Versatility of phonemic pitch in affective iconicity and perceptual reorganisation

Zheng, T.

Chapter 4
Adaptive significance of tonemes for affective iconicity in Standard Chinese

A shorter version of this chapter is published as: Zheng, T., Levelt, C. C., & Chen, Y. (2024). *The adaptive value of Mandarin tones for affective iconicity*. Proc. Speech Prosody 2024, 562–566. https://doi.org/10.21437/SpeechProsody.2024-1

Abstract: Iconicity is prevalent across languages, yet its underlying mechanisms remain debated. While previous studies have explained iconicity through perceptual analogy (see a review in Sidhu & Pexman, 2018), recent research proposes an adaptation perspective. Specifically, Adelman et al. (2018) found that the first phoneme of a word predicted its valence better than subsequent phonemes in Indo-European languages. Additionally, word-initial phonemes that are produced more rapidly tend to convey negativity rather than positivity, especially in English and Spanish. This led Adelman et al. to suggest that iconicity may have evolved due to the adaptive advantage of negative emotions in survival and social interactions. However, this view is largely based on the affective iconicity of segmental phonemes in Indo-European languages and focuses only on valence. What has left open is whether emotional arousal, a more ancient and innate dimension of emotion (Darwin, 1998), may also play a role. Additionally, many languages of the world use not only segments but also suprasegmental lexical tones to distinguish word meanings. The primary perceptual cue for tone is pitch variation, which expresses emotions as well. Previous studies have demonstrated affective iconicity in Standard Chinese lexical tones (Zheng et al., 2025, under revision), but little is known about their adaptive significance. Using three Standard Chinese corpora with valence and arousal ratings for bisyllabic words, this study employed hierarchical linear regression models to examine whether lexical tones in the first and second syllables predict a word's emotional ratings differently. Results revealed that valence is predicted by first-syllable tone, with positive words more likely to carry a rising tone than a falling or high-level tone. This positional effect of tone for valence in a Sinitic tonal language conceptually replicates the adaptive significance of the first phoneme

for valence found in segments of an Indo-European language. Arousal, in contrast, is predicted by the tones of both syllables, with high-arousing words more likely to contain a falling tone than a rising or low-dipping tone. This distributed effect of tone on arousal suggests that the spread of crucial information, particularly through iconicity, may have evolved due to the amplified reactions and evaluations associated with high-arousing stimuli. This extends the adaptive significance of valence to encompass arousal, highlighting its role in shaping emotional and cognitive processes. These findings lend further support to the proposal that (affective) iconicity is an adaptation providing communication efficiency in speech.

Keywords: Affective iconicity, Standard Chinese, Emotional arousal, Emotional valence, Adaptation, First syllable tone, Second syllable tone

4.1 Introduction

Iconicity refers to the subjective and non-arbitrary association between a signal's form and meaning as perceived by language users, a phenomenon found across languages and linguistic components (Adelman et al., 2018; Aryani et al., 2018; Perniss et al., 2010; Winter et al., 2023; Yap et al., 2014; Zheng et al., 2025, under revision). Among its various dimensions, affective iconicity, where linguistic forms convey emotional meaning, has been widely studied across languages (e.g., German, English, Spanish, Polish, and Dutch in Adelman et al., 2018; German in Aryani et al., 2018; English in Whissell, 2003; Standard Chinese Yao et al., 2013; Standard Chinese and English in Yu et al., 2021).

Emotional expression is essential for survival and social interaction, guiding threat detection (e.g., recognising danger), adaptive responses (e.g., fight-or-flight responses), and social communication (Kashima et al., 2020). Psycholinguistic studies show that emotionally-valenced words (positive or negative) are processed more efficiently than neutral words, with faster reaction times and enhanced neural responses in lexical decision tasks (e.g., Kissler et al., 2007; Kousta et al., 2009). This processing advantage may reflect an adaptive mechanism where phonetic features help encode emotional meaning.

Supporting this, Adelman et al. (2018) demonstrated that across Indo-European languages (e.g., English, Spanish, Dutch, German, and Polish), phonemes in the general lexicon predict the emotional valence of words. Notably, in English and German, phonemes with shorter articulation times were found to be more likely to initiate negative words, while those with longer articulation times correlated better with

positive words. This form–affect association, termed emotional sound symbolism by Adelman et al. (2018), suggests that iconicity may have evolved to enhance communication efficiency, akin to alarm calls in non-human species. They argued that the phoneme-based structure reflects the adaptive significance of emotional sound symbolism, which enhances communication efficiency by prioritising negative signals. This efficiency further reinforces the emergence and development of iconicity in speech.

In this chapter, we use the term "affective iconicity" to refer to emotional sound symbolism (Adelman et al., 2018) with two important disclaimers. First, unlike traditional sound symbolism, which links specific phonetic features to particular meanings (e.g., the bouba/kiki sound-shape symbolism Ćwiek et al., 2021), we use "iconicity" to encompass broader forms, such as speech, sign, and word (Winter et al., 2023). Second, following the embodied view of language, we use "affect" to encompass behavioural and neural emotional responses across species and cultures (Aryani et al., 2019). Although this study does not directly examine neural responses, insights from neuroscience inform our understanding of affective iconicity's adaptive functions. See more details in the following paragraphs.

Existing theories or hypotheses on the emergence of iconicity or sound symbolism often attribute it to general cognitive faculties, such as analogy and generalisation (e.g., Monaghan et al., 2014; Sidhu & Pexman, 2018). The statistical co-occurrence hypothesis, for instance, suggests that sound-meaning correspondences emerge because certain features frequently co-occur (see an overview in Sidhu & Pexman, 2018). This is supported by findings that the articulatory muscles involved in phoneme production overlap with those used in emotional expression (e.g., Körner & Rummer, 2023). While cognitive analogies

explain why affective iconicity is possible, Adelman et al. (2018) argue that it fails to explain the positional effect of phonemes, especially how the temporal characteristics of phonemes influence the words' valence.

However, the adaptation account requires further refinement to clarify its scope and generality. First, while it provides a compelling explanation for valence-based sound symbolism, it treats emotional arousal (the degree of activation, another core dimension of emotion) as a control variable, rather than examining its independent role in shaping sound-affect associations (Adelman et al., 2018). Specifically, Adelman et al. examined how phonemes contribute to valence ratings after accounting for arousal. Yet, valence and arousal are both fundamental to emotional processing and may play distinct yet complementary roles in adaptation. Arousal, considered more ancient and evolutionarily fundamental than valence (Darwin, 1998), which reflects immediate and automatic physiological activation (Russell, 1980), whereas valence involves cognitive appraisal and contextual interpretation (Citron, 2012).

Despite these differences, valence and arousal are interrelated. They often form a V-shaped relationship, where both positive and negative emotions are typically high-arousing, while neutral emotions tend to be low-arousing (Kuppens et al., 2013). Their interaction plays a crucial role in evaluative processing (Robinson et al., 2004), suggesting that ignoring arousal's distinct role risks overlooking key aspects of affective iconicity. To fully understand the adaptive functions of affective iconicity, it is essential to consider both arousal and valence as independent yet interacting factors.

Second, prior research on affective iconicity primarily focused on segmental phonemes in Indo-European languages, while ignoring that suprasegmental features, such as lexical tone, play a similar role in

distinguishing word meanings, and they may also carry adaptive significance in emotional communication. Lexical tones in many languages rely on pitch variations, which are also a primary cue for emotional intonation, to distinguish lexical meaning, similar to phonemes. Research shows that pitch at the prosodic level correlates with emotional arousal and valence expression. Specifically, high-arousing emotions (e.g., fear, excitement) generally correlate with a higher pitch level and average pitch height, a wider pitch range, and a steeper pitch slope, in comparison to low-arousing emotions (Bänziger & Scherer, 2005; Laukka et al., 2005; Scherer et al., 2003). Positive emotions generally correlate with a higher pitch level and average pitch height and a wider pitch range, while negative emotions show the opposite (Belyk & Brown, 2014; Kamiloğlu et al., 2020; Laukka et al., 2005; c.f., Scherer & Oshinsky, 1977).

Studies on Standard Chinese (SC) suggest that lexical tones evoke similar emotional valence and arousal effects due to their phonetic similarities to emotion-induced pitch modulations at the sentence level, despite the conventional view that lexical tone is distinct from affective meaning (Zheng et al., 2025). For example, falling-falling tone sequence (i.e., T4T4) in both general lexicon and nonce words is associated with higher arousal than rising-rising tones (i.e., T2T2); and in nonce words, falling-falling tone sequence is associated with negative valence, while rising-rising tone sequence with positive valence. These findings suggest that pitch variations, despite their phonemic function at the lexical level in tone languages, remain iconic to convey affective meaning.

Zheng et al. (2025) argue that the tone–affect associations align with the frequency code hypothesis (Gussenhoven, 2016; Ohala, 1984), which posits that high or rising *fo* signals friendliness, excitement, or

non-threat, which can promote social bonding. On the other hand, low or falling *f0* signals authority, dominance, or threat, aiding conflict resolution or warning. These associations are also evolutionarily meaningful and may suggest an adaptive significance of tones similar to phonemes in Indo-European languages, lending support to the idea that iconicity in speech emerges and persists due to its evolutionary advantages. This study explored the positional effects of lexical tones in affective iconicity and further extended the current findings on valence to arousal.

Using three Standard Chinese corpora (CAWS, NORM, DCAWS; Y. Wang et al., 2008; X. Xu et al., 2022; Zheng et al., 2025), we analysed how the first- and second-syllable tones of bisyllabic words predict their emotional arousal and valence ratings. We included all four lexical tones (also known as tonemes): High-level tone (T1), Rising tone (T2), Low-dipping tone (T3), and Falling tone (T4) for each syllable. To minimise confounding effects, we incorporated factors such as word frequency (Brysbaert et al., 2018), onset consonant category (Louwerse & Qu, 2017), and part-of-speech (Perry et al., 2015) into our model analyses (as done in Zheng et al., 2025). Additionally, we examined the relationship between arousal and valence across each lexical tone and syllable position to further clarify their potential interactions in affective iconicity.

In this exploratory study, we hypothesised that lexical tone contributes to signalling both valence and arousal. Furthermore, we predicted that pitch features, including pitch level, average pitch height, pitch range and slope, and pitch contour, contribute to predicting arousal and valence ratings, in line with the effects of pitch in non-tonal languages to express emotion (Bänziger & Scherer, 2005; Frick, 1985;

Laukka et al., 2005) and prior research on tone–arousal and tone–valence iconicity (Zheng et al., 2025, under revision).

Importantly, we propose that the adaptive significance of tone–affect iconicity manifests differently for valence and arousal, given their distinct cognitive and neural processing characteristics. Specifically, we predicted a positional effect, such that the tone of the first syllable would have a stronger influence on valence, given that valence is extracted earlier than arousal in word and picture processing (Gianotti et al., 2008) and prioritised in phonemes at the beginning of words (Adelman et al., 2018). Moreover, lexical tones with higher overall pitch level, higher average pitch height, wider pitch ranges, and/or steeper slopes are more likely to predict negative valence. In contrast, lexical tones with lower overall pitch level, lower average pitch height, narrower pitch ranges, and/or more gradual slopes are more likely to predict positive valence. For example, typically exhibits sharper, more rapid pitch changes over a short temporal period, which may contribute to its association with negative affect. These patterns would reflect the adaptive significance of negative valence in survival and communication.

In contrast, we predict a distributed effect for arousal, as arousal-related information unfolds over a longer time course. ERP studies show that higher-arousal emotions elicit distinct neural responses, notably the Early Posterior Negativity (EPN), which reflects enhanced attention to emotional stimuli, and the Late Positive Component (LPC), which indicates more sustained and detailed emotional processing (Fischler & Bradley, 2006; Herbert et al., 2008; Kissler et al., 2007, 2009; Schacht & Sommer, 2009). This suggests that arousal-related effects of lexical tone may not be limited to the initial syllable but rather extend throughout the word (i.e., distributed effect). Arousal has been

shown to amplify reactions, leading to intensified evaluations and enhanced long-term memory for events. This suggests that the arousal dimension conveys information about the urgency or importance of stimuli (Storbeck & Clore, 2008). Given these effects, it is plausible that arousal, particularly high-arousing stimuli, contributes to the adaptive significance and emergence of iconicity, as its benefits in facilitating communication are well-documented. Furthermore, lexical tones with a higher average pitch height and more dynamic pitch variations, such as a wider range and steeper slopes, tend to be linked to high arousal, whereas tones with a lower average pitch height and less dynamic pitch variations tend to be associated with low arousal. For example, T4 is commonly linked to high arousal.

4.2 Method

4.2.1 Corpus datasets

We analysed three corpus datasets, wherein all bisyllabic words were annotated with their underlying lexical tones.

The Chinese Affective Word System (CAWS) dataset (Y. Wang et al., 2008) comprises arousal and valence ratings for 1,492 bisyllabic words. A total of 64 undergraduate students (aged 18–21 years) rated arousal and valence on a 9-point Likert scale. A retest was subsequently conducted with 30 undergraduates (aged 18–21 years). Arousal values ranged from low (1, e.g., peaceful) to high (9, e.g., alert), while valence ratings spanned from negative (1, e.g., unhappy) to positive (9, e.g., happy).

The Affective Norms for Chinese Words dataset (X. Xu et al., 2022), hereafter NORM, contains arousal and valence ratings of 9,573 bisyllabic words in Standard Chinese. Arousal ratings were collected

from 1,189 participants, and valence ratings from 1,232 participants (aged 18–62 years). Most participants (96.9%) received a college-level education or higher. The NORM dataset employed a 5-point Likert scale for arousal (0–4) and a 7-point scale for valence (-3 to +3).

The Developed Chinese Affective Word System (DCAWS) dataset (Zheng et al., 2025) includes arousal and valence ratings for 1,500 bisyllabic words. Arousal ratings were obtained from 51 participants (aged 18–27 years), while valence ratings were provided by 53 participants (aged 18–24 years). Ratings for both dimensions were recorded using a 1–100 slider scale with 1-point increments, ranging from low to high or negative to positive.

4.3 Data analysis

We first visually inspected the graphs displaying emotional arousal and valence ratings for all words to examine the relationship between the ratings of arousal and valence in each dataset.

We then explored how lexical tone in different positions (first vs. second syllable) predicts the emotional arousal and valence of bisyllabic words, separately in the three datasets. To observe how lexical tone contributes to account for the arousal and valence variances, we conducted a set of hierarchical linear regression analyses (HLM) in each dataset (R Core Team, 2023). The HLM helps to determine whether lexical tone explains a statistically significant amount of variance in the arousal and valence after accounting for all other variables, such as word frequency, onset consonant, and part-of-speech. The HLM consisted of three models: the first model includes factors such as word frequency, onset consonant, and part-of-speech, the second model added an individual tone (either on the 1st or 2nd syllable) beyond the factors in the first model, the third model includes

tones of both syllables along with the other factors in the first model. The dependent variables were the aggregated emotional arousal and valence rating scores. The HLM results highlighted the influence of individual tones on emotional ratings. Pairwise multiple comparisons, adjusted with Bonferroni correction using the *emmeans* package (Lenth, 2023), were employed to examine differences across individual tones.

Additionally, generalised linear mixed models (GLMMs) were conducted only for the DCAWS dataset, given that this is the only corpus with individual rating data points available. GLMMs for both arousal and valence consist of individual lexical tones as the fixed effect, and by-subject and by-item intercepts as the random effects to account for potential heterogeneity among raters and items.

4.4 Results

4.4.1 V-shaped relationship between arousal and valence

Figures 4.1–4.3 illustrate the relationship between emotional arousal and valence in CAWS, NORM, and DCAWS datasets, respectively. The numerical values on the axes correspond to Likert/Slider scale ratings. A distinctive V-shaped relationship emerged between arousal and valence in all datasets, confirming that positive and negative valence tend to associate with high arousal, while neutral valence aligns with low arousal. In each panel of the figure, “H” is used for high-level tone (T1), “R” for rising tone (T2), “L” for low-dipping tone (T3), and “F” for falling tone (T4).

Figure 4.1

The V-shaped relationship between emotional arousal and valence in the CAWS dataset.

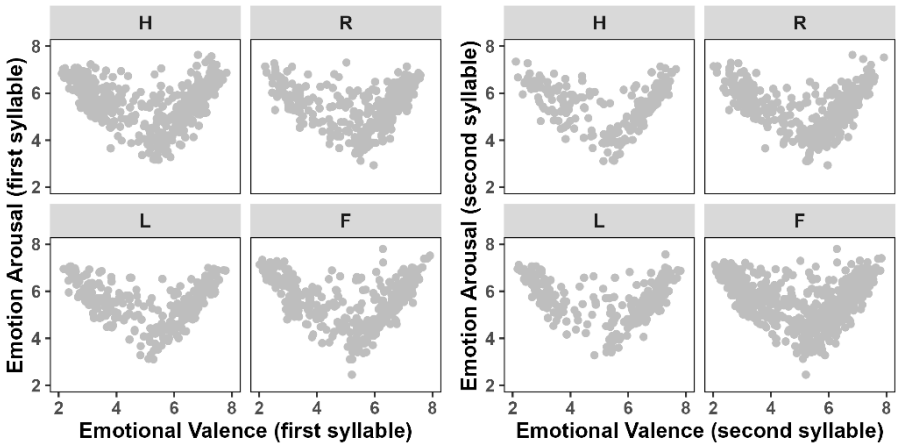


Figure 4.2

The V-shaped relationship between emotional arousal and valence in the NORM dataset.

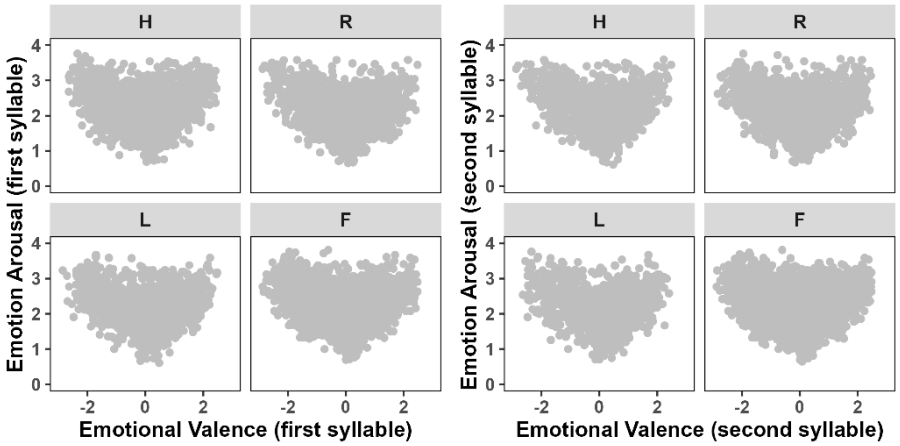
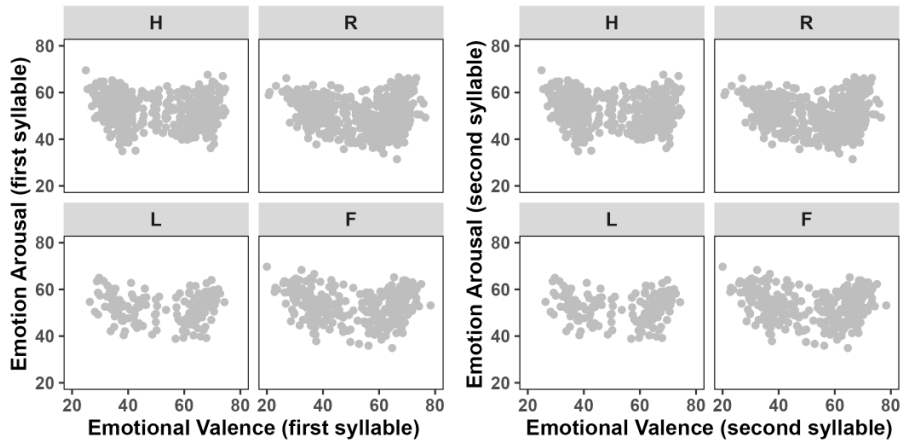


Figure 4.3

The V-shaped relationship between emotional arousal and valence in the DCAWS dataset.



4.4.2 Effects of individual lexical tone on arousal and valence ratings

Regression diagnostics of HLMs confirmed the fulfilment of linearity, homoscedasticity, and normality assumptions, with no identified collinearity issues. No violations of independence existed for the average scores of arousal and valence ratings across participants. GLMMs diagnostics also showed no violation of linear regression assumptions.

4.4.2.1 Arousal ratings

Regarding the arousal ratings, individual tones in both the first and second syllables significantly impacted the unique variance across the three datasets, showing a distributed effect. In the CAWS datasets, the model with first-syllable tones demonstrated a significant predictive effect on emotional arousal ratings (F = 5.10, p < 0.01). Notably, the first-syllable tone accounted for a significant portion of unique

variance in the arousal ratings (ΔR^2 = 0.88%, p < .001), with the falling tone (M = 5.75) predicting higher arousal levels than the rising tone (M = 5.50, p < .001). Likewise, the model with second-syllable tones demonstrated a significant predictive effect on emotional arousal ratings (F = 5.56, p < 0.001). The results indicated a significant influence of the second-syllable tone on the unique variance in arousal ratings (ΔR^2 = 0.96%, p < .001). Further comparisons revealed the larger arousing effect of the falling tone (M = 5.72) in comparison to the rising tone (M = 5.50, p < .01). The third model showed consistent effects of monosyllabic tones on emotional arousal ratings, with first and second syllable tones combined predicting emotional arousal ratings better than single-tone models.

The same HLM analyses were conducted for the NORM corpus and revealed significant predictive effects of both first and second-syllable tones on emotional arousal ratings (F = 5.06, p < 0.01; F = 5.56, p < 0.001). The first-syllable tone and second-syllable tone accounted for a considerable portion of unique variance in the arousal ratings (ΔR^2 = 0.15%, p < .001; ΔR^2 = 0.22%, p < .001). Pairwise comparisons revealed higher arousal levels for the falling tone (M = 2.07) compared to the rising tone (M = 2.02, p < .01) and the low tone (M = 2.02, p < .05) at the first syllable. Similarly, at the second syllable, the falling tone (M = 2.07) showed a larger arousing effect compared to the rising tone (M = 2.01, p < .001) and the low tone (M = 2.03, p < .05). The third model demonstrated consistent effects of monosyllabic tones on emotional arousal ratings, and the model with tones at both first and second syllable together were preferred in model comparisons.

HLM analyses for the DCAWS corpus revealed a significant predictive effect of both first- and second-syllable tones on emotional arousal ratings (F = 4.42, p < 0.01; F = 3.33, p < 0.05). Notably, both

first-syllable and second-syllable tones accounted for a significant portion of the unique variance in the arousal ratings (ΔR^2 = 0.84%, p < .001; ΔR^2 = 0.64%, p < .001). Pairwise comparisons revealed higher arousal levels for the falling tone (M = 52.7) compared to the rising tone (M = 51.0, p < .01) at the first syllable. Although pairwise comparisons showed no significant differences among second-syllable tones, the falling tone had the highest arousal rating score (M = 52.1), while the rising tone had the lowest (M = 50.7) at the second syllable.

GLMM analysis for the DCAWS corpus revealed a similar predictive effect of both first- and second-syllable tones on emotional arousal ratings. Comparisons between the null model and the interest model provided compelling evidence that the first-syllable tone significantly predicted emotional arousal ratings (F = 21.73, p < .001). Pairwise multiple comparisons revealed that both the falling tone (M = 53.0) and the high-level tone (M = 52.5) predicted higher arousal ratings than the rising tone (M = 51.1, p < .001 and p < 0.01, respectively). Similarly, the second syllable tone significantly predicted emotional arousal ratings (F = 13.39, p < .01). Pairwise multiple comparisons showed that the falling tone (M = 52.5) yielded higher arousal ratings than both the rising tone (M = 50.7, p < .05) and the low tone (M = 51.3, p < .05). Again, the model with two syllable tones showed a better fitness than single-tone models.

4.4.2.2 Valence ratings

Regarding the valence ratings, only the lexical tone of the first syllable, demonstrating a positional effect, showed a noteworthy influence within the NORM (F = 3.76, p < 0.05, ΔR^2 = 0.12%) and DCAWS (F = 3.60, p < 0.05, ΔR^2 = 0.68%) datasets. Specifically, the rising tone (M = 0.13) was perceived as more positive compared to both the high-level

tone (M = 0.04, p < .05) and the falling tone (M = 0.04, p < .05) in the NORM dataset. Similarly, in the DAWS dataset, the rising tone (M = 53.2) was perceived as more positive compared to both the high-level tone (M = 50.4, p < .05) and falling tone (M = 50.3, p = 0.058).

Figures 4.4–4.6 illustrate further individual tones' different contributions to the emotional arousal and valence ratings in the three corpora, respectively. In these plots, lexical tones are organised in ascending order based on their average rating scores for each syllable (first syllable on the left) and emotional dimension in each panel (arousal rating scores on the top). The mean rating scores for each tone are connected to show the overall trends. Take the left-upper panel of Figure 4.4 as an example. For the first syllable, the rising tone is associated with the lowest mean arousal rating score (score 5.50), while the falling tone with the highest mean arousal rating score (score 5.75).

Figure 4.4

Arousal and valence rating scores for each tone at the 1st and 2nd syllable in the CAWS dataset.

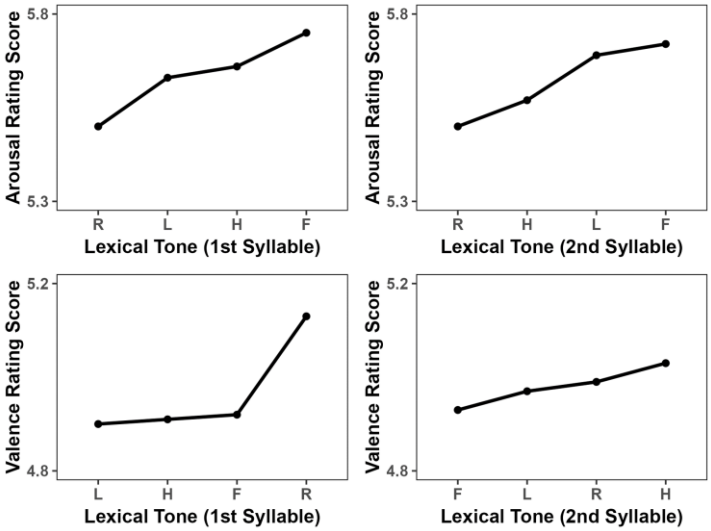


Figure 4.5

Arousal and valence rating scores for each tone at the 1st and 2nd syllable in the NORM dataset.

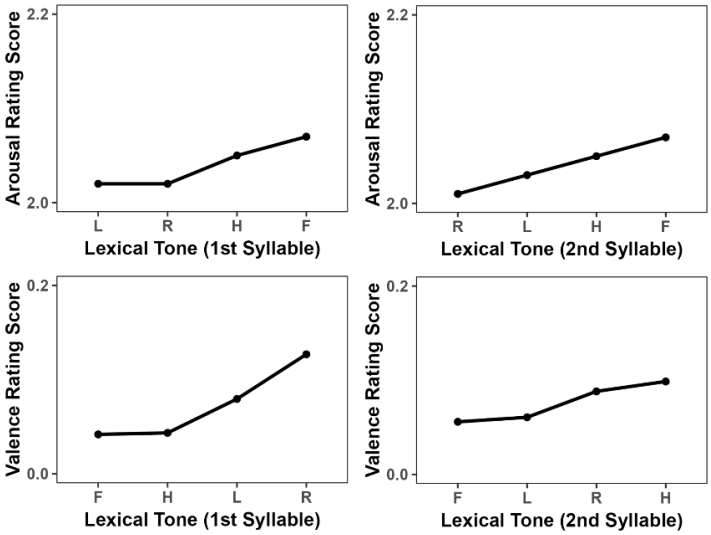
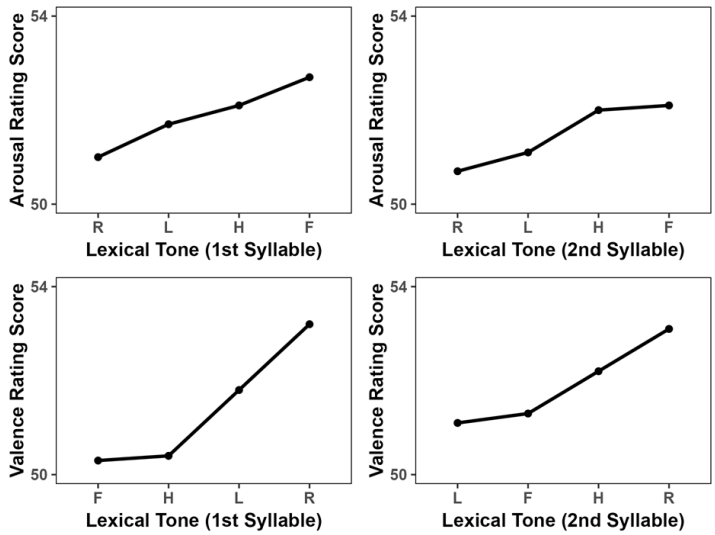


Figure 4.6

Arousal and valence rating scores for each tone at the 1st and 2nd syllable in the DCAWS dataset.



4.5 Discussion

This study explored whether lexical tones exhibit adaptive significance for both emotional arousal and valence in affective iconicity. Specifically, we investigated whether positional effects exist for valence and distributed effects for arousal when using lexical tones to predict emotional arousal and valence ratings of bisyllabic Standard Chinese words in three corpora (CAWS, NORM, DCAWS; Y. Wang et al., 2008; X. Xu et al., 2022; Zheng et al., 2025). Our results confirmed the commonly observed V-shaped relationship between arousal and valence (Citron et al., 2014; Kuppens et al., 2013), reinforcing their interdependent roles in emotional processing. This suggests that both arousal and valence should be taken into account in the investigation of affective iconicity to provide a comprehensive understanding of nuanced emotional expressions.

By incorporating key lexical factors (e.g., word frequency, part of speech, onset-consonant category), we found that individual tones in the first and second syllables of SC bisyllabic words significantly influenced arousal ratings, demonstrating a distributed effect of tone on arousal. Specifically, the falling tone (T4) correlated with high-arousing words, whereas the rising and low-dipping tones (T2 and T3) correlated with low-arousing words. In contrast, only first-syllable tones influenced valence ratings, indicating a positional effect; words beginning with a falling or high-level tone (T1 or T4) were more likely to be associated with negative valence. Although the magnitude of these tonal effects on arousal and valence was relatively small, it is comparable to the effects of phonemes on valence ratings found in prior research (Adelman et al., 2018), suggesting a subtle yet consistent role of sound in encoding emotional meaning across languages.

The positional effect of lexical tones on valence parallels findings from phoneme-based research, where the first phoneme of a word critically influences valence (Adelman et al., 2018). Adelman et al. (2018) proposed the concept of *negative priority*, suggesting that negative information is given preferential attention due to its survival relevance. This leads to faster pronunciation of the first phoneme in negative words, as they signal urgent or threatening information. However, in Standard Chinese, lexical tone *fo* contrasts are predominantly realised within the rhyme of a syllable, especially for syllables with a voiceless onset. So, the negative priority account attributed to temporal features of segmental affective iconicity may have limited applicability to lexical tones. Instead, the pitch characteristics of tones, particularly the falling contour (T4), may better explain the tendency for negative words to begin with falling tones. This aligns with the frequency code hypothesis (Ohala, 1994),

which posits that falling pitch contours signal threat or danger. By initiating a falling contour, speakers can efficiently convey urgency or negative valence, enhancing communication efficiency in contexts where rapid transmission of critical information is vital.

The contrastive effect of rising (T2) and falling (T4) contours also aligns with the conceptual metaphor of "Good is up, bad is down" (Ley-Flores et al., 2022; Yap et al., 2014). Worth noting is that the high-level tone (T1) is not as positively valenced as the rising tone (T2), which may be taken to emphasise the importance of rising contour in conveying positive emotional meaning (Laukka et al., 2005; cf. Scherer & Oshinsky, 1977). While these findings highlight the role of pitch direction in valence iconicity, further research is needed to disentangle the effects of average pitch height and pitch range, as well as the direction of pitch contour on valence ratings. This will provide a more nuanced understanding of how tonal (pitch) features contribute to affective iconicity.

For arousal, tones of both syllables influenced ratings, with the falling tone (T4) consistently associated with higher arousal. This association may be attributed to the pitch characteristics of this tone: the falling tone in SC is realised with a higher average pitch height, greater pitch range, and steeper slope compared to rising and low-dipping tones. These features likely contribute to the perception of urgency or intensity, aligning with the adaptive role of high-arousing stimuli in signalling importance or threat. The adaptive significance of high-arousing stimuli is well-documented, as such stimuli tend to capture attention and enhance memory, likely due to their role in signalling urgency or emotionally significant information (Storbeck & Clore, 2008). The association of T4 with high-arousing words may reflect an evolutionary mechanism where high-arousing tonal features

amplify the salience and memorability of urgent or emotionally significant messages.

Moreover, the link between arousal and the social transmission of information further supports the adaptive role of tonal effects. Berger (2011) demonstrated that participants experiencing high emotional arousal, whether positive or negative, were more likely to share neutral media content (e.g., an article or video) encountered afterwards. For our results, the correlation between falling tones (T4) and high-arousing words may facilitate the dissemination of emotionally charged or urgent information within a social group. This reinforces the communicative efficiency of tonal iconicity, as arousal plays a crucial role in attention and sustained processing. It is plausible that arousal information has been evolutionarily preserved in iconicity to optimise social communication, ensuring that emotionally significant messages are prioritised and retained.

Some may argue that arousal reflects the level of emotional activation, often correlating with stimulus intensity and increasing attentional engagement with more intense signals. Under this view, certain tonal patterns, such as high tones, may capture attention due to their acoustic salience, independent of emotional content (see insights from consonants in Baroni, 2014). However, when investigating the mechanisms underlying the relationship between tonal patterns and arousal, it is crucial to distinguish between arousal and attention, which are intertwined yet distinct concepts, with each engaging different neural substrates (Paus, 2000).

Arousal typically refers to the physiological and psychological changes influencing the overall state of alertness or readiness to respond to stimuli. In contrast, attention involves the selective focus on specific aspects of stimulation. Arousal can impact attentional

processes by increasing alertness and enhancing stimulus salience. Elevated arousal can lead to increased vigilance, manifested by the modulation of neural activity in sensory cortices, which, in turn, facilitates attention and influences subsequent cognitive performance (Asutay & Västfjäll, 2017). In our data, a falling tone seems to evoke higher arousal than a rising tone, which could lead to more engaged attention and conscious processing (but see (Lialiou et al., 2024) for an attention-orienting effect of a rising tone in German compared to a falling tone). Further research is necessary to fully understand the processing of pitch contour in relation to arousal and valence and its potential impact on cognitive processing.

It is important to note that the perceptual and cognitive analogy view of iconicity (e.g., Sidhu & Pexman, 2018) and the adaptive view (e.g., Adelman et al., 2018) are not mutually exclusive. Adaptive significance might have driven the co-occurrence of muscle movements involved in both the articulation of specific sounds and the expression of emotions, enhancing communicative efficiency and offering survival advantages.

In summary, this study extends the adaptation view of affective iconicity in three key ways: from segmental to suprasegmental features, from valence-only to both valence and arousal, as well as from Indo-European to Sinitic languages. Our findings emphasise the role of linguistic pitch in emotional communication, the distinct yet complementary contributions of emotional arousal and valence to emotional communication, and the potential cross-linguistic and adaptive or evolutionary significance of affective iconicity. However, the sources or acoustic correlates of the adaptive significance may differ between phonemes and tonemes. Further investigation is necessary to explore the positional influence of both segmental

phonemes and lexical tones on affective iconicity. It is plausible that there exist two distinct mechanisms for the positional effects of segmental and suprasegmental tonal iconicity, or a more general mechanism that extends and further refines the negative priority developed for segments (as defined in Adelman et al., 2018).