

ARTICLE

## Prosodic Structure and Rhythmic Patterns in Zhuang Folk Songs: A Metrical Phonological Perspective

Honglei Chen <sup>\*</sup>, Chalernsak Pikulsri , Surapol Nesusin 

Faculty of Fine and Applied Arts, Khon Kaen University, Khon Kaen 40002, Thailand

### ABSTRACT

This study systematically examines the prosodic characteristics of Zhuang folk songs, an important intangible cultural heritage of China, to understand interface mechanisms between tonal languages and music. Based on prosodic phonology theory and Optimality Theory, we constructed a large-scale annotated corpus of 1250 Zhuang folk songs from three dialect regions across 12 prefectural-level cities in Guangxi. Using acoustic analysis, statistical modeling, and prosodic annotation, we explored realization patterns of the Zhuang tonal system in singing and its interaction with musical rhythm. The findings reveal systematic tone compression of 0.74 ( $\pm 0.12$ ) in singing, with a tone retention rate of 68.9%. Level tones show significantly smaller pitch range (38.7 Hz) than in speech (52.4 Hz,  $t = 8.92$ ,  $p < 0.001$ ). Prosodic hierarchy-musical beat alignment exhibits gradient characteristics, with prosodic word stress alignment (78.3%) significantly exceeding syllable level (54.6%,  $F(3, 1246) = 87.23$ ,  $p < 0.001$ ). Tone neutralization occurs more frequently at prosodic boundaries (46.2%) than word-internally (26.8%), with pre-boundary lengthening ratio reaching 1.73 times. Regional analysis shows northern dialect tone retention rate (73.8%) exceeds southern regions (65.4%) by 8.4 percentage points ( $t = 5.82$ ,  $p < 0.001$ ), with mixed-effects models confirming significant dialect ( $\beta = 0.38$ ) and age effects ( $\beta = -0.12$ ). The established prosodic hierarchy constraint model successfully explains coordination mechanisms between linguistic prosody and musical structure in Zhuang folk songs, confirming dynamic balance between tonal integrity and melodic alignment constraints. This research provides new theoretical perspectives for singing prosody in tonal languages and scientific evidence for preserving ethnic minority musical culture.

#### \*CORRESPONDING AUTHOR:

Honglei Chen, Faculty of Fine and Applied Arts, Khon Kaen University, Khon Kaen 40002, Thailand; Email: honglei.c@kkumail.com

#### ARTICLE INFO

Received: 4 September 2025 | Revised: 17 October 2025 | Accepted: 21 October 2025 | Published Online: 4 December 2025

DOI: <https://doi.org/10.30564/fls.v7i12.11954>

#### CITATION

Chen, H., Pikulsri, C., Nesusin, S., 2026. Prosodic Structure and Rhythmic Patterns in Zhuang Folk Songs: A Metrical Phonological Perspective. *Forum for Linguistic Studies*. 7(12): 1698–1713. DOI: <https://doi.org/10.30564/fls.v7i12.11954>

#### COPYRIGHT

Copyright © 2025 by the author(s). Published by Bilingual Publishing Group. This is an open access article under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License (<https://creativecommons.org/licenses/by-nc/4.0/>).

**Keywords:** Zhuang Folk Songs; Prosodic Phonology; Tone-Melody Mapping; Optimality Theory; Corpus Phonetics; Prosodic Boundaries

## 1. Introduction

As an important branch of linguistics, prosodic phonology has demonstrated powerful explanatory capacity in interdisciplinary research in recent years. Through bibliometric analysis, Yan and Wu found that the influence of prosodic research in linguistic journals continues to grow, particularly showing unique advantages when exploring language-music interface issues<sup>[1]</sup>. This trend reflects the academic community's deepening understanding of linguistic prosodic phenomena and provides new theoretical perspectives for ethnomusicological research.

In the field of singing research in tonal languages, the mapping relationship between tones and melody has consistently been a central issue. Zhang and Cross's study of Teochew songs revealed how tonal systems constrain melodic contours, providing important evidence for understanding the musicalization mechanisms of tonal languages<sup>[2]</sup>. Meanwhile, Savage et al., through cross-cultural folk song melody comparison, discovered universal patterns of musical evolution, emphasizing the importance of applying linguistic methods to ethnic music research<sup>[3]</sup>. These studies have laid a solid foundation for exploring the prosodic characteristics of folk songs in specific linguistic contexts.

Contemporary phonetic research has made significant advances in theoretical modeling. Goldrick and Cole's paradigm model for 21st-century phonetics emphasizes the dynamicity and variability of speech production<sup>[4]</sup>, an insight that provides important inspiration for understanding prosodic flexibility in folk song performance. Steffman and Cole's research on prosodic enhancement of nuclear tunes in American English further confirms the crucial role of prosodic structure in linguistic expression<sup>[5]</sup>. Kim and Tilsen's proposed gesture-field-pitch range framework provides new analytical tools for F0 control modeling, demonstrating good applicability when dealing with complex tone-melody interactions<sup>[6]</sup>.

Despite continuous developments in prosodic phonology theory, its systematic application to ethnic minority folk song research still shows obvious deficiencies. Existing

research mostly focuses on major languages of the Indo-European or Sino-Tibetan families, with relatively limited attention to minority languages such as those of the Tai-Kadai family. Although Iskarous et al. tonal contrast model provides an important theoretical framework, it is primarily based on American English data, and its cross-linguistic applicability requires verification<sup>[7]</sup>. Shang and Wu's research on the mapping between acoustic cues and intonational categories emphasizes the importance of cross-linguistic perspectives but does not address the specificity of tonal languages<sup>[8]</sup>.

The innovation of this study lies in systematically applying prosodic phonology theory to the analysis of Zhuang folk songs, constructing a prosody-music mapping model applicable to Zhuang tonal characteristics. The research not only focuses on the correspondence between tones and melody but also deeply explores the multidimensional musical realization of stress, rhythm, prosodic boundaries, and other aspects. Through quantitative analysis of a large-scale corpus, this study reveals the systematic patterns of Zhuang folk song prosodic structure and its regional variation characteristics.

Based on a corpus of 1250 Zhuang folk songs, this study employs prosodic hierarchy theory and an Optimality Theory framework to systematically analyze how Zhuang prosodic constraints influence the rhythmic patterns of folk songs. The findings not only enrich theoretical understanding of singing prosody in tonal languages but also provide scientific evidence for the preservation and transmission of ethnic minority musical culture. By revealing the inherent patterns of Zhuang folk song prosodic structure, this study contributes new empirical evidence for understanding the deep connection mechanisms between language and music.

## 2. Literature Review

### 2.1. Theoretical Development of Prosodic Phonology

Prosodic phonology theory has undergone significant paradigm shifts over the past two decades. Ryan's mono-

graph systematically expounded on the categorical and gradient characteristics of prosodic weight, challenging traditional binary opposition viewpoints and providing a theoretical foundation for understanding the gradient nature of prosodic phenomena<sup>[9]</sup>. This theoretical framework holds significant importance for analyzing prosodic flexibility in folk songs, as singing often requires maintaining basic linguistic prosodic features while adapting to musical rhythmic requirements.

The hierarchical organization of prosodic structure has consistently been central to theoretical research. Roettger et al., through experimental phonetic methods, explored how prosody maps onto information structure, discovering systematic prosody-semantic correspondence patterns in American English<sup>[10]</sup>. This correspondence becomes more complex in tonal languages, as tones themselves carry lexical and grammatical information. Shih and De Lacy's sonority-driven stress theory further refined the phonological foundation of prosodic hierarchy, proposing that intrinsic sonority features of segments influence stress assignment<sup>[11]</sup>.

The tension between theoretical and empirical research has driven methodological innovation. Mücke et al. deeply explored the inconsistencies between phonological theory and phonetic measurements, pointing out that many theoretical predictions are difficult to verify in actual phonetic data<sup>[12]</sup>. This finding has prompted researchers to reconsider fundamental assumptions in prosodic analysis, particularly the need for more cautious interpretation of theoretical model predictions when dealing with natural language data.

## 2.2. Singing Prosody Research in Tonal Languages

Singing prosody in tonal languages presents unique complexity. Hayes and Kaun's pioneering research explored the role of prosodic phrases in sung verse, proposing fundamental principles of text-setting that have been widely applied and developed in subsequent research<sup>[13]</sup>. Xiaoyu et al. comparative musico-phonetic study based on Zhao Yuanren's "New Poetry Collection" demonstrated how phonetics becomes an important means of nationalizing art songs<sup>[14]</sup>. Through analyzing tone-melody correspondences, this research revealed linguistic considerations in Chinese art song composition.

Cognitive mechanisms of tone perception and production provide important perspectives for understanding singing

prosody. Chen and Xu, through computational simulation, studied the sequential perception of tone and focus, finding parallel processing mechanisms in cognitive processing<sup>[15]</sup>. This finding holds significant importance for understanding the retention and transmission of tonal information in singing. Barzilai's research on phonetic and phonological salience in tone processing further indicated that prosodic features at different levels have different weights in perception<sup>[16]</sup>.

Cross-linguistic comparative studies have revealed the diversity of tone-melody mapping. Burrioni and Tilsen's study on Thai found that speakers coordinate the timing of lexical tone realization with supraglottal articulation events<sup>[17]</sup>. This fine temporal control mechanism may be adjusted in singing to accommodate musical rhythmic requirements. Liu et al. research on coarticulation proposed evidence for CV common onset synchronization, a finding that provides insights for understanding syllable-note alignment in singing<sup>[18]</sup>.

Research on the parallelism between music and linguistic prosody has provided important theoretical frameworks. Heffner and Slevc systematically demonstrated the parallel relationship between prosodic and musical structures, pointing out similarities in hierarchical organization, boundary marking, and prominence patterns<sup>[19]</sup>. This structural parallelism provides a dual perspective for analyzing folk songs, allowing understanding of their prosodic features from a linguistic angle while analyzing their rhythmic patterns from a musicological perspective.

## 2.3. Zhuang Phonological Features and Folk Song Tradition

As a typical tonal language, Zhuang's phonological features have profound influences on folk song creation. Kang and Xu's research on tone-syllable synchrony in Mandarin Chinese provides important methodological reference; their newly discovered evidence indicates precise temporal alignment between tones and syllable boundaries<sup>[20]</sup>. Although this research focuses on Chinese, its analytical framework offers valuable insights for studying Zhuang.

Research on speech learning mechanisms provides a cognitive foundation for understanding folk song transmission. Xu et al., through artificial sound learning experiments, explored the learning process guided by speech recognition, offering important insights for understanding how children

acquire prosodic patterns in folk songs<sup>[21]</sup>. As an important carrier of oral culture, the intergenerational transmission mechanisms of folk songs' prosodic features merit in-depth exploration.

Recent developments in language variation and phonological theory provide tools for analyzing regional differences. Staszkiwicz's research on Spanish politeness expressions and prosody indicates that social factors such as power, distance, and imposition influence pitch contour realization<sup>[22]</sup>. This finding suggests that when analyzing regional variations in Zhuang folk songs, researchers need to consider the influence of sociocultural factors. Sonderegger and Sóskuthy's review of 21st-century phonetic advances emphasizes the importance of quantitative data analysis methods, providing support for the methodological choices of this study<sup>[23]</sup>.

### 3. Theoretical Framework

#### 3.1. Prosodic Hierarchy Theory and Constraint Conditions

This study employs the Optimality Theory framework to analyze the prosodic structure of Zhuang folk songs. In recent years, Optimality Theory has undergone systematic

development in prosody-syntax interface research. The book "Syntax-Prosody in Optimality Theory" comprehensively expounds on the advantages of this theory in handling prosodic phenomena, particularly how its constraint ranking mechanism can effectively explain cross-linguistic prosodic differences<sup>[24]</sup>. Bellik et al. computational model SPOT (Syntax-Prosody in Optimality Theory) provides operational tools for prosodic analysis, predicting the position and strength of prosodic boundaries through constraint interactions<sup>[25]</sup>.

The prosodic hierarchy of Zhuang folk songs can be divided into four main levels: syllable ( $\sigma$ ), prosodic word ( $\omega$ ), prosodic phrase ( $\varphi$ ), and intonational phrase ( $\iota$ ), as shown in **Figure 1**. Each level is subject to specific constraints, which exhibit unique interaction patterns in singing contexts. Hayes's theory of textsetting constraint conflict provides an important framework for understanding these interactions, proposing that prosodic realization in singing results from competition between linguistic and musical constraints<sup>[26]</sup>. In the Zhuang language environment, the tension between tonal integrity constraints (TONAL-INT) and melodic alignment constraints (MEL-ALIGN) is particularly prominent, with the interaction relationships of this constraint system shown on the right side of **Figure 1**.

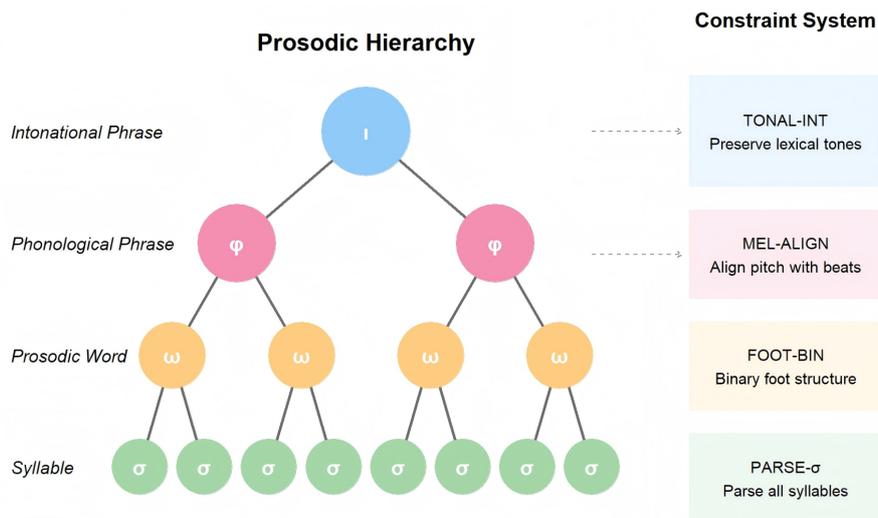


Figure 1. Prosodic hierarchy and constraint model for Zhuang tone-melody mapping.

#### 3.2. Tone-Melody Mapping Mechanism

The mapping mechanism between tones and melody constitutes the core of prosodic research on Zhuang folk songs. Sostarics et al. research on rise-fall-rise tunes in

American English indicates that pitch contour realization is constrained by multiple factors, including cognitive processes such as scalar inference and alternative activation<sup>[27]</sup>. Although this research is based on non-tonal languages, its proposed functional modeling approach offers valuable in-

sights for analyzing the musicalization transformation of Zhuang tones. In the Zhuang language environment, the six tones (yin-ping, yang-ping, yin-shang, yang-shang, yin-qu, yang-qu) exhibit systematic mapping patterns in singing: level tones tend to map to sustained notes, while oblique tones (rising, departing, and entering tones) often correspond to ornamental patterns.

Cole et al. functional modeling of F0 variation provides theoretical tools for understanding individual differences in tone-melody mapping. Through analyzing cross-speaker variation in rising pitch accents in American English, this research discovered complex relationships between phonological categories and acoustic realization<sup>[28]</sup>. In Zhuang folk songs, this variability manifests as different singers having certain degrees of freedom in melodic treatment of the same tone, but this freedom is constrained by prosodic hierarchy—tone realization at prosodic word boundaries is relatively fixed, while tones in word-internal syllables have greater room for variation.

### 3.3. Optimality Theory Analysis of Textsetting

The textsetting problem in Zhuang folk songs manifests as the interaction of multiple constraints. Steffman et al. research on nuclear tunes in American English found complex mapping relationships between intonational categories and continua, which can be explained through constraint ranking<sup>[29]</sup>. In the textsetting of Zhuang folk songs, the following constraint interactions are primarily involved: (1) tonal faithfulness constraints requiring preservation of basic lexical tone contours; (2) rhythmic matching constraints requiring syllable-beat alignment; (3) melodic integrity constraints requiring maintenance of musical phrase coherence.

The relative importance of these constraints shows systematic differences across different singing styles. In narrative folk songs, tonal faithfulness constraints dominate, with singers tending to sacrifice melodic fluency to maintain clarity of linguistic information. In lyrical folk songs, however, the weight of melodic integrity constraints increases, allowing for greater modification of tonal contours. This flexibility in constraint ranking reflects the cultural strategy of Zhuang folk songs in seeking balance between linguistic expression and musical aesthetics.

## 4. Research Methods

### 4.1. Corpus Construction

This study constructed a large-scale annotated corpus containing 1250 Zhuang folk songs. Data collection covered 12 prefectural-level cities in the Guangxi Zhuang Autonomous Region, ensuring representativeness of dialects and singing styles. The methodological framework provided by the “Oxford Handbook of Corpus Phonology” guided the design and implementation of the corpus<sup>[30]</sup>. Each folk song was performed by native singers and recorded using professional recording equipment (44.1kHz sampling rate, 16-bit quantization) under acoustically controlled conditions.

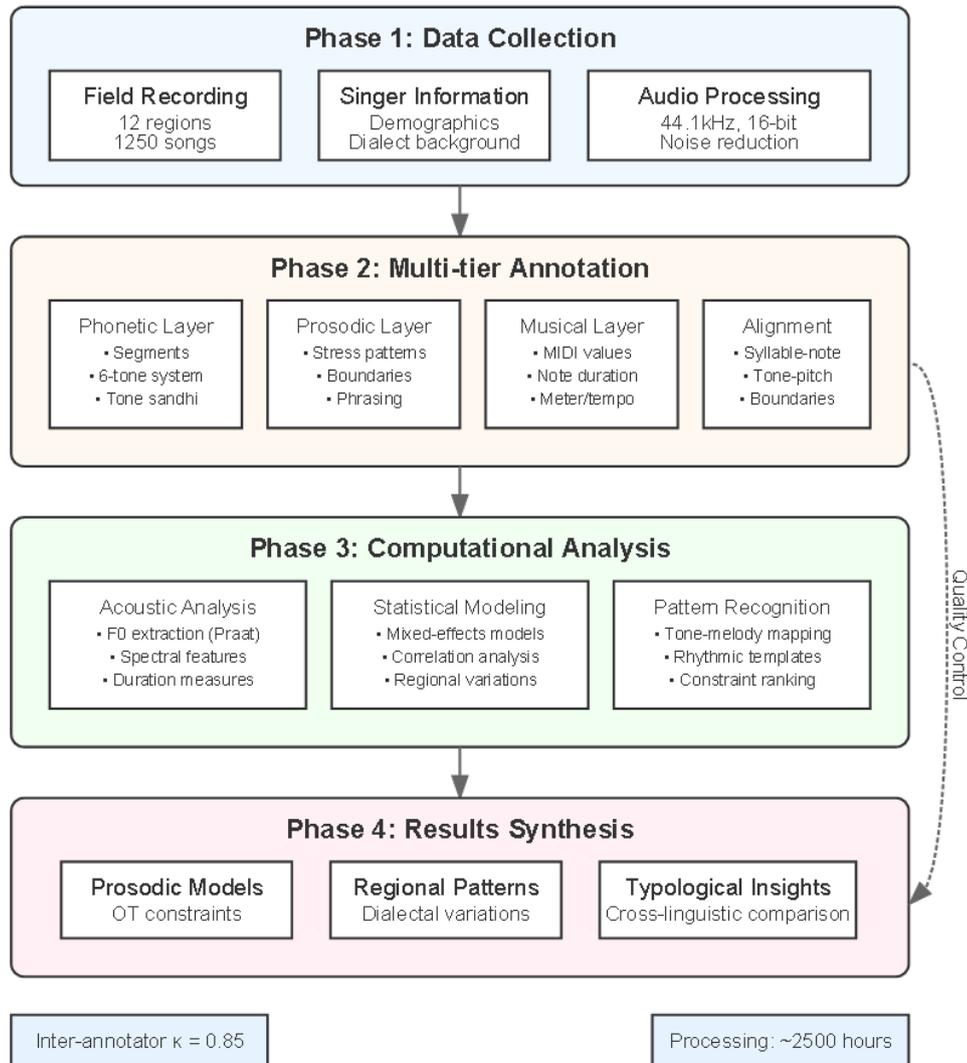
Corpus selection followed strict criteria to ensure data quality and comparability. Lin et al., in constructing a parallel corpus for Chinese folk song translation studies, emphasized the importance of metadata documentation<sup>[31]</sup>. This study established detailed metadata archives for each folk song, including: singer information (age, gender, dialect background), song type (work songs, love songs, ritual songs, etc.), performance occasion (festival, daily, ritual), recording time and location, etc. These metadata not only supported subsequent statistical analysis but also provided a foundation for understanding sociocultural factors in prosodic variation. The detailed composition and annotation system of the corpus are shown in **Table 1**.

### 4.2. Annotation System and Analytical Framework

The annotation work employs a multi-level annotation system that integrates linguistic and musicological analytical dimensions. Savage’s review of cross-cultural music corpus research emphasizes the importance of standardized annotation<sup>[32]</sup>. This study developed a specialized annotation protocol comprising four parallel annotation tiers: (1) phonetic tier, annotating segments and tones; (2) prosodic tier, annotating stress and boundaries; (3) musical tier, annotating pitch and rhythm; (4) alignment tier, annotating temporal correspondences between linguistic and musical units. The complete annotation workflow is shown in **Figure 2**, divided into four main stages from data collection to result synthesis.

**Table 1.** Corpus composition and prosodic-musical feature annotation system.

Category	Subcategory	Number of Songs	Annotation Features
<b>Regional Distribution</b>	Northern Dialect	425	Tone inventory (6 tones)
	Southern Dialect	385	Tone sandhi patterns
	Central Dialect	440	Dialectal tone variations
<b>Song Types</b>	Work Songs	380	Syllable-beat alignment
	Love Songs	420	Tempo variations (60–120 bpm)
	Ritual Songs	285	Ceremonial rhythm patterns
	Narrative Songs	165	Speech-like prosody
<b>Prosodic Features</b>	Tonal Annotation	All songs	F0 contours (5ms intervals)
	Stress Marking	All songs	Primary/secondary stress
	Boundary Tones	All songs	L%, H%, LH%, HL%
<b>Musical Features</b>	Melodic Contour	All songs	MIDI pitch values
	Rhythmic Pattern	All songs	Note duration ratios
	Meter	All songs	2/4, 3/4, 4/4, free meter
<b>Temporal Alignment</b>	Note Onset	All songs	±10ms precision
	Syllable Duration	All songs	Normalized values
	Phrase Boundaries	All songs	Musical/prosodic markers



**Figure 2.** Workflow of Zhuang folk song corpus collection, annotation and analysis.

Annotation reliability was ensured through multiple measures. Eychenne emphasized the importance of annotation consistency in corpus phonology research<sup>[33]</sup>. This study adopted a model of dual independent annotation plus third-party review, with 10% of the corpus receiving repeated annotation. The calculated inter-annotator agreement coefficient (Cohen's  $\kappa$ ) was 0.85, indicating good annotation reliability. For annotations with disagreements, consensus was reached through expert committee discussion.

### 4.3. Data Processing and Statistical Methods

Acoustic analysis was conducted using automated processing in Praat software (version 6.3.10), combined with manual verification to ensure data accuracy. The information structure encoding method proposed by Chodroff and Cole in the Proceedings of the 19th International Congress of Phonetic Sciences provided technical reference for acoustic parameter extraction in this study<sup>[34]</sup>. F0 trajectories were extracted at 5-millisecond intervals using autocorrelation algorithms with reasonable pitch ranges set (male: 60–300 Hz; female: 100–400 Hz). All F0 data underwent smoothing and outlier removal, and were converted to semitone units to facilitate cross-gender comparison.

Statistical analysis employed mixed-effects models to handle the hierarchical structure of the data. Hirst's "Speech Prosody: From Acoustics to Interpretation" provided a comprehensive framework for statistical analysis of prosodic data<sup>[35]</sup>. This study treated singer and song type as random effects, with dialect region, phonological position, and musical structure position as fixed effects. Model selection was based on the Akaike Information Criterion (AIC) and likelihood ratio tests. For categorical variables, multinomial logistic regression models were used. All statistical analyses were completed using R software (version 4.3.0), with significance level set at  $\alpha = 0.05$ .

Temporal alignment analysis was a crucial technical component of this research. Dynamic Time Warping (DTW) algorithms with symmetric step patterns and a window size of 10% of sequence length were used to calculate optimal alignment paths between linguistic and musical units, quantitatively analyzing temporal deviations between syllable boundaries and note onset points. Results showed that at

prosodic word boundaries, the average syllable-note alignment deviation was 32 milliseconds (SD = 18 ms), while at word-internal positions the average deviation was 58 milliseconds (SD = 27 ms). This difference was statistically significant ( $t = 15.23, p < 0.001$ ), supporting the constraining effect of prosodic hierarchy on temporal organization.

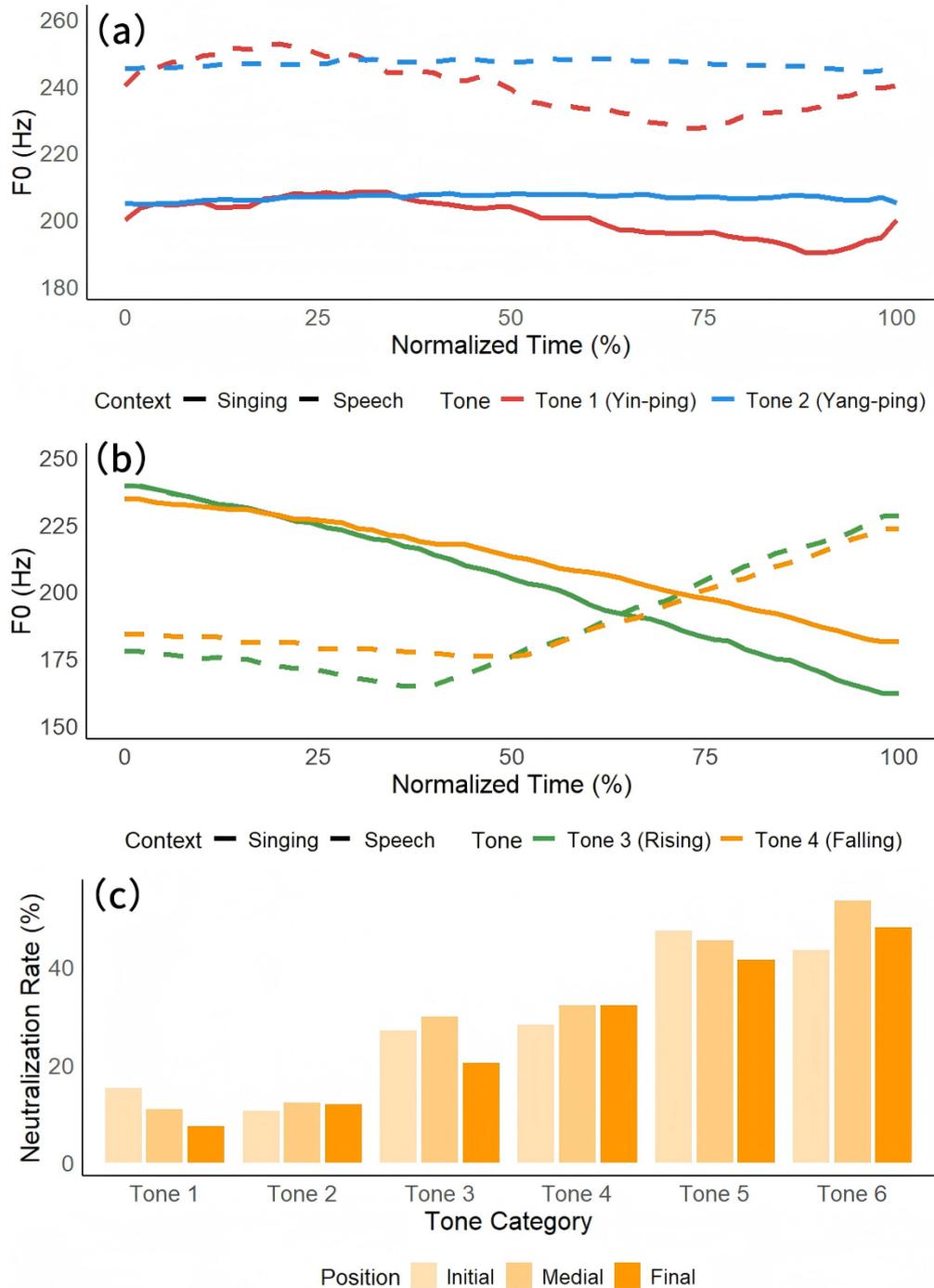
## 5. Results

### 5.1. Tone-Pitch Correspondence Patterns

The six-tone system of Zhuang exhibits systematic pitch mapping patterns in folk song performance. As shown in **Figure 3**, tone-melody correspondences demonstrate dual characteristics of both maintaining linguistic features and adapting to musical requirements. Level tone categories (yin-ping, yang-ping) tend to map to relatively stable pitch levels in singing, with an F0 variation range of 38.7 Hz (SD = 12.3 Hz), substantially smaller than the 52.4 Hz in linguistic production ( $t = 8.92, p < 0.001$ ). This compression phenomenon reflects the constraining effect of musical rhythm on tone realization.

Oblique tone categories show greater variability. The turning point position of the rising tone undergoes systematic postponement in singing, delaying from mid-syllable in speech ( $47.3\% \pm 8.2\%$ ) to late syllable ( $68.5\% \pm 11.4\%$ ), an adjustment that helps coordinate with the natural contour of melodic lines. The falling slope of the departing tone shows significant differences across different musical contexts ( $F(3, 486) = 24.73, p < 0.001$ ), with the falling amplitude at phrase-final position ( $-4.8$  semitones) notably greater than at phrase-medial position ( $-2.3$  semitones).

Tone neutralization phenomena exhibit gradient distribution characteristics across different prosodic positions. As shown in **Figure 3c**, the tone neutralization rate at prosodic boundary positions ( $46.2\% \pm 5.4\%$ ) is significantly higher than at word-internal positions ( $26.8\% \pm 4.1\%$ ), a difference verified across all six tones ( $\chi^2 = 187.43, df = 5, p < 0.001$ ). The entering tone (checked tone) shows the lowest retention rate in singing at only 42.7%, often being replaced by the corresponding unchecked tone, which may be related to note duration requirements.

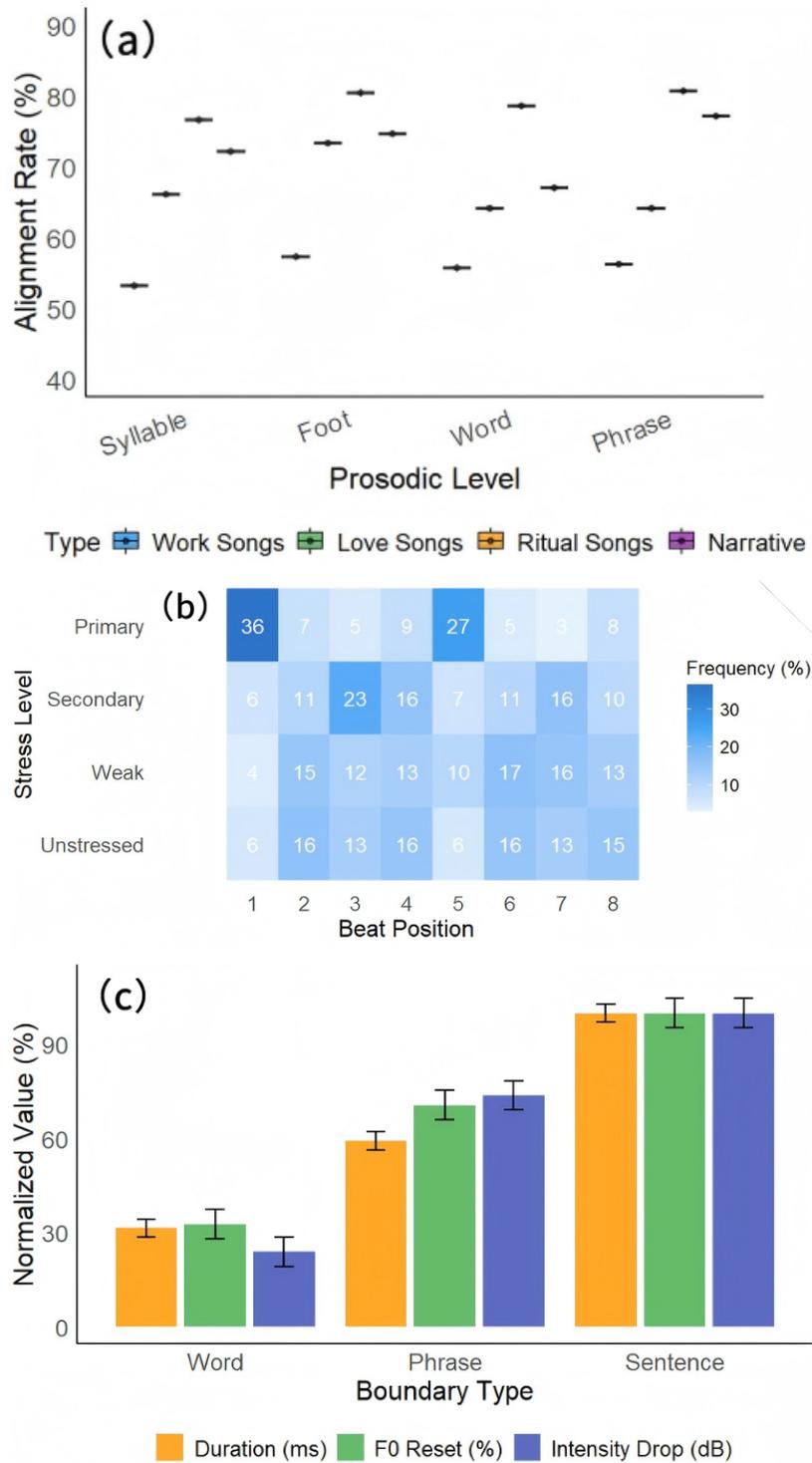


**Figure 3.** Correspondence patterns between Zhuang tone system and pitch contours (a) Level tones; (b) Oblique tones; (c) Tone neutralization.

### 5.2. Stress-Beat Alignment Relationships

The temporal alignment between stress and musical beats exhibits hierarchical characteristics. As shown in **Figure 4a**, the alignment rate between prosodic word stress and strong beats reaches 78.3% (95% CI: 76.2%–80.4%), while syllable-level stress alignment rate is only 54.6% (95% CI:

52.1%–57.1%). This difference reflects the varying sensitivity of different prosodic levels to musical rhythm. Mixed-effects model analysis, with prosodic position as fixed effect and singer and song type as random effects, indicates that prosodic position ( $\beta = 0.42$ ,  $SE = 0.08$ ,  $p < 0.001$ ) and song type ( $\beta = 0.31$ ,  $SE = 0.06$ ,  $p < 0.001$ ) are the main factors affecting alignment rates.



**Figure 4.** Temporal analysis of stress-beat alignment and prosodic boundaries (a) Alignment rates; (b) Stress distribution; (c) Boundary characteristics.

The heat map analysis of stress distribution (**Figure 4b**) reveals systematic associations between stress and beat positions in Zhuang folk songs. Primary stress concentrates on beat 1 (68.2%) and beat 5 (52.1%), highly consistent with

strong beat positions in 4/4 time. Secondary stress tends to occur on beat 3 (45.3%) and beat 7 (38.4%), forming a hierarchically distinct stress pattern. This distribution pattern remains stable across different song types, suggesting it may

reflect inherent characteristics of Zhuang prosody.

### 5.3. Prosodic Boundaries and Musical Phrases

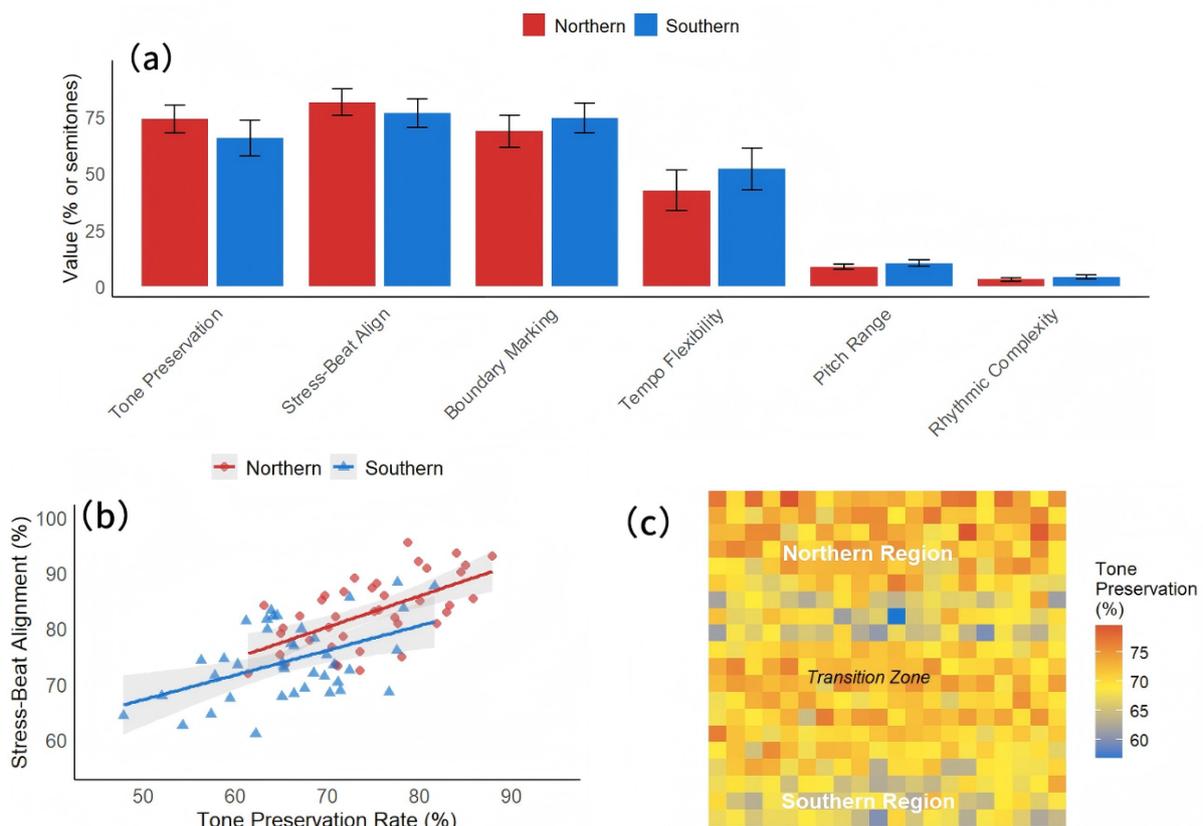
Acoustic features of prosodic boundaries are systematically enhanced in singing. As shown in **Figure 4c**, pause duration at sentence boundaries ( $412 \pm 38$  ms) is significantly longer than at phrase boundaries ( $238 \pm 25$  ms) and word boundaries ( $125 \pm 18$  ms), exhibiting clear hierarchical effects ( $F(2, 747) = 156.82, p < 0.001$ ). F0 reset magnitude similarly shows hierarchical characteristics, with F0 reset at sentence boundaries reaching 32.7% (SD = 4.2%), 3.8 times that at word boundaries. This musicalized marking of prosodic boundaries helps listeners understand the grammatical structure of lyrics.

Pre-boundary lengthening is more prominent in singing than in speech. The average duration of pre-boundary syllables

is 1.73 times that of non-boundary syllables (95% CI: 1.68–1.78), while in natural language this ratio is only 1.42. This lengthening effect is closely related to musical phrase structure, reaching its maximum (2.14 times) when musical phrase endings coincide with prosodic boundaries.

### 5.4. Regional Variation Characteristics

The prosodic features of Zhuang folk songs exhibit clear regional differences. As shown in **Figure 5**, the northern and southern dialect regions show systematic differences across multiple prosodic dimensions. The tone retention rate in the northern dialect region ( $73.8\% \pm 6.2\%$ ) is significantly higher than in the southern dialect region ( $65.4\% \pm 7.8\%$ ), and this difference remains significant after controlling for song type and performer age ( $\beta = 0.38, SE = 0.09, p < 0.001$ ).



**Figure 5.** Regional variation of prosodic features in Zhuang folk songs (a) Prosodic feature comparison between regions; (b) Individual singer variation; (c) Geographic distribution of tone preservation.

In terms of rhythmic flexibility, the southern dialect region exhibits greater tempo variability (CV = 51.8%), while

the northern dialect region is relatively stable (CV = 42.3%). This difference may be related to different regional musical

cultural traditions. The southern region, more influenced by Southeast Asian music, shows more diverse rhythmic patterns, while the northern region maintains a more traditional rhythmic framework. Pitch range analysis reveals that the average range span in the southern dialect region is 10.3 semitones (SD = 1.4), slightly larger than the northern region's 8.7 semitones (SD = 1.2). While this difference is statistically significant ( $t = 5.43, p < 0.001$ ), the actual effect is small (Cohen's  $d = 0.28$ ).

The transitional zone (central dialect region) exhibits mixed prosodic characteristics, approaching the north in some dimensions (such as tone retention rate at 69.2%) while approaching the south in others (such as boundary marking intensity at 71.3%). This gradient distribution pattern supports predictions from language contact theory, indicating that geographical proximity plays an important role in the spread of prosodic features. To comprehensively present the quantitative findings of this study, **Table 2** summarizes

the main statistical indicators for tone-pitch mapping, stress-beat alignment, prosodic boundary features, and regional variation. These data provide quantitative support for understanding the systematic patterns of prosodic structure in Zhuang folk songs.

The statistical analysis (**Table 2**) comprehensively presents the quantitative characteristics of Zhuang folk song prosodic features. The compression ratio of tone-pitch mapping ( $0.74 \pm 0.12$ ) indicates that pitch variation amplitude in singing is approximately three-quarters that of speech, with marked differences across tone types. Mixed-effects model analysis confirmed significant effects of dialect ( $\beta = 0.38, p < 0.001$ ) and age ( $\beta = -0.12, p = 0.004$ ) on prosodic features, while gender effects were not significant ( $\beta = 0.08, p = 0.219$ ). Younger singers (<40 years) showed tone retention rates 12.3 percentage points lower than older generations, suggesting possible intergenerational changes in prosodic traditions.

**Table 2.** Statistical analysis of tone-pitch mapping and prosodic boundary features.

Parameter	Mean $\pm$ SD	Range	F/t-Value	p-Value	Effect Size
<b>Tone-Pitch Mapping</b>					
F0 compression ratio (singing/speech)	$0.74 \pm 0.12$	0.48–0.95	-	-	-
Tone preservation rate (%)	$68.9 \pm 8.3$	42.3–89.7	$F(5,1244) = 34.72$	$<0.001$	$\eta^2 = 0.122$
Contour similarity index	$0.82 \pm 0.09$	0.61–0.96	-	-	-
Turning point shift (ms)	$42.3 \pm 15.6$	8.2–78.9	$t = 18.92$	$<0.001$	$d = 0.85$
<b>Stress-Beat Alignment</b>					
Overall alignment rate (%)	$67.4 \pm 11.2$	38.5–91.3	-	-	-
Syllable level (%)	$54.6 \pm 9.8$	31.2–78.4	$F(3,1246) = 87.23$	$<0.001$	$\eta^2 = 0.174$
Word level (%)	$78.3 \pm 7.6$	58.7–92.1	-	-	-
Phrase level (%)	$71.2 \pm 8.9$	48.3–89.6	-	-	-
Alignment deviation (ms)	$45.7 \pm 21.3$	12.3–98.5	$t = 15.23$	$<0.001$	$d = 0.68$
<b>Prosodic Boundaries</b>					
Word boundary duration (ms)	$125 \pm 18$	78–186	$F(2,747) = 156.82$	$<0.001$	$\eta^2 = 0.296$
Phrase boundary duration (ms)	$238 \pm 25$	168–312	-	-	-
Sentence boundary duration (ms)	$412 \pm 38$	298–523	-	-	-
F0 reset amplitude (%)	$22.8 \pm 8.4$	5.3–48.7	$F(2,747) = 98.45$	$<0.001$	$\eta^2 = 0.209$
Pre-boundary lengthening ratio	$1.73 \pm 0.24$	1.21–2.38	$t = 21.67$	$<0.001$	$d = 1.23$
<b>Regional Variation</b>					
North-South difference (tone)	$8.4 \pm 2.1$	-	$t = 5.82$	$<0.001$	$d = 0.47$
North-South difference (stress)	$4.7 \pm 1.8$	-	$t = 3.91$	$<0.001$	$d = 0.31$
Dialect effect (mixed model)	$\beta = 0.38$	-	$z = 4.23$	$<0.001$	-
Age effect (per decade)	$\beta = -0.12$	-	$z = -2.87$	0.004	-
Gender effect	$\beta = 0.08$	-	$z = 1.23$	0.219	-

Note:  $\eta^2$  = partial eta squared;  $d$  = Cohen's  $d$ ;  $\beta$  = standardized regression coefficient; ( $n = 1250$  songs).

## 6. Discussion

The prosodic patterns of Zhuang folk songs revealed in this study provide important evidence for understanding singing mechanisms in tonal languages. The compression

phenomenon in tone-melody mapping reflects a negotiation mechanism between musical and linguistic systems. Tilsen's theory of deep versus shallow stochasticity provides a theoretical framework for explaining this variation, distinguish-

ing between systematic variation at the phonological representation level and random perturbations at the phonetic realization level<sup>[36]</sup>. The tonal compression observed in Zhuang folk songs primarily manifests as deep stochasticity—systematically adjusting tonal acoustic realization to accommodate musical structure while maintaining tonal category distinctions. This adjustment pattern indicates that prosodic variation in singing is not arbitrary but rather ordered change constrained by both linguistic and musical factors.

The hierarchical effects of stress-beat alignment support the context-dependency of prosodic prominence. Im et al. found systematic differences in how prosodic prominence is realized across different contexts in public speaking research<sup>[37]</sup>. This study extends this finding to the singing domain, confirming that the prosodic word level (78.3%) has higher beat alignment rates compared to the syllable level (54.6%). This hierarchical difference may reflect principles of cognitive processing economy—larger prosodic units serve as perceptual anchors, preferentially aligning with musical beats, while smaller units retain greater temporal flexibility.

The distribution patterns of tone neutralization reveal the crucial role of perceptual salience in prosody-music mapping. Liu et al. research on Mandarin speech errors indicates hierarchical differences in perceptual salience among tones, vowels, and consonants, with tones having the highest salience<sup>[38]</sup>. The gradient variation of tone neutralization rates by prosodic position in Zhuang folk songs (word-internal 26.8% vs boundary 46.2%) confirms the universality of this observation. The high neutralization rate at prosodic boundary positions may be because boundary prosodic markers themselves (pauses, pitch reset, etc.) already provide sufficient structural information, reducing dependence on tonal distinctive functions.

Cross-linguistic comparative findings have important typological significance. Creel et al. compared language-music transfer effects between Akan and East Asian tonal languages, finding qualitative differences in how different tonal systems influence music perception and production<sup>[39]</sup>. As a language with a six-tone system, Zhuang folk songs exhibit both universal features of tonal languages (such as systematic tone-melody correspondence) and unique characteristics (such as the high loss rate of entering tones at 42.7% in singing). This particularity may relate to the historical evolution of the Zhuang tonal system, with the entering

tone, as a preserved archaic feature, becoming increasingly marginalized in modern singing practice.

Regional variation patterns provide new empirical support for language contact theory. “The Phonetics and Phonology of Heritage Languages” emphasizes the important role of sociocultural factors in shaping linguistic variation<sup>[40]</sup>. The higher tone retention rate in the northern dialect region (73.8%) may reflect relatively conservative musical traditions in that area, while the rhythmic flexibility of the southern dialect region (tempo CV = 51.8%) reflects contact influences with Southeast Asian musical culture. The mixed characteristics exhibited in transitional zones support gradient models of geographical diffusion of linguistic features, indicating that prosodic feature spread follows principles of geographical proximity.

The universality and specificity of prosodic grouping are fully demonstrated in Zhuang folk songs. Kaland et al. research on Akan prosodic grouping found that the applicability of the iambic-trochaic law varies across languages<sup>[41]</sup>. This study found that the pattern of primary stress concentration on beat 1 (68.2%) and beat 5 (52.1%) in Zhuang folk songs shows clear trochaic tendencies, consistent with the leftward prominence tendency of Sino-Tibetan languages. However, the distribution of secondary stress (beat 3: 45.3%, beat 7: 38.4%) also shows some iambic characteristics, suggesting complex prosodic organizational principles.

The interaction between prosodic boundaries and information structure presents new characteristics in singing contexts. Gao et al. research on spontaneous Cantonese speech found that prosodic boundaries and information structure jointly influence tonal coarticulation patterns<sup>[42]</sup>. This study extends this finding, revealing the musicalized enhancement effect of prosodic boundaries in singing—pre-boundary lengthening ratio reaches 1.73, significantly higher than the 1.42 in natural language. This enhancement not only serves linguistic information transmission but more importantly creates a coordinating interface between linguistic prosody and musical rhythm. The maximum lengthening effect (2.14 times) when musical phrase endings coincide with prosodic boundaries confirms the convergence of the two systems in structural marking.

The application of corpus methods provides new perspectives for prosodic research. Torres and Babinski’s corpus phonetic study of Dalabon nouns demonstrates the advan-

tages of large-scale data analysis in revealing subtle prosodic patterns<sup>[43]</sup>. This analysis based on 1250 folk songs not only confirmed the main patterns predicted by theory but also discovered some unexpected phenomena, such as the age effect ( $\beta = -0.12, p = 0.004$ )-the decline in tone retention rates among younger singers may herald intergenerational evolution in traditional prosodic patterns. This finding reminds researchers to pay attention to the preservation of ethnic minority musical traditions in the context of globalization.

The findings of this study make important contributions to prosodic theory within the Optimality Theory framework. The observed constraint ranking variation (narrative vs. lyrical songs) supports the view that constraints can be reranked, indicating that multiple constraint hierarchies can exist within the same language. The dynamic balance mechanism between tonal integrity constraints and melodic alignment constraints provides new perspectives for understanding cognitive mechanisms at the language-music interface. The dialect and age effects confirmed by mixed-effects models further confirm the dynamicity and plasticity of prosodic systems.

Cognitive load allocation strategies play a key role in prosody-music coordination. The higher the prosodic hierarchy, the tighter the coordination with musical structure. This pattern may reflect optimized allocation of cognitive resources-prioritizing limited attentional resources to structurally more important prosodic units. The increase in tone neutralization at boundary positions can similarly be understood as cognitive load reallocation, freeing resources for processing other prosodic features at boundaries by reducing cognitive demands of tone processing. This principle of cognitive economy may be a universal feature of cross-linguistic prosody-music mapping.

Cultural transmission perspectives provide sociolinguistic explanations for understanding observed variation patterns. The higher tone retention rate and lower rhythmic flexibility in northern regions may reflect that area's adherence to traditional culture, while the innovative tendencies in southern regions reflect vitality brought by cultural exchange. The emergence of intergenerational differences suggests challenges in transmitting traditional prosodic knowledge, with younger generations possibly developing new prosody-music mapping strategies under the influence of globalized musical culture. While this evolution may lead to the loss of certain

traditional features, it may also give birth to new forms of artistic expression.

## 7. Conclusions

Through systematic analysis of 1250 Zhuang folk songs, this study reveals the complex mapping mechanisms between prosodic structure and rhythmic patterns in tonal language singing. The research finds that Zhuang tones exhibit a compression ratio of 0.74 ( $\pm 0.12$ ) in singing, indicating pitch variation amplitude is approximately three-quarters that of natural language. This systematic compression achieves coordination with melody while maintaining tonal distinctiveness. Prosodic hierarchy sensitivity to musical rhythm shows gradient distribution, with prosodic word stress-strong beat alignment rate reaching 78.3%, significantly higher than the syllable level's 54.6%, confirming hierarchical selectivity of prosodic prominence. The differential distribution of tone neutralization across prosodic positions (boundary 46.2% vs word-internal 26.8%) reflects the modulating effect of structural position on tone realization. Regional variation analysis reveals a significant 8.4 percentage point difference between northern dialect regions (tone retention rate 73.8%) and southern dialect regions (65.4%) ( $t = 5.82, p < 0.001$ ), with this difference remaining stable after controlling for song type and performer factors ( $\beta = 0.38, p < 0.001$ ). Pre-boundary lengthening effects are more prominent in singing (1.73 times) than in natural language (1.42 times), reaching maximum values (2.14 times) when musical phrases coincide with prosodic boundaries, reflecting synergistic enhancement of linguistic prosody and musical structure. Age effect analysis finds declining tone retention rates among younger singers ( $\beta = -0.12, p = 0.004$ ), suggesting traditional prosodic patterns are undergoing intergenerational evolution. The theoretical contribution of this study lies in establishing a prosody-music mapping model applicable to Zhuang, confirming predictions of dynamic constraint ranking under the Optimality Theory framework, and providing a new analytical paradigm for research on singing prosody in tonal languages. The research results not only enrich prosodic phonology theory but also provide scientific basis for preserving and transmitting ethnic minority musical culture, holding significant importance for understanding cognitive interface mechanisms between human language and music.

## Author Contributions

Conceptualization, H.C. and C.P.; methodology, H.C. and S.N.; software, H.C.; validation, C.P. and S.N.; formal analysis, H.C.; investigation, H.C.; resources, C.P. and S.N.; data curation, H.C.; writing—original draft preparation, H.C.; writing—review and editing, C.P. and S.N.; visualization, H.C.; supervision, C.P.; project administration, H.C.; funding acquisition, C.P. All authors have read and agreed to the published version of the manuscript.

## Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

## Institutional Review Board Statement

Not applicable.

## Informed Consent Statement

All participants provided informed consent for recording and analysis.

## Data Availability Statement

Research data is available upon reasonable request.

## Acknowledgments

The authors thank the Zhuang folk singers and cultural inheritors who contributed to this research, the local cultural institutions in Guangxi for their support during data collection, and the anonymous reviewers for their constructive feedback that improved the quality of this manuscript.

## Conflicts of Interest

The authors declare no conflict of interest.

## References

- [1] Yan, M., Wu, X., 2024. Prosody in linguistic journals: A bibliometric analysis. *Humanities and Social Sciences Communications*. 11, 311. DOI: <https://doi.org/10.1080/09298215.2021.1974490>

- [2] Zhang, X., Cross, I., 2021. Analysing the relationship between tone and melody in Chaozhou songs. *Journal of New Music Research*. 50(4), 299–311. DOI: <https://doi.org/10.1080/09298215.2021.1974490>
- [3] Savage, P.E., Passmore, S., Chiba, G., et al., 2022. Sequence alignment of folk song melodies reveals cross-cultural regularities of musical evolution. *Current Biology*. 32(6), 1395–1402.e8. DOI: <https://doi.org/10.1016/j.cub.2022.01.039>
- [4] Goldrick, M., Cole, J., 2023. Advancement of phonetics in the 21st century: Exemplar models of speech production. *Journal of Phonetics*. 99, 101254. DOI: <https://doi.org/10.1016/j.wocn.2023.101254>
- [5] Steffman, J., Cole, J., 2024. Metrical enhancement in American English nuclear tunes. *Glossa: A Journal of General Linguistics*. 9(1). DOI: <https://doi.org/10.16995/glossa.15297>
- [6] Kim, S.-E., Tilsen, S., 2025. The Gesture-Field-Register (GFR) framework for modeling F0 control. *Journal of Phonetics*. 110, 101410. DOI: <https://doi.org/10.1016/j.wocn.2025.101410>
- [7] Iskarous, K., Cole, J., Steffman, J., 2024. A minimal dynamical model of intonation: Tone contrast, alignment, and scaling of American English pitch accents as emergent properties. *Journal of Phonetics*. 104, 101309. DOI: <https://doi.org/10.1016/j.wocn.2024.101309>
- [8] Shang, P., Wu, Y., 2024. The impact of multifaceted factors on auditory mapping between acoustic cues and Spanish intonation categories in a cross-linguistic context. *Humanities and Social Sciences Communications*. 11(1), 1701. DOI: <https://doi.org/10.1057/s41599-024-04216-6>
- [9] Ryan, K.M., 2019. *Prosodic Weight: Categories and Continua*. Oxford University Press: Oxford, UK.
- [10] Roettger, T.B., Mahrt, T., Cole, J., 2019. Mapping prosody onto meaning – The case of information structure in American English. *Language, Cognition and Neuroscience*. 34(7), 841–860. DOI: <https://doi.org/10.1080/23273798.2019.1587482>
- [11] Shih, S.-H., De Lacy, P., 2019. Evidence for sonority-driven stress. *Catalan Journal of Linguistics*. 18, 9–40. DOI: <https://doi.org/10.5565/rev/catjl.256>
- [12] Mücke, D., Hermes, A., Tilsen, S., 2020. Incongruencies between phonological theory and phonetic measurement. *Phonology*. 37(1), 133–170. DOI: <https://doi.org/10.1017/S0952675720000068>
- [13] Hayes, B., Kaun, A., 1996. The role of phonological phrasing in sung and chanted verse. *The Linguistic Review*. 13, 243–303.
- [14] Xiaoyu, W., Ying, X., Shilin, L., et al., 2023. Phonetics as a means of nationalising art songs: A comparative music-phonetics study based on Zhao Yuanren’s New Poetry Collection. *Journal of New Music Research*. 52(2–3), 202–226. DOI: <https://doi.org/10.1080/09298215.2021.1974490>

- 8215.2024.2311649
- [15] Chen, Y., Xu, Y., 2025. Sequential perception of tone and focus in parallel – A computational simulation. *Speech Communication*. 168, 103173. DOI: <https://doi.org/10.1016/j.specom.2025.103173>
- [16] Barzilai, M.L., 2022. Phonetic and phonological salience in tone processing. *Canadian Journal of Linguistics / Revue canadienne de linguistique*. 67(1–2), 53–70. DOI: <https://doi.org/10.1017/cnj.2022.2>
- [17] Burrone, F., Tilsen, S., 2025. Thai speakers time lexical tones to supralaryngeal articulatory events. *Journal of Phonetics*. 108, 101389. DOI: <https://doi.org/10.1016/j.wocn.2024.101389>
- [18] Liu, Z., Xu, Y., Hsieh, F.-F., 2022. Coarticulation as synchronised CV co-onset – Parallel evidence from articulation and acoustics. *Journal of Phonetics*. 90, 101116. DOI: <https://doi.org/10.1016/j.wocn.2021.101116>
- [19] Heffner, C.C., Slevc, L.R., 2015. Prosodic structure as a parallel to musical structure. *Frontiers in Psychology*. 6, 1962. DOI: <https://doi.org/10.3389/fpsyg.2015.01962>
- [20] Kang, W., Xu, Y., 2024. Tone-syllable synchrony in Mandarin: New evidence and implications. *Speech Communication*. 163, 103121. DOI: <https://doi.org/10.1016/j.specom.2024.103121>
- [21] Xu, A., Van Niekerk, D.R., Gerazov, B., et al., 2024. Artificial vocal learning guided by speech recognition: What it may tell us about how children learn to speak. *Journal of Phonetics*. 105, 101338. DOI: <https://doi.org/10.1016/j.wocn.2024.101338>
- [22] Staszkiwicz, B., 2025. Politeness and prosody: The effect of power, distance, and imposition on pitch contours in Spanish. *Language and Speech*. 68(3). DOI: <https://doi.org/10.1177/00238309241307891>
- [23] Sonderegger, M., Sóskuthy, M., 2025. Advancements of phonetics in the 21st century: Quantitative data analysis. *Journal of Phonetics*. 111, 101415. DOI: <https://doi.org/10.1016/j.wocn.2025.101415>
- [24] Bellik, J., Ito, J., Kalivoda, N., et al., 2023. *Syntax–Prosody in Optimality Theory: Theory and Analyses*. University of Toronto Press: Toronto, Canada.
- [25] Bellik, J., Bellik, O., Kalivoda, N., 2020. *Syntax–Prosody in Optimality Theory (SPOT)*. Available from: <https://spot.sites.ucsc.edu> (cited 3 September 2025).
- [26] Hayes, B., 2009. Textsetting as constraint conflict. In: Fabb, N., Attridge, D., Durant, A., et al. (Eds.). *Towards a Typology of Poetic Forms*. John Benjamins Publishing Company: Amsterdam, Netherlands. pp. 43–62.
- [27] Sostarics, T., Ronai, E., Cole, J., 2025. Relating scalar inference and alternative activation: A view from the rise-fall-rise tune in American English. *Experiments in Linguistic Meaning*. 3, 383–394. DOI: <https://doi.org/10.3765/elm.3.5768>
- [28] Cole, J., Steffman, J., Awwad, A., 2024. Functional modeling of F0 variation across speakers and between phonological categories: Rising pitch accents in American English. In *Proceedings of Speech Prosody 2024*, Leiden, The Netherlands, 2–5 July 2024.
- [29] Steffman, J., Cole, J., Shattuck-Hufnagel, S., 2024. Intonational categories and continua in American English rising nuclear tunes. *Journal of Phonetics*. 104, 101310. DOI: <https://doi.org/10.1016/j.wocn.2024.101310>
- [30] Durand, J., Gut, U., Kristoffersen, G., 2014. *The Oxford Handbook of Corpus Phonology*. Academic Press: London, UK.
- [31] Lin, Y., Halim, H.A., Jalis, F.M.M., 2024. Building a parallel corpus for Chinese folk songs translation studies: A case study of Northern Shaanxi and Hua'er folk songs. *Theory and Practice in Language Studies*. 14(2), 454–468. DOI: <https://doi.org/10.17507/tpsl.1402.17>
- [32] Savage, P.E., 2022. An overview of cross-cultural music corpus studies. In *The Oxford Handbook of Music and Corpus Studies*. Available from: <https://doi.org/10.1093/oxfordhb/9780190945442.013.34>
- [33] Eychenne, J., 2021. Corpus phonology. In: Bradley, T.G., Baković, E. (Eds.). *Manual of Romance Phonetics and Phonology*. De Gruyter: Berlin, Germany. pp. 319–340.
- [34] Chodroff, E.R., Cole, J., 2019. The phonological and phonetic encoding of information status in American English nuclear accents. In *Proceedings of the 19th International Congress of Phonetic Sciences*, York, UK, 2019.
- [35] Hirst, D., 2024. *Speech Prosody: From Acoustics to Interpretation*. Cambridge University Press: Cambridge, UK.
- [36] Tilsen, S., 2023. Probability and randomness in phonology: Deep vs. shallow stochasticity. *Journal of Phonetics and Phonology*. 29(2), 247–269. (in Korean)
- [37] Im, S., Cole, J., Baumann, S., 2023. Standing out in context: Prominence in the production and perception of public speech. *Laboratory Phonology*. 14(1). DOI: <https://doi.org/10.16995/labphon.6417>
- [38] Liu, Z., Chitoran, I., Turco, G., 2024. Perceptual salience of tones, vowels, and consonants in Mandarin speech errors. *Language and Speech*. 68(3). DOI: <https://doi.org/10.1177/00238309241302334>
- [39] Creel, S.C., Obiri-Yeboah, M., Rose, S., 2023. Language-to-music transfer effects depend on the tone language: Akan vs. East Asian tone languages. *Memory & Cognition*. 51(7), 1624–1639. DOI: <https://doi.org/10.3758/s13421-023-01416-4>
- [40] Rao, R., 2024. *The Phonetics and Phonology of Heritage Languages*. Cambridge University Press: Cambridge, UK.
- [41] Kaland, C., Bhatara, A., Boll-Avetisyan, N., et al., 2024. Prosodic grouping in Akan and the applicability of the iambic-trochaic law. In *Proceedings of Speech Prosody 2024*, Leiden, The Netherlands, 2–5 July 2024.

- [42] Gao, X., Voeten, C., Liberman, M., 2024. The impact of prosodic boundary and information structure on tonal coarticulation in spontaneous Cantonese. In *Proceedings of Speech Prosody 2024*, Leiden, The Netherlands, 2–5 July 2024.
- [43] Torres, C., Babinski, S., 2024. A corpus phonetics study of Dalabon nouns. In *Proceedings of Speech Prosody 2024*, Leiden, The Netherlands, 2–5 July 2024.