This is an Accepted Manuscript of an article published by Taylor & Francis in *Cognition and Emotion* on July 7, 2024, available at: https://doi.org/10.1080/02699931.2024.2371088

Semantic effects on the perception of emotional prosody in native and non-native Chinese speakers

Cheng Xiao[1], Jiang Liu[1,2]

Linguistics Program, University of South Carolina, Columbia, SC, USA

Department of Language, Literatures and Cultures, University of South Carolina, Columbia, SC, USA

Correspondence concerning this article should be addressed to Cheng Xiao, Linguistics Program, University of South Carolina, Columbia, SC, USA, 29208. Email: cxiao@email.sc.edu

Semantic effects on the perception of emotional prosody in native and non-native Chinese speakers

ABSTRACT

While previous research has found an in-group advantage (IGA) favouring native speakers in emotional prosody perception over non-native speakers, the effects of semantics on emotional prosody perception remain unclear. This study investigated the effects of semantics on emotional prosody perception in Chinese words and sentences for native and non-native Chinese speakers. The critical manipulation was the congruence of prosodic (positive, negative) and semantic (positive, negative, and neutral) valence. Participants listened to a series of audio clips and judged whether the emotional prosody was positive or negative for each utterance. The results revealed an IGA effect: native speakers perceived emotional prosody more accurately and quickly than non-native speakers in Chinese words and sentences. Furthermore, a semantic congruence effect was observed in Chinese words, where both native and non-native speakers recognized emotional prosody more accurately in the semantic-prosody congruent condition than in the incongruent condition. However, in Chinese sentences, this congruence effect was only present for non-native speakers. Additionally, the IGA effect and semantic congruence effect on emotional prosody perception were influenced by prosody valence. These findings illuminate the role of semantics in emotional prosody perception, highlighting perceptual differences between native and non-native Chinese speakers.

Keywords: Emotional prosody; semantics; emotion perception; in-group advantage; Mandarin Chinese

1. Introduction

Emotions, positive and negative, expressed through language, are a critical part of communication in daily life. In spoken language, emotional information is primarily conveyed via two channels: semantics and emotional prosody (Kotz & Paulmann, 2007; Lin et al., 2020; Pell et al., 2011). Semantics is the meaning of words and sentences. Emotional prosody refers to the ways in which the tone of voice can be modulated to convey positive or negative emotions, feelings, and attitudes (Kemmerer, 2014). Previous research on emotional prosody perception has corroborated Elfenbein and Ambady's (2002) in-group advantage (IGA) hypothesis: emotional prosody is more accurately perceived when expressed by members of their own cultural group, although emotional prosody can be recognized at better-than-chance level universally for both native and non-native speakers (e.g., Cowen et al., 2019; Chronaki et al., 2018; Laukka & Elfenbein, 2021). A crucial aspect that remains unexplored is the influence of semantics on emotional prosody perception. For instance, when listening to an utterance spoken with a *happy* emotional prosody, it is likely that the perception of emotional prosody in a positive semantic sentence "*she is wise*" would be different from the perception in a negative semantic sentence "*she is arrogant.*" It is unclear whether such semantic effects differ between native and non-native speakers on their perception of emotional prosody.

Furthermore, few studies have considered whether the IGA hypothesis, originally found in non-tonal languages (e.g., English), can be generalized to tonal languages (e.g., Chinese). In addition to encoding emotional prosody, the tone of voice in Chinese is also used to differentiate lexical meaning (Xu, 2005; Yip, 2002). Chinese has four lexical tones (indicated by superscripted numbers), which contain semantic information. For example, ma^1 means *mother*, ma^2 means *hemp*, ma^3 means *horse*, and ma^4 means *to scold.* The coexistence of lexical tone and emotional prosody (Shen, 1990) suggests an interplay between linguistic cues (semantics) and paralinguistic cues (emotional prosody) in Chinese. Therefore, the current study investigates the effects of semantics on emotional prosody perception for native and non-native speakers in Chinese words and sentences.

1.1 Semantic knowledge and semantic information

Semantics exerts its influence on emotional prosody perception in two ways: the semantic knowledge of participants and the semantic information of stimuli. Previous findings comparing native speakers with non-native speakers without semantics knowledge (i.e., naïve listeners) were inconsistent with those comparing native speakers with non-native speakers with semantics knowledge (i.e., second language (L2) learners), revealing that the semantic knowledge of participants significantly affected emotional prosody perception.

Some studies found that native speakers outperform non-native speakers without semantic knowledge in recognizing emotional prosody (Chronaki et al., 2018; Cowen et al., 2019; Paulmann & Uskul, 2014), which supported the IGA hypothesis. However, other studies indicated that non-native speakers with semantic knowledge (L2 learners) recognized emotional prosody equally well as native speakers (Dromey et al., 2005; Paone & Frontera, 2019), and surprisingly in tonal languages, non-native speakers with semantic knowledge (i.e., L2 Chinese learners) even better performed than native Chinese speakers (Zhu, 2013), thus challenging the IGA hypothesis. These findings suggested that native speakers' in-group advantage may be attenuated or offset by the effect of semantic knowledge of non-native speakers. Furthermore, how the semantic knowledge of non-native speakers affects their perception of emotional prosody in their second languages also remains elusive. While some studies found non-native speakers with semantic knowledge outperformed those without L2 semantic knowledge in the perception of emotional prosody in their L2 (Shochi et al., 2016; Zhu, 2013), others found that L2 English learners with higher English proficiency were less accurate in recognizing positive emotional prosody in their L2 English compared to those with lower English proficiency (Bhatara et al., 2016).

Moreover, semantic information of stimuli plays a role in emotional prosody perception. Firstly, the absence or presence of semantic information in stimuli (pseudo-word vs. real-word) influences emotional prosody perception. Castro and Lima (2010) found that native speakers recognized emotional prosody in pseudo-word sentences more quickly compared to real-word sentences in Portuguese. Furthermore, the semantic information of stimuli modulates the IGA effects in a tonal language as Chinese. When perceiving emotional prosody in pseudo-word sentences, Paulmann and Uskul (2014) found that native Chinese speakers outperformed non-native Chinese speakers in the perception of Chinese emotional prosody. However, when perceiving emotional prosody in real-word sentences, a contrasting result emerged from Zhu's (2013) study which found that non-native Chinese speakers recognized Chinese emotional prosody on par with or even better than native speakers. Additionally, Pell and Kotz (2011) discovered that stimuli length (i.e., amount of semantic information) influences the perception of emotional prosody such that native speakers recognized emotional prosody in English with greater accuracy as they heard more of an utterance.

Collectively, these findings highlight the impact of semantic knowledge and semantic information on emotional prosody perception. However, there is a lack of research that has systematically controlled for semantic knowledge in both native and non-native speakers. Few studies have examined whether the IGA hypothesis holds true for both real-word and pseudo-word Chinese stimuli. Importantly, few studies have explored the interaction between semantic effects and emotional prosody perception in Chinese. Therefore, the present study aims to investigate the emotional prosody perception of native and non-native speakers with comparable semantic knowledge in both real-word and pseudo-word Chinese stimuli.

1.2 Semantic valence and congruence effect

When people perceive the emotional prosody in words and sentences, their perceptions are also affected by the semantic valence of those words and sentences (Barrett et al., 2007; Itkes & Kron, 2019). The semantic valence of words (e.g., *wise* has a positive semantic valence, and *arrogant* has a negative semantic valence) plays a significant role in emotion perception (e.g., Jackson et al., 2019).

Some researchers found a semantic-prosody congruence effect that native English speakers recognized emotional prosody more accurately in a congruent condition (e.g., "*she is wise*" spoken in a positive prosody) than in an incongruent condition, and this congruence effect can be modulated by prosody valence (Cho & Dewaele, 2021; Min & Schirmer, 2011; Pell et al., 2011). Recently, a similar semantic-prosody congruence effect was observed in Chinese for native Chinese speakers (Lin et al., 2020). These studies have shown that the congruence of semantic and prosodic valence can facilitate native speakers' perception of emotional prosody.

For non-native speakers with semantic knowledge (i.e., L2 learners), the simultaneous processing of semantics and emotional prosody can be challenging, as speech signals need to be integrated within a fraction of a second (Kao & Zhang, 2020). A few studies have explored the congruence effect on emotional prosody perception with L2 English learners, but the findings are mixed. While Min and Schirmer (2011) observed a semantic-prosody congruence effect in L1-Chinese L2-English learners, Cho and Dewaele (2021) found no congruence effect in L1-Korean L2-English learners.

Additionally, it is important to note that prosody valence, on its own, can also influence the perception of emotional prosody. Converging evidence revealed that negative emotional prosody is generally perceived more accurately than positive emotional prosody (e.g., Liu & Pell, 2012; Pell et al., 2009; Sauter et al., 2010). The effect of prosody valence interacts with the IGA effect on emotional prosody perception, as suggested in Laukka and Elfenbein (2021) which found that native speakers showed a greater in-group advantage in identifying positive emotional prosody compared to negative emotional prosody.

Taken together, semantic valence and prosody valence can influence the perception of emotional prosody for both native speakers and non-native speakers. Previous studies have found that native English and native Chinese speakers recognize emotional prosody more accurately in semantic-prosody congruent conditions, but interestingly, such congruence effect was not observed in L2 English learners. However, it remains unclear if the semantic congruence effect exists in L2 Chinese learners' emotional prosody perception, and if the effect of semantic congruence differs between native and non-native Chinese speakers, and how this congruence effect interacts with prosodic valence and the IGA effect in tonal languages. Therefore, it is crucial to investigate the semantic congruence effect and its interaction with prosody valence in Chinese emotional prosody perception.

2. The current study

The current study investigates the effects of semantics and its interaction with prosody valence[1] on emotional prosody perception for native speakers and non-native speakers in Chinese words and sentences within the framework of the IGA hypothesis (Elfenbein & Ambady, 2002). In this

---

[1] The terms "positive" and "negative" prosody valence in this study were used to refer *happy* and *sad* emotional prosody, respectively.

study, the non-native group are native English speakers who have semantic knowledge in L2 Chinese, namely, L1-English L2-Chinese learners. To explore semantic effects, we control for the semantic knowledge of participants, ensuring that both native and non-native Chinese speakers have similar prior semantic knowledge of the stimuli. We include both real-word and pseudo-word words and sentences as stimuli: real-word stimuli carry positive or negative semantic valence, allowing examination of the semantic congruence effect and its interaction with prosody valence; pseudo-word stimuli, with neutral semantic valence, serve as a semantic baseline considering neither native nor non-native speakers have the semantic knowledge of these novel stimuli. We also divide the stimuli into two tasks to control for semantic information. Task 1 investigates semantic effects on emotional prosody perception in Chinese words, and task 2 examines in Chinese sentences.

We propose four hypotheses to examine the interplay between semantics and emotional prosody in Chinese emotional prosody perception. First, we predict an in-group advantage (IGA) effect (Elfenbein and Ambady, 2002), where native speakers will perceive emotional prosody more accurately and quickly than L2 learners. Second, we hypothesize a semantic congruence effect where native speakers will perceive emotional prosody more accurately and quickly when the semantic and prosodic valence are congruent compared to when they are incongruent (Lin et al., 2020), and we expect differences in the congruence effect between native and non-native speakers (Cho & Dewaele, 2021). Third, we anticipate the effect of prosody valence will interact with IGA effect and semantic congruence effect (Laukka & Elfenbein, 2021; Min & Schirmer, 2011). Finally, we predict a similar IGA effect and semantic congruence effect in both tasks, with an overall improved perception of emotional prosody in the sentence task due to longer duration of the stimuli (Pell & Kotz, 2011).

3. Method

3.1 Participants

Based on a closely related study (Lin et al., 2020), a total of 33 participants were included in the analysis: 17 native Chinese speakers (native group, 10 males; mean age = 26.71; SD of age = 4.20), 16 L1-English L2-Chinese learners (learner group, 11 males; mean age = 19.63; SD of age = 0.10). The non-native speakers were native English speakers enrolled in an intermediate level Chinese course[2], and none were heritage speakers of Mandarin Chinese or any other tonal language. To ensure comparable semantic knowledge across native and non-native groups, we conducted a pre-experiment vocabulary screening test on the real-word stimuli. Only L2 Chinese learners scoring above 75% (group mean accuracy = 92%, indicating good semantic knowledge) were included. Neither group had prior semantic knowledge of the novel pseudo-word stimuli. All participants had normal or corrected-to-normal vision and reported no hearing problems. They received course credit or monetary compensation for their participation. All aspects of the study were approved by the Institutional Review Board of the University of South Carolina.

3.2 Stimuli

There were 128 disyllabic Chinese words (Task 1) and 128 Chinese sentences (Task 2). The critical manipulation of the stimuli was the valence of emotional prosody (positive, negative) and semantics (positive, negative, and neutral). A male native Chinese speaker recorded the stimuli with both positive (i.e., happy) and negative (i.e., sad) emotional prosody (Cho & Dewaele, 2021; Lin et al., 2020; Pell et al., 2011).[3] The stimuli included real words (i.e., disyllables) and

[2] L2 Chinese learners at this instructional level were chosen to ensure they have acquired the basic knowledge of syllable-tone combinations and sentence structures in Chinese.

[3] Supplemental Figures 1 and 2 show the density plots of emotional prosody ratings for words and sentences, respectively. Both native and non-native speaker groups rated the intended positive prosody stimuli positively, and the negative stimuli negatively.

sentences with positive and negative semantic valence, as well as pseudo-words and sentences with neutral semantic valence (see example stimuli in Supplemental Table 1). The semantic congruence of the stimuli was further manipulated: "congruent" for the real-word stimuli with matched semantic and prosody valence; "incongruent" for the real-word stimuli with mismatched semantic and prosody valence; and "neutral" for the pseudo-word stimuli with neutral semantic valence.

In Task 1, we matched the real words with positive (e.g., rui^{4}zhi^4, *wise*) and negative (e.g., ao^{4}man^4, *arrogant*) semantic valence in frequency and syllable length (Sun et al., 2018). We created the novel pseudo-word stimuli (e.g., chun^{4}pou^4, no meaning) using the lexical gaps (Ryu et al., 2019). A lexical gap occurs when a syllable can only be combined with specific tones but not all four tones. For example, the syllable "chun" can be combined with tone 1, tone 2, and tone 3, but not tone 4, rendering chun^4 a lexical gap in Chinese. In Task 2, the sentence stimuli were constructed by embedding the same word stimuli from Task 1 into a simple carrier sentence "ta^1 hen^3 (*She is very*) _____," minimizing syntax parsing and contextual information for L2 learners (Cho & Dewaele, 2021). Additionally, we ensured the distribution of four lexical tone categories was consistent across the six stimuli conditions (2 emotional prosody x 3 semantics) in both tasks. A summary of relevant acoustic parameters of the stimuli can be found in Supplemental Table 2.

3.3 Procedure

All participants came to the language laboratory to participate in a four-task study in the following order: an emotion judgment task for words (Task 1), an emotion judgment task for sentences (Task 2), an untimed emotion perception questionnaire, and an untimed emotion rating task. In total, the study took approximately 60 minutes. In each emotion judgment task, participants wore headphones and performed the task on a computer with instructions given on the screen. Participants listened to a series of audio clips, one at a time, and then judged whether the emotional prosody for each utterance was positive or negative as accurately and quickly as possible. Auditory stimuli from six conditions were randomly presented, and the duration of each trial was the same (5000ms). A new trial started immediately after the participant made a response. Participants' accuracy and reaction time (RT) were recorded for each trial. After the emotion judgment task, all participants answered an online emotion perception questionnaire to provide language background information, and they also completed an untimed emotional rating task online where they listened to each utterance and then rated the valence of emotional prosody on a 7-point Likert scale from -3 *very negative* to 3 *very positive*.

3.4 Analyses

Accuracy was binary coded: "1" for correctly judgment of the emotional prosody and "0" for incorrect judgments. We excluded RTs of incorrect judgments or those greater than three standard deviations from the participant's mean, resulting in 7667 valid trials. Additionally, we also performed a log transformation on RTs to address the positive skewness in perceptual experiments (Baayen & Milin, 2010). We derived the ratings from each participant's responses for each item on a 7-point Likert scale[4].

We used logistic mixed-effects models (Jaeger, 2008) to analyse the accuracy, and utilized linear mixed-effects models (Bates et al., 2015) to analyse RT. The full model for accuracy and RT analyses are the same in Task 1 and 2. Each model included three fixed factors: (1) group with two levels (native group, learner group), (2) semantic congruence with three levels (congruent, incongruent, and neutral), and (3) emotion prosody valence with two levels

[4] The 7-point Likert scale used in this study: -3 = *very negative*, -2 = *negative*, -1 = *slightly negative*, 0 = *neither negative nor positive*, 1 = *slightly positive*, 2 = *positive*, 3 = *very positive*.

(positive, and negative). The native group, neutral semantic congruence, and positive prosody valence were treated as the reference levels, respectively. The random factors were the individual ratings of each subject for each item collected from the emotion rating task. We utilized a maximal mixed-effect model that considered all potential main effects and interactions (Barr et al., 2013). When there was a significant main effect or interaction, Tukey's post hoc tests were performed using the *emmeans* package (Lenth, 2020). All analyses were performed in R (R Core Team, 2022).

4. Results

4.1 Emotional prosody perception in Chinese words

As shown in Figure 1, both the learner group (mean accuracy (SD) = 0.850 (0.357); mean RT (SD) = 1.585 (0.700)s) and the native group (mean accuracy(SD) = 0.936(0.245); mean RT (SD) = 1.300 (0.530)s) recognized emotional prosody in Chinese words with a high accuracy and fast speed.

A) Accuracy B) Reaction Time

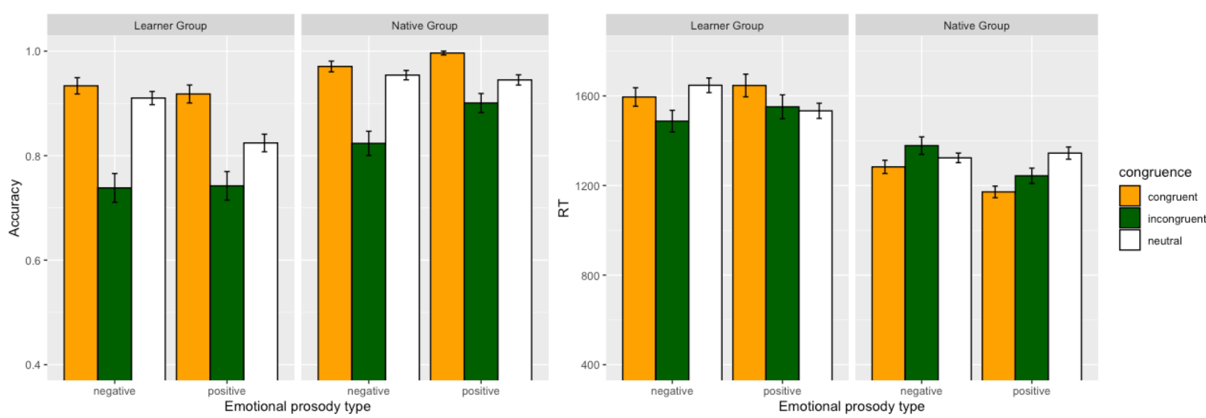


Figure 1. Task 1 mean (A) accuracy and (B) reaction time (in milliseconds) of the native and learner groups' emotional prosody perception in the semantic congruent (orange bars), incongruent (green bars), and neutral (white bars) conditions in words. Error bars represent 95% confidence intervals.

Mixed-effects regression models (Supplemental Tables 3-4) revealed a main effect of group such that the learner group recognized emotional prosody less accurately ($\beta = -1.171$, $SE = 0.224$, $z = -5.234$, $p < 0.001$) and more slowly ($\beta = 0.123$, $SE = 0.025$, $t = 5.005$, $p < 0.001$) compared to the native group; a main effect of semantic congruence, where both groups perceived emotional prosody more accurately ($\beta = 2.798$, $SE = 1.011$, $z = 2.767$, $p < 0.01$) and more quickly ($\beta = -0.107$, $SE = 0.028$, $t = -3.866$, $p < 0.001$) in the semantic congruent condition than in the neutral condition, and less accurately in the semantic incongruent condition compared to the neutral condition ($\beta = -0.597$, $SE = 0.277$, $z = -2.158$, $p < 0.05$). We also found that negative emotional prosody was recognized slower than the positive emotional prosody ($\beta = 0.131$, $SE = 0.041$, $t = 3.213$, $p < 0.01$) in Chinese words.

Moreover, there were significant two-way interactions on RTs between group and semantic congruence ($\beta = 0.168$, $SE = 0.041$, $t = 4.103$, $p < 0.001$), and between semantic congruence and prosody valence ($\beta = 0.084$, $SE = 0.039$, $t = 2.139$, $p < 0.05$). We further subset the semantic neutral condition (pseudo-words) and found that native Chinese speakers demonstrated comparable accuracy rates and RTs for recognizing positive and negative emotional prosody. But interestingly, L2 Chinese learners exhibited distinct patterns of perception for positive and negative prosody in pseudo-words: their accuracy was higher for negative emotional prosody perception, while RT was quicker for positive emotional prosody perception (Supplemental Figure 3). Post-hoc analyses indicated that the effect of prosody valence was only significant for L2 learners, but not for native speakers (Supplemental Table 5).

Additionally, a significant three-way interaction of the group, semantic congruence, and prosody valence on RTs ($\beta = -0.152$, $SE = 0.058$, $t = -2.644$, $p < 0.01$) was observed in Chinese words. When focusing on positive emotional prosody, a striking contrast emerged: native

Chinese speakers recognized the emotional prosody in real words, whether semantic congruent or incongruent, faster than in pseudo-words. In contrast, L2 Chinese learners recognized pseudo-words faster than real words, even in the semantic congruent condition (Supplemental Figure 4).

4.2 Emotional prosody perception in Chinese sentences

As shown in Figure 2, both the learner group (mean accuracy (SD) = 0.897 (0.303); mean RT (SD) = 1.602 (0.635) s) and the native group (mean accuracy (SD) = 0.988 (0.109); mean RT (SD) = 1.297 (0.558) s) recognized emotional prosody at a very high accuracy and fast speed in the perception of emotional prosody in Chinese sentences.

As predicted, compared to the word task, both groups were more accurate in perceiving emotional prosody in the sentence task (Figure 2A). The native group had an increase in their accuracy from 93.6% in words to 98.8% in sentences, and the learner group showed an increase from 85% to 89.7%. Notably, the native group showed a ceiling effect in the accuracy of their emotional prosody perception regardless of semantic congruence. However, the RTs for both groups were similar across words and sentences.

A) Accuracy B) Reaction Time

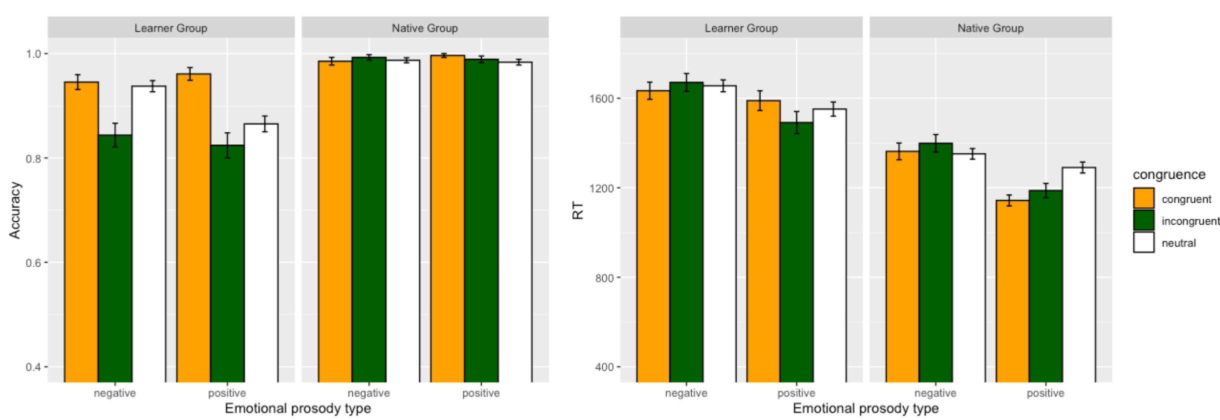


Figure 2. Task 2 mean (A) accuracy and (B) reaction time (in milliseconds) of the native and learner groups' emotional prosody perception in the semantic congruent (orange bars), incongruent (green bars), and neutral (white bars) conditions in sentences. Error bars represent 95% confidence intervals.

In the sentence task, in terms of accuracy, there was only a significant main effect of group (Supplemental Table 6): the learner group recognized emotional prosody in Chinese sentences less accurately compared to the native group ($\beta = -1.961$, $SE = 0.362$, $z = -5.420$, $p < 0.001$). No other main effects or interactions were found. Interestingly, post-hoc analyses revealed a difference between the native group and the learner group: although no significant effect was observed in the native group, significant main effects of semantic congruence ($\beta = 0.991$, $SE = 0.354$, $z = 2.799$, $p < 0.01$), prosody valence ($\beta = 1.116$, $SE = 0.292$, $z = 3.823$, $p < 0.001$), along with their interactions ($\beta = -1.144$, $SE = 0.488$, $z = -2.343$, $p < 0.05$) were observed in the learner group (Supplemental Table 7).

In terms of RT (Supplemental Table 8), there were a main effect of group, where the learner group recognized emotional prosody more slowly than the native group ($\beta = 0.160$, $SE = 0.026$, $t = 6.037$, $p < 0.001$); a main effect of semantic congruence, with the congruent condition recognized faster than in the neutral condition ($\beta = -0.072$, $SE = 0.031$, $t = -2.344$, $p < 0.05$); significant two-way interactions between the group and semantic congruence ($\beta = 0.131$, $SE = 0.045$, $t = 2.930$, $p < 0.01$), and between semantic congruence and prosody valence ($\beta = 0.092$, $SE = 0.043$, $t = 2.126$, $p < 0.05$) ; and a significant three-way interaction of group, semantic congruence, and prosody valence ($\beta = -0.145$, $SE = 0.063$, $t = -2.310$, $p < 0.05$).

As in word task, the effect of prosody valence in semantic baseline (i.e., pseudo-word sentences) was examined in the sentence task: native Chinese speakers demonstrated comparable accuracy rates and RTs for positive and negative prosody, whereas L2 learners showed a higher accuracy but slower RT for negative prosody compared to positive prosody (Supplemental Figure 3). The significant effect of prosody valence was observed for the perception of emotional prosody in L2 learners only (Supplemental Table 5). Furthermore, in perceiving positive prosody in sentences, native speakers recognized real-word sentences faster than pseudo-word sentences, whereas L2 learners recognized pseudo-word sentences more quickly than real-word sentences (Supplemental Figure 4), which mirrored the patterns in the Chinese words.

5. Discussion

In the present study, we investigated the effects of semantics and its interaction with prosody valence on emotional prosody perception in Chinese words (Task 1) and sentences (Task 2) for both native and non-native (i.e., L2 learners) speakers. [5] Participants were instructed to judge whether the emotional prosody of the utterance was positive or negative and then rated the valence of emotional prosody for each stimulus using a 7-point Likert scale. To ensure the reliability of the semantic effects, we controlled participants' semantic knowledge such that both groups knew the semantic meaning of the real words, and neither group knew the semantics of the pseudo-words. Additional controls were implemented on lexical tone, word frequency, and syllable length of the stimuli. The key findings are discussed below.

Firstly, in both tasks, the in-group advantage (IGA) effect was observed: native Chinese speakers demonstrated significantly higher accuracy and faster recognition of emotional prosody compared to non-native speakers (i.e., L2 Chinese learners), although both groups recognized emotional prosody well above chance levels. The findings provided clear evidence in support of Elfenbein and Ambady's (2002) IGA hypothesis in a tonal language (i.e., Chinese). Specifically, when neither group (native and non-native) had semantic knowledge of the pseudo- words and sentences, native Chinese speakers recognized emotional prosody with a higher accuracy and

[5] Previous studies have reported significant effects of acoustic parameters on emotional prosody perception (e.g., Kao et al., 2021). Interested readers can find the additional acoustic analyses of the current study in the Supplemental Tables 9-10.

faster reaction times compared to non-native speakers, which aligns with previous research (e.g., Cowen et al., 2019; Paulmann & Uskul, 2014). Our results extend the examinations of IGA effect in pseudo-word stimuli from comparisons between native speakers and non-native speakers without semantic knowledge to the comparisons between native speakers and non-native speakers with semantic knowledge.

We also found evidence in favor of the IGA effect in real-word stimuli. When both groups had semantic knowledge of real Chinese words and sentences, our results showed native Chinese speakers still recognized emotional prosody more accurately and faster than non-native speakers. This contradicts with Zhu's (2013) finding on L2 Chinese learners outperforming native Chinese speakers. The discrepancy between our findings and Zhu (2013) highlights the pivotal influence of semantics in Chinese emotional prosody perception (Lin et al., 2020). Notably, it is crucial to control the semantic knowledge when conducting cross-linguistic comparisons of emotional prosody perception in a tonal language. Non-native speakers without semantic knowledge may be more sensitive to the paralinguistic aspect of tone of voice, while native speakers, possessing semantic knowledge, may encounter interference from the linguistic aspect of tone of voice. The asymmetry of semantic knowledge between native speakers and non-native speakers interacts with the IGA effect on emotional prosody perception in Chinese. Our study provides insights to the understanding of emotion prosody perception in tonal languages and offers robust evidence supporting the IGA hypothesis in Chinese words and sentences.

Secondly, as predicted, semantics plays a significant role in the perception of emotional prosody in Chinese words and sentences. We found significant effects of semantic congruence on emotional prosody perception for both native and non-native speakers. In the word task, both native and non-native Chinese speakers were more accurate at recognizing emotional prosody when the valences of semantics and prosody were congruent, compared to the incongruent condition. In the sentence task, the semantic congruence effect was only observed in non-native Chinese speakers. Native Chinese speakers showed high accuracy in their emotional prosody perception regardless of semantic congruence.

The differences in semantic congruence effect between native and non-native speakers could further be explained by the accessibility model (Robinson & Clore, 2002), which proposed that people preferentially use the specific sources of emotional information at hand to judge emotions, and when these sources are inaccessible, they then access general semantic knowledge about emotions. In our word task, when the emotional information from prosodic channel was vague in a very short stimulus (disyllables), both native and non-native speakers needed to access semantic knowledge to make judgments, thus showing a semantic congruence effect. But in the sentence task, differences between native and non-native speakers emerged: when specific emotional information (emotional prosody) was clear enough for native speakers to make judgments without accessing semantic valence, they showed no significant semantic congruence effect, consistent with the prosody dominance effect found in Lin et al. (2020). However, for non-native speakers, the emotional prosody in a four-syllable short sentence may not be sufficient for them to make judgments in their second language, leading them to still access the semantic knowledge and thus, still showing a significant congruence effect. Our findings suggest that non-native speakers may be more attuned to the effect of semantic valence and its interaction with prosody valence compared to native speakers.

Meanwhile, native and non-native speakers' emotional prosody perception also differed in the semantic neutral condition. Across two tasks, native speakers showed similar accuracy and reaction time in perceiving positive and negative prosody, whereas non-native speakers perceived negative prosody with a higher accuracy but a slower speed than positive prosody. When perceiving emotional prosody in pseudo-words and pseudo-word sentences, prosody valence significantly influences emotional prosody perception for non-native speakers, but not for native speakers. This suggests that native and non-native speakers may have different semantic baselines of their emotional prosody perception.

Thirdly, we found significant three-way interactions of the group, semantic congruence, and prosody valence on reaction times of emotional prosody perception in both tasks. Following the previous debate (Bhatara et al., 2016; Shochi et al., 2016), we further examined whether semantics facilitates or interferes with the perception of native and non-native speakers. Comparing the perception in real-word and pseudo-word stimuli, native speakers recognized positive prosody in real-word stimuli (even in an incongruent condition) faster than in pseudo-word stimuli, suggesting a semantic facilitation. However, non-native speakers were slower to identify emotional prosody in real-word stimuli (even in a congruent condition) compared to pseudo-word stimuli, indicating a semantic interference.

Such divergence of semantic effects (i.e., facilitation for native and interference for non-native) on the reaction times can be attributed to their processing preferences regarding real-word and pseudo-word stimuli. The pseudo-word stimuli were novel and were created using lexical gaps, potentially activating a larger pool of phonological neighbours for native speakers, but not for non-native speakers. This likely requires more effort for native speakers to exclude competitors associated with the novel pseudo-words (Ziegler et al., 2003), and thus the native group showed a faster perception of real-word stimuli compared to pseudo-word stimuli. However, due to the reduced processing automaticity of non-native speakers (Thoma & Baum, 2019), retrieving the semantics of real-word stimuli in their second languages can be more demanding, and they thus showed a faster identification of pseudo-word stimuli compared to real-word stimuli.

Our study was not without limitations. First, our investigation was highly controlled due to the manipulation of semantic-prosody congruence and only focused on the perception of Chinese emotional prosody in two categorical emotions: happy (positive), and sad (negative). It would be interesting to explore the semantic congruence and IGA effects across a broader range of emotions and in more naturalistic settings. Moreover, the current study found that native speakers showed a ceiling effect of semantic congruence in the sentence task, suggesting a need for future researchers to utilize more refined measures to examine the nuances of emotional prosody perception among native speakers in Chinese sentences. Furthermore, while our samples of intermediate-level L2 learners and focus on Chinese may limit the generalizability of the findings, this study contributes to the broader landscape of cross-cultural research on emotional prosody perception. To better understand the gradient nature of semantic knowledge, future research needs to investigate emotional prosody perception in L2 learners with varying levels of proficiency or adopt a longitudinal study design. In addition, the effects of semantic facilitation or interference on reaction times were only observed in positive emotional prosody, but not in negative ones. Future research is needed to address this question.

6. Conclusion

This study examined the effects of semantics and its interaction with prosody valence on Chinese emotional prosody perception for both native and non-native Chinese speakers. Native Chinese speakers consistently outperformed non-native Chinese speakers in perceiving emotional prosody perception in both real- and pseudo- Chinese words and sentences, supporting the IGA hypothesis. Moreover, the semantic congruence effect was observed for both native and non-native Chinese speakers in Chinese words, whereas such congruence effect was only found for non-native speakers in Chinese sentences. Furthermore, in terms of semantic baseline, native Chinese speakers showed no difference in perceiving positive emotional prosody and negative emotional prosody, while non-native speakers exhibited distinct patterns of perceiving positive and negative emotional prosody. Additionally, the effects of semantics on emotional prosody perception diverged in positive emotional prosody perception, with the facilitation effect in native speakers and the interference effect in non-native speakers. In summary, this study investigated the interplay of linguistic and paralinguistic information on Chinese emotional prosody perception, highlighting differences in native and non-native speakers from a cross-cultural perspective.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the University of South Carolina SPARC Graduate Research Grant (no. 120700-22-59283) awarded to CX.

Data availability statement

The data that support the findings of this study are available from the corresponding author, CX, upon reasonable request.

Supplemental material

The supplemental material for this study is openly available via Figshare at https://doi.org/10.6084/m9.figshare.25992550.

References

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International journal of psychological research, 3*(2), 12-28. https://doi.org/10.21500/20112084.807

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., 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*, 1-48. https://doi.org/10.18637/jss.v067.i01

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

Castro, S. L., & Lima, C. F. (2010). Recognizing emotions in spoken language: A validated set of Portuguese sentences and pseudosentences for research on emotional prosody. *Behavior Research Methods, 42*(1), 74-81. https://doi.org/10.3758/BRM.42.1.74

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

Chronaki, G., Wigelsworth, M., Pell, M. D., & Kotz, S. A. (2018). The development of cross-cultural recognition of vocal emotion during childhood and adolescence. *Scientific reports, 8*(1), 1-17. https://doi.org/10.1038/s41598-018-26889-1

Cowen, A. S., Laukka, P., Elfenbein, H. A., Liu, R., & Keltner, D. (2019). The primacy of categories in the recognition of 12 emotions in speech prosody across two cultures. *Nature human behaviour*, *3*(4), 369-382. https://doi.org/10.1038/s41562-019-0533-6

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

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological bulletin*, *128*(2), 203-235. https://doi.org/10.1037/0033-2909.128.2.203

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., ... & 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., Sera, M. D., & Zhang, Y. (2022). Emotional speech processing in 3-to 12-month-old infants: Influences of emotion categories and acoustic parameters. *Journal of Speech, Language, and Hearing Research, 65*(2), 487-500. https://doi.org/10.1044/2021_JSLHR-21-00234

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. https://doi.org/10.4324/9781315764061

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

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

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2020). emmeans: Estimated marginal means, aka least-squares means (R package, Version 1.5.3) [Computer software].

Liu, P., & Pell, M. D. (2012). Recognizing vocal emotions in Mandarin Chinese: A validated database of Chinese vocal emotional stimuli. *Behavior research methods, 44*(4), 1042-1051. https://doi.org/10.3758/s13428-012-0203-3

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

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

Paone, E., & Frontera, M. (2019). Emotional prosody perception in Italian as a second language. *ExLing 2019, 25*, 161. https://doi.org/10.36505/ExLing-2019/10/0040/000402

Paulmann, S., & Uskul, A. K. (2014). Cross-cultural emotional prosody recognition: Evidence from Chinese and British listeners. *Cognition & emotion*, *28*(2), 230-244. https://doi.org/10.1080/02699931.2013.812033

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

Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior*, *33*(2), 107-120. https://doi.org/10.1007/s10919-008-0065-7

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: evidence for an accessibility model of emotional self-report. *Psychological bulletin, 128*(6), 934. https://psycnet.apa.org/doi/10.1037/0033-2909.128.6.934

Ryu C., Mandarin Tone Perception & Production Team, and Michigan State University Libraries. Tone Perfect: Multimodal Database for Mandarin Chinese. Accessed 1 January 2019. 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

Shen, X. S. (1990). Tonal coarticulation in Mandarin. *Journal of Phonetics, 18*(2), 281-295. https://doi.org/10.1016/S0095-4470(19)30394-8

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

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

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

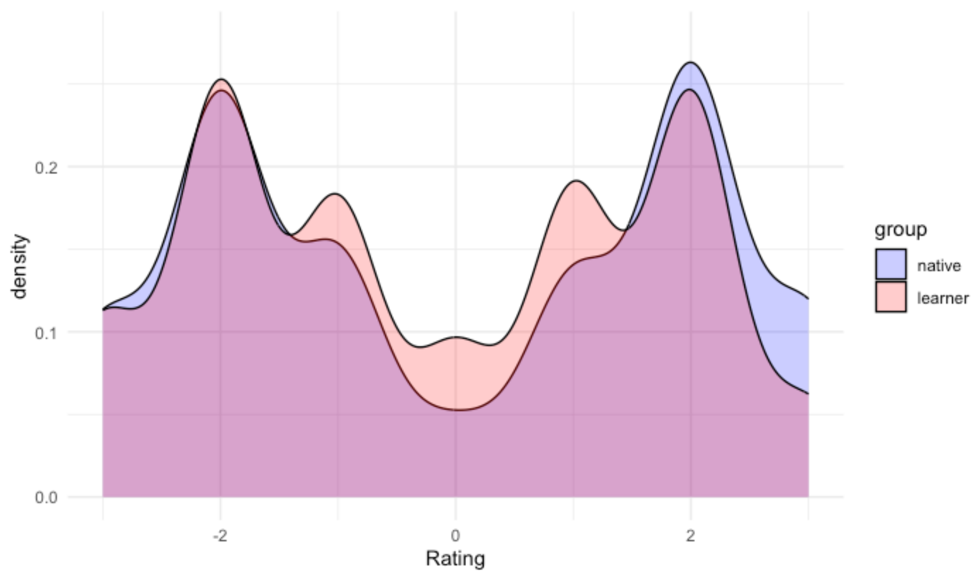
Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication, 46*(3–4), 220–251. https://doi.org/10.1016/j.specom.2005.02.014

Yip, M. (2002). *Tone*. Cambridge University Press. https://doi.org/10.1017/CBO9781139164559

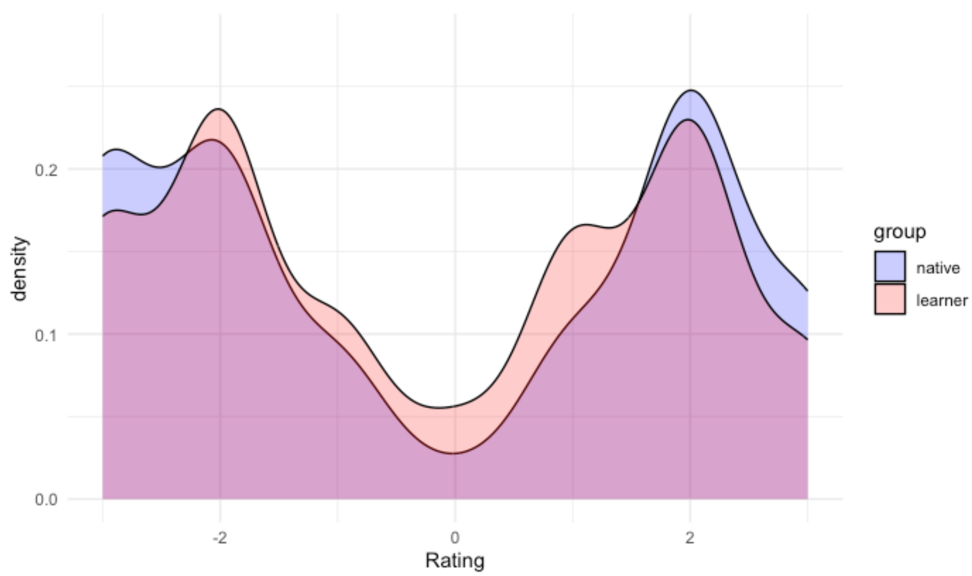
Zhu, Y. (2013). Which is the best listener group?: Perception of Chinese emotional prosody by Chinese natives, naïve Dutch listeners and Dutch L2 learners of Chinese. *Dutch Journal of Applied Linguistics, 2*(2), 170-183. https://doi.org/10.1075/dujal.2.2.03zhu

Ziegler, J. C., Muneaux, M., & Grainger, J. (2003). Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory and Language, 48*(4), 779-793. https://doi.org/10.1016/S0749-596X(03)00006-8

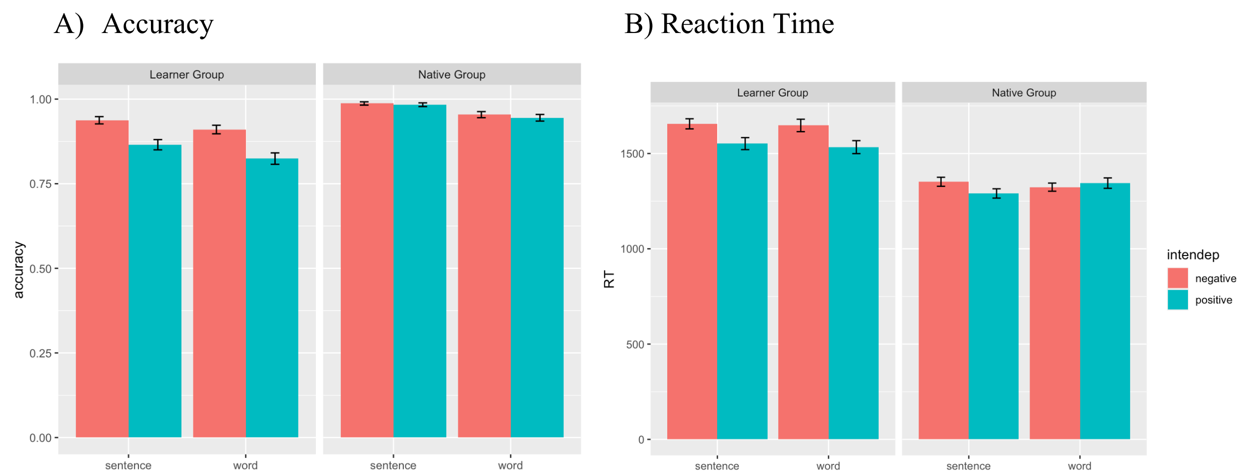
This document contains supplemental material to accompany the manuscript *Semantic effects on the perception of emotional prosody in native and non-native Chinese speakers*
by Cheng Xiao and Jiang Liu



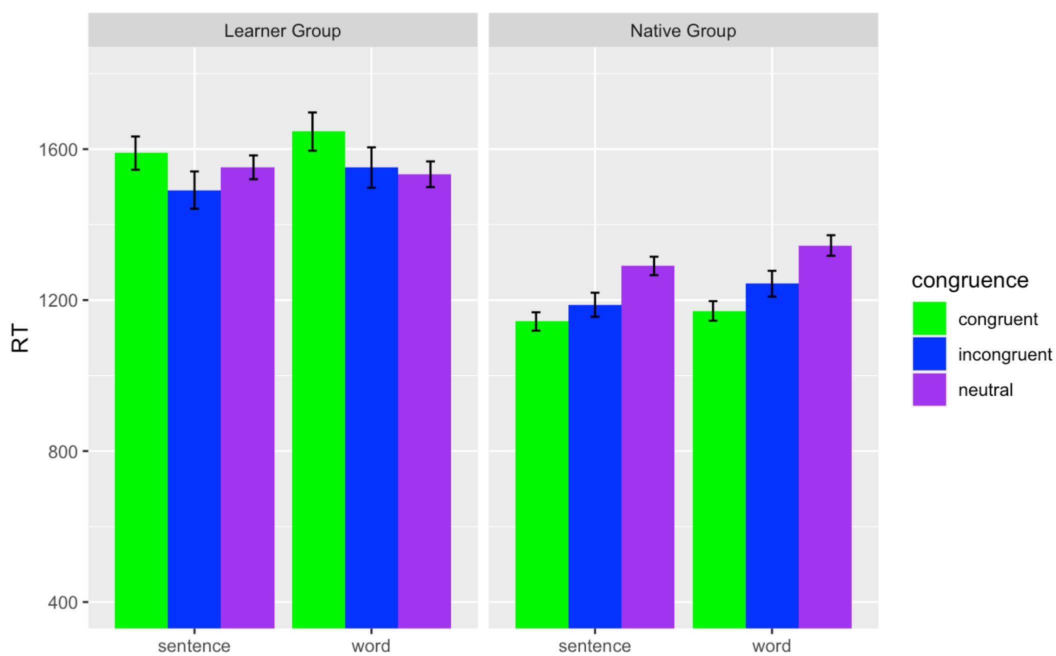
Supplemental Figure 1. The density plot of emotional prosody ratings in Chinese words on a 7-point Likert scale (-3 = *very negative*, -2 = *negative*, -1 = *slightly negative*, 0 = *neither negative nor positive*, 1 = *slightly positive*, 2 = *positive*, 3 = *very positive*).



Supplemental Figure 2. The density plot of emotional prosody ratings in Chinese sentences on a 7-point Likert scale (-3 = *very negative*, -2 = *negative*, -1 = *slightly negative*, 0 = *neither negative nor positive*, 1 = *slightly positive*, 2 = *positive*, 3 = *very positive*).



Supplemental Figure 3. Mean (A) accuracy and (B) reaction time (in milliseconds) of emotional prosody perception in the semantic baseline conditions (i.e., pseudowords and pseudoword sentences) of the native and learner groups for Chinese words and sentences. “intendep” indicates the intended positive emotional prosody (teal bars) and intended negative emotional prosody (red bars). Error bars represent 95% confidence intervals.



Supplemental Figure 4. Mean reaction time (in milliseconds) for positive emotional prosody perception in the native and learner groups for Chinese words and sentences across three semantic congruence conditions: congruent (green bars), incongruent (blue bars), and neutral (purple bars). Error bars represent 95% confidence intervals.

Supplemental Table 1. Examples of word stimuli (Task 1) with (1a) positive, (1b) negative, and (1c) neutral semantic valence, and sentence stimuli (Task 2) with (2a) positive, (2b) negative, and (2c) neutral semantic valence.

	Positive semantics	Negative semantics	Neutral semantics
Word task	(1a) rui^{4}zhi^4 wise 'wise'	(1b) ao^4 man^4 arrogant 'arrogant'	(1c) chun^{4}pou^4 chunpou 'chunpou
Sentence task	(2a) ta^1 hen^3 rui^{4}zhi^4 She very wise 'She is very wise.'	(2b) ta^1 hen^3 ao^{4}man^4 She very arrogant 'She is very arrogant.'	(2c) ta^1 hen^3 chun^4 pou^4 She very chunpou 'She is very chunpou.'

Supplemental Table 2. The means and standard deviations (in parentheses) of five acoustic parameters of positive and negative emotional prosody stimuli.

Emotion	**Task**	**Pitch (Hz)**	**Intensity (dB)**	**Duration (ms)**	**HNR (dB)**	**Spectral Centroid (Hz)**
Positive (happy)	Word	251.49 (50.88)	75.34 (2.26)	708.00 (54.20)	11.50 (3.31)	788.46 (195.80)
	Sentence	247.43 (32.22)	75.19 (1.63)	960.91 (49.63)	10.28 (1.86)	812.32 (113.14)
Negative (sad)	Word	203.40 (46.56)	74.99 (2.42)	924.15 (91.10)	14.29 (3.61)	714.62 (294.94)
	Sentence	205.06 (37.23)	76.94 (1.51)	1308.18(74.72)	13.32 (2.57)	678.44 (124.73)

Supplemental Table 3. Logistic mixed-effects model results for accuracy of emotional prosody perception for the native and learner group in Chinese words.

Fixed effects	***β***	**SE**	**z**	***p***	
(Intercept)	2.856	0.304	9.395	< 2e-16	***
Group: learner	-1.171	0.224	-5.234	0.000	***
Congruence: congruent	2.798	1.011	2.767	0.006	**
Congruence: incongruent	-0.597	0.277	-2.158	0.031	*
Prosody: negative	0.114	0.328	0.348	0.728	
learner x congruent	-1.909	1.043	-1.831	0.067	
learner x incongruent	0.105	0.334	0.315	0.753	
learner x negative	0.514	0.341	1.506	0.132	
congruent x negative	-2.421	1.092	-2.217	0.027	*
incongruent x negative	-0.673	0.382	-1.764	0.078	
learner x congruent x negative	1.776	1.160	1.531	0.126	
learner x incongruent x negative	-0.067	0.475	-0.141	0.887	
Random effect	**σ^2**	**SD**			
Rating	0.035	0.578			

Note. Model formula: *glmer* (accuracy ~ group * congruence * prosody + (1|rating), family = binomial, control = glmerControl (optimizer = "bobyqa")). ***p < .001 **p < .01. *p < .05.

Supplemental Table 4. Linear mixed-effects model results for reaction times (in seconds) of emotional prosody perception for the native and learner group in Chinese words.

Fixed effects	***β***	**SE**	**df**	**t**	***p***	
(Intercept)	0.168	0.045	8.635	3.708	0.005	**
Group: learner	0.123	0.025	3697.973	5.005	0.000	***
Congruence: congruent	-0.107	0.028	3691.714	-3.866	0.000	***
Congruence: incongruent	-0.069	0.029	3692.449	-2.396	0.017	*
Prosody: negative	0.131	0.041	338.935	3.213	0.001	**
learner x congruent	0.168	0.041	3691.812	4.103	0.000	***
learner x incongruent	0.074	0.043	3693.661	1.713	0.087	
learner x negative	0.063	0.034	3697.975	1.859	0.063	
congruent x negative	0.084	0.039	3692.306	2.139	0.032	*
incongruent x negative	0.048	0.042	3695.208	1.150	0.250	
learner x congruent x negative	-0.152	0.058	3691.667	-2.644	0.008	**
learner x incongruent x negative	-0.167	0.062	3693.745	-2.716	0.007	**
Random effects	σ^2	**SD**				
Rating	0.011	0.103				
residual	0.134	0.366				

Note. Model formula: *lmer* (log (RT/1000) ~group*congruence*prosody+(1|rating)). ***p < .001 **p < .01. *p < .05.

Supplemental Table 5. Effects of prosody valence on accuracy and reaction times (in seconds) of the emotional prosody perception in Chinese pseudo-words and pseudo-sentences.

		Accuracy				***RT***			
		β	**SE**	**z**	***p***	***β***	**SE**	**t**	***p***
Native Group	Word Task	0.192	0.278	0.691	0.49	0.013	0.022	0.566	0.571
	Sentence Task	0.001	0.4	0.003	0.998	0.048	0.026	1.868	0.062
Learner Group	Word Task	1.071	0.310	3.457	**0.001**	0.082	0.027	3.021	**0.003**
	Sentence Task	0.719	0.237	3.033	**0.002**	0.087	0.028	3.259	**0.001**

Supplemental Table 6. Logistic mixed-effects model results for accuracy of emotional prosody perception for the native and learner group in Chinese sentences.

Fixed effects	*β*	SE	z	*p*	
(Intercept)	3.822	0.486	7.871	0.000	***
Group: learner	-1.961	0.362	-5.420	0.000	***
Congruence: congruent	1.350	1.044	1.293	0.196	
Congruence: incongruent	0.297	0.667	0.446	0.656	
Prosody: negative	0.666	0.543	1.226	0.220	
learner x congruent	-0.318	1.100	-0.289	0.773	
learner x incongruent	-0.739	0.702	-1.053	0.292	
learner x negative	0.445	0.555	0.801	0.423	
congruent x negative	-1.661	1.216	-1.367	0.172	
incongruent x negative	0.387	1.041	0.372	0.710	
learner x congruent x negative	0.558	1.306	0.427	0.669	
learner x incongruent x negative	-1.045	1.093	-0.956	0.339	
Random effect	σ^2	**SD**			
Rating	0.796	0.893			

Note. Model formula: *glmer* (accuracy ~ group * congruence * prosody + (1|rating), family = binomial, control = glmerControl (optimizer = "bobyqa")). ***p < .001 **p < .01. *p < .05.

Supplemental Table 7. Logistic mixed-effect model results for accuracy of emotional prosody perception for the learner group in Chinese sentences.

Fixed effects	*β*	SE	z	*p*	
(Intercept)	2.030	0.481	4.218	0.000	***
Congruence: congruent	0.991	0.354	2.799	0.005	**
Congruence: incongruent	-0.470	0.221	-2.124	0.034	*
Prosody: negative	1.116	0.292	3.823	0.000	***
congruent x negative	-1.144	0.488	-2.343	0.019	*
incongruent x negative	-0.687	0.341	-2.014	0.044	*
Random effects	σ^2	**SD**			
Rating	1.398	1.182			

Note. Model formula: *glmer* (accuracy ~ congruence * prosody + (1|rating), family = binomial, control = glmerControl (optimizer = "bobyqa")). ***p < .001 **p < .01. *p < .05.

Supplemental Table 8. Linear mixed-effects model results for reaction times (in seconds) of emotional prosody perception for the native and learner group in Chinese sentences.

Fixed effects	*β*	SE	df	t	*p*	
(Intercept)	0.190	0.055	8.081	3.427	0.009	**
Group: learner	0.160	0.026	3941.245	6.037	0.000	***
Congruence: congruent	-0.072	0.031	3943.116	-2.344	0.019	*
Congruence: incongruent	-0.064	0.031	3939.366	-2.082	0.037	*
Prosody: negative	0.076	0.041	959.917	1.849	0.065	
learner x congruent	0.131	0.045	3939.458	2.930	0.003	**

learner x incongruent	0.034	0.046	3939.547	0.731	0.465	
learner x negative	0.044	0.037	3941.518	1.185	0.236	
congruent x negative	0.092	0.043	3941.657	2.126	0.034	*
incongruent x negative	0.068	0.043	3939.693	1.562	0.118	
learner x congruent x negative	-0.145	0.063	3939.203	-2.310	0.021	*
learner x incongruent x negative	-0.014	0.065	3939.765	-0.215	0.830	
Random effects	σ^2	**SD**				
Rating	0.017	0.131				
residual	0.166	0.407				

Note. Model formula: *lmer* (log (RT/1000) ~group*congruence*prosody+(1|rating)). ***p < .001 **p < .01. *p < .05.

Supplemental Table 9. Effects of five acoustic features on positive and negative emotional prosody in word and sentence task.

Fixed	Word task				Sentence task			
	positive prosody		negative prosody		positive prosody		negative prosody	
	accuracy	RT	accuracy	RT	accuracy	RT	accuracy	RT
pitch	N.S.	N.S.	**$p < 0.01^{+}$**	**$p < 0.001^{-}$**	N.S.	N.S.	N.S.	**$p < 0.5^{-}$**
intensity	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.
duration	**$p < 0.05^{+}$**	**$p < 0.001^{+}$**	**$p < 0.001^{+}$**	**$p < 0.001^{+}$**	N.S.	**$p < 0.01^{+}$**	N.S.	**$p < 0.05^{+}$**
HNR	**$p < 0.001^{+}$**	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.
spectral centroid	N.S.	**$p < 0.01^{-}$**	N.S.	N.S.	**$p < 0.05^{+}$**	N.S.	N.S.	N.S.

Note. The superscripted plus and minus signs indicate positive and negative estimated coefficients respectively.

Supplemental Table 10. The interaction between semantic congruence effect with five acoustic parameters in word and sentence task.

Fixed effects	Word task		Sentence task	
	accuracy	RT	accuracy	RT
congruence	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.05$
pitch	N.S.	N.S.	N.S.	N.S.
intensity	$p < 0.05$	$p < 0.05$	N.S.	N.S.
duration	N.S.	$p < 0.001$	N.S.	$p < 0.001$
HNR	N.S.	N.S.	$p < 0.05$	$p < 0.01$
spectral centroid	N.S.	N.S.	N.S.	N.S.
pitch*congruence	N.S.	N.S.	N.S.	N.S.
intensity*congruence	**$p < 0.05$**	**$p < 0.01$**	N.S.	N.S.
duration*congruence	N.S.	N.S.	**$p < 0.01$**	N.S.
HNR*congruences	N.S.	**$p < 0.01$**	N.S.	**$p < 0.05$**
spectral centroid*congruence	N.S.	N.S.	N.S.	**$p < 0.05$**