Bilingualism: Language and Cognition

cambridge.org/bil

Knowing and feeling: The role of semantic valence in emotional prosody perception for L2 Chinese learners

Cheng Xiao[1,2], Jiang Liu[2,3], Svetlana V. Shinkareva[4], Xuan Yang[1] and Rutvik H. Desai[1,4]

[1]Department of Psychology, University of South Carolina, Columbia, USA; [2]Linguistics Program, University of South Carolina, USA; [3]Department of Languages, Literatures and Cultures, University of South Carolina, USA and [4]Institute for Mind and Brain, University of South Carolina, USA

Research Article

Cite this article: Xiao, C., Liu, J., Shinkareva, S.V., Yang, X. and Desai, R.H. (2026). Knowing and feeling: The role of semantic valence in emotional prosody perception for L2 Chinese learners. *Bilingualism: Language and Cognition* 1–12. https://doi.org/10.1017/S1366728926101813

Received: 8 May 2025 UTC
Revised: 17 June 2026 UTC
Accepted: 29 July 2026 UTC

Keywords:
semantic valence; emotional prosody; semantic knowledge; bilingual emotion processing; Mandarin Chinese

Corresponding author:
Cheng Xiao;
Email: cxiao@mailbox.sc.edu

This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

Abstract

In Mandarin Chinese, the same acoustic feature, pitch, conveys both semantic valence and emotional prosody. The effects of semantic valence on bilingual emotional prosody perception are mixed, with some studies reporting semantic facilitation and others showing interference. To clarify this, we extend the conceptual account of semantic valence to second language acquisition by examining its two components: learners' semantic knowledge and the inherent semantic valence of stimuli. We investigated the independent and interactive effects of these components of semantic valence on emotional prosody perception for L1-English L2-Chinese learners in Chinese words and sentences. We found that semantic valence facilitated L2 learners' perception when congruent with emotional prosody, but interfered when incongruent. Notably, these semantics-prosody (in)congruence effects were significantly predicted by learners' semantic knowledge. Our results suggest an interaction between linguistic knowledge and paralinguistic perception in bilingual emotion processing.

Highlights

- Linguistic and paralinguistic cues interact in bilingual emotion processing.
- For bilinguals, semantic valence consists of semantic knowledge and stimulus valence as distinct components.
- Regardless of learners' semantic knowledge, semantics-prosody congruence aids emotion perception, whereas incongruence hinders it.
- Greater semantic knowledge amplifies semantics-prosody (in)congruence effects in L2 emotion processing.
- Semantic valence influences the accuracy and reaction time of emotional prosody perception in distinct ways for L2 learners.

1. Introduction

Emotional information in spoken language is conveyed through multiple channels, most prominently semantic valence and emotional prosody (Kotz & Paulmann, 2007; Nygaard & Tzeng, 2021; Pell et al., 2011; Perlovsky, 2013; Seydell-Greenwald et al., 2020; Xiao; 2025). Semantic valence refers to the stored semantic knowledge about whether an event, object or concept is positive or negative (Itkes et al., 2017; Itkes & Kron, 2019). Emotional prosody refers to the modulation of the tone of voice to express positive or negative emotions, attitudes, and feelings (Kemmerer, 2014). When processing emotional speech, listeners decode not only the semantic valence of an utterance (what is said), but also the speaker's emotional prosody (how it is said) (Cutler et al., 1997; Wilson & Wharton, 2006). For example, a listener may perceive an utterance with positive semantic valence (e.g., *she is wise*) differently from an utterance with negative semantic valence (e.g., *she is arrogant*), even if both are spoken with happy emotional prosody. Previous research has focused on the effects of semantic valence on word recognition (Estes & Adelman, 2008; Gao et al., 2022; Kim & Sumner, 2017; Kuperman et al., 2014; Yap & Seow, 2014) and emotion judgments (Hauser & Schwarz, 2016; Lin et al., 2020; Xiao & Liu, 2025, 2026) among native speakers in their first language (L1). However, much less is known about how second language (L2) learners process semantic valence and emotional prosody.

In tonal languages such as Mandarin Chinese, the interplay between semantic valence and emotional prosody is even more complex because both rely on the same acoustic feature – pitch (Shen, 1985; Xu, 2005). In other words, the modulation of pitch does not merely convey emotional prosody; it also functions as lexical tone to differentiate word meanings. There are four lexical tones in Mandarin. For instance, the syllable "ma" can represent four distinct words depending on the lexical tone applied: *mother* (Tone 1), *hemp* (Tone 2), *horse* (Tone 3) and *to scold* (Tone 4) (Yip, 2002). When processing emotional speech in tonal languages, semantic valence and emotional prosody coexist and interact (Chang et al., 2023; Ip & Cutler, 2020; Ouyang & Kaiser, 2015), raising an important question for English-Chinese bilinguals: Does greater knowledge of semantic valence influence the perception of emotional prosody? Therefore, this study investigates the role of semantic valence in Chinese emotional prosody perception for L1-English L2-Chinese learners.

1.1 Semantic knowledge and semantic valence

The Theory of Constructed Emotion proposes that language plays a fundamental role in supporting individuals in acquiring emotion categories initially and using this knowledge later to interpret emotion perception (Barrett, 2006; Barrett, 2017; Barrett et al., 2007; Lindquist et al., 2015; Lindquist & Gendron, 2013). Within this framework, semantic knowledge acts as the conceptual glue for emotion perception, binding otherwise variable sensorimotor and affective states into recognizable emotion categories (Barrett et al., 2007). This reliance on semantic knowledge is evident in facial emotion perception. For example, when an emotion word was semantically satiated (e.g., by repeatedly saying *ange*r to temporarily disrupt its meaning), participants became slower and less accurate in categorizing facial expressions into emotions (Lindquist et al., 2006). Furthermore, when semantic knowledge is chronically impaired, such as in patients with semantic dementia, individuals perceive emotional facial expressions merely as broad affective dimensions (e.g., pleasant or unpleasant) rather than as specific, discrete emotions (e.g., anger, sadness) (Lindquist et al., 2014). Despite these findings in facial emotions, the role of semantic knowledge in perceiving vocal emotions, such as emotional prosody, is largely underexplored. This study therefore examines the effect of semantic knowledge in emotional prosody perception. Here, semantic knowledge refers narrowly to lexico-semantic knowledge, namely, the knowledge of individual word meanings.

The theoretical link between emotional prosody and semantic knowledge can be grounded in Russell's Circumplex model (Russell, 1980), which maps emotional states along two primary dimensions: valence (positive versus negative) and arousal (excited versus calm). Although both dimensions are theoretically relevant, the present study focuses on the dimension of valence, building on a conceptual framework of valence that distinguishes between the knowledge of valence and the feeling of valence (Itkes et al., 2017; Itkes & Kron, 2019). The knowledge of valence (semantic valence) is the stored semantic knowledge about whether an object or event is positive or negative, whereas the feeling of valence (affective valence) represents the immediate, subjective, and sensory-driven affective reaction elicited by a stimulus (Itkes et al., 2017; Itkes & Kron, 2019). Furthermore, they found that semantic valence and affective valence dissociate in both self-reported experience and physiological responses (Itkes et al., 2017; Itkes et al., 2019; Olteanu et al., 2023). Specifically, knowledge-based perception of semantic valence showed less attenuation with repeated exposures, lower sensitivity to perceptual changes, and a higher likelihood of eliciting mixed emotions. In contrast, feeling-based perception of affective valence showed greater attenuation with repeated exposures to pleasant and unpleasant images, greater sensitivity to the perceptual details of the stimuli, and a lower likelihood of eliciting mixed emotions.

The work of Itkes and colleagues represents a notable advancement in the field of emotion perception by demonstrating that the effect of semantics, semantic valence specifically, is not a passive byproduct but a distinct system with its own behavioral profile. However, their research, and emotion perception research more broadly, has largely been conducted in native speakers, examining how semantic valence varies across contexts without considering how participants' underlying semantic knowledge may itself vary across individuals. Notably, L2 learners' semantic knowledge is often incomplete or variable. For example, an L2 English learner unfamiliar with the intended semantically positive word *sleek* might mistakenly interpret it as having a negative valence, as evidenced by phonaesthemes, which suggest that certain sound clusters (e.g., sl-) are often linked to particular affective associations (Bergen, 2004; Hutchins, 1998).

Taken together, prior research shows that semantic knowledge provides a conceptual scaffold for perceiving and interpreting emotions (Barrett, 2006). Importantly, semantic knowledge of valence has been shown to operate independently from the affective feeling of valence (Itkes et al., 2017). In an L2 context, however, such semantic knowledge may be less stable and more variable than in an L1. Because of this inherent variability, L2 learners' semantic knowledge cannot be treated as random variance. We propose that participants' semantic knowledge, particularly their lexico-semantic knowledge, can also modulate emotional prosody perception in addition to the semantic valence of the stimuli. Therefore, this study extends the established research of semantic valence into the L2 domain by conceptualizing it as two components: the semantic knowledge of participants and the inherent semantic valence of stimuli. We expect that both components play distinct roles in bilinguals' perception of emotional prosody.

1.2 Semantic valence in second language acquisition

L2 learners can identify a wide range of emotional prosody with above-chance accuracy at syllable, word, and sentence levels (e.g., Alm & Llorà, 2006; Altrov, 2013; Dromey et al., 2005; Laukka & Elfenbein, 2021; Wei et al., 2022; Xiao, 2025; Xiao & Liu, 2026; Zhu, 2013). However, the effects of semantic valence on emotion perception in an L2 is distinct from that of an L1. For example, L2 learners generally perceive the emotionality in words and sentences (e.g., *I love you*) as having less emotional weight and greater emotional distance compared to their native language (Dewaele, 2008; Keysar et al., 2012; Pavlenko, 2008, 2012). Moreover, a learner's linguistic and cultural background also influences their interpretation of L2 semantic valence. Jackson et al. (2019) found that the semantic valence of stimuli varies significantly across language families. For example, the meaning of *anxiety* is closely associated with *fear* in Tai-Kadai languages, while it is associated with *grief* and *regret* in Austroasiatic languages. Therefore, it is necessary to systematically investigate the role of semantic valence in L2 emotion perception.

Current debates in the field of L2 emotional prosody perception have centered on the two components of semantic valence: the semantic knowledge of the participants and the inherent semantic valence of the stimuli. The first debate explores whether L2 learners' semantic knowledge facilitates or interferes with their perception of emotional prosody in an L2. Some studies found a facilitation effect of semantic knowledge. For example, non-native speakers with semantic knowledge (i.e., L2 learners) outperformed non-native speakers without such knowledge (i.e., naïve listeners) in recognizing emotional prosody (Altrov, 2013; Xiao & Liu, 2026; Zhu, 2013).

https://doi.org/10.1017/S1366728926101813

Additionally, L2 Japanese learners with more learning experience recognized emotional prosody more accurately compared to those with less learning experience (Shochi et al., 2016). However, Bhatara et al. (2016) found that L2 English learners with higher proficiency were less accurate in recognizing positive emotional prosody in their L2 English compared to those with lower proficiency, and they suggested this is due to top-down semantic interference, where highly proficient learners paid more attention to the lexical and semantic meaning rather than the speaker's emotional tone.

Notably, in previous studies, researchers have typically measured L2 semantic knowledge by general proficiency levels (e.g., Bhatara et al., 2016; Shochi et al., 2016), assuming that advanced learners possess more semantic knowledge (i.e., knowing more words). This general proficiency is a broad composite construct that serves as an indirect proxy for semantic knowledge. Meanwhile, other researchers have also demonstrated that vocabulary screening tests directly quantify lexico-semantic knowledge, a factor critical to emotion perception (Brysbaert et al., 2017; Ferré et al., 2024; Streubel et al., 2020). Therefore, we first measured learners' semantic knowledge by using general proficiency in the pilot experiment, aiming to contribute evidence from a tonal language to the existing debate on the semantic knowledge effect (e.g., Bhatara et al., 2016; Shochi et al., 2016). In the main experiment, we then used a vocabulary screening test to assess participants' semantic knowledge, examining whether greater knowledge of semantic valence facilitates the perception of emotional prosody in L2 Chinese.

The second debate addresses how the semantic valence of the stimuli influences emotional prosody perception in a second language. Building on evidence that native speakers perceive emotional prosody faster and more accurately in congruent conditions than in incongruent ones (e.g., Lin et al., 2020; Pell et al., 2011), L2 research has also begun to explore this semantics-prosody congruence effect, but has yielded mixed results. Some researchers reported a congruence effect, where L2 Chinese learners recognized emotional prosody more accurately and quickly when the semantic valence and emotional prosody of the stimuli were congruent (Xiao & Liu, 2025). But others did not replicate this effect in L2 English learners (Cho & Dewaele, 2021). One limitation of this line of work is that it only considers the semantic valence of stimuli, without accounting for the variability in L2 learners' semantic knowledge. It is possible that the congruence effect emerges only in L2 learners with specific levels of semantic knowledge. To address this, the present study manipulated both components of semantic valence, namely, L2 learners' semantic knowledge and the inherent semantic valence of stimuli, aiming to clarify the independent and interactive effects of the two components of semantic valence in L2 emotional prosody perception.

1.3 Semantic valence in L2 Chinese learners

As demonstrated above, previous studies on emotional prosody perception in L2 learners have yielded contrasting results regarding the role of semantic valence, specifically the effects of semantic knowledge (e.g., Bhatara et al., 2016; Shochi et al., 2016; Xiao & Liu, 2026) and semantics-prosody congruence (e.g., Cho & Dewaele, 2021; Xiao & Liu, 2025). These conflicting findings also suggest that the perception of emotional prosody may differ between tonal languages (e.g., Mandarin Chinese) and non-tonal languages (e.g., English). Specifically, tonal languages constrain the use of pitch for signaling emotional prosody because pitch is already occupied by lexical tone (Ross et al., 1986). Consequently, Chinese listeners may face a higher cognitive demand when integrating these overlapping pitch cues for semantic and prosodic information (Lin et al., 2020; Xiao, 2025; Xiao & Liu, 2025), or they may compensate by adopting secondary acoustic cues, such as intensity and speech rate, to differentiate emotional prosody (Wang et al., 2018).

Mandarin Chinese, as a tonal language, uses tone of voice to encode both semantics and emotional prosody (Xu, 2005). It has four lexical tones that differentiate word meanings: 1: high-level tone; 2: rising tone; 3: low-rising tone; 4: falling tone (Yip, 2002). For example, ma^1 means *mother*, ma^2 means *hemp*, ma^3 means *horse*, and ma^4 means *to scold.* Semantic cues and prosodic cues coexist within the shared acoustic feature – pitch (Shen, 1985), suggesting an interaction between semantic valence and emotional prosody in Chinese words and sentences (Lin et al., 2020; Xiao & Liu, 2025). Therefore, processing emotional prosody in Mandarin Chinese can be challenging for L2 learners, as it requires rapid integration of linguistic (semantic valence) and paralinguistic (emotional prosody) cues within a fraction of a second (Kao & Zhang, 2020; Taguchi, 2008).

Prior research on Chinese emotional prosody perception has mainly compared L2 Chinese learners with native Chinese speakers from a cross-cultural perspective (e.g., Xiao & Liu, 2025, 2026; Zhu, 2013). Notably, these studies found that L2 Chinese learners, even with a brief learning experience (e.g., less than six months), achieved accuracy on par with native speakers in identifying emotional prosody in Chinese words and sentences, with performance improving as syllable length increases (Xiao & Liu, 2026; Zhu, 2013). Furthermore, much like native speakers, L2 Chinese learners demonstrated a significant semantics-prosody congruence effect, recognizing emotional prosody more accurately and faster when the semantic valence and emotional prosody of the stimuli were congruent (Xiao & Liu, 2025). Interestingly, L2 Chinese learners exhibited an even stronger reliance on the semantic valence of stimuli than native Chinese speakers: while both groups showed this congruence effect when prosodic information was limited in disyllable words, only L2 learners showed the influence of semantic valence when prosodic cues were ample in full sentences (Xiao & Liu, 2025).

However, in L2 emotional prosody perception, few studies have systematically manipulated the two components of semantic valence, namely, the semantic knowledge of participants and the semantic valence of the stimuli (Bhatara et al., 2016; Shochi et al., 2016; Xiao & Liu, 2025). Existing research has either examined the effects of the stimuli's semantic valence while overlooking learners' varying semantic knowledge (Cho & Dewaele, 2021; Xiao & Liu, 2025) or manipulated the learners' semantic knowledge without controlling for the stimuli's semantic valence (Bhatara et al., 2016; Shochi et al., 2016). To clarify the role of semantic valence in L2 Chinese, the current study manipulated both learners' semantic knowledge and the semantic valence of the stimuli to investigate the perception of emotional prosody in Chinese words and sentences. The findings of this study will provide insights into how L2 learners simultaneously process linguistic knowledge and paralinguistic information in a tonal language.

2. The current study

The overarching aim of this study was to examine the role of semantic valence in bilingual emotion processing. While semantic valence is traditionally defined as the stored semantic knowledge

https://doi.org/10.1017/S1366728926101813

about whether an event, object, or concept is positive or negative (Itkes et al., 2017), we argue that it must be decoupled into two distinct components for bilinguals: the semantic knowledge of participants and the semantic valence of the stimuli. To isolate the independent and interactive effects of these two components, participants' semantic knowledge was operationalized through learning experience (in the pilot experiment) and a vocabulary test (in the main experiment); the semantic valence of the stimuli was also manipulated across both experiments. This study investigated whether greater knowledge of semantic valence facilitated the perception of emotional prosody for L1-English L2-Chinese learners in Chinese words and sentences.

Specifically, we conducted a pilot experiment to examine whether novice and advanced learners differ in their perception of Chinese emotional prosody, contributing new evidence from a tonal language to the existing debate (Bhatara et al., 2016; Shochi et al., 2016). In the main experiment, we manipulated the two components of semantic valence (i.e., learners' semantic knowledge and stimuli's semantic valence) to investigate their specific effects and interactions during Chinese emotional prosody perception, extending the account of semantic valence (Itkes et al., 2017) to the field of second language acquisition.

We used G*Power (Faul et al., 2009) to determine the a priori sample size. For both experiments, 40 participants were required to obtain 0.89 power to detect a medium effect size of 0.22 (obtained from Lin et al., 2020) at the standard 0.05 alpha error probability. All materials, code, data, and anonymized participant details are available on the Open Science Framework (OSF) repository at https://osf.io/fs4py/overview. Data were analyzed using R (R Core Team, 2024). To build the model, we used backward elimination starting with a maximal model that included all potential main effects and interactions, and we removed the non-significant variables one at a time based on model comparison (Barr et al., 2013).

3. The pilot experiment

To examine the contrasting results regarding semantic knowledge in non-tonal languages (Bhatara et al., 2016; Shochi et al., 2016), the pilot experiment extended the investigation in a tonal language, Mandarin Chinese. We investigated whether greater semantic knowledge affects the perception of emotional prosody in Mandarin Chinese words and sentencesby comparing novice and advanced L1-English L2-Chinese learners. To isolate the specific impact of semantic knowledge, we controlled the semantic valence of the stimuli to be neutral (e.g., shou^{1}yin^{1}ji^{1}, *radio*). We hypothesized that L2 learners' semantic knowledge would significantly influence their perception of emotional prosody in Chinese words and sentences. Additionally, we predicted that the effect of semantic knowledge would interact with both emotional prosody and syllable length based on prior cross-cultural findings (e.g., Pell & Kotz, 2011; Xiao & Liu, 2026).

3.1 Methods

3.1.1 Participants

Participants were L1-English L2-Chinese learners enrolled in Mandarin courses at a public U.S. university. We excluded 12 heritage speakers of Mandarin Chinese or other tonal languages, resulting in a total of 40 participants in the current analysis. Following prior studies (Bhatara et al., 2016; Shochi et al., 2016), we categorized participants based on their length of L2 Chinese learning experience: 20 novice L1-English L2-Chinese learners (novice learners, learning experience < 6 months; 9 females; mean age = 19, *SD* = 0.65) and 20 advanced L1-English L2-Chinese learners (advanced learners, learning experience > 3 years; 13 females; Mean age = 21.05, *SD* = 0.94). The novice learners were also included in a study that addressed a different research question (Xiao & Liu, 2026). All participants had normal or corrected-to-normal vision, reported no hearing problems, and received course credit or monetary compensation. The study was approved by the Institutional Review Board of the University of South Carolina.

Table 1. Example stimuli from the pilot experiment, categorized by syllable length

	Monosyllable	Disyllable	Trisyllable	Sentence
Pinyin	xiu^{1}	shou^{1}yin^{1}	zhang^{1} zhong^{1}bin^{1}	zhang^{1} zhong^{1}bin^{1} xing^{1}qi^{1}tian^{1} xiu^{1} shou^{1}yin^{1}ji^{1}
Chinese character(s)	修	收音	张中斌	张中斌星期天修收音机
English translation	"repair"	"Receive sound"	"Zhang Zhongbin"	"Zhang Zhongbin repairs the radio on sunday."

Note: Pinyin represents the official romanization system used to transcribe Mandarin Chinese pronunciation, with numbers indicating the specific lexical tone. Chinese characters are the logographic symbols representing the distinct word for each syllable-tone combination, followed by their English translations.

3.1.2 Stimuli

We selected and adapted a total of 144 Chinese stimuli from Shen (1985), controlling for lexical tone and neutral semantic valence (see Table 1 for examples). A female native Chinese speaker recorded these stimuli in four emotional prosodies: neutral, joy, anger, and sadness. Utterances were segmented and normalized using Praat (Boersma & Weenink, 2023), and the intended emotional prosody of each stimulus was further validated by six native Chinese speakers prior to the experiment. Based on the combination of emotional prosody and lexical tone, the stimulus set included 16 monosyllabic words, 64 disyllabic words, 48 trisyllabic words, and 16 sentences. The acoustic parameters of the stimuli can be found in the supplementary materials (Supplementary Table S6).

3.1.3 Procedure

Participants first completed a brief language background questionnaire regarding age, gender, native languages, and L2 Chinese learning experience, and then were randomly assigned to one of the four versions of the emotional prosody judgment task. There are four blocks in each task: monosyllable, disyllable, trisyllable and sentence blocks. The four versions were used to counterbalance the order of the block across participants. Within each block, the order of emotional prosody was also counterbalanced. The emotional prosody judgment task was untimed, where participants were instructed to sit in a quiet environment, listen to a series of utterances, one at a time, and then judge the intended emotional prosody for each utterance using a 4-alternative forced-choice format (i.e., neutral, joy, anger, and sadness). Six filler utterances were included to monitor participant attention. Participants' emotional prosody recognition responses for each trial were recorded. The experiment lasted about 30 minutes.

3.1.4 Analysis

The total number of trials in the data analysis was 5,760. For each trial, participants received a score of 1 if they recognized the emotional prosody correctly; they received a score of 0 if their judgment mismatched the intended emotional prosody of the

https://doi.org/10.1017/S1366728926101813

utterance. The accuracy (ACC) of the participants' responses was analyzed using a logistic mixed-effects model (Jaeger, 2008). The model included three fixed factors: semantic knowledge group with two levels (novice learners, advanced learners); emotion prosody type with four levels (neutral, joy, anger, and sadness); and syllable length with four levels (monosyllable, disyllable, trisyllable, and sentence). The advanced group, neutral prosody, and monosyllable were treated as the reference levels, respectively. Dummy coding was used, and we entered intercepts for items and subjects, as well as a by-item random slope for the effect of emotion prosody type (Cunnings, 2012).

3.2 Results

Both novice learners and advanced learners demonstrated very high ACC in recognizing emotional prosody, averaged across four types of emotional prosody (Figure 1). The mean (SD) ACC for novice learners was 95.9% (0.199), and the mean (SD) ACC for advanced learners was 91.8% (0.275). Overall, the mean ACC of emotional prosody judgment of L2 Chinese learners increased with syllable length: 87.5% for monosyllables, 93.0% for disyllables, 96.1% for trisyllables, and 96. 3% for sentences, respectively. However, the SD decreased with syllable length: 0.331 for monosyllables, 0.254 for disyllables, 0.193 for trisyllables and 0.190 for sentences. The mean ACC of the four types of emotional prosody was 95.7% for neutral, 90.9% for joy, 94.3% for anger, and 94.4% for sadness.

As shown in Supplementary Table S1, a logistic mixed-effects model showed a significant main effect of emotion prosody type, where both novice and advanced learners showed lower ACC for joy ($\beta = -1.789$, SE = 0.396, $z = -4.521$, $p < .001$), anger ($\beta = -1.104$, SE = 0.340, $z = -3.250$, $p < .01$), and sadness ($\beta = -0.947$, SE = 0.296, $z = -3.197$, $p < .01$) compared to neutral emotional prosody. There was a significant main effect of syllable length such that both groups showed a higher ACC for disyllables ($\beta = 0.859$, SE = 0.325, $z = 2.642$, $p < .01$), trisyllables ($\beta = 1.455$, SE = 0.351, $z = 4.143$, $p < .001$) and sentences ($\beta = 1.303$, SE = 0.445, $z = 2.927$, $p < .01$) compared to monosyllables. This indicated that L2 Chinese learners, regardless of their semantic knowledge, perceived neutral prosody more accurately than positive and negative prosody, and their ACC of emotional prosody perception increased with syllable length.

We also observed significant interactions between semantic knowledge group and emotional prosody, and between semantic knowledge group and syllable length, in L2 emotional prosody processing. Specifically, novice learners perceived emotional prosody more accurately compared to advanced learners when processing joy ($\beta = 1.527$, SE = 0.377, $z = 4.052$, $p < .001$), anger ($\beta = 1.127$, SE = 0.399, $z = 2.827$, $p < .01$) and sadness ($\beta = 1.235$, SE = 0.405, $z = 3.049$, $p < .01$) compared to processing neutral prosody. Furthermore, a significant interaction between group and syllable length was also observed, where novice learners' advantage over advanced learners was reduced when recognizing emotional prosody in disyllables ($\beta = -0.712$, SE = 0.356, $z = -2.002$, $p < .05$) compared to monosyllables. We further subset the data to include neutral prosody only, and we found that there was no significant difference in emotional prosody perception between novice and advanced learners. The results revealed that novice learners had an advantage over advanced learners in recognizing positive and negative prosody in Chinese words and sentences, but not in identifying neutral prosody.

To sum up, both emotional prosody and syllable length significantly influenced L2 emotional prosody perception, with learners' semantic knowledge modulating these effects. Specifically, novice learners showed higher accuracy than advanced learners in identifying positive and negative prosody, whereas both groups performed similarly on neutral prosody. This advantage for novice learners was mitigated by syllable length, impacting shorter stimuli more than longer ones. These findings suggested a potential semantics-prosody incongruence effect where conflict between semantic valence and emotional prosody interferes with perception, and since advanced learners possessed more semantic knowledge, they may experience greater interference than novice learners. Importantly, while no congruence effect emerged when semantic valence and emotional prosody were both neutral, this effect may still occur in stimuli with positive or negative semantic valence. Because this pilot experiment only used stimuli with neutral semantic valence, we conducted the second experiment (the main experiment) to further investigate how the stimuli's semantic valence interacts with learners' semantic knowledge in L2 emotional prosody perception.

4. The main experiment

The main experiment investigated the independent and interactive effects of two components of semantic valence, namely, the

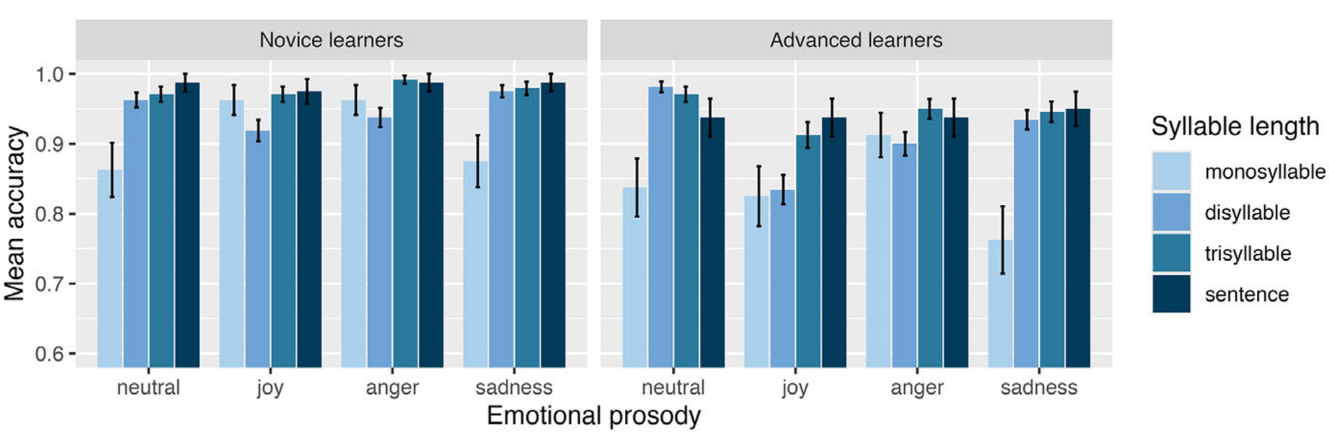

Figure 1. Mean accuracy of emotional prosody judgments across four types of emotional prosody and syllable length in novice and advanced groups. The black vertical lines show standard errors.

https://doi.org/10.1017/S1366728926101813

participants' semantic knowledge and the semantic valence of the stimuli, on the perception of emotional prosody in L2 Chinese learners. Building on the pilot study findings, we refined our measure of semantic knowledge based on learners' performance on a target vocabulary test. To systematically manipulate the semantic valence of the stimuli, we selected words with positive, negative, and neutral semantic valence, matched for syllable length, lexical tone, and frequency. Furthermore, as the pilot study found no group differences for neutral prosody, the current experiment focused exclusively on positive and negative emotional prosody. Finally, because the pilot study showed that syllable length interacted with semantic knowledge, we divided the stimuli into word and sentence tasks.

Based on the pilot experiment and previous research (Bhatara et al., 2016; Lin et al., 2020; Shochi et al., 2016; Xiao & Liu, 2025), we anticipated an effect of semantic knowledge on emotional prosody perception in L2 Chinese. We hypothesized that L2 Chinese learners would demonstrate a semantics-prosody congruence effect regardless of their semantic knowledge levels. Furthermore, we predicted an interaction between semantic knowledge and semantics-prosody congruence, such that learners with high semantic knowledge would show a larger congruence effect than those with low semantic knowledge.

4.1 Method

4.1.1 Participants

We recruited L1-English L2-Chinese learners enrolled in intermediate-level Chinese courses. These learners had studied Chinese for one to three years and were familiar with basic syllable-tone combinations and simple sentence structures in Mandarin Chinese, indicating a similar baseline of general proficiency. All learners were taught the meanings of the real target words used as experimental stimuli in class three times per week for two weeks. Notably, the experimental stimuli also included another subset of novel pseudowords, which were explicitly excluded from instruction to ensure they remained unknown to the participants. During this period, learners completed two practice vocabulary tests.

Based on previous research demonstrating that vocabulary tests can quantify learners' semantic knowledge (Brysbaert et al., 2017; Ferré et al., 2024; Streubel et al., 2020), we administered a timed vocabulary test immediately prior to the experiment. This vocabulary test allowed us to categorize learners of similar general proficiency into two groups based on their semantic knowledge of the target stimuli. Learners who scored above 90% on the final vocabulary test were categorized as having more semantic knowledge, while those who scored below 75% were categorized as having less semantic knowledge. We excluded two participants whose scores fell between 75% and 90%, as well as six participants who were heritage speakers of Mandarin Chinese or other tonal languages.

A total of forty participants were included in the analysis: 23 L1-English L2-Chinese learners with more semantic knowledge (mean vocabulary score = 98%; 12 females; mean age = 19.35, *SD* = 0.99) and 17 L1-English L2-Chinese learners with less semantic knowledge (mean vocabulary score = 57%; 5 females; mean age = 19.39, *SD* = 1.08). Twelve participants in the more semantic knowledge group were also included in a previous study that addressed a different research question (Xiao & Liu, 2025). All participants had normal or corrected-to-normal vision, reported no hearing problems, and received course credit or monetary compensation for their participation. The study were approved by the Institutional Review Board of the University of South Carolina.

4.1.2 Stimuli

The stimuli were divided into word and sentence tasks. The word task included 128 disyllabic Chinese words, and the sentence task included 128 Chinese sentences. The critical manipulation of the stimuli was the congruence of semantic valence (positive, negative, and neutral) and emotional prosody (positive and negative). All words with positive, negative, and neutral semantic valence were matched for syllable length and lexical tone. In the word task, real disyllables with positive and negative semantic valence were matched in word frequency (Sun et al., 2018). We also created novel pseudo-disyllables with neutral semantic valence using lexical gaps. A lexical gap in Chinese occurs when a syllable can combine with specific tones but not all four tones (Ryu et al., 2022). For example, chun^4 represents a lexical gap in Mandarin Chinese, as the syllable "chun" can combine with tone 1 (chun^1 means *spring*), tone 2 (chun^2 means *pure*), and tone 3 (chun^3 means *stupid*), but not tone 4 (no real word exists in chun^4). In the sentence task, we embedded the word stimuli from the word task into a simple carrier sentence "ta^1 hen^3 (*She is very*) ______" to minimize syntax parsing and contextual information for L2 learners (Cho & Dewaele, 2021). Table 2 provides example stimuli with positive semantic valence (1a, 2a), negative semantic valence (1b, 2b), and neutral semantic valence (1c, 2c).

A male native Chinese speaker recorded all stimuli for the word and sentence tasks with both positive (i.e., happy) and negative (i.e., sad) emotional prosody in a soundproof studio (Cho & Dewaele, 2021; Lin et al., 2020; Pell & Kotz, 2011). Background noise was removed, and peak volume was normalized using Audacity. The stimuli were segmented and normalized with the same amplitude using Praat (Version 6.3.10). The acoustic parameters of the stimuli

Table 2. Example stimuli for the word task and sentence task in the main experiment, categorized by semantic valence

	Positive semantic valence		Negative semantic valence		Neutral semantic valence	
Word	(1a)	rui^{4}zhi^4	(1b)	ao^{4}man^4	(1c)	chun^{4}pou^4
		睿智		傲慢		chunpou
		"wise"		"arrogant"		"chunpou" (pseudo-word)
Sentence	(2a)	ta^{1}hen^{3}rui^{4}zhi^4	(2b)	ta^{1}hen^{3}ao^{4}man^4	(2c)	ta^{1}hen^{3}chun^{4}pou^4
		她很睿智		她很傲慢		她很chunpou
		"She is very wise."		"She is very arrogant."		"She is very chunpou."

Note: Pinyin (Row 2 and 5) represents the official romanization system used to transcribe Mandarin Chinese pronunciation, with numbers indicating the specific lexical tone. Chinese characters (Row 3 and 6) are the logographic symbols representing the distinct word for each syllable-tone combination, followed by their English translations (Row 4 and 7).

https://doi.org/10.1017/S1366728926101813

can be found in the supplementary materials (Supplementary Table S7).

4.1.3 Procedure

All participants completed the final vocabulary test immediately prior to the emotional prosody judgment tasks in the language lab. The vocabulary test was administered online with a 10-minute time limit under a Chinese instructor's supervision. For each target word, participants heard it spoken in neutral prosody and selected its correct meaning using a three-alternative forced-choice format. Following the vocabulary test, participants completed two emotional prosody judgment tasks, performing the word task before the sentence task. For each trial, participants listened to an audio clip and judged whether the emotional prosody of each utterance was positive or negative as quickly as possible without making any mistakes. Auditory stimuli were presented in a randomized order using the E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). Each trial lasted 5000 ms, and a new trial began immediately after the participant made a response. Participants' ACC and reaction time (RT) were recorded for each trial. After completing the word and sentence tasks, participants provided language background information and completed a prosody rating task. As the present study focuses on semantic valence, the prosody rating data are not reported here. The experiment took approximately 30 minutes to complete.

4.1.4 Analysis

In the main experiment, we analyzed both ACC and RT. We excluded RTs of incorrect judgments or those greater than three standard deviations calculated based on each participant's mean, resulting in 8,786 valid trials. We used logistic mixed-effects models (Jaeger, 2008) to analyze ACC and linear mixed-effects models (Bates et al., 2015) to analyze RT. The full model for ACC and RT analyses was the same in both tasks. Each model included three fixed factors: (1) semantic knowledge group with two levels (learners with more semantic knowledge, learners with less semantic knowledge), (2) semantics-prosody congruence with three levels (congruent, incongruent,and neutral) and (3) emotional prosody with two levels (positive, negative). The more knowledge group, neutral semantics-prosody congruence, and positive prosody valence were treated as the reference levels, respectively. Dummy coding was used, and we entered intercepts for items and subjects (Cunnings, 2012). In addition, we examined the RT distributions in the two tasks separately using the diffusion model analysis (Brysbaert et al., 2017; Myers et al., 2022).

4.2 Results

4.2.1 ACC results

Overall, L2 Chinese learners exhibited a similar accuracy in perceiving emotional prosody in both word task (learners with more semantic knowledge: M = 84.4%, SD = 0.36; learners with less semantic knowledge: M = 84.6%, SD = 0.36) and sentence task (learners with more semantic knowledge: M = 89.1%, SD = 0.31; learners with less semantic knowledge: M = 89.8%, SD = 0.30). Compared to the neutral condition, both learner groups demonstrated higher accuracy in the semantics-prosody congruent condition (Word Task: Mean diff = 6.3%; Sentence Task: Mean diff = 3.3%) and lower accuracy in the incongruent condition (Word Task: Mean diff = 9.4%; Sentence Task: Mean diff = 5.8%) (Figure 2).

For the word task (Supplementary Table S2), a logistic mixed-effects regression model indicated that learners with more semantic knowledge and learners with less semantic knowledge both showed semantics-prosody (in)congruence effects: compared to the neutral condition, their ACC of emotional prosody perception was significantly higher in the congruent condition (β = 1.200, SE = 0.202, z = 5.932, $p < .001$) and significantly lower than in the incongruent condition (β = −0.982, SE = 0.155, z = −6.331, $p < .001$). Furthermore, there were significant interactions between group and congruence, where learners with less semantic knowledge showed a smaller congruence effect (β = −0.864, SE = 0.251, z = −3.439, $p < .001$) and a smaller incongruence effect (β = 0.421, SE = 0.200, z = 2.111, $p < .05$) compared to learners with more semantic knowledge. In addition, a significant interaction between group and emotion prosody was found such that learners with less semantic knowledge were more accurate in recognizing positive emotional prosody compared to learners with more semantic knowledge (β = 0.770, SE = 0.177, z = 4.333, $p < .001$).

Similar patterns were also found in the sentence task (Supplementary Table S4). Both learner groups recognized

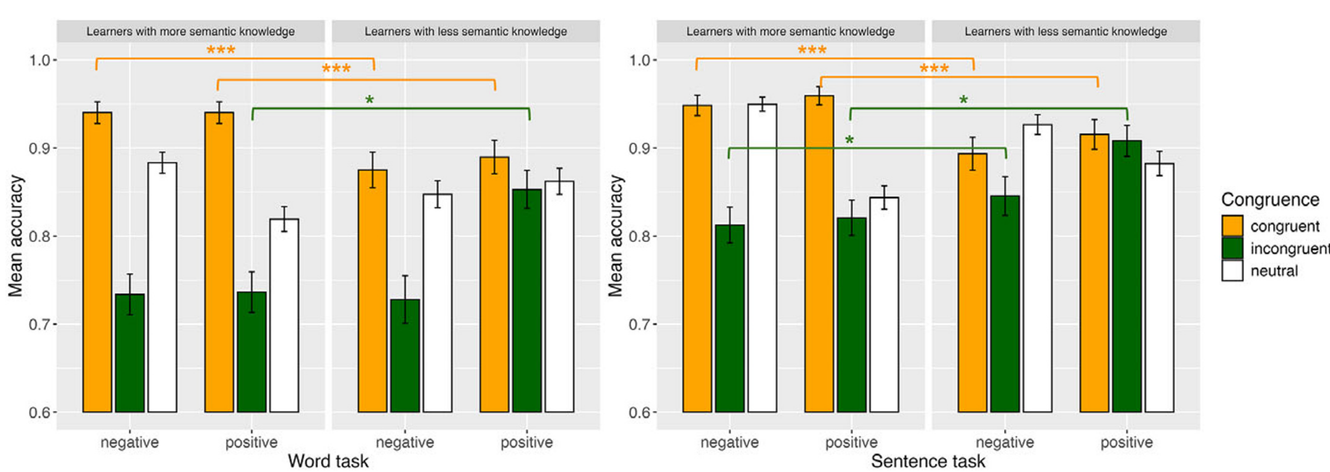

Figure 2. Mean accuracy of emotional prosody judgments in the word task (left) and the sentence task (right). The black vertical lines show the standard errors. Orange brackets indicate significant between-group differences in the congruent condition, and green brackets indicate significant between-group differences in the incongruent condition. ***$p < .001$, **$p < .01$, *$p < .05$.

https://doi.org/10.1017/S1366728926101813

emotional prosody perception more accurately in the congruent condition (β = 1.096, SE = 0.227, z = 4.822, p < .001), and less accurately in the incongruent condition (β = −1.026, SE = 0.175, z = −5.874, p < .001). Furthermore, there were significant interactions between group and congruence such that the effects of congruence (β = −1.117, SE = 0.296, z = −3.781, p < .001) and incongruence (β = 0.611, SE = 0.246, z = 2.484, p < .05) were reduced on the perception of emotional prosody for learners with less semantic knowledge. Additionally, we also found a significant interaction between group and emotion prosody, where learners with less semantic knowledge recognized positive prosody more accurately than those with more semantic knowledge (β = 0.763, SE = 0.217, z = 3.519, p < .001).

4.2.2 RT results

As shown in Figure 3, learners with more semantic knowledge were slower in their recognition of emotional prosody than learners with less semantic knowledge in both the word task (Mean diff = 194 ms) and the sentence task (Mean diff = 186 ms).

Linear mixed-effect models revealed that in the word task (Supplementary Table S3), there was a significant interaction between group and emotional prosody, indicating that learners with less semantic knowledge were quicker in recognizing positive emotional prosody than learners with more semantic knowledge (β = −132.945, SE = 36.806, t = −3.612, p < .001). However, no significant main effects of congruence, group, or their interaction were found on the RTs of emotional prosody perception in Chinese words.

In the sentence task (Supplementary Table S5), both learner groups recognized positive prosody significantly faster than negative prosody (β = −97.159, SE = 23.501, t = −4.134, p < .001) in Chinese sentences. Interestingly, both groups recognized emotional prosody in the congruent condition significantly slower compared to the neutral condition (β = 73.324, SE = 33.575, t = 2.184, p < .05). In addition, there was a significant interaction between group and congruence: learners with less semantic knowledge perceived emotional prosody in the congruent condition significantly faster than those with more semantic knowledge (β = −104.687, SE = 48.593, t = −2.154, p < .05).

We modeled the word and sentence tasks separately for the diffusion model analyses. For each participant, three parameters were estimated: boundary separation, indexing response caution; non-decision time, indexing encoding and motor execution; and drift rate, indexing information processing efficiency. Across both word and sentence tasks, we found no significant group differences in boundary separation (p > 0.19 in both tasks), indicating comparable response caution between learners with more and less semantic knowledge. Furthermore, significant effects were observed for the drift rate. Both learner groups showed a significantly reduced drift rate on the incongruent trials, suggesting less efficient integration of semantic and prosodic cues. Moreover, there were significant interactions between group and congruence: learners with less semantic knowledge exhibited reduced susceptibility to the semantics-prosody incongruence effects compared to learners with more semantic knowledge in the word task (B = 0.284, p = .033) and sentence task (B = 0.313, p = .005). Finally, in the sentence task only, learners with less semantic knowledge showed a significantly faster non-decision time (p = .017), suggesting more rapid sensory encoding and/or motor execution. Collectively, these results indicated that the observed group differences were primarily attributable to their distinct information processing efficiencies, rather than response caution.

5. Discussion

The influence of semantic valence, defined as the stored semantic knowledge about whether an object or event is positive or negative, on the perception of emotional prosody is well-established in native speakers (Itkes et al., 2017; Itkes & Kron, 2019). However, this account of semantic valence assumes a stable and relatively uniform baseline of semantic knowledge among participants, which is insufficient to account for emotion processing in L2 learners. Recent research in L2 processing has yielded divergent findings (Bhatara et al., 2016; Cho & Dewaele, 2021; Shochi et al., 2016; Xiao & Liu, 2025), sparking a debate regarding whether semantic valence facilitates emotional prosody perception by providing multiple channels of information or interferes with it by increasing the processing load. This investigation into the role of semantic valence in emotional prosody perception is particularly relevant in a tonal language such as Mandarin Chinese, where the same acoustic feature

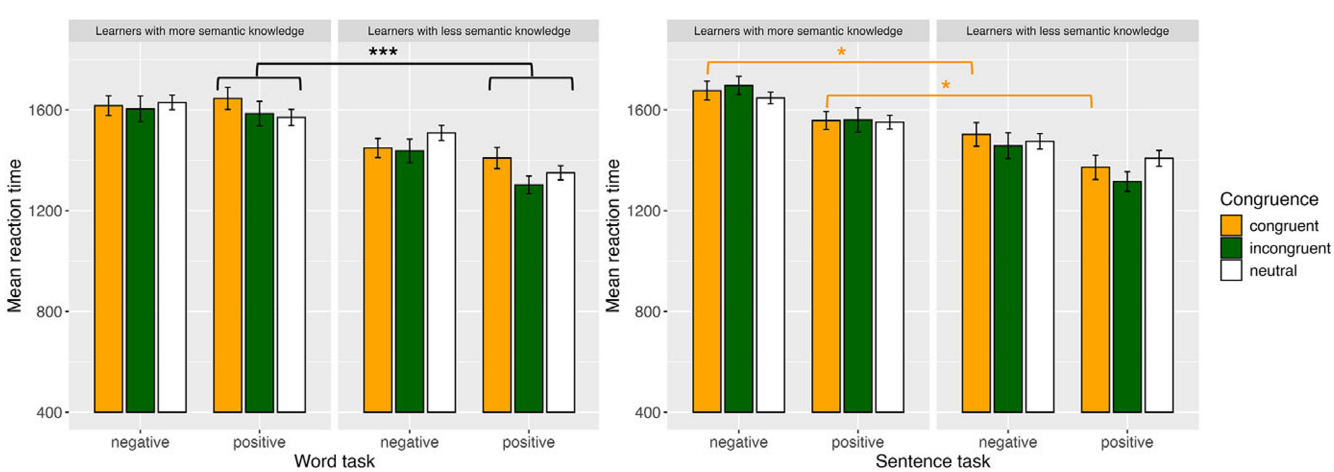

Figure 3. Mean reaction time in milliseconds for emotional prosody judgments in the word task (left) and the sentence task (right). The black vertical lines show the standard errors. Orange brackets indicate significant between-group differences in the congruent condition, and black brackets indicate significant between-group differences in the positive prosody. ***p < .001, **p < .01, *p < .05.

https://doi.org/10.1017/S1366728926101813

(i.e., pitch) encodes both semantic and prosodic information (Xu, 2005).

Therefore, the current study investigated the role of semantic valence in emotional prosody perception in L2 Chinese learners. In doing so, we extend the established account of semantic valence (Itkes et al., 2017; Itkes & Kron, 2019) to the field of second language acquisition and propose that in the L2 domain, semantic valence should be examined in terms of two distinct components: learners' semantic knowledge and the inherent semantic valence of the stimuli.

We first conducted a pilot experiment to examine how learners' semantic knowledge affects their perception of emotional prosody in Chinese words and sentences. Following prior research (Shochi et al., 2016; Xiao & Liu, 2026), we used learning experience of L2 learners as an estimate of semantic knowledge and categorized them into novice and advanced learners. Building on this, the main experiment investigated the independent and interactive effects of both the learners' semantic knowledge and the stimuli's semantic valence on emotional prosody perception in Chinese words and sentences. To improve the measure of semantic knowledge, we trained and tested learners of similar general proficiency on the real words from the target stimulus set and then categorized them as learners with more and less semantic knowledge based on their vocabulary scores (Brysbaert et al., 2017; Ferré et al., 2024; Streubel et al., 2020). We also manipulated the congruence of semantic valence and emotional prosody in the stimuli. Overall, our results from the two experiments indicated that L2 Chinese learners could recognize emotional prosody with high accuracy and rapid response time, regardless of their level of semantic knowledge. The following section discusses the role of semantic valence in emotional prosody perception in L2 Chinese learners.

First, our results underscored the critical role of semantic knowledge in the perception of emotional prosody in L2 contexts. Specifically, in the pilot experiment, novice learners outperformed advanced learners in recognizing positive emotional prosody. This pattern was replicated in the main experiment: learners with less semantic knowledge showed higher accuracy than those with more semantic knowledge in perceiving positive emotional prosody in both word and sentence tasks. These findings converged with prior research evidence that L2 Chinese learners outperformed native Chinese speakers in identifying positive emotional prosody (Xiao & Liu, 2026; Zhu, 2013) and that L2 English learners with lower proficiency outperformed those with higher proficiency in their positive emotional prosody perception (Bhatara et al., 2016). One interpretation is that this advantage for positive prosody in learners with less semantic knowledge reflects an asymmetry in bilingual emotion processing. Negative prosody often functions as a warning signal and tends to be acoustically more distinct and readily recognizable universally, whereas positive prosody is typically more subtle or culturally variable (Laukka & Elfenbein, 2021; Sauter et al., 2010). When prosodic cues are less salient, as in positive emotions, learners with more semantic knowledge may weight semantic valence more heavily, which can bias their judgments if semantic and prosodic cues are incongruent. Learners with less semantic knowledge, by contrast, may rely more consistently on prosodic cues, resulting in a relative advantage in recognizing positive emotional prosody. These findings also provide novel support for the Theory of Constructed Emotion (Barrett, 2006, 2017; Barrett et al., 2007; Lindquist et al., 2006, 2015; Lindquist & Gendron, 2013) in a bilingual context, demonstrating that L2 learners actively draw on their semantic knowledge to interpret emotional prosody, and, importantly, that these interactions vary with specific type of emotional prosody.

We also found that the semantic valence of the stimuli, specifically, the (in)congruence between semantic valence and emotional prosody, significantly influenced the perception of emotional prosody for L2 Chinese learners. Building on previous evidence from native speakers (e.g., Cho & Dewaele, 2021; Lin et al., 2020; Min & Schirmer, 2011; Xiao & Liu, 2025), we explored the semantics-prosody (in)congruence effects on the perception of emotional prosody in learners with more semantic knowledge and learners with less semantic knowledge. We found that both groups showed significant congruence and incongruence effects in perceiving emotional prosody in Chinese words and sentences. Specifically, when semantic valence and emotional prosody were congruent (e.g., semantically positive word rui^{4}zhi^4, *wise*, spoken in a joyful tone), semantic valence facilitated emotional prosody perception. However, when semantic valence and emotional prosody were incongruent (e.g., semantically negative word ao^{4}man^4, *arrogant*, spoken in a joyful tone), semantic valence interfered with emotional prosody perception. These findings indicate robust semantics-prosody (in)congruence effects in L2 learners, regardless of their levels of semantic knowledge.

Moreover, the semantic (in)congruence effects had a greater impact on L2 learners with more semantic knowledge than on those with less semantic knowledge. In the pilot experiment, when the semantic valence of the stimuli was controlled to be neutral, novice learners recognized positive and negative emotional prosody (i.e., semantic-prosody incongruent) even more accurately than advanced learners. We therefore speculated that the semantics-prosody incongruence may hinder emotional prosody perception more for advanced learners than for novice learners, which suggests an interaction between the two components of semantic valence: the learners' semantic knowledge and the semantic valence of the stimuli. To confirm this hypothesis, in the main experiment, we further examined the interactions between the two components of semantic valence. We found that when semantic valence was congruent with emotional prosody, learners with more semantic knowledge received more facilitation from semantic valence than those with less semantic knowledge. However, increased knowledge of linguistic information did not always enhance the paralinguistic perception of emotional prosody: learners with more semantic knowledge also experienced greater interference than those with less semantic knowledge when semantic valence was incongruent with emotional prosody. The findings of (in)congruence effects can be further explained through the three-stage working model of emotional prosody processing (Schirmer & Kotz, 2006). This model posits three stages: an early sensory processing of low-level acoustic properties, an integration of emotionally significant acoustic cues, and a cognition stage involving evaluative judgments and semantic processing. The stronger congruence and incongruence effects observed in learners with more semantic knowledge may suggest their greater engagement of this final cognition stage. Namely, when semantic and prosodic cues are congruent, joint evaluation enhances emotion judgments, resulting in higher accuracy. When cues conflict, integration imposes additional cognitive demands to resolve competing signals, reducing accuracy. This computational cost is also reflected in the RT data: learners with more semantic knowledge are consistently slower in recognizing emotional prosody than those with less semantic knowledge.

How do semantic and prosodic cues interact in bilingual emotion processing? The observed group differences may reflect two possible accounts: a competition account (Lin et al., 2020), where semantics-prosody congruent cues facilitate while incongruent cues interfere with processing, or an integration account (Schirmer & Kotz, 2006; Taguchi, 2008), where L2 learners, depending on their semantic

https://doi.org/10.1017/S1366728926101813

knowledge, attempt to simultaneously integrate multiple cues. While the current study was not designed to adjudicate between these two accounts, we note that semantic information was processed implicitly in our emotion judgment task, as participants were not explicitly instructed to attend to word meaning. Furthermore, our results showed that learners with more semantic knowledge showed slower responses than those with less semantic knowledge regardless of congruence, and they also received greater facilitation and interference from the semantics-prosody (in)congruence effects. These findings may align better with the integration account, where L2 listeners with more semantic knowledge may be more automatic, albeit more effortful, in integrating semantic and prosodic cues during emotion processing than those with less semantic knowledge.

Furthermore, our findings showed that semantic valence influences the ACC and RT of the perception of emotional prosody in distinct ways for L2 learners. In terms of ACC, learners' semantic knowledge can either facilitate or interfere with accurate emotional prosody recognition, depending on the congruence between semantic valence and emotional prosody in the stimuli. Intriguingly, in terms of RT, L2 learners with less semantic knowledge consistently recognized emotional prosody in words and sentences more quickly than those with more semantic knowledge. This divergence between ACC and RT suggests that semantic valence may act as a double-edged sword: enhancing or hindering ACC, while also introducing a processing cost when learners must integrate both linguistic and paralinguistic cues simultaneously. Subsequent diffusion model analyses further confirmed that the group differences were driven by differences in information processing efficiencies, rather than a simple speed-accuracy trade-off.

While learners with less semantic knowledge may bypass or activate semantic cues when judging emotional prosody, those with more semantic knowledge are likely to reach the third stage and thus be more attuned to L2 semantic information, as predicted by the three-stage model (Schirmer & Kotz, 2006). But why do L2 learners with more semantic knowledge show a slower response in their perception of emotional prosody? One explanation for the observed RT slowdown can be attributed to reduced automaticity in L2 semantic processing (Thoma & Baum, 2019). For bilinguals with more developed semantic knowledge, the effortful retrieval of lexical meaning and its integration with prosodic information increases cognitive load (Bylund et al., 2023; Schmidtke, 2014), resulting in slower emotion prosody recognition. However, this is not the only explanation. According to the Revised Hierarchical Model (Kroll et al., 2010; Kroll & Tokowicz, 2001), learners with more semantic knowledge are expected to show slower RTs to access their L2 semantic information, as deeper semantic processing leads to longer response times. This holds true even after a relatively brief L2 exposure (Ferré et al., 2006; Talamas et al., 1999). Another alternative explanation draws on the Bilingual Interactive Activation plus model (Dijkstra & van Heuven, 2002), which proposes that bilingual semantic processing involves the non-selective, parallel activation of linguistic representations. As learners with more semantic knowledge successfully access L2 semantic information alongside prosodic cues, they need to resolve the competition between these parallel streams, resulting in an RT slowdown. Notably, the results of the current study cannot adjudicate which account better explains the semantic processing in L2 emotional prosody perception, and future research manipulating these specific cognitive mechanisms is therefore needed.

In summary, this study reveals significant effects of learners' semantic knowledge, stimuli's semantic valence, as well as their interactions, on the perception of emotional prosody in Chinese words and sentences. Our findings indicate that while the two components of semantic valence are separable, they jointly modulate how L2 learners perceive emotional prosody. More broadly, this research highlights that individual variability in bilingual semantic knowledge is not merely background noise, but an important factor in affective speech perception. By investigating the role of semantic valence in L2 contexts, we provide new empirical evidence from a tonal language on how knowledge-based perception and feeling-based perception can be integrated in bilingual emotion processing.

Finally, several limitations of the present study warrant consideration. First, the use of pseudowords may inadvertently evoke affective impressions for L2 Chinese learners who are still actively mapping tones to meanings. Future research should systematically manipulate lexical tones and emotional prosody in pseudowords to disentangle these effects. Second, we only examined the perception of basic emotional prosody in L1-English L2-Chinese learners, which may limit the generalizability of the findings. Future investigations are needed to explore the role of semantic valence in learners from diverse L1 backgrounds (e.g., tonal and non-tonal languages) as well as in more nuanced emotional prosody types. Third, this study operationalized semantic knowledge using learning experience and a vocabulary test, yielding empirical evidence on the impact of learners' semantic knowledge on emotional prosody perception. Building on these findings, future work may benefit from adopting more granular measures (e.g., vocabulary depth, valence-specific or arousal-specific semantic knowledge) to explore which dimensions of semantic knowledge drive these effects in bilingual emotion processing.

6. Conclusions

This study examined two components of semantic valence: learners' semantic knowledge and the semantic valence of the stimuli, and their interactions in the perception of emotional prosody in a tonal language. Although L2 Chinese learners demonstrated high accuracy and speed in recognizing emotional prosody at word and sentence levels, semantic valence significantly modulated their performance. Specifically, learners' semantic knowledge facilitated emotional prosody perception when the semantic valence and emotional prosody of the stimuli were congruent but interfered with emotional prosody perception under incongruent conditions. These semantics-prosody (in)congruent effects were more pronounced among learners with more semantic knowledge compared to those with less semantic knowledge. Interestingly, greater semantic knowledge was also associated with slower recognition of positive emotional prosody, suggesting a potential processing cost in bilingual emotion processing. The findings highlight that semantic valence influences emotional prosody perception, suggesting a dynamic link between linguistic knowledge and paralinguistic perception in an L2. This study advances our understanding of how bilinguals simultaneously process semantic and prosodic cues in a tonal language, and offers implications for L2 instruction and cross-cultural communication.

Data availability statement. The data that support the findings of this study are openly available through the Open Science Framework (OSF) repository at https://osf.io/fs4py/overview.

Funding statement. The research was supported by NSF grant BCS-2234914 and NIH/NIDCD grant R01DC017162.

Competing interests. The authors declare no competing interests.

References

Alm, C. O., & **Llorà, X.** (2006). Evolving emotional prosody. *Interspeech*, **2006**. https://doi.org/10.21437/interspeech.2006-504

Altrov, R. (2013). Aspects of cultural communication in recognizing emotions. *Trames*, **17**(2), 159–174. https://doi.org/10.3176/tr.2013.2.04

Barr, D. J., **Levy, R.**, **Scheepers, C.**, & **Tily, H. J.** (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, **68**(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Barrett, L. F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review*, **10**(1), 20–46. https://doi.org/10.1207/s15327957pspr1001_2

Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, **12**(1), 1–23. https://doi.org/10.1093/scan/nsw154

Barrett, L. F., **Lindquist, K. A.**, & **Gendron, M.** (2007). Language as context for the perception of emotion. *Trends in Cognitive Sciences*, **11**(8), 327–332. https://doi.org/10.1016/j.tics.2007.06.003

Bates, D., **Mächler, M.**, **Bolker, B.**, & **Walker, S.** (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bergen, B. K. (2004). The psychological reality of phonaesthemes. *Language*, **80**(2), 290–311. https://doi.org/10.1353/lan.2004.0056

Bhatara, A., **Laukka, P.**, **Boll-Avetisyan, N.**, **Granjon, L.**, **Anger Elfenbein, H.**, & **Bänziger, T.** (2016). Second language ability and emotional prosody perception. *PLoS One*, **11**(6), e0156855. https://doi.org/10.1371/journal.pone.0156855

Boersma, P. & **Weenink, D.** (2023). *Praat: doing phonetics by computer [Computer program]*. Version 6.3.10. http://www.praat.org/

Brysbaert, M., **Lagrou, E.**, & **Stevens, M.** (2017). Visual word recognition in a second language: A test of the lexical entrenchment hypothesis with lexical decision times. *Bilingualism: Language and Cognition*, **20**(3), 530–548. https://doi.org/10.1017/S1366728916000353

Bylund, E., **Antfolk, J.**, **Abrahamsson, N.**, **Olstad, A. M. H.**, **Norrman, G.**, & **Lehtonen, M.** (2023). Does bilingualism come with linguistic costs? A meta-analytic review of the bilingual lexical deficit. *Psychonomic Bulletin & Review*, **30**(3), 897–913. http://doi.org/10.3758/s13423-022-02136-7

Chang, H. S., **Lee, C. Y.**, **Wang, X.**, **Young, S. T.**, **Li, C. H.**, & **Chu, W. C.** (2023). Emotional tones of voice affect the acoustics and perception of Mandarin tones. *PLoS One*, **18**(4), e0283635. https://doi.org/10.1371/journal.pone.0283635

Cho, C. M., & **Dewaele, J. M.** (2021). A crosslinguistic study of the perception of emotional intonation. Influence of the pitch modulations. *Studies in Second Language Acquisition*, **43**(4), 870–895. https://doi.org/10.1017/S0272263120000674

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, **28**(3), 369–382. https://doi.org/10.1177/0267658312443651

Cutler, A., **Dahan, D.**, & **Van Donselaar, W.** (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, **40**(2), 141–201. https://doi.org/10.1177/002383099704000203

Dewaele, J. M. (2008). The emotional weight of I love you in multilinguals' languages. *Journal of Pragmatics*, **40**(10), 1753–1780. https://doi.org/10.1016/j.pragma.2008.03.002

Dijkstra, T., & **Van Heuven, W. J.** (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, **5**(3), 175–197. https://doi.org/10.1017/S1366728902003012

Dromey, C., **Silveira, J.**, & **Sandor, P.** (2005). Recognition of affective prosody by speakers of English as a first or foreign language. *Speech Communication*, **47**(3), 351–359. https://doi.org/10.1016/j.specom.2004.09.010

Estes, Z., & **Adelman, J. S.** (2008). Automatic vigilance for negative words is categorical and general. *Emotion*, **8**(4), 453–457. https://doi.org/10.1037/a0012887

Faul, F., **Erdfelder, E.**, **Buchner, A.**, & **Lang, A. G.** (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, **41**(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Ferré, P., **Guasch, M.**, **Stadthagen-González, H.**, **Hinojosa, J. A.**, **Fraga, I.**, **Marín, J.**, & **Pérez-Sánchez, M. Á.** (2024). What makes a word a good representative of the category of "emotion"? The role of feelings and interoception. *Emotion*, **24**(3), 745–758. https://doi.org/10.1037/emo0001300

Ferré, P., **Sánchez-Casas, R.**, & **Guasch, M.** (2006). Can a horse be a donkey? Semantic and form interference effects in translation recognition in early and late proficient and nonproficient Spanish-Catalan bilinguals. *Language Learning*, **56**(4), 571–608. https://doi.org/10.1111/j.1467-9922.2006.00389.x

Gao, C., **Shinkareva, S. V.**, & **Peelen, M. V.** (2022). Affective valence of words differentially affects visual and auditory word recognition. *Journal of Experimental Psychology: General*, **151**(9), 2144–2159. https://doi.org/10.1037/xge0001176

Hauser, D. J., & **Schwarz, N.** (2016). Semantic prosody and judgment. *Journal of Experimental Psychology: General*, **145**(7), 882–896. https://doi.org/10.1037/xge0000178

Hutchins, S. S. (1998). *The psychological reality, variability, and compositionality of English phonesthemes*. Emory University.

Ip, M. H. K., & **Cutler, A.** (2020). Universals of listening: Equivalent prosodic entrainment in tone and non-tone languages. *Cognition*, **202**, 104311. https://doi.org/10.1016/j.cognition.2020.104311

Itkes, O., **Eviatar, Z.**, & **Kron, A.** (2019). Semantic and affective manifestations of ambi (valence). *Cognition and Emotion*, **33**(7), 1356–1369. https://doi.org/10.1080/02699931.2018.1564249

Itkes, O., **Kimchi, R.**, **Haj-Ali, H.**, **Shapiro, A.**, & **Kron, A.** (2017). Dissociating affective and semantic valence. *Journal of Experimental Psychology: General*, **146**(7), 924–942. https://doi.org/10.1037/xge0000291

Itkes, O., & **Kron, A.** (2019). Affective and semantic representations of valence: A conceptual framework. *Emotion Review*, **11**(4), 283–293. https://doi.org/10.1177/1754073919868759

Jackson, J. C., **Watts, J.**, **Henry, T. R.**, **List, J. M.**, **Forkel, R.**, **Mucha, P. J.**, **Greenhill, S. J.**, **Gray, R. D.**, & **Lindquist, K. A.** (2019). Emotion semantics show both cultural variation and universal structure. *Science*, **366**(6472), 1517–1522. https://doi.org/10.1126/science.aaw8160

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, **59**(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Kao, C., & **Zhang, Y.** (2020). Differential neurobehavioral effects of cross-modal selective priming on phonetic and emotional prosodic information in late second language learners. *Journal of Speech, Language, and Hearing Research*, **63**(8), 2508–2521. https://doi.org/10.1044/2020_jslhr-19-00329

Kemmerer, D. (2014). *The cognitive neuroscience of language: An introduction*. Psychology Press.

Keysar, B., **Hayakawa, S. L.**, & **An, S. G.** (2012). The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological Science*, **23**(6), 661–668. https://doi.org/10.1177/0956797611432178

Kim, S. K., & **Sumner, M.** (2017). Beyond lexical meaning: The effect of emotional prosody on spoken word recognition. *The Journal of the Acoustical Society of America*, **142**(1), EL49–EL55. https://doi.org/10.1121/1.4991328

Kotz, S. A., & **Paulmann, S.** (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Research*, **1151**, 107–118. https://doi.org/10.1016/j.brainres.2007.03.015

Kroll, J. F., & **Tokowicz, N.** (2001). The development of conceptual representation for words in a second language. In J. L. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 49–71). Blackwell Publishers.

Kroll, J. F., **Van Hell, J. G.**, **Tokowicz, N.**, & **Green, D. W.** (2010). The revised hierarchical model: A critical review and assessment. *Bilingualism: Language and Cognition*, **13**(3), 373–381. https://doi.org/10.1017/S136672891000009X

Kuperman, V., **Estes, Z.**, **Brysbaert, M.**, & **Warriner, A. B.** (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General*, **143**(3), 1065–1081. https://doi.org/10.1037/a0035669

Laukka, P., & **Elfenbein, H. A.** (2021). Cross-cultural emotion recognition and in-group advantage in vocal expression: A meta-analysis. *Emotion Review*, **13**(1), 3–11. https://doi.org/10.1177/1754073919897295

Lin, Y., **Ding, H.**, & **Zhang, Y.** (2020). Prosody dominates over semantics in emotion word processing: Evidence from cross-channel and cross-modal

Stroop effects. *Journal of Speech, Language, and Hearing Research*, **63**(3), 896–912. https://doi.org/10.1044/2020_JSLHR-19-00258

Lindquist, K. A., **Barrett, L. F.**, **Bliss-Moreau, E.**, & **Russell, J. A.** (2006). Language and the perception of emotion. *Emotion*, **6**(1), 125–138. http://doi.org/10.1037/1528-3542.6.1.125

Lindquist, K. A., & **Gendron, M.** (2013). What's in a word? Language constructs emotion perception. *Emotion Review*, **5**(1), 66–71. https://doi.org/10.1177/1754073912451351

Lindquist, K. A., **Gendron, M.**, **Barrett, L. F.**, & **Dickerson, B. C.** (2014). Emotion perception, but not affect perception, is impaired with semantic memory loss. *Emotion*, **14**(2), 375–387. https://doi.org/10.1037/a0035293

Lindquist, K. A., **Satpute, A. B.**, & **Gendron, M.** (2015). Does language do more than communicate emotion? *Current Directions in Psychological Science*, **24**(2), 99–108. https://doi.org/10.1177/0963721414553440

Min, C. S., & **Schirmer, A.** (2011). Perceiving verbal and vocal emotions in a second language. *Cognition & Emotion*, **25**(8), 1376–1392. https://doi.org/10.1080/02699931.2010.544865

Myers, C. E., **Interian, A.**, & **Moustafa, A. A.** (2022). A practical introduction to using the drift diffusion model of decision-making in cognitive psychology, neuroscience, and health sciences. *Frontiers in Psychology*, **13**, 1039172. https://doi.org/10.3389/fpsyg.2022.1039172

Nygaard, L. C., & **Tzeng, C. Y.** (2021). Perceptual integration of linguistic and non-linguistic properties of speech. In *The handbook of speech perception* (pp. 398–427). Wiley. https://doi.org/10.1002/9781119184096.ch15

Olteanu, L., **Salama, K.**, **Kimchi, R.**, & **Kron, A.** (2023). A divergent effect of stimulus perceptual details on affective and semantic representations of valence. *Emotion*, **23**(5), 1236–1253. https://doi.org/10.1037/emo0001148

Ouyang, I. C., & **Kaiser, E.** (2015). Prosody and information structure in a tone language: An investigation of mandarin Chinese. *Language, Cognition and Neuroscience*, **30**(1–2), 57–72. https://doi.org/10.1080/01690965.2013.805795

Pavlenko, A. (2008). Emotion and emotion-laden words in the bilingual lexicon. *Bilingualism: Language and Cognition*, **11**(2), 147–164. https://doi.org/10.1017/S1366728908003283

Pavlenko, A. (2012). Affective processing in bilingual speakers: Disembodied cognition? *International Journal of Psychology*, **47**(6), 405–428. https://doi.org/10.1080/00207594.2012.743665

Pell, M. D., **Jaywant, A.**, **Monetta, L.**, & **Kotz, S. A.** (2011). Emotional speech processing: Disentangling the effects of prosody and semantic cues. *Cognition & Emotion*, **25**(5), 834–853. https://doi.org/10.1080/02699931.2010.516915

Pell, M. D., & **Kotz, S. A.** (2011). On the time course of vocal emotion recognition. *PLoS One*, **6**(11), e27256. https://doi.org/10.1371/journal.pone.0027256

Perlovsky, L. (2013). Language and cognition—Joint acquisition, dual hierarchy, and emotional prosody. *Frontiers in Behavioral Neuroscience*, **7**, 123. https://doi.org/10.3389/fnbeh.2013.00123

R Core Team (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ross, E. D., **Edmondson, J. A.**, & **Seibert, G. B.** (1986). The effect of affect on various acoustic measures of prosody in tone and non-tone languages: A comparison based on computer analysis of voice. *Journal of Phonetics*, **14**(2), 283–302. https://doi.org/10.1016/S0095-4470(19)30669-2

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, **39**(6), 1161–1178. https://doi.org/10.1037/h0077714

Ryu C., **Mandarin Tone Perception & Production Team, and Michigan State University Libraries**. (2022). Tone Perfect: Multimodal Database for Mandarin Chinese. Accessed 1 January 2022. https://tone.lib.msu.edu/

Sauter, D. A., **Eisner, F.**, **Ekman, P.**, & **Scott, S. K.** (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences*, **107**(6), 2408–2412. https://doi.org/10.1073/pnas.0908239106

Schirmer, A., & **Kotz, S. A.** (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, **10**(1), 24–30. https://doi.org/10.1016/j.tics.2005.11.009

Schmidtke, J. (2014). Second language experience modulates word retrieval effort in bilinguals: Evidence from pupillometry. *Frontiers in Psychology*, **5**, 137. https://doi.org/10.3389/fpsyg.2014.00137

Seydell-Greenwald, A., **Chambers, C. E.**, **Ferrara, K.**, & **Newport, E. L.** (2020). What you say versus how you say it: Comparing sentence comprehension and emotional prosody processing using fMRI. *NeuroImage*, **209**, 116509. https://doi.org/10.1016/j.neuroimage.2019.116509

Shen, J. (1985). Beijinghua shengdiao de yinyu he yudiao [Pitch range of tone and intonation in Beijing dialect]. In T. Lin & L. Wang (Eds.), *BeijingYuyin Shiyanlu* (pp. 73–130). Beijing University Press. (in Chinese)

Shochi, T., **Brousse, A.**, **Guerry, M.**, **Erickson, D.**, & **Rilliard, A.** (2016). Learning effect of social affective prosody in Japanese by French learners. *Speech Prosody 2016*, https://doi.org/10.21437/speechprosody.2016-199

Streubel, B., **Gunzenhauser, C.**, **Grosse, G.**, & **Saalbach, H.** (2020). Emotion-specific vocabulary and its contribution to emotion understanding in 4- to 9-year-old children. *Journal of Experimental Child Psychology*, **193**, 104790. https://doi.org/10.1016/j.jecp.2019.104790

Sun, C. C., **Hendrix, P.**, **Ma, J.**, & **Baayen, R. H.** (2018). Chinese lexical database (CLD). *Behavior Research Methods*, **50**(6), 2606–2629. https://doi.org/10.3758/s13428-018-1038-3

Taguchi, N. (2008). The role of learning environment in the development of pragmatic comprehension: A comparison of gains between EFL and ESL learners. *Studies in Second Language Acquisition*, **30**(4), 423–452. https://doi.org/10.1017/S0272263108080716

Talamas, A., **Kroll, J. F.**, & **Dufour, R.** (1999). From form to meaning: Stages in the acquisition of second-language vocabulary. *Bilingualism: Language and Cognition*, **2**(1), 45–58. https://doi.org/10.1017/S1366728999000140

Thoma, D., & **Baum, A.** (2019). Reduced language processing automaticity induces weaker emotions in bilinguals regardless of learning context. *Emotion*, **19**(6), 1023–1034. https://doi.org/10.1037/emo0000502

Wang, T., **Lee, Y. C.**, & **Ma, Q.** (2018). Within and across-language comparison of vocal emotions in mandarin and English. *Applied Sciences*, **8**(12), 2629. https://doi.org/10.3390/app8122629

Wei, H., **He, Y.**, **Kauschke, C.**, **Scharinger, M.**, & **Domahs, U.** (2022). An EEG-study on L2 categorization of emotional prosody in German. *Speech Prosody 2022*, 629–633. https://doi.org/10.21437/SpeechProsody.2022-128

Wilson, D., & **Wharton, T.** (2006). Relevance and prosody. *Journal of Pragmatics*, **38**(10), 1559–1579. https://doi.org/10.1016/j.pragma.2005.04.012

Xiao, C. (2025). The Processing of Emotional Prosody and Semantics in Native and Non-NativeChinese Speakers. (Doctoral dissertation). Retrieved from https://scholarcommons.sc.edu/etd/8628

Xiao, C., & **Liu, J.** (2025). Semantic effects on the perception of emotional prosody in native and non-native Chinese speakers. *Cognition and Emotion*, **39**(7), 1542–1552. https://doi.org/10.1080/02699931.2024.2371088

Xiao, C., & **Liu, J.** (2026). The perception of emotional prosody in Mandarin Chinese words and sentences. *Second Language Research*, **42**(1), 3–30. https://doi.org/10.1177/02676583241286748

Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication*, **46**(3–4), 220–251. https://doi.org/10.1016/j.specom.2005.02.014

Yap, M. J., & **Seow, C. S.** (2014). The influence of emotion on lexical processing: Insights from RT distributional analysis. *Psychonomic Bulletin & Review*, **21**(2), 526–533. http://doi.org/10.3758/s13423-013-0525-x

Yip, M. (2002). *Tone*. Cambridge University Press.

Zhu, Y. (2013). Which is the best listener group?: Perception of Chinese emotional prosody by Chinese natives, naïve Dutch listeners and Dutch L2 learners of Chinese. *Dutch Journal of Applied Linguistics*, **2**(2), 170–183. https://doi.org/10.1075/dujal.2.2.03zhu