Acoustic and Perceptual Study of Zaiwa Tones: Postprint

Authors: Lu Yao, Kong Jiangping

Date: 2018-11-29T00:00:00+00:00

Abstract

This study employs acoustic and perceptual methods to investigate the tone values of the Zaiwa language. Acoustic analysis reveals that the language possesses 2 level tones, level tone(s), level tone(s), 3 falling tones, falling tone(s), and 1 rising tone.

Full Text

Preamble

An Acoustic and Perceptual Study of Tones in Zaiwa

LU Yao, KONG Jiangping

Abstract This study investigates the tonal values of Zaiwa using acoustic and perceptual methods. Acoustic analysis reveals that Zaiwa possesses two level tones, three falling tones, and one rising tone, with tonal values of 55, 44, 35, 51, 31, and 21. Building upon this acoustic analysis, we conducted a perceptual study examining three combinations of tone and phonation types: 1) different tones with both vowels in modal voice; 2) different tones with both vowels in pressed voice; and 3) different tones with different vowel phonation types. Both acoustic and perceptual findings demonstrate that fundamental frequency (F0) and phonation type contribute to tone perception in Zaiwa, both carrying linguistic significance. When phonation type is absent, subjects exhibit "F0 perceptual neutrality."

[**Keywords**] Zaiwa language, phonation type, acoustic analysis, tone perception

Research on speech perception is crucial for understanding the psychological reality of phonemes and establishing phonemic categories. Beginning in the 1970s, tone perception research gained widespread attention in linguistics. Although F0 cannot be completely equated with tone, scholars generally agree that the pitch height and contour of tones are primarily determined by F0 variation. Consequently, current tone perception research predominantly employs listening experiments that manipulate F0.

Wang (1976) synthesized 11 stimulus continua for Mandarin Yinping-Yangping by altering F0, using /i/ as the carrier syllable. Identification and discrimination experiments revealed that native Mandarin (Beijing dialect) speakers exhibit categorical perception of the Yinping-Yangping contrast. Abramson (1977) used the Haskins Laboratories parallel-resonance synthesizer to vary F0, creating 8 stimulus continua spanning three level tones (high, mid, and low) in Central Thai. Perception experiments with 33 native speakers found that tone perception was continuous rather than categorical. Abramson proposed that this discrepancy with Wang (1976) might stem from differences in stimulus design: Wang' s continua transitioned from a level to a rising tone, whereas Abramson' s involved only level tones. He suggested that if discrimination experiments were conducted on rising-falling continua in Central Thai (which also possesses a rising and a falling tone), the divergent results might be reconciled. Additionally, Abramson noted that Wang used the simple monophthong /i/, while his own study employed the more complex syllable /kha:/, which could also account for the different outcomes. Francis et al. (2003) further interpreted these findings, arguing that Wang (1976) used tones involving both pitch contour and pitch height changes, whereas Abramson' s Thai tones involved only pitch height variation. To test this hypothesis, Francis examined native speakers' perception of three level tones, level-contour tone pairs, and contour-contour tone pairs in Cantonese. The results showed that perception of the three level tones resembled that of Central Thai—continuous perception—while perception of level-contour tone pairs was categorical, matching Wang' s (1976) findings. Francis' s results thus demonstrated the importance of tonal contour variation for categorical perception, a conclusion supported by numerous other studies: native speakers typically perceive tones with identical contours (e.g., among level tones or among rising tones) continuously, while tones with different contours—especially opposite directions (e.g., level vs. rising, level vs. falling, rising vs. falling)—are perceived categorically. However, whether this pattern applies universally across all tonal languages remains to be investigated.

Recent advances in acoustic, physiological, and psychological research have revealed that F0 alone cannot fully describe tones or explain perception results, particularly when languages exhibit special phonation types. For instance, Zhang and Kong (2014) found that phonation type also contributes to tone perception in the Yuzhou dialect of Henan, with minimal contribution when F0 differences are large but substantial contribution when F0 differences are small. Liu and Zhang (2016) discovered that the distinction between the low-level tone (T5: 22) and low-falling tone (T6: 31) in Yuliang Miao results from the combined effect of phonation and F0.

In several minority languages of China, tonal contrasts involve not only F0 opposition but also phonation type opposition. Whether native perception relies primarily on F0 or phonation type, and what constitutes the "distinctive features" of tones, requires further investigation. This study therefore selects Zaiwa, which exhibits both F0 and phonation type contrasts, as its research object. First, we determine its tonal values and phonation types through acoustic analysis. Second, we conduct perceptual experiments on three tone-phonation combinations: "different tones, both modal voice," "different tones, both pressed voice," and "different tones with different phonation types," to explore the respective contributions of F0 and phonation type to native perception. The findings carry important theoretical implications for discussing and distinguishing tonal phonemic categories in Zaiwa.

2. Acoustic Analysis of Zaiwa Tones

Zaiwa is spoken by the Zaiwa branch of the Jingpo ethnic group in China, belonging to the Sino-Tibetan language family, Tibeto-Burman group, Burmish branch, locally known as "Xiaoshan speech." According to *A Brief Description of the Zaiwa Language* (Xu Xijian & Xu Guizhen, 1984), Zaiwa has three tones: 21 (some scholars suggest 22), 55, and 51. Modal voice vowels can combine with all initials, whereas pressed voice vowels combine only with unaspirated stops, voiced fricatives, nasals, and laterals. When the initial is an unaspirated stop (except voiceless fricatives) and the vowel is modal, the actual tonal value of 55 is 15 (some suggest 35) (He Lela, 2016). Tones 21 and 55 have both checked and unchecked variants depending on whether they carry a final stop. Dai Qingxia (1989), through comparative analysis with related languages (Burmese, Achang, Hani), detailed the origins and differentiation of Zaiwa tones, noting two classification methods: the current method groups checked and unchecked tones based on similar tonal values, resulting in both 55 and 21 containing both variants; the alternative method separates them into five tones—three unchecked and two checked. Dai observed that the three-tone classification is convenient for describing tonal changes, but separating checked and unchecked tones better facilitates research on tonal development. However, previous Zaiwa tone research relied primarily on fieldwork impressionistic listening and transcription, without acoustic analysis.

To accurately clarify Zaiwa' s tonal system, we recorded four native speakers from Huyu Township, Ruili City, Dehong Dai-Jingpo Autonomous Prefecture: two females (ages 19 and 48) and two males (ages 20 and 36). All were native Zaiwa speakers who also spoke Mandarin. We extracted mean F0 values for male and female speakers. International tone and intonation research employs the semitone method (Liu Fu 1924; Kong Jiangping 2015:64-65). Relative to pure physical F0 values, semitone values better correspond to human auditory perception because semitone relationships reflect perceptual relationships. We therefore converted F0 to semitone values using the formula: semitone = 12 $\times$ log (f /f), where f is the minimum value in the pitch range and f is the measured F0 at each point. When f equals the maximum F0 in the range, the resulting semitone value represents the speaker' s pitch range. In this study, the maximum F0 was 248 Hz and the minimum 146 Hz, yielding a pitch range of approximately 9 semitones. After obtaining semitone values, we converted them to five-point scale values using: five-point value = [(lgf - lgf)/(lgf - lgf) × 4] + 1. The resulting F0 patterns and five-point values for Zaiwa tones are shown in Figure 1 [Figure 1: see original paper].

Figure 1: F0 and Five-Point Values of Zaiwa Tones

Figure 1 reveals significant discrepancies from traditional transcription in two main respects:

First, checked and unchecked variants of the same tone show substantial differences. The checked 55 tone is approximately 1 degree higher than the unchecked 55, with values of about 55 vs. 44. The same applies to the 21 tone: checked is approximately 31, while unchecked is about 21. Additionally, these two tones differ greatly in duration. We therefore argue that separating checked and unchecked tones better aligns with human perception.

Second, traditional transcription merges the 15 and 55 tones as a single phoneme, considering them complementary due to differences in phonation type and initial. According to Kong Jiangping' s (2001:188) distinctive feature matrix for voice quality, Zaiwa phonation types can be classified as modal voice and pressed voice, where "lax" phonation is actually modal voice. Kong (2001:56) notes that measuring the amplitude difference between the first (H1) and second (H2) harmonics (in dB) is a common method for analyzing voice quality, with larger H2-H1 values indicating tenser vocal folds. Using harmonic analysis, we examined stable portions of modal [a] and pressed [a], as shown in Figure 2 [Figure 2: see original paper], confirming the tense-lax contrast. However, the actual tonal value of modal voice is 35, substantially different from 55. We therefore argue against merging them into a single tone category.

Figure 2: Power Spectrum of [a]-[a] Vowel

In summary, based on F0 patterns, we propose that Zaiwa has six tones: two level tones, one rising tone, and three falling tones, with tonal values of 55, 44, 35, 51, 31, and 21.

3. Perceptual Experiments on Zaiwa Tones

To investigate the effects of F0 and phonation type on tone perception, we selected three groups of Zaiwa monosyllabic tones as experimental stimuli: /po21/ "frog"and /po51/ "cheap"; /i21/ "urine"and /i51/ "wine"; and /mau35/ "strange" and /mau44/ "to deceive." These six words are common in Zaiwa. In the first two groups, the syllables share the same initial and final but differ in tone, with identical phonation types (both pressed or both modal). The third group shares initial and final but differs in both tone and phonation type—one modal, one pressed—representing a contrast in both F0 and phonation. The original syllables were produced by a 20-year-old male native speaker, with average duration of 500 ms. Recording took place in a quiet room using Cool Edit software, a Sony microphone, ThinkPad laptop, external sound card, and mixer, with sampling rate of 22,050 Hz and 16-bit precision.

3.1 Experimental Stimuli

We synthesized experimental stimuli using PSOLA (pitch synchronous overlap-add). Taking the "po21-po51" pair as an example: we first extracted 11 equidistant F0 measurement points from both po21 and po51 using Praat scripts, then interpolated to calculate 9 intermediate F0 values between these 11 points. Using po21 as the base, we used its original F0 contour as the first stimulus, then progressively modified its F0 according to the 9 calculated values to create a continuum of 9 stimuli, finally synthesizing the 11th stimulus on the po21 base using po51' s actual F0. Since the choice of base stimulus affects perception, we repeated the process using po51 as the base, yielding another 11 stimuli. In total, we created 3 pairs, 6 groups, with 11 stimuli per group, totaling 66 stimuli. The F0 contours of the stimulus sets are shown in Figure 3 [Figure 3: see original paper].

Figure 3: Design of Three Tone Synthesis Sets

The left panel shows the "po21-po51, pressed voice" continuum, with F0 ranging from 108.31 Hz to 160.12 Hz; the middle panel shows the "i21-i51, modal voice" continuum, ranging from 107.41 Hz to 162.5 Hz; the right panel shows the "mau35-mau44" continuum, with phonation type contrast, ranging from 115.69 Hz to 138.49 Hz.

3.2 Experimental Subjects

Twenty-eight subjects (12 male, 16 female) were selected from four villages in Huyu Village, Huyu Township, Ruili City, Dehong Prefecture: Yinshan, Guangpa, Manglong, and Huyu. Ages ranged from 16 to 52. All subjects were native Zaiwa speakers born and raised in the village, used Zaiwa as their primary daily language, had normal hearing and vision, and had at least primary school education.

3.3 Experimental Procedure

All subjects completed classic identification and discrimination tasks. Stimuli were presented binaurally through Sony MDR-7506 headphones at 72 dB SPL. Experiments and data collection were conducted using E-PRIME software. Before formal testing, subjects could practice to familiarize themselves with the equipment and procedure. During formal testing, subjects could press the spacebar to pause and rest at any time.

(1) Identification Task

The six synthesized stimulus groups were presented randomly. Each of the 11 stimuli in a group appeared twice ($2 \times 11 = 22$ trials), with each stimulus played twice consecutively. During playback, two corresponding Zaiwa words with Chinese translations appeared on screen. After playback, subjects had 5 seconds to make a forced-choice decision about which word they heard.

(2) Discrimination Task

We employed an AX discrimination paradigm. To maximize differentiation between within-category and between-category differences (Pisoni, 1973), stimulus pairs were separated by 500 ms, and subjects had 5 seconds to judge whether the pair was "same" or "different." Within each stimulus group, different pairs were separated by two steps and presented in both forward and reverse orders. The six groups yielded $18 \times 6 + 9 \times 6 = 162$ sample pairs. Each pair was presented twice, requiring $162 \times 2 = 324$ responses, with all pairs randomized.

3.4 Data Analysis

We analyzed each subject' s identification and discrimination results to explore the contributions of F0 and phonation type and determine perception patterns for different tonal contrasts. Statistical analysis was conducted using Excel 2010, IBM SPSS Statistics 20.0, and R (The R Project for Statistical Computing).

(1) Identification Functions and Boundaries

Identification results for each stimulus were calculated as the percentage of subjects identifying it as each lexical item. We constructed binary logistic regression models from the identification data: log (P /(1-P)) = b + b x (Xu et al., 2006). For each continuum, P represents identification results, x is the step number, and b and b are regression coefficients. The slope is calculated from b , and the identification boundary corresponds to the x-value where identification probability equals 0.5: b + b x_cb = log (0.5/(1-0.5)) = 0 → x_cb = -b /b . Boundary width is the linear distance between the 25% and 75% identification points (Xu et al., 2006). Steeper boundaries with narrower widths indicate higher categorization.

(2) Discrimination Scores and Peaks

Discrimination rates were calculated as: P = P("S" |S) × P(S) + P("D" |D) × P(D) (Xu et al., 2006), where P("S" |S) is the probability of judging identical pairs as same, P("D" |D) is the probability of judging different pairs as different, and P(S) and P(D) are the proportions of same and different pairs (1/3 and 2/3, respectively, in this experiment). To locate discrimination peaks, we conducted One-way ANOVA and Tukey HSD post-hoc tests to identify which sample pairs showed significantly higher discrimination rates than others.[1]

[1] Common pairwise comparison methods include LSD, Scheffe, Dunnett, and Tukey. Tukey' s Honestly Significant Difference is appropriate for equal sample sizes across groups, as in this study.

3.5 Experimental Results

(1) i21-i51 Tonal Contrast

Figure 4 [Figure 4: see original paper]: Perception Results for i21-i51 Contrast

The left panel shows results with i21 as the base. Identification boundary and width were 4.22 and 3.04, respectively, with the boundary near the continuum' s midpoint and relatively steep. The right panel shows results with i51 as the base, yielding boundary and width of 4.44 and 2.99, also near the midpoint and steep. The two boundaries largely coincide. For discrimination, One-way ANOVA and Tukey HSD tests revealed significant between-group differences among the nine sample pairs ($F(8,243) = 5.758$, $p = 0.000$) and ($F(8,243) = 4.927$, $p = 0.000$). Pairwise comparisons divided the nine pairs into three subsets with no significant differences within subsets but significant differences between subsets. However, as shown in Table 1 , the three subsets overlap substantially, preventing identification of a clear discrimination peak.

Table 1: Homogeneous Subsets of Discrimination Rates for i21-i51 Contrast

(2) po21-po51 Tonal Contrast

Figure 5 [Figure 5: see original paper]: Perception Results for po21-po51 Contrast

The left panel shows results with po21 as the base, with identification boundary and width of 6.14 and 2.79, respectively, located slightly right of center with relatively steep slope. The right panel shows results with po51 as the base, yielding boundary and width of 4.46 and 3.91, also near the midpoint and steep. Base stimulus had minimal effect on boundary location. For discrimination, the left panel showed significant between-group differences ($F(8,243) = 3.827$, $p = 0.000$), but as shown in Table 2 , pairwise comparisons yielded three overlapping subsets, preventing peak identification. The right panel showed no significant between-group differences ($F(8,243) = 1.574$, $p = 0.133$), indicating no discrimination peak.

Table 2: Homogeneous Subsets of Discrimination Rates for po21-po51 (po21) and mau35-mau44 (mau35) Contrasts

(3) mau35-mau44 Tonal Contrast

Figure 6 [Figure 6: see original paper]: Perception Results for mau35-mau44 Contrast

The left panel shows results with mau35 (modal voice) as the base. Identification boundary and width were 10.21 and 5.86, respectively, with an extremely flat boundary near the final stimulus and wide width. Subjects identified the first stimulus as mau35 nearly 100% of the time and as mau44 nearly 0%. As F0 progressively changed from 35 to 44, mau35 identification decreased but never reached 0%, while mau44 identification increased but never reached 100%. At stimuli 10 and 11, identification rates for both tones were approximately 50% each.

The right panel shows results with mau44 (pressed voice) as the base. Identification boundary and width were 4.82 and 3.04, respectively, near the continuum' s midpoint. Compared to the left panel, perception differed significantly, with a much narrower width and steeper boundary. mau35 identification decreased from 100% to 0%, while mau44 identification increased from 0% to 100%.

For discrimination, the left panel showed significant between-group differences ($F(8,243) = 3.537$, $p = 0.001$), but as shown in Table 2, pairwise comparisons yielded three overlapping subsets, preventing peak identification. The right panel showed no significant between-group differences ($F(8,243) = 2.250$, $p = 0.025$).

The results show that for the "i21-i51" and "po21-po51" contrasts, identification boundaries were relatively steep, with high-falling and low-falling tones clearly distinguished as two separate phonemic categories. Within each category, F0 variation was perceived as the same tone, while crossing the category boundary triggered perception of the other tone. Boundaries were consistent across different base stimuli, indicating that F0 primarily drives the distinction between these two tone pairs. While "clear identification boundaries" and "discrimination peaks corresponding to boundaries" are widely accepted criteria for categorical perception, these results show clear boundaries but no statistically significant discrimination peaks. The behavioral perception pattern is therefore continuous. Previous research suggests that native speakers perceive same-contour tones (e.g., both falling) continuously and different-contour tones categorically. Although both 21 and 51 are falling tones, their slopes differ substantially, representing both contour and pitch height differences. Yet native perception was not categorical, contradicting previous findings and requiring further investigation.

More intriguingly, the mau35-mau44 contrast showed significant differences depending on the base stimulus.

In the left panel of Figure 6, with modal voice as the base, the identification boundary was extremely flat, and subjects could hardly distinguish mau35 and mau44 as separate phonemic categories. Why? First, with modal voice as the base, even when F0 changed from 35 to 44, subjects could not identify stimuli as mau44 without the pressed voice quality. Second, considering Zaiwa phonology, although the syllable mau44 does not exist, the combination "initial (aspirated or voiceless fricative) + modal voice + high level tone 44" does exist, as shown in Table 4 with /khau44/ "cousin." When the base was modal voice, subjects became confused by stimuli approaching mau44, causing confusion with mau44 and preventing complete identification as either mau35 or mau44 (both around 50%). Thus, lacking phonation type, F0 did not contribute to perception, resulting in "F0 perceptual neutrality."

Table 4: Combination of Initials and Finals for High Level and High

Rising Tones in Zaiwa

Initial	Unaspirated (except voiceless fricatives)	Voiceless fricatives/Aspirated	All initials	Voiceless fricatives/Aspirated	All initials
Final	Modal + stop final	Pressed + stop final	/mau/ strange	/khau/ cousin	/mau/ deceive
			/ ut/ wrong	/tsut/ lung	

In the right panel of Figure 6, with pressed voice as the base, identification boundaries were clear, with two distinct phonemic categories. When F0 approached 44, subjects clearly identified mau44 using both "F0" + "pressed voice" cues. When F0 approached 35, since Zaiwa lacks the combination "pressed voice final + tone 35," subjects ignored the anomalous phonation and identified the stimulus as mau35 based on F0 alone. These results demonstrate that both F0 and phonation type contribute significantly to perception of pressed-voice tones. However, due to phonemic complementation, phonation type lost its role in modal voice identification, with subjects identifying pressed-voice stimuli as modal-voice syllables. Although identification boundaries were relatively steep, the inability to identify discrimination peaks indicates that the behavioral perception pattern remains continuous.

In summary, these perceptual results demonstrate that the high-level (44) and high-rising (35) tones in Zaiwa are distinguished by both F0 and phonation type as distinctive features, which jointly play important roles in tone perception.

This study first determined through acoustic analysis that Zaiwa has six tones: two level tones (55, 44), three falling tones (51, 31, 21), and one rising tone (35). We then examined perception of high-level vs. high-rising tones and low-falling vs. high-falling tones. Results indicate that native speakers perceive tonal contrasts with identical phonation types continuously. For contrasts differing in both tone and phonation type, perception was continuous when the base was pressed voice; when the base was modal voice, subjects could not distinguish phonemic categories. This demonstrates that both F0 and phonation type contribute to tone perception in Zaiwa, carrying linguistic significance. When phonation type is absent, subjects exhibit "F0 perceptual neutrality."

References

Zhang, Ruifeng, & Kong, Jiangping. (2014). An acoustic and perceptual study of tones in the Yuzhou dialect of Henan. *Fangyan*, (3), 206-214.

Liu, Wen, & Zhang, Ruifeng. (2016). An acoustic and perceptual study of low-level and low-falling tones in Yuliang Miao. *Linguistic Essays*, (2), 197-212.

Xu, Xijian, & Xu, Guizhen. (1984). *A brief description of Jingpo languages (Zaiwa)*. Minzu Press.

Dai, Qingxia. (1989). A study of Zaiwa tones. *Journal of Minzu University of China: Philosophy and Social Sciences Edition*, (1), 61-66.

He, Lela. (2016). *A phonological study of Zaiwa* (Unpublished master' s thesis). Shanghai Normal University.

Liu, Fu. (1924). *Record of experiments on the four tones*. Qunyi Book Company.

Kong, Jiangping. (2015). *Fundamentals of experimental phonetics*. Peking University Press.

Kong, Jiangping. (2001). *On linguistic phonation*. Minzu University Press.

Zhang, Wentong, & Dong, Wei. (2004). *Advanced SPSS statistical analysis*.

Jia, Junping. (2006). *Statistics*. Tsinghua University Press.

Yu, Qian. (2016). *A study on the influence of dialectal background on Mandarin tone categorical perception* (Unpublished doctoral dissertation). Peking University.

Wang, W. S. Y. (1976). Language change. *Annals of the New York Academy of Sciences*, 280(1).

Abramson, A. S. (1977). Noncategorical perception of tone categories in Thai. *The Journal of the Acoustical Society of America*, 61(S1), S66-S66.

Francis, A. L., Ciocca, V., & Ng, B. K. C. (2003). On the (non) categorical perception of lexical tones. *Perception & Psychophysics*, 65(7), 1029-1044.

Francis, A. L., & Ciocca, V. (2003). Stimulus presentation order and the perception of lexical tones in Cantonese. *The Journal of the Acoustical Society of America*, 114(3), 1611-1621.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 13(2), 253-260.

Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America*, 120(2), 1063-1074.

Author Information

LU Yao (100871 Beijing, Department of Chinese Language and Literature, Peking University / Renmin University of China Library, luyiru2012@163.com)

KONG Jiangping (100871 Beijing, Center for Chinese Linguistics, Peking University / Department of Chinese Language and Literature, jpkong@pku.edu.cn)

Note: Figure translations are in progress. See original paper for figures.