exhibited the lowest accuracy rate (M = 0.29, SD = 0.1). The accuracy rates for T1/T2 (mean = 0.52, SD = 0.1) and T2/T3 (mean = 0.61, SD = 0.2) were slightly higher compared to T1/T3.

These findings align with the previous results from the production data, indicating that native Naxi speakers had difficulties in distinguishing the Bai tones, including T1, T2, and T3 in perception. For T4/T5, although native Naxi speakers could not distinguishe these two tones in their production, they still had a relatively higher accuracy rate than T1/T2/T3.

3.3. EGG parameters

The tone pairs with similar pitch patterns are selected to see if there is a phonation difference. The results of SSANOVA on the EGG parameters are plotted in Figures 5, 6, 7.

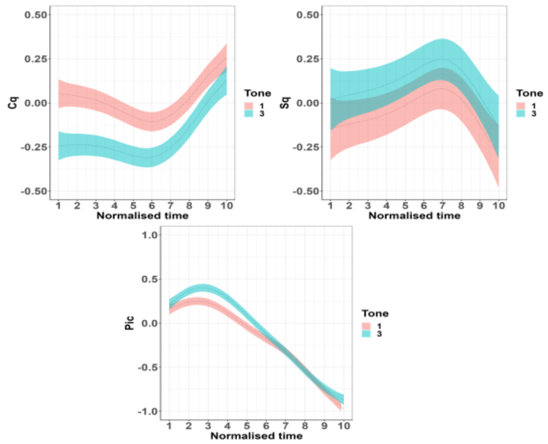

Figure 5. *SSANOVA plots of EGG parameters (α=0.5) between T1 and T3.*

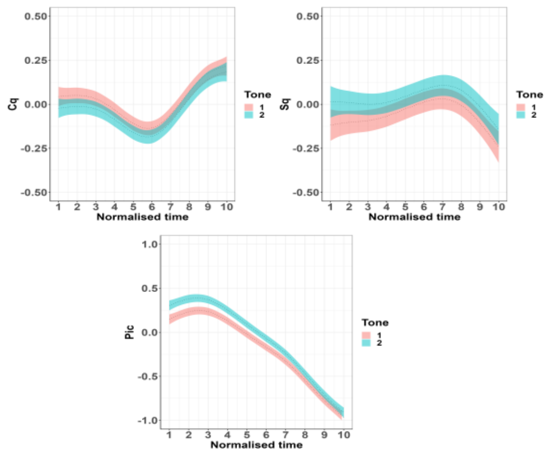

Figure 6. *SSANOVA plots of EGG parameters (α=0.5) between T1 and T2.*

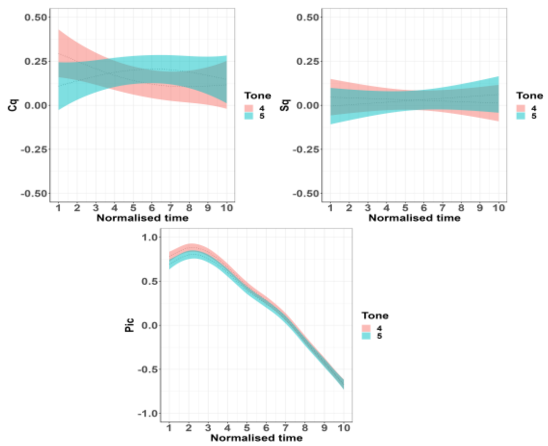

Figure 7. *SSANOVA plots of EGG parameters (α=0.5) between T4 and T5.*

The figures show that only the phonation parameters of T1 and T3 are significantly different, and there is no significant difference between T1 and T2, T4 and T5. The CQ of T3 is smaller than T1, suggesting a smaller glottal constriction in T3, i.e., breathier. For the other two tone pairs, the results indicate that native Naxi speakers do not have phonation differences as well, similar to the pitch findings.

4. Discussion

Consistent with previous research, our study demonstrated that L1 tonal experience plays a crucial role in L2 tone acquisition. Native Naxi speakers, who have a relatively simple tone system in their L1, encountered difficulties in acquiring the more complex tone patterns of Bai. Our results on F0 demonstrated that L1 tonal experience indeed influences L2 tone acquisition. Tones with similar F0 values were more challenging to differentiate, such as T1/T2 and T4/T5. T6, which had a relatively higher F0 and was distinct from the other tones, was the most easily distinguishable tone. This finding suggests that the phonetic similarity in L2 speech system has an influence on the discrimination of phonological categories in L2 learners.

This finding aligns with the predictions of SLM, which emphasizes the influence of prior tone experience on L2 tone production. Specifically, Bai T1 and T2 are similar to the low-level tone in Naxi, while T4 and T5 in Bai are close to the mid-level tone in Naxi. As a result, native speakers of Naxi find it difficult to distinguish between these tones. The transfer of tone patterns from the L1 to the L2, as evidenced by the merging of certain Bai tones in the participants' production, supports the idea that learners rely on their existing tone categories when acquiring a new tone system.

Furthermore, our study found that native Naxi speakers exhibited phonation differences in their production of some Bai tones, although not all tense tones demonstrated this distinction. Based on the observation, the glottal constriction was smaller and the glottal closure was less abrupt in the T3 of native Naxi speakers, which implies that T3 may be breathier than T1. Thus, this suggests that native Naxi speakers may have, to some extent, acquired the phonation patterns present in Bai tones during the process of acquisition, but not necessarily acquiring the entire phonation patterns of Bai tones. This finding highlights that L2 learners can also acquire the phonation contrast of the tones with such constrastive cues.

Interestingly, the findings of our study revealed a novel and intriguing aspect of L2 tonal acquisition that extends beyond previous research. In addition to tones with similar pitch contours being challenging to distinguish, our study found that tones with similar phonation types also presented difficulties for the learners. This observation was unexpected and adds a new dimension to our understanding of L2 tone perception.

In the perception tasks conducted in our study, the tone pair T1/T3 was particularly challenging for the learners to distinguish. Surprisingly, the accuracy rate for this tone pair was even lower than that of tone pairs with similar pitch contours, such as T1/T2. This unexpected result suggests that the perceptual assimilation process, as proposed by the Perceptual Assimilation Model (PAM), plays a crucial role in L2 tone acquisition. According to the PAM, learners rely on their existing phonetic categories to perceive and categorize L2 tones. When encountering tones with similar phonation types, learners may experience difficulties in distinguishing between them due to the overlap or confusion in their existing phonetic categories. In the case of the T1/T3 tone pair, the similarities in phonation types may have led to perceptual assimilation, where learners assimilated these tones into a single category, resulting in lower accuracy in distinguishing between them.

However, the present study also challenges the existing models in predicting L2 tone acquisition. The findings suggest that when multiple perceptual cues are present, various similarities may exist, and relying solely on similarity in one aspect may not accurately predict which tones are more challenging to distinguish. In our study, we observed that the discrimination accuracy of the T1/T2 tone pair was relatively higher than that of the T1/T3 pair. This unexpected result suggests that the phonation cue may outweigh the pitch cue in the perception of native Naxi speakers. Consequently, they assimilated these two tones into a single category and encountered difficulties in distinguishing them during production.

This observation highlights the significance of incorporating phonation contrasts into models of L2 tonal acquisition. Existing models have primarily focused on pitch-related cues, while neglecting the role of phonation patterns as influential factors. However, our findings suggest that phonation cues can play a crucial role in tonal perception, potentially outweighing the influence of pitch cues in certain contexts. The omission of phonation contrasts in current models may limit their ability to accurately predict and explain the complexities of L2 tonal acquisition. Therefore, further refinement and integration of perceptual models are necessary to account for the multifaceted nature of L2 tonal acquisition, encompassing both pitch and phonation cues.

5. Conclusions

In conclusion, the findings of this study contribute to our understanding of L2 tone acquisition by investigating the L2 acquisition of the complicated Bai tones with multiple perceptual cues. The present study examined the acquisition of Bai tones by native Naxi speakers and highlighted the influence of L1 tonal experience on L2 tonal acquisition. The findings revealed that native speakers of Naxi faced difficulties in distinguishing between certain Bai tones due to the similarity in pitch contours or phonation types. Additionally, partial acquisition of phonation cues in Bai tones was observed in the production of native Naxi speakers. The results also emphasize the limitations of existing models and the importance of considering learners' perceptual performance in predicting and facilitating L2 tone acquisition. Given that the study has a relatively small sample size, the interpretation of the results can only be tentative. We are collecting data from further speakers to enhance the study. Future research should continue to explore and refine our understanding of the factors that influence L2 tone acquisition.

6. Acknowledgements

The research was supported by Hong Kong PhD Fellowship Scheme awarded to the first author. We thank the two anonymous reviewers for the helpful comments.

7. References

[1] X. Li and F. Wang, “Phonation variation and strategy of tone distinction: The case of Meiba Bai,” *Journal of Chinese Linguistics*, vol. 48, no. 2, pp. 379–401, 2020, doi: 10.1353/jcl.2020.0009.

[2] 汪 鋒 (Wang, Feng), “VARIATIONS OF LARYNGEAL FEATURES IN JIANCHUAN BAI/劍川白語的嗓音變異,” *Journal of Chinese Linguistics*, pp. 434–452, 2015.

[3] C. T. Best, “The emergence of native-language phonological influences in infants: A perceptual assimilation model,” *The development of speech perception: The transition from speech sounds to spoken words*, vol. 167, no. 224, pp. 233–277, 1994.

[4] J. E. Flege, “Second language speech learning: Theory, findings, and problems,” *Speech Perception and Linguistic Experience: Issue in Cross-Language Research*, vol. 92, pp. 233–277, 1995.

[5] J. E. Flege and O.-S. Bohn, “The Revised Speech Learning Model (SLM-r),” in *Second Language Speech Learning*, 1st ed., R. Wayland, Ed., Cambridge University Press, 2021, pp. 3–83. doi: 10.1017/9781108886901.002.

[6] C. K. So and C. T. Best, “Cross-language Perception of Non-native Tonal Contrasts: Effects of Native Phonological and Phonetic Influences,” *Lang Speech*, vol. 53, no. 2, pp. 273–293, Jun. 2010, doi: 10.1177/0023830909357156.

[7] P. P. K. Mok, D. Zuo, and P. W. Y. Wong, “Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese,” *Language Variation and Change*, vol. 25, no. 3, pp. 341–370, Oct. 2013, doi: 10.1017/S0954394513000161.

[8] A. C. L. Yu, C. W. T. Lee, C. Lan, and P. P. K. Mok, “A New System of Cantonese Tones? Tone Perception and Production in Hong Kong South Asian Cantonese,” *Lang Speech*, vol. 65, no. 3, pp. 625–649, Sep. 2022, doi: 10.1177/00238309211046030.

[9] Y.-L. Shue, P. Keating, and C. Vicenik, “VOICESAUCE: A program for voice analysis.,” *The Journal of the Acoustical Society of America*, vol. 126, no. 4_Supplement, pp. 2221–2221, Oct. 2009, doi: 10.1121/1.3248865.

[10] C. Ge, W. Xu, W. Gu, and P. P. K. Mok, “The change in breathy voice after tone split: A production study of Suzhou Wu Chinese,” *Journal of Phonetics*, vol. 98, p. 101239, May 2023, doi: 10.1016/j.wocn.2023.101239.

[11] P. Rose, “Considerations in the normalisation of the fundamental frequency of linguistic tone,” *Speech Communication*, vol. 6, no. 4, pp. 343–352, Dec. 1987, doi: 10.1016/0167-6393(87)90009-4.