Original Article

The perception of emotional prosody in Mandarin Chinese words and sentences

Cheng Xiao [iD] and Jiang Liu
University of South Carolina, USA

Abstract
Emotional prosody refers to the ways in which the tone of voice can be modulated to convey emotions, feelings, and attitudes. Previous studies have explored the perception of emotional prosody and whether native speakers (L1) have an in-group advantage in recognizing the emotional prosody of their own cultural groups over non-native speakers. However, little is known about whether these findings in non-tonal languages can be generalized to tonal languages. Mandarin Chinese uses the tone of voice to encode word meanings in addition to emotional prosody. This study investigates the perception of emotional prosody in Mandarin Chinese using an emotion judgment task, focusing on the effects of emotion type (e.g., neutral, joy, anger, sadness) and syllable length (e.g., monosyllable, disyllable, trisyllable, and sentence). Three groups were included, consisting of 20 native Chinese speakers (native group), 20 L1-English L2-Chinese learners (second language group), and 20 native English speakers without Chinese learning experience (non-native group). The results revealed that all three groups can identify emotional prosody well above the chance level in Mandarin Chinese words and sentences. Moreover, the native group and the second language (L2) group showed an in-group advantage in recognizing emotional prosody compared to the non-native group, highlighting the impact of linguistic experience in addition to cultural backgrounds on the perception of emotional prosody. Notably, the effects of emotion type and syllable length differed across the three groups in terms of their perception of emotional prosody. The native group had difficulty identifying positive emotional prosody, whereas both the L2 group and the non-native group showed a pattern of improved accuracy as syllable length increased, with an interaction effect with emotion type.

Keywords
emotional prosody, in-group advantage, L2 Chinese learners, Mandarin Chinese, cross-cultural communication

Corresponding author:
Cheng Xiao, Linguistics Program, University of South Carolina, 616 Welsh Humanities Building, 1620 College Street, Columbia, SC 29208, USA
Email: cxiao@email.sc.edu

I Introduction

Emotion, such as joy, sadness, or anger, is a critical aspect of communication in our daily life (Cutler et al., 1997; Wilson and Wharton, 2006). A long-standing debate is whether the perception of emotions is universal or culturally specific (e.g., Brooks et al., 2019; Ekman et al., 1969; Elfenbein and Ambady, 2003; Gendron et al., 2018; Jack et al., 2012; Matsumoto, 1988). In early research, psychologists focused on facial expressions of emotions (e.g., Ekman and Friesen, 1986; Russell, 1994). Russell and Barrett (1999) defined prototypical emotional episodes, which include happiness, sadness, disgust, anger, fear, and surprise. Since then, a growing body of research has begun to investigate differences in emotion recognition in a cross-cultural context. Elfenbein and Ambady (2002a) conducted a meta-analysis of 97 cross-cultural studies, and proposed the in-group advantage (IGA) hypothesis: emotions can be more accurately perceived when expressed by members of one's own cultural group, while emotions are recognized at a better-than-chance level universally. However, humans communicate emotions not only through facial expressions but also through verbal expressions, such as emotional prosody.

Emotional prosody refers to the ways in which the tone of voice can be modulated to convey emotions, feelings, and attitudes (Kemmerer, 2014). In the field of emotional prosody perception, previous research has been categorized into three types of comparisons (Paulmann and Uskul, 2014). The first type involves listeners from different cultural groups judging emotional prosody expressed by speakers from a single cultural group (e.g., Scherer et al., 2001). The second type involves listeners from a single cultural group judging emotional prosody expressed by speakers from different cultural groups (e.g., Chronaki et al., 2018; Pell et al., 2009). The third type involves listeners from different cultural groups judging emotional prosody expressed by speakers from different cultural groups (e.g., Paulmann and Uskul, 2014). In these studies, the critical manipulation is the cultural backgrounds of both emotional prosody expressors (i.e., speakers) and perceivers (i.e., listeners). When speakers and listeners belong to the same cultural group, the listeners are typically considered native speakers; otherwise, they are non-native speakers.

Previous researchers often utilized an emotion judgment task to compare the emotional prosody perception between native and non-native speakers. They asked voice actors to portray the stimuli with various types of emotional prosody so that the intended emotional prosody type for each stimulus was known to the experimenters (e.g., Beier and Zautra, 1972; Chronaki et al., 2018; Pell et al., 2009; Scherer et al., 2001; Van Bezooijen et al., 1983). In the emotion judgment task, participants listened to the auditory stimuli and judged the intended emotion for each utterance in a forced-choice identification question where a list of predefined response alternatives was given (e.g., neutral, joy, anger, sadness), and their accuracy rates in recognizing emotional prosody were then measured. The results from previous studies revealed that native speakers showed an advantage compared to non-native speakers, although both groups were capable of recognizing emotional prosody (Elfenbein, 2013; Juslin and Laukka, 2003; Laukka and Elfenbein, 2021). These empirical findings extend Elfenbein and Ambady's (2002a) IGA hypothesis to the study of emotional prosody perception.

Few studies considered whether these findings in non-tonal languages (e.g., English) can be generalized to tonal languages (e.g., Mandarin Chinese). In Mandarin Chinese, the tone of voice can be used to differentiate lexical meaning in addition to encoding

emotional prosody (Xu, 2005), and such lexical and prosodic cues coexist (Ip and Cutler, 2020; Ouyang and Kaiser, 2015) and interact (Chang et al., 2023). There are four lexical tone categories in Mandarin Chinese, namely, tone 1: high level, tone 2: rising, tone 3: fall-rising, and tone 4: falling (Yip, 2002). For example, ma^1 means 'mother', ma^2 means 'hemp', ma^3 means 'horse', and ma^4 means 'to scold' (the superscripted numbers indicate different lexical tones). Thus, this dual function of the tone of voice (i.e., emotional prosody and lexical tone) raises questions about how emotional prosody is perceived in Mandarin Chinese words and sentences by both native and non-native Chinese speakers, and whether native Chinese speakers have an in-group advantage over non-native Chinese speakers in recognizing Chinese emotional prosody, and whether having second language (L2) Chinese learning experience improves non-native speakers' perception of emotional prosody in Mandarin Chinese.

There have been limited attempts to explore the perception of emotional prosody in Mandarin Chinese within the framework of the IGA hypothesis, which has yielded inconsistent findings. Some researchers primarily utilized pseudo-words and pseudo-sentences, and they found that native Chinese speakers recognized emotional prosody more accurately than non-native Chinese speakers, supporting the IGA hypothesis (Cowen et al., 2019; Liu and Pell, 2012; Liu et al., 2021; Paulmann and Uskul, 2014). However, when tested with real Chinese words and sentences, Zhu (2013) found that L2 Chinese learners outperformed native Chinese speakers, contradicting the IGA hypothesis. Furthermore, since the lexical tones were not controlled in these studies, it is unclear whether these results can inform the question of whether the IGA hypothesis holds true in tonal languages. Considering the coexistence and interaction of lexical tone and emotional prosody, as well as the scarcity of research on real Chinese words and sentences, it is crucial to examine the perception of emotional prosody using real Chinese words and sentences with controlled lexical tones.

Therefore, the present study utilizes real Chinese words and sentences to investigate the perception of emotional prosody in both native (L1) and non-native Chinese speakers. Specifically, the study examines how emotional prosody is perceived in Mandarin Chinese words and sentences by three groups of speakers: native Chinese speakers (native group), L1-English L2-Chinese learners (L2 group), and native English speakers without Chinese learning experience (non-native group). Furthermore, this study explores the effects of emotion type and syllable length on Chinese emotional prosody perception and further examines whether these effects are the same or different for the three groups. This study has pedagogical implications for L2 Chinese learners and language educators, and it also provides insights into the role of emotional prosody in cross-cultural communication.

II The perception of emotional prosody

1 Effects of emotion type and syllable length

In the literature on emotional prosody perception, researchers have often attributed the observed in-group advantage to the differences between in-group members and out-group members based on their cultural backgrounds. Specifically, when both the expressor and

the perceiver share the same cultural background (i.e., in-group members), their accuracy in perceiving emotional prosody is higher. In contrast, when they come from different cultural groups (i.e., out-group members), the accuracy tends to be lower. Previous studies have examined Elfenbein and Ambady's (2002a) IGA hypothesis in various cultural contexts, and they have also shown that, in addition to cultural backgrounds, factors such as emotion type and stimuli length also have an impact on the perception of emotional prosody (Laukka and Elfenbein, 2021; Laukka et al., 2016).

First, the perception of emotional prosody is influenced by different types of emotions. Cross-cultural comparisons have revealed a negative correlation between the in-group advantage and the accuracy of emotion expressions (Elfenbein and Ambady, 2002b; Juslin and Laukka, 2003). Sauter et al. (2010) found that positive emotions, such as achievement and relief, were not recognized bidirectionally by both English and Himba listeners, whereas negative emotions such as anger and disgust were well recognized across cultures. They ascribed these differences to different social functions of positive and negative emotions: positive emotions facilitate social cohesion within in-group members which may not be shared with out-group members, whereas negative emotions are more closely linked to biological reactions and less affected by cultural learning. Laukka and Elfenbein's (2021) meta-analysis found that, across different cultures, positive emotional prosody perception showed a greater in-group advantage between native and non-native speakers than negative emotional prosody, despite being recognized less accurately overall. However, it remains unclear how emotion type affects emotional prosody perception in tonal languages, considering that tone of voice can encode both lexical and emotional information.

In addition to the emotion type effect, syllable length also influences emotional prosody perception. Scherer (1986) first proposed that vocal emotion expressions exhibit emotion-specific acoustic patterns across different emotional states. While most studies have limited the scope of acoustic measures to f0 (e.g., Cho and Dewaele, 2021), intensity (e.g., Bachorowski and Owren, 1995), and speech rate (e.g., Koolagudi and Krothapalli, 2011), there is evidence that duration (e.g., syllable length) plays a role in the perception of emotional prosody. Blicher et al. (1990) indicated that the increase in syllable length can enhance the detectability of the tone of voice in Mandarin Chinese. Furthermore, Pell and Kotz (2011) constructed the auditory 'gates' by increasing the number of syllables, examining how much vocal information is needed for native English listeners to recognize basic emotions (e.g., neutral, happiness, anger, sadness). Their results showed that participants' accuracy improved as the syllable length increased: 12.6% accuracy rate at the first syllable and 87.1% at the seventh syllable. Additionally, they also found an interaction between emotion type and syllable length in the perception of emotional prosody in English. For example, shorter utterances showed higher accuracy in recognizing specific emotional prosodies (e.g., sadness, and neutral), while longer utterances resulted in better recognition of positive emotions (e.g., happiness). In the studies of Chinese emotional prosody perception, previous researchers either used sentences (Liu and Pell, 2012; Paulmann and Uskul, 2014) or did not control for the syllable length (Lin et al., 2020; Zhu, 2013). Thus, it is unknown whether emotional prosody in Chinese words and sentences can be both perceived by native and non-native Chinese speakers, and to what extent syllable length influences emotional prosody perception, and whether this effect of syllable length interacts with emotion type on their perception.

2 Second language experience effect

The existing studies have mainly compared the perception of emotional prosody between native speakers and non-native speakers without L2 learning experience, focusing on cultural backgrounds while ignoring the potential effects of linguistic experience. This leads to further questions regarding the role of linguistic experience in emotional prosody perception and, more specifically, how non-native speakers with language learning experience (i.e., L2 learners) perceive emotional prosody in Mandarin Chinese. In the field of second language acquisition (SLA), three main inquiries have been raised regarding the perception of emotional prosody within the framework of the IGA hypothesis.

The first question is to what extent second language learners can perceive emotional prosody in their second language. Alm and Llorà (2006) found that L1-Swedish L2-English learners and L1-Spanish L2-English learners can distinguish different emotional prosodies in L2 English even in a one-word utterance. Wei et al. (2022) showed that L1-Chinese L2-German learners can recognize emotional prosody in disyllabic German words with an above-chance-level accuracy. In addition, multiple studies indicated that L2 learners can recognize emotional prosody in sentences (e.g., Altrov, 2013; Bhatara et al., 2016; Dromey et al., 2005; Zhu, 2013). The results demonstrated that L2 learners are capable of accurately recognizing emotional prosody in their L2 at both word and sentence levels.

Building upon the findings of L2 learners' abilities to perceive emotional prosody, the second question is whether native speakers maintain an in-group advantage compared to L2 learners. Some researchers believe that native speakers have an in-group advantage of emotional prosody perception compared to L2 learners. For example, Altrov (2013) found that native Estonian speakers can recognize Estonian emotional prosody better than L1-Russian L2-Estonian learners. Similarly, Graham et al. (2001) showed native English speakers outperformed both L1-Japanese L2-English learners and L1-Spanish L2-English learners in perceiving English emotional prosody. Other researchers claim there are no significant differences between native speakers and L2 learners in terms of emotional prosody perception. For example, Dromey et al. (2005) reported no differences in English emotional prosody perception between native English speakers and L2 English learners. Min and Schirmer (2011) also found that the performance of emotional prosody perception was comparable between native English and L1-Chinese L2-English speakers. However, the most surprising results were found in a tonal language, where L2 Chinese learners outperformed native Chinese speakers in Chinese emotional prosody perception. Zhu (2013) observed that L1-Dutch L2-Chinese learners recognized emotional prosody in Mandarin Chinese more accurately than native Chinese speakers, and native Dutch speakers without L2 Chinese learning experience recognized emotional prosody in Mandarin Chinese as well as native Chinese speakers. Zhu further interpreted these unexpected findings in Chinese emotional prosody perception in light of the differences in the mechanisms of processing tone of voice between native and non-native Chinese speakers. Specifically, as speakers of tonal languages, native Chinese speakers tend to prioritize the linguistic function of the tone of voice (e.g., lexical tone) over its paralinguistic role (e.g., emotional prosody), resulting in less accurate recognition of paralinguistic cues (e.g.,

emotional prosody) compared to speakers of non-tonal languages. In addition, such perception differences between native speakers and L2 learners have been found to interact with emotion type in the perception of emotional prosody. For example, Paone and Frontera (2019) found that native Italian speakers showed comparable performance with L1-Russian L2-Italian speakers in identifying negative emotional prosodies such as anger and sadness, while native Italian speakers showed an in-group advantage at recognizing positive emotional prosody such as joy compared to L2 learners.

Given the inconsistent findings in the field of SLA, the third question is to determine whether second language experience facilitates or interferes with the perception of emotional prosody. Comparing emotional prosody perception among non-native speakers with different levels of L2 learning experience, some studies have revealed that the L2 learning experience can contribute to L2 learners' perception of emotional prosody in their second languages. For example, Zhu (2013) found that native Dutch speakers with L2 Chinese learning experience outperformed those without L2 Chinese learning experience in the perception of Chinese emotional prosody. Similarly, Shochi et al. (2016) found that native French speakers with more L2 Japanese learning experience were able to recognize Japanese emotional prosody more accurately compared to those with less L2 Japanese learning experience. On the contrary, other researchers have argued that one's second language experience may interfere with their perception of emotional prosody. For instance, Bhatara et al. (2016) found that L1-French L2-English learners with higher English proficiency were less accurate in recognizing positive emotional prosody in English compared to those with lower English proficiency. They argued that the interference effect may arise from semantics, as L2 learners with higher English proficiency may have focused more on the lexical meaning of the sentence rather than its emotional prosody, compared to those with lower English proficiency.

While previous literature demonstrates that L2 learners are capable of perceiving emotional prosody in their L2, there is still no consensus on whether native speakers have an in-group advantage over L2 learners and whether the L2 learning experience enhances or hinders L2 learners' emotional prosody perception. Additionally, it is important to note that not only is the systematic teaching of emotional prosody vastly neglected in L2 classrooms (Lengeris, 2012), but also the available curriculum and study materials (which are predominantly emotion-neutral do not teach L2 learners to perceive emotional prosody in their L2 (Dewaele, 2005; Kaneko and Yamane, 2022). Therefore, further investigation into emotional prosody perception in SLA is crucial to address these inadequacies and will have pedagogical implications for both L2 learners and language educators.

3 Some methodological limitations in previous research

The field of SLA has witnessed a growing body of research that investigates how L2 learners perceive paralinguistic information, such as emotional prosody. However, a closer examination of previous studies reveals some limitations in the experimental design. One noticeable limitation is the lack of inclusion of all three groups of speakers: native speakers (native group), non-native speakers with L2 learning experience (L2 group), and non-native speakers without L2 learning experience (non-native group).

Most studies have limited their comparisons to two of the three groups of speakers: native group vs. L2 group (e.g., Altrov, 2013); native group vs. non-native group (e.g., Paulmann and Uskul, 2014); L2 group vs. non-native group (e.g., Shochi et al., 2016). Including all three groups of speakers would elucidate the effects of cultural backgrounds and linguistic experiences on the perception of emotional prosody.

Furthermore, there has been a scarcity of studies investigating emotional prosody perception in tonal languages. In Mandarin Chinese, Zhu (2013) examined emotional prosody perception among three groups of speakers (i.e., native, L2, and non-native groups), yielding the unexpected results that L1-Dutch L2-Chinese learners showed higher accuracy in recognizing Chinese emotional prosody than native Chinese speakers. However, Zhu's experimental design is problematic in three ways. First, Zhu did not consider the lexical tone effect on emotional prosody perception. She not only used phrases such as *shi*^{4}*ni*^{3} 'it is you' with falling tones but also included sentences such as *jin*^{1}*tian*^{1} *xia*^{4}*wu*^{3} *ta*^{1} *bu*^{4}*neng*^{2} *lai*^{2} *can*^{1}*jia*^{1} *zhe*^{4}*ge*^{4} *hui*^{4} 'He cannot attend the meeting this afternoon' with all four lexical tones. The distribution of the four lexical tones in Zhu's stimuli was not controlled. Given that lexical tones can interfere with the perception of emotional prosody in Mandarin Chinese (Ross et al.,1986), the results of comparing emotional prosody perception between native speakers and L2 learners may be confounded by the effect of lexical tone. Second, Zhu also did not control the syllable length of the stimuli. The stimuli were only six sentences which ranged from the two-syllable sentence *shi*^{4}*ni*^{3} 'it is you' to the 13-syllable sentence *jin*^{1}*tian*^{1} *xia*^{4}*wu*^{3} *ta*^{1} *bu*^{4}*neng*^{2} *lai*^{2} *can*^{1}*jia*^{1} *zhe*^{4}*ge*^{4} *hui*^{4} 'He cannot attend the meeting this afternoon.' Zhu's comparisons between native and L2 Chinese speakers did not take into account the potential modulation effect of syllable length on emotional prosody perception, although previous studies have shown that an increase in syllable length can improve the accuracy of emotional prosody perception (Blicher et al., 1990; Pell and Kotz, 2011). Lastly, Zhu did not control the semantic valence of stimuli, treating a sentence with positive semantics such as *xie*^{4}*xie*^{4}*ni*^{3} 'thank you' the same as a sentence with negative semantics such as *jin*^{1}*tian*^{1} *xia*^{4}*wu*^{3} *ta*^{1} *bu*^{4}*neng*^{2} *lai*^{2} *can*^{1}*jia*^{1} *zhe*^{4}*ge*^{4} *hui*^{4} 'He cannot attend the meeting this afternoon.' As a result, the lower accuracy rate shown by native Chinese speakers in Zhu's study may be due to their experiencing greater interference from the semantics of the stimuli compared to L2 Chinese learners (Cho and Dewaele, 2021; Lin et al., 2020). Thus, the current study further examines the perception of emotional prosody in Mandarin Chinese with a more systematic control of the stimuli.

To sum up, in the field of SLA, several studies have explored the perception of emotional prosody between native speakers and L2 learners using Elfenbein and Ambady's (2002a) IGA hypothesis as a framework. While L2 learners have shown to be capable of recognizing emotional prosody, the findings from previous research have been inconsistent (Altrov, 2013; Dromey et al., 2005; Graham et al., 2001; Min and Schirmer, 2011; Zhu, 2013). Moreover, one previous study compared emotional prosody perception among native, L2, and non-native groups in a tonal language (Zhu, 2013), but it did not control for confounding factors (e.g., lexical tone, syllable length, and semantic valence). Hence, the present study investigates how L2 Chinese learners perceive emotional prosody in Chinese words and sentences, and whether having L2 Chinese learning experience improves non-native speakers' perception of emotional prosody in Mandarin Chinese.

III The current study

In light of previous research, the current study extends the psycholinguistic account of emotional prosody perception to the field of SLA specifically in a tonal language. This study investigates how native Chinese speakers (native group), L1-English L2-Chinese learners (L2 group), and native English speakers without Chinese learning experience (non-native group) perceive emotional prosody in Mandarin Chinese words and sentences within the framework of the IGA hypothesis (Elfenbein and Ambady, 2002a). Furthermore, the study explores the effects of emotion type (neutral, joy, anger, and sadness) and syllable length (monosyllable, disyllable, trisyllable, and sentence) on emotional prosody perception in Mandarin Chinese. Therefore, the present study addresses the following research questions:

> Research question 1: Does the In-Group Advantage (IGA) hypothesis hold true in Mandarin Chinese words and sentences?
> a. Does the native group show an advantage in recognizing emotional prosody in Mandarin Chinese over the non-native group?
> b. Does the native group show an advantage in recognizing emotional prosody in Mandarin Chinese over the L2 group?
> c. Does the L2 group show an advantage in recognizing emotional prosody in Mandarin Chinese over the non-native group?
> Research question 2: To what extent do emotion type and syllable length affect emotional prosody perception in Mandarin Chinese among the three groups?

We have made the following predictions in accordance with each research question. First, if the IGA hypothesis stands true in Mandarin Chinese (Elfenbein and Ambady, 2002a), we predict an effect of group such that the native group would have an advantage in recognizing emotional prosody over the non-native group. However, considering the inconsistent findings in previous studies, it remains unclear if the native group would maintain an advantage in recognizing emotional prosody compared to the L2 group (Paulmann and Uskul, 2014; Zhu, 2013); and if the L2 group would have an advantage in recognizing emotional prosody over non-native group (Bhatara et al., 2016; Shochi et al., 2016; Zhu, 2013). Moreover, we predict an effect of emotion type such that negative emotional prosody will be perceived more accurately compared to positive emotional prosody in Mandarin Chinese (Laukka and Elfenbein, 2021; Sauter et al., 2010). We also anticipate both an effect of syllable length such that the accuracy of emotional prosody perception improves as the syllable length increases, and an interaction between syllable length and emotion type in the perception of Chinese emotional prosody (Pell and Kotz, 2011). Additionally, we anticipate an interaction between emotion type and group (Bhatara et al., 2016; Paone and Frontera, 2019) in Mandarin Chinese words and sentences.

IV Methods

1 Participants

Based on a closely related study (Zhu, 2013), a total of 60 participants were included in the analysis: 20 native Chinese speakers (native group: 10 male, 10 female; mean

age = 24.7; SD of age = 2.45; age range = 22–30), 20 L1-English L2-Chinese learners (L2 group: 7 male, 13 female; mean age = 19; SD of age = 0.65; age range = 18–22), and 20 native English speakers without Chinese learning experience (non-native group: 4 male, 16 female, mean age = 21.1; SD of age = 1.74; age range = 19–24).[1] At the time of their participation, all native Chinese speakers were in China and indicated Mandarin Chinese as their native language. All native English speakers were in the United States and indicated English as their native language. All L2 Chinese learners were enrolled in their second semester of Mandarin course (mean L2 Chinese learning experience = 6.8 months)[2] at a public US university, and no L2 Chinese learners were heritage speakers of Mandarin Chinese or any other tonal language. All participants had normal hearing. All participants were tested remotely online and received class credit or $10 for their participation. All aspects of the study were approved by the Institutional Review Board (IRB) of the first author's university.

2 Stimuli

We adapted from Shen (1985) to create the word and sentence stimuli with controlled lexical tones and neutral semantic valence.[3] We selected words and sentences that were not typically found in L2 learners' textbooks to minimize the influence of semantic knowledge on their judgments of emotional prosody. Given the prevalence of relatively simple syllable structures, with only approximately 400 distinct syllables in Chinese (Duanmu, 2007), such construction of the stimuli allows for a comparison among the three groups: the native group (familiar with both phonology and semantics), the L2 group (familiar with phonology but not semantics), and the non-native group (unfamiliar with either semantics or phonology).

Furthermore, based on previous research (Paulmann and Uskul, 2014; Zhu, 2013), we manipulated the syllable length and emotion type of the stimuli, ensuring a similar distribution of four lexical tone categories across different syllable lengths and emotion types. Specifically, to explore the effect of syllable length, we included monosyllables, disyllables, trisyllables, and sentences. To probe into the effect of emotion type, we asked a professional female voice actress to record all the words and sentences in four types of emotional prosody: neutral, joy, anger, and sadness. After the collection of sound files, we used Praat (Boersma and Weenink, 2023) to segment the recorded utterances. In addition, we asked six native Chinese speakers to validate these recorded utterances by classifying the emotional prosody of each utterance in a four-alternative forced-choice format, and we only used the utterances that received unanimous agreement in the current experiment (144 out of 288 utterances). Thus, there were 144 stimuli (i.e., 16 monosyllabic words, 64 disyllabic words, 48 trisyllabic words, and 16 sentences) in the emotion judgment task. Table 1 provides examples of these stimuli, and Table 2 presents the acoustic parameters of the stimuli across syllable lengths and emotion types.

In the emotion judgment task, the stimuli were presented in four blocks: monosyllable block, disyllable block, trisyllable block, and sentence block. A cross-block Latin square design was used to counterbalance the presentation order of blocks, and thus four versions of the emotion judgment task were created using Qualtrics survey. Furthermore, within each block, the order of stimuli with different emotion types was

Table 1. Example stimuli of Chinese words and sentences.

Example	Monosyllabic word	Disyllabic word	Trisyllabic word
Pinyin	xiū	shōu yīn	zhāng zhōng bīn
IPA	ɕəu¹	ʂəu¹ jin¹	tʂan¹ tʂʷuŋ¹ pʲin¹
Chinese Character(s)	修	收 音	张 中 斌
English translation	repair	Receive sound	Zhang Zhongbin

Example	Sentence		
Pinyin	zhāng zhōng bīn	xīng qī tiān xiū	shōu yīn jī
IPA	tʂan¹ tʂʷuŋ¹ pʲin¹	ɕəŋ¹ tɕʰiː¹ tʰʲæn¹ ɕəu¹	ʂəu¹ jin¹ tɕiː¹
Chinese characters	张 中 斌	星 期 天 修	收 音 机
English translation	Zhang Zhongbin repairs radio on Sunday.		

Note. Superscript numbers indicate the distinct lexical tones in Mandarin Chinese.

Table 2. The means and standard deviations (in parentheses) of three acoustic parameters of the stimuli.

	F0 (Hz)	Intensity (dB SPL)	Duration (ms)
Syllable length:			
monosyllable	268.377 (56.206)	54.322 (4.655)	619.269 (140.522)
disyllable	292.143 (59.764)	54.246 (3.744)	717.842 (151.506)
trisyllable	295.402 (60.607)	55.005 (3.293)	898.905 (231.538)
sentence	277.690 (46.567)	55.906 (3.516)	2461.395 (412.817)
Emotion type:			
neutral	239.988 (32.043)	52.212 (2.660)	1050.859 (552.316)
joy	346.544 (43.131)	55.546 (2.440)	897.286 (543.131)
anger	327.620 (29.362)	58.650 (2.895)	719.172 (454.840)
sadness	241.779 (22.770)	52.359 (2.327)	1,176.570 (680.702)

also counterbalanced using a Latin square design. Additionally, six filler utterances were used in the experiment to check participants' attention.

3 Procedure

In this study, participants first completed the language background questionnaire and were randomly assigned to one version of the online emotion judgment task using Qualtrics in their respective native languages. In the language background questionnaire, participants were asked to provide information about their native languages and L2 Chinese learning experience (if any) prior to the emotion judgment task. The emotion judgment task was self-paced, and the participants were instructed to listen to a series of utterances, one at a time, and then judge the intended emotional prosody for each utterance in a four-alternative forced-choice format (i.e., neutral, joy, anger, and sadness). Participants' responses from the language background questionnaires and emotion judgment tasks were recorded. After the emotion judgment task, we asked L1-English–L2-Chinese learners to report if they

knew the meanings of the target words and sentences used in the experiment. The post-experiment reports showed that L2 learners only had limited knowledge of the semantics of the stimuli, suggesting that semantics had little influence on emotional prosody perception for the L2 group.

4 Analysis

In the emotion judgment task, the total number of trials in data analysis was 8,640 (2,880 from the native group, 2,880 from the L2 group, and 2,880 from the non-native group). For each trial, participants' judgments of emotional prosody were recorded and collected using Qualtrics. Participants received a score of '1' if they recognized the emotional prosody correctly, as their judgment matched the intended emotional prosody of the utterance; they received a score of '0' if their judgment mismatched the intended emotional prosody of the utterance. The raw scores (coded as 1 and 0) were averaged across the participants to calculate their accuracy rate.

Moreover, to test the IGA hypothesis in Mandarin Chinese, a logistic mixed-effects model (Jaeger, 2008) was conducted using the *glmer* function in R (R Core Team, 2022). We used the judgment of emotional prosody as the dependent variable, coded as 1 for a correct judgment and 0 for an incorrect judgment. The model included three fixed factors: (1) group with three levels (native group, L2 group, and non-native group); (2) emotion type with four levels (neutral, joy, anger, and sadness); and (3) syllable length with four levels (monosyllable, disyllable, trisyllable, and sentence). Sum coding was used, and item and participant (coded as ID) were entered as random factors for intercepts (Cunnings, 2012). To build the model, we used backward elimination starting with a maximal model that included all potential effects and interactions and removed the non-significant variables one at a time based on model comparison (Barr et al., 2013). When there was a significant effect or interaction, Tukey's post hoc tests were performed using the *emmeans* package (Lenth, 2020). Assumption checks for outliers and multicollinearity in the logistic regression model were conducted, revealing no influential outliers and low generalized variance inflation factor values, indicating that the assumptions are met.

V Results

1 Descriptive statistical results

In Figure 1, the confusion matrix shows that both native and L2 groups had higher overall accuracy rates than that of the non-native group (native group: 94.7%, L2 group: 95.9%; non-native group: 78.7%). Even with a lower accuracy rate, the non-native group's accuracy rate was still well above the chance level. The native and L2 groups showed higher accuracy rates than the non-native group across four emotion types.

A notable observation in Figure 1 is that, in the 'joy' condition, the native group showed a lower accuracy rate compared to the L2 group (native group: 89%; L2 group: 95%). Further analysis of error patterns revealed that native Chinese speakers had a higher tendency to mistake the emotion of 'joy' for 'neutral', with 38.8% (59 errors out
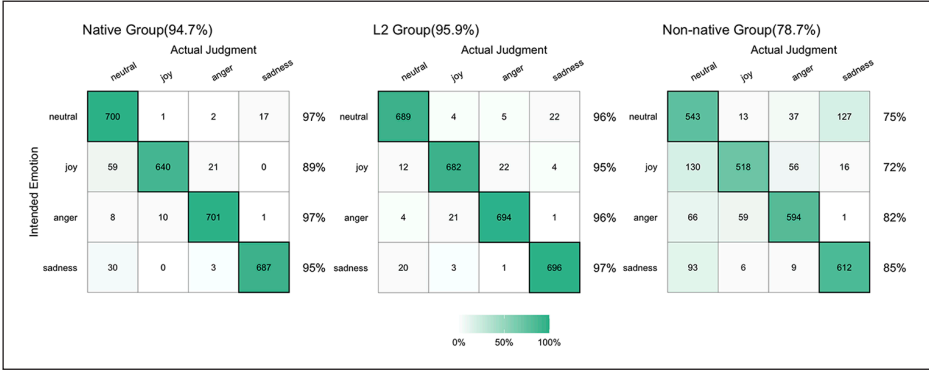
Figure 1. Confusion matrixes and mean accuracy rates (%) of emotional prosody judgments in three groups.
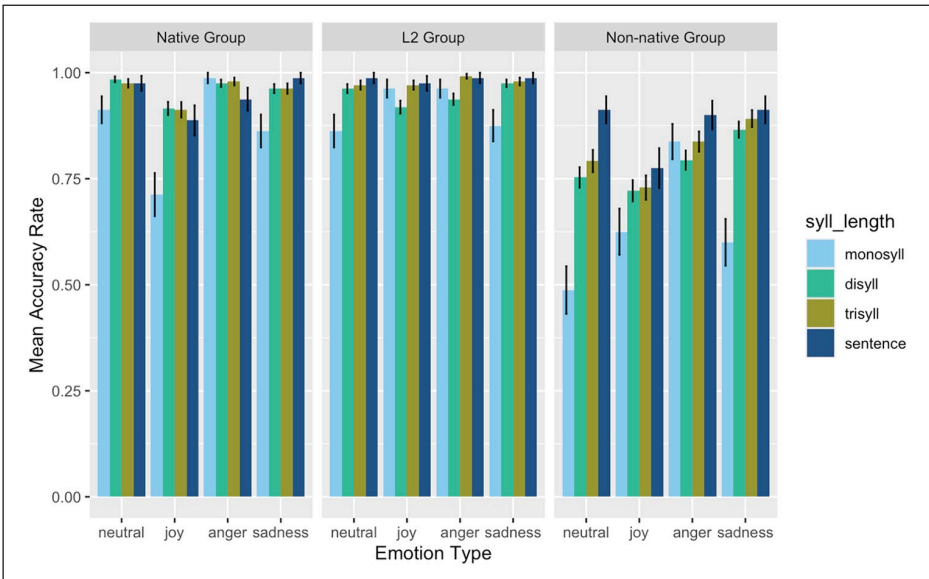

Figure 2. Mean accuracy rates of emotional prosody judgments across four emotion types and syllable lengths in three groups. The black vertical lines show the standard error.

of a total of 152 errors) of their errors involving this specific misjudgment. However, the same pattern for the L2 group and the non-native group accounted for only 10.1% and 21.2% of their errors respectively, which showed that they were less likely to confuse 'joy' for 'neutral'.

Moreover, Figure 2 illustrates the mean accuracy rates for three groups (native group, L2 group, and non-native group) across four emotion types (neutral, joy, anger, and sadness) and four syllable lengths (monosyllable, disyllable, trisyllable, and sentence). The
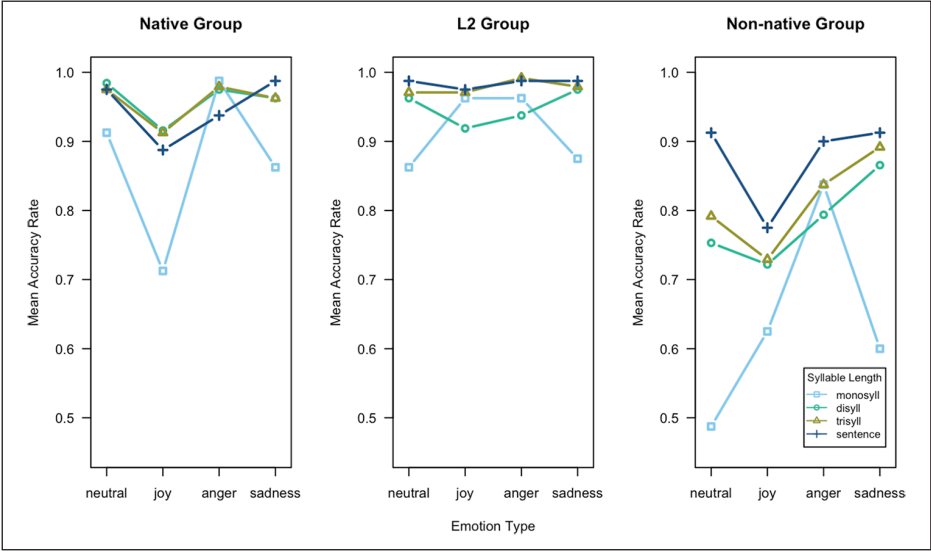
Figure 3. Plots of interaction between emotion type and syllable length for three groups in terms of mean accuracy rate.

native and L2 groups consistently outperformed the non-native group in all four emotion types and syllable lengths.

Interestingly, as shown in Figure 2, in the 'joy' condition, the native group showed a lower accuracy rate compared to the L2 group, particularly in the 'monosyllable' condition (native group: 71.3%; L2 group: 96.3%). However, for the other three emotion types (i.e., neutral, anger, and sadness), the native and L2 groups had comparable accuracy rates.

In addition, Figure 3 shows the interaction between emotion type and syllable length on the accuracy of emotional prosody perception for the three groups. In each group, the mean accuracy rates of disyllables, trisyllables, and sentences were higher than the monosyllables. The mean accuracy of the three participant groups was 80.7% in the monosyllable condition, and 89.7%, 91.6%, and 93.5% in the disyllable, trisyllable, and sentence conditions, respectively.

As shown in Figure 3, the accuracy rate in the monosyllable condition (represented by the light blue line) shows a larger fluctuation compared to other syllable length conditions for all three groups. While all three groups had the lowest accuracy rate in the monosyllable condition, the specific emotion type associated with this lowest accuracy varied across three groups: the native group had the lowest accuracy in the 'joy' condition, the L2 group showed the lowest accuracy in the 'neutral' and 'sadness' conditions, and the non-native group exhibited the lowest accuracy in the 'neutral' condition.

2 Inferential statistical results

As shown in Table 3, there was an effect of group: both the native group ($\beta = 0.660$, $p < .01$) and the L2 group ($\beta = 0.740$, $p < .001$) demonstrated significantly higher accuracy of

Table 3. Mixed-effects logistic regression model for the accuracy of the emotional prosody perception in three groups: native group, L2 group, and non-native group.

Fixed effects:	Estimate	SE	z	Pr (>|z|)
(Intercept)	2.999	0.170	17.624	< .001***
Group: native	0.660	0.211	3.121	.002**
Group: L2	0.740	0.213	3.475	.001***
Emo_type: joy	−0.709	0.085	−8.288	< .001***
Emo_type: anger	0.414	0.109	3.786	< .001***
Emo_type: sadness	0.179	0.105	1.705	.088
Syll_length: monosyllable	−0.900	0.171	−5.265	< .001***
Syll_length: disyllable	−0.006	0.117	−0.047	.963
Syll_length: trisyllable	0.255	0.127	2.012	.044*
joy × monosyllable	0.232	0.136	1.701	.089
anger × monosyllable	0.905	0.180	5.027	< .001***
sadness × monosyllable	−0.593	0.149	−3.990	< .001***
joy × disyllable	0.081	0.101	0.803	.422
anger × disyllable	−0.384	0.119	−3.222	.001**
sadness × disyllable	0.239	0.124	1.917	.055
joy × trisyllable	−0.004	0.111	−0.033	.974
anger × trisyllable	−0.116	0.135	−0.859	.390
sadness × trisyllable	0.150	0.138	1.091	.275
native × joy	−0.499	0.107	−4.674	< .001***
L2 × joy	0.317	0.114	2.777	.006**
native × anger	0.256	0.144	1.773	.076
L2 × anger	−0.214	0.134	−1.603	.109
native × sadness	−0.242	0.129	−1.882	.060
L2 × sadness	0.019	0.134	0.144	.885

Random effects	Variance	SD		
ID	1.082	1.040		
Item	0.128	0.357		

Notes. Model formula: glmer(accuracy ~ group + emo_type + syll_length + emo_type * syll_length + emo_type * group + (1 | ID) + (1 | item), control = glmerControl(optimizer = 'bobyqa'). All predictors were sum-coded, and the intercept represents the grand mean. *$p < .05$; **$p < .01$; ***$p < .001$.

emotional prosody perception than the grand mean. Moreover, effects of emotion type were observed such that 'joy' was recognized less accurately than the grand mean ($\beta = -0.709$, $p < .001$), whereas 'anger' was recognized more accurately than the grand mean ($\beta = 0.414$, $p < .001$). As for the effect of syllable length, while emotional prosody in 'monosyllable' was recognized less accurately than the grand mean ($\beta = -0.900$, $p < .001$), emotional prosody in 'trisyllable' was recognized more accurately than the grand mean ($\beta = 0.255$, $p < .05$). Furthermore, significant interactions between group and emotion type 'joy' were observed. Specifically, the significant interaction between 'native' and 'joy' was found ($\beta = -0.499$, $p < .001$), where the negative coefficient reflected that native speakers' advantage (relative to the grand mean, i.e., the simple effect of 'native') in perceiving emotional prosody was

Table 4. Post-hoc analysis results comparing the mean accuracy for three participant groups.

Group contrast	diff	lwr	upr	P adj
L2–non-native	0.172	0.153	0.190	<.001***
Native–non-native	0.160	0.142	0.178	<.001***
Native–L2	−0.011	−0.030	0.007	.298

Notes. diff = difference between group means; lwr = lower bound of 95% confidence interval; upr = upper bound of 95% confidence interval; P adj = adjusted *p*-value after correction for multiple comparisons. *p < .05; **p < .01; ***p < .001.

reduced on the emotion type 'joy.' In contrast, the significant interaction between 'L2' and 'joy' suggested that L2 learners' advantage (i.e., the simple effect of 'L2') in emotional prosody perception was enhanced for the emotion type 'joy (β = 0.317, p < .01). Additionally, three interaction terms between emotion type and syllable length were found to be significant, highlighting that specific emotion-syllable combinations impact the accuracy of emotional prosody perception differently. We thus examined these interactions for each group respectively.

To address research question 1 of whether the IGA hypothesis holds in Mandarin Chinese, we used Tukey's test from the *emmeans* package (Lenth, 2020) to conduct pairwise comparisons on groups. As seen in Table 4, our results indicated that the native group recognized emotional prosody more accurately than the non-native group (Mean diff = 0.160, 95% CI [0.142, 0.178], p < .001), indicating that the native group showed an in-group advantage at recognizing emotional prosody in Mandarin Chinese words and sentences over the non-native group. A critical finding is that the L2 group also outperformed the non-native group (Mean diff = 0.172, 95% CI [0.153, 0.190], p < .001) and showed no difference from the native group (Mean diff = −0.011, 95% CI [−0.030, 0.007], p = .298). This suggested that L2 Chinese learning experience gave the L2 group an in-group advantage in perceiving emotional prosody in Mandarin Chinese words and sentences compared to the non-native group. As there was a significant interaction between group and emotion type 'joy' in Table 3, we subset the 'joy' condition and found that the native group recognized positive emotional prosody (i.e., joy) less accurately compared to the L2 group in Mandarin Chinese words and sentences (see Figure 1).

Furthermore, significant effects and interactions in the omnibus model warranted us to conduct separate analyses for each group to address research question 2, namely, to what extent emotion type and syllable length affect emotional prosody perception. As shown in Table 5, for the native group, the emotional prosody of 'joy' was recognized less accurately (β = −1.255, p < .001) compared to the grand mean, whereas 'anger' was recognized more accurately (β = 0.663, p < .01). Emotional prosody in 'monosyllable' was recognized less accurately (β = −0.794, p < .01) compared to the grand mean. A significant interaction between emotion type 'anger' and syllable length 'monosyllable' was also observed in the native group (β = 1.548, p < .01): the positive coefficient indicated that the simple effect of 'anger' (higher accuracy compared to the grand mean) was made even more positive when presented in monosyllables. Table 6 presents the post-hoc pairwise comparisons, indicating that the native group perceived the emotional prosody of 'joy' less accurately than that of other emotion types. Furthermore, the native group recognized the emotional prosody in

Table 5. Mixed-effects logistic regression model for the accuracy of emotional prosody perception in the native group.

	Estimate	SE	z	Pr(>|z|)
(Intercept)	3.581	0.294	12.168	< .001***
Emo_type: joy	−1.255	0.164	−7.651	< .001***
Emo_type: anger	0.663	0.255	2.596	.009**
Emo_type: sadness	0.140	0.240	0.583	.560
Syll_length: monosyllable	−0.794	0.283	−2.810	.005**
Syll_length: disyllable	0.362	0.200	1.807	.071
Syll_length: trisyllable	0.267	0.210	1.274	.203
joy × monosyllable	−0.423	0.295	−1.433	.152
anger × monosyllable	1.548	0.598	2.591	.009**
sadness × monosyllable	−0.701	0.360	−1.947	.052
joy × disyllable	0.187	0.227	0.825	.409
anger × disyllable	−0.325	0.334	−0.974	.330
sadness × disyllable	−0.250	0.307	−0.816	.414
joy × trisyllable	0.231	0.240	0.964	.335
anger × trisyllable	−0.041	0.364	−0.113	.910
sadness × trisyllable	−0.162	0.323	−0.503	.615

Random effects	Variance	SD
ID	1.118	1.057
Item	0.121	0.348

Notes. Model formula: glmer (accuracy ~ emo_type + syll_length + emo_type * syll_length + (1|ID) + (1|item), control = glmerControl(optimizer = 'bobyqa'). All predictors were sum-coded, and the intercept represents the grand mean. *$p<.05$; **$p<.01$; ***$p<.001$.

Table 6. Pairwise comparisons for the accuracy of emotional prosody perception in the native group.

Contrast	Estimate	SE	z	p
Emotion type:				
joy–anger	−1.918	0.343	−5.583	< .001***
joy–sadness	−1.395	0.320	−4.363	< .001***
joy–neutral	−1.707	0.294	−5.804	< .001***
anger–sadness	0.523	0.423	1.236	.604
anger–neutral	0.210	0.404	0.521	.954
sadness–neutral	−0.313	0.384	−0.814	.848
Syllable length:				
monosyllable–disyllable	−1.156	0.392	−2.947	.017*
monosyllable–trisyllable	−1.061	0.402	−2.642	.041*
monosyllable–sentence	−0.958	0.513	−1.867	.242
disyllable–trisyllable	0.095	0.285	0.332	.987
disyllable–sentence	0.198	0.430	0.460	.968
trisyllable–sentence	0.103	0.438	0.234	.996

Note. *$p<.05$; **$p<.01$; ***$p<.001$.

Table 7. Mixed-effects logistic regression model for the accuracy of emotional prosody perception in the second language (L2) group.

	Estimate	SE	z	Pr(>|z|)
(Intercept)	4.128	0.379	10.898	< .001***
Emo_type: joy	−0.210	0.228	−0.921	.357
Emo_type: anger	0.372	0.276	1.348	.178
Emo_type: sadness	0.033	0.252	0.133	.894
Syll_length: monosyllable	−0.897	0.325	−2.761	.006**
Syll_length: disyllable	−0.413	0.235	−1.761	.078
Syll_length: trisyllable	0.518	0.280	1.849	.064
joy × monosyllable	0.949	0.405	2.342	.019*
anger × monosyllable	0.367	0.434	0.845	.398
sadness × monosyllable	−0.710	0.350	−2.028	.043*
joy × disyllable	−0.414	0.268	−1.545	.122
anger × disyllable	−0.690	0.314	−2.194	.028*
sadness × disyllable	0.655	0.321	2.042	.041*
joy × trisyllable	−0.204	0.339	−0.602	.547
anger × trisyllable	0.515	0.468	1.099	.272
sadness × trisyllable	−0.092	0.373	−0.246	.806

Random effects	Variance	SD
ID	1.700	1.304
Item	0.323	0.568

Notes. Model formula: glmer (accuracy ~ emo_type + syll_length + emo_type * syll_length + (1|ID) + (1|item), control = glmerControl(optimizer = 'bobyqa'). All predictors were sum-coded, and the intercept represents the grand mean. *$p < .05$; **$p < .01$; ***$p < .001$.

'monosyllable' less accurately than in 'disyllable' or 'trisyllable', while there was no difference among the other three syllable length conditions.

For the L2 group, as shown in Table 7, emotional prosody in 'monosyllable' was recognized less accurately compared to the grand mean ($\beta = -0.897$, $p < .01$). Significant interactions of syllable length and emotion type were also observed for the L2 learners. The significant interaction between 'monosyllable' and 'joy' showed that the simple effect of 'monosyllable' (lower accuracy than the grand mean) was reduced (made less negative) for the emotion type 'joy' ($\beta = 0.949$, $p < .05$). Furthermore, the significant interaction between 'monosyllable' and 'sadness' indicated that the simple effect of 'monosyllable' was enhanced (made even more negative) for the emotion type 'sadness' ($\beta = -0.710$, $p < .05$). In addition, two interaction terms between emotion type and 'disyllable' were also found to be significant. In Table 8, the post-hoc pairwise comparison indicated that the L2 group recognized the emotional prosody of short stimuli (i.e., monosyllables and disyllables) less accurately than that of longer stimuli (i.e., trisyllables or sentences). No significant effect of emotion type on the perception of emotional prosody was found among L2 learners.

For the non-native group, as shown in Table 9, the emotional prosody of 'joy' ($\beta = -0.519$, $p < .001$) was recognized less accurately compared to the grand mean, while

Table 8. Pairwise comparisons for the accuracy of emotional prosody perception in the second language (L2) group.

Contrast	Estimate	SE	z	p
Emotion type:				
joy–anger	−0.582	0.415	−1.402	.498
joy–sadness	−0.244	0.383	−0.636	.921
joy–neutral	−0.014	0.375	−0.038	1.000
anger–sadness	0.339	0.442	0.767	.869
anger–neutral	0.568	0.435	1.306	.559
sadness–neutral	0.229	0.404	0.568	.942
Syllable length:				
monosyllable–disyllable	−0.484	0.426	−1.137	.667
monosyllable–trisyllable	−1.416	0.478	−2.959	.016*
monosyllable–sentence	−1.690	0.657	−2.572	.049*
disyllable–trisyllable	−0.931	0.357	−2.610	.045*
disyllable–sentence	−1.205	0.575	−2.097	.154
trisyllable–sentence	−0.274	0.614	−0.446	.970

Note. $*p < .05$; $**p < .01$; $***p < .001$.

Table 9. Mixed-effects logistic regression model for the accuracy of emotional prosody perception in the non-native group.

	Estimate	SE	z	$Pr(>|z|)$
(Intercept)	1.641	0.240	6.828	< .001***
Emo_type: joy	−0.519	0.101	−5.128	< .001***
Emo_type: anger	0.394	0.118	3.335	.001***
Emo_type: sadness	0.351	0.118	2.981	.003**
Syll_length: monosyllable	−0.894	0.213	−4.204	< .001***
Syll_length: disyllable	−0.031	0.145	−0.215	.830
Syll_length: trisyllable	0.169	0.156	1.086	.277
joy × monosyllable	0.389	0.188	2.071	.038*
anger × monosyllable	0.806	0.222	3.637	< .001***
sadness × monosyllable	−0.608	0.197	−3.094	.002**
joy × disyllable	0.085	0.133	0.635	.525
anger × disyllable	−0.356	0.149	−2.383	.017*
sadness × disyllable	0.285	0.156	1.833	.067
joy × trisyllable	−0.092	0.144	−0.641	.521
anger × trisyllable	−0.248	0.163	−1.519	.129
sadness × trisyllable	0.320	0.172	1.854	.064

Random effects	Variance	SD
ID	0.898	0.947
Item	0.191	0.437

Notes. Model formula: glmer (accuracy ~ emo_type + syll_length + emo_type * syll_length + (1|ID) + (1|item), control = glmerControl(optimizer = 'bobyqa'). All predictors were sum-coded, and the intercept represents the grand mean. $*p < .05$; $**p < .01$; $***p < .001$.

Table 10. Pairwise comparisons for the accuracy of emotional prosody perception in the non-native group.

Contrast	Estimate	SE	z	p
Emotion type:				
joy–anger	−0.913	0.177	−5.147	< .001***
joy–sadness	−0.870	0.177	−4.921	< .001***
joy–neutral	−0.293	0.171	−1.716	.315
anger–sadness	0.043	0.196	0.220	.996
anger–neutral	0.620	0.192	3.232	.007**
sadness–neutral	0.576	0.191	3.016	.014*
Syllable length:				
monosyllable–disyllable	−0.863	0.290	−2.976	.016*
monosyllable–trisyllable	−1.063	0.301	−3.530	.002**
monosyllable–sentence	−1.650	0.388	−4.250	< .001***
disyllable–trisyllable	−0.200	0.206	−0.974	.764
disyllable–sentence	−0.787	0.319	−2.465	.066
trisyllable–sentence	−0.587	0.330	−1.780	.283

Note. *$p < .05$; **$p < .01$; ***$p < .001$.

'anger' ($\beta = 0.394$, $p < .001$) and 'sadness' ($\beta = 0.351$, $p < .01$) were recognized more accurately than the grand mean. Furthermore, emotional prosody in 'monosyllable' was recognized less accurately compared to the grand mean ($\beta = −0.894$, $p < .001$). Significant interactions showed that the simple effect of 'monosyllable' (lower accuracy compared to the grand mean) was reduced (made less negative) for emotion types 'joy' ($\beta = 0.389$, $p < .05$) and 'anger' ($\beta = 0.806$, $p < .001$). But this simple effect of 'monosyllable' was enhanced (made more negative) for the emotion type 'sadness' ($\beta = −0.608$, $p < .01$). Additionally, a significant interaction between 'anger' and 'disyllable' suggested that the simple effect of 'anger' (higher accuracy compared to the grand mean) was reduced in disyllables ($\beta = −0.356$, $p < .05$). The post-hoc pairwise comparisons indicated that the accuracy of negative emotional prosody (i.e., 'anger', and 'sadness') was significantly higher than in 'joy' or 'neutral' conditions (Table 10). Furthermore, the non-native group recognized the emotional prosody in the 'monosyllable' condition less accurately than the other three syllable length conditions.

VI Discussion

In this study, we investigated emotional prosody perception in Mandarin Chinese for three groups of speakers (native group, L2 group, and non-native group) across four emotion types (neutral, joy, anger, sadness) and four syllable lengths (monosyllable, disyllable, trisyllable, and sentence) using an emotion judgment task within the framework of In-Group Advantage Hypothesis (Elfenbein and Ambady, 2002a). The study contributed to the existing literature on emotional prosody perception in tonal languages by utilizing real Chinese words and sentences as stimuli while manipulating the effects

of emotion type and syllable length. Furthermore, our study extended the psycholinguistic account of emotional prosody perception to the field of second language acquisition, providing insights into how L2 learners perceive paralinguistic information, such as emotional prosody, in their second language.

Overall, our study indicated that native Chinese speakers (native group) and L1-English L2-Chinese learners (L2 group) recognized emotional prosody in Mandarin Chinese at a very high accuracy rate (native group: 94.7%; L2 group: 95.9%). In contrast, native English speakers without Chinese learning experience (non-native group) recognized emotional prosody in Mandarin Chinese less accurately (non-native group: 78.7%) but still well above the chance level. These results showed that although the non-native group demonstrated the ability to perceive emotional prosody in an unfamiliar tonal language at both the word and sentence levels, the native group had an in-group advantage in recognizing emotional prosody in Mandarin Chinese words and sentences compared to the non-native group. The findings provide support for Elfenbein and Ambady's (2002a) IGA hypothesis in the context of tonal languages.

In addition to the in-group advantage demonstrated by the native group, we also found that the L2 group, who had only a short period of Chinese language learning, was able to recognize Chinese emotional prosody more accurately than the non-native group, even though both groups belonged to the same cultural group (i.e., native English speakers). Our results indicated that the L2 group showed an advantage in recognizing emotional prosody over the non-native group in Mandarin Chinese words and sentences. This finding can be explained by the phonological familiarity gained through L2 Chinese learners' linguistic experience. Notably, Mandarin Chinese features a relatively small set of distinct syllables (approximately 400) and just over 1,300 unique syllable-tone combinations (Duanmu, 2007). Therefore, despite their limited experience with the Chinese language, L2 Chinese learners may have already gained a certain degree of phonological familiarity with many syllables and syllable-tone combinations. This phonological familiarity has been shown to improve linguistic processing for L2 learners (e.g., Kaushanskaya et al., 2013; Liu and Wiener, 2020). Our findings suggest this facilitation effect of linguistic experience can be extended to paralinguistic processing, thereby potentially compensating for L2 learners' disadvantages of not being a native speaker in their perception of emotional prosody.

Taken together, our study revealed that individuals with linguistic experience, including both native group and L2 group, outperformed those without such experience (non-native group) in the perception of emotional prosody, which aligned with previous studies' findings (Paulmann and Uskul, 2014; Zhu, 2013). However, Elfenbein and Ambady's (2002a) IGA hypothesis only predicts an in-group advantage based on cultural backgrounds where native speakers (culturally in-group members) have an in-group advantage over non-native speakers (culturally out-group members), but they do not explicitly address what emotional prosody perception looks like for L2 learners. This raises an important question in the framework of the IGA hypothesis, that is, how should we define and measure 'in-groupness' when including L2 learners in studies? While cultural background is indeed a contributing factor to the in-group advantage, it is not necessarily the only one. Our results found that native English speakers with L2 Chinese learning experience demonstrated significantly better perception of Chinese emotional

prosody compared to those without L2 Chinese learning experience. This finding high-lights the pivotal role of the second language experience in shaping emotional prosody perception, alongside cultural background. Therefore, we suggest that future research considers both cultural background and language experience when investigating emo-tional prosody perception involving non-native learners.

Interestingly, our study found that L2 Chinese learners showed a comparable perfor-mance with native Chinese speakers in perceiving emotional prosody in Mandarin Chinese words and sentences, consistent with prior studies (Dromey et al., 2005; Min and Schirmer, 2011). We also found that L2 Chinese learners recognized positive emo-tional prosody (i.e., joy) more accurately than native Chinese speakers particularly in monosyllabic words (native group: 71.3%; L2 group: 96.3%). This finding aligns with previous findings (Zhu, 2013) and indicates an interaction between emotion type and linguistic experience in a tonal language. Our results can be explained by the precedence of tone of voice as linguistic cues over paralinguistic cues among tonal language speak-ers (Zhu, 2013) coupled with an asymmetric perception of emotional prosody (Laukka and Elfenbein, 2021). Neural studies indicated that native Chinese speakers, as tonal language speakers, exhibited greater sensitivity to the task-irrelevant linguistic cues (Liu et al., 2015) and experienced more interference from lexical tones in speech perception (Yu and Zhang, 2018) compared to non-tonal language speakers. Meanwhile, prior stud-ies revealed a notable asymmetry where negative emotional prosody is generally more readily identified than positive emotional prosody (Laukka and Elfenbein, 2021; Liu and Pell, 2012). Negative emotional prosody often serves as warning signals, thus evolving to be more distinct and recognizable (Sauter et al., 2010), whereas positive emotional prosody is usually perceived across multiple channels alongside contextual meanings or facial expressions (Chang et al., 2023; Pell et al., 2009). Therefore, in our study, native Chinese speakers, as tonal language speakers, may be more susceptible to task-irrelevant linguistic cues (i.e., lexical tones), receiving more interference from lexical tones in per-ceiving positive emotional prosody compared to L2 Chinese learners. However, due to the innate salience of negative emotional prosody, native Chinese speakers may experi-ence minimal interference from lexical tones, resulting in high accuracy in identifying negative emotional prosody (mean accuracy of anger and sadness=96%), similar to L2 Chinese learners.

Another possible explanation is the influence of semantics on the perception of emo-tional prosody. Recent research shows that the semantic valence of the stimuli and the semantic knowledge of the participants can affect emotional prosody perception for both native and L2 speakers. For example, Cho and Dewaele (2021) indicated that semantic valence facilitated the perception of English emotional prosody for native and L2 English speakers in an emotion-congruent condition. Bhatara et al. (2016) found the semantic knowledge of participants interfered with emotional prosody perception for L2 English learners. Ben-David et al. (2016) discovered that semantics had an impact on the percep-tion of emotional prosody for native English speakers, even if it is task-irrelevant. Importantly, recent studies in Mandarin Chinese found a semantic-prosody congruency effect on the perception of Chinese emotional prosody for native and L2 Chinese speak-ers (Lin et al., 2020; Xiao and Liu, 2024). In our study, although we controlled the semantic valence of the stimuli, L2 Chinese learners had limited semantic knowledge of

the stimuli, whereas native Chinese speakers did know the meanings of the stimuli. This semantic knowledge interfered with emotional prosody perception, manifesting in specific error patterns: native Chinese speakers were more likely to mistake 'joy' as 'neutral' with a confusion rate of 38.8%, leading to significantly lower accuracy in their perception of positive emotional prosody; in contrast, the confusion rate for L2 Chinese learners was only 10.1%. Our research findings indicated that native Chinese speakers were more biased by semantics and thus confused the emotional prosody of 'joy' with 'neutral', especially when the encoded prosodic information was subtle and limited (e.g., 'joy' in monosyllables). Conversely, L2 Chinese learners had limited semantic knowledge of Chinese words and sentences and thus experienced less semantic interference in their perception of emotional prosody. As a result, L2 learners may have focused more on prosodic cues rather than semantic cues in stimuli, enabling them to recognize positive emotional prosody better than native speakers.

Just as there is an effect of emotion type, syllable length also has an impact on emotional prosody perception in Mandarin Chinese. Few studies have specifically investigated the effect of syllable length on emotional prosody perception in tonal languages. In the study, we discovered that native Chinese speakers, L1-English L2-Chinese speakers, and native English speakers without Chinese learning experience can perceive emotional prosody above chance level in Chinese words and sentences, and the recognition of emotional prosody improved as syllable length increased for all three groups. Furthermore, there were group differences in perception of emotional prosody: the native group showed the lowest accuracy in recognizing 'joy' in the monosyllable condition, while for the L2 and non-native groups, the lowest accuracy was associated with recognizing the 'neutral' emotion in the monosyllable condition. Evidence from cognitive neuroscience has shown that the brain state in 'neutral' serves as a central hub within the network of emotions (Kragel et al., 2022). Thus, we speculate that this central role of recognizing neutral emotions could present greater challenge for non-native speakers (both L2 and non-native groups) in establishing a baseline of emotion perception in an unfamiliar or second language, especially when emotional information is limited (e.g., monosyllables).

The current study was not without limitations. First, the experiment was conducted remotely. In future studies, it is essential to control the acoustic environments during emotional judgment tasks since the remote setup may result in diverse perception environments (Yan et al., 2022). Second, previous studies have reported significant gender and age effects on emotional prosody perception (e.g., Hunter et al., 2010; Lin et al., 2021a, 2021b; Sen et al., 2018). Although our results showed no significant effects of gender and age across groups, the imbalance of gender representation in the non-native group raises a potential concern. Future studies can manipulate the gender and age factors, exploring potential interactions with linguistic experience in the perception of emotional prosody. Furthermore, the current study found that native Chinese speakers with semantic knowledge of the stimuli showed a lower accuracy in the perception of positive emotional prosody compared to the L2 Chinese learners without such semantic knowledge. It would be interesting to examine how semantics influence emotional prosody perception when both native and L2 speakers have the semantic knowledge of stimuli. In

addition, our study provided evidence that second language learning experience can aid in paralinguistic processing for non-native speakers in a tonal language. To elucidate the scope and mechanisms of this facilitation effect, it is necessary to examine the perception of emotional prosody in L2 learners with different stages of language proficiency, including elementary, intermediate, and advanced levels. Such investigations have pedagogical implications for L2 education, where emotional cues are often ignored in language learning and language teaching.

VII Conclusions

The present study examined emotional prosody perception in Mandarin Chinese words and sentences for three groups: the native group, the L2 group, and the non-native group. The results showed that native Chinese speakers had an advantage in recognizing emotional prosody in Mandarin Chinese compared to native English speakers without Chinese learning experience, which supports the IGA hypothesis in a tonal language. L1-English L2-Chinese learners also recognized Chinese emotional prosody more accurately than native English speakers without Chinese learning experience, indicating that linguistic learning experience plays a significant role in emotional prosody perception. Interestingly, our study also revealed an interaction between emotion type and language experience: L2 Chinese learners outperformed native Chinese speakers in the perception of positive emotional prosody. We argued that the emotional prosody perception of native Chinese speakers was more biased by linguistic cues (such as lexical tone and semantics), compared to L2 Chinese learners.

Furthermore, we found emotion type and syllable length have impacts on the perception of emotional prosody in Mandarin Chinese. Negative emotional prosody was perceived more accurately than positive emotional prosody in Chinese words and sentences. Although all three groups demonstrated the ability to perceive emotional prosody in a single syllable (monosyllables), the accuracy of emotional prosody perception was found to be positively correlated with the syllable length. Additionally, there was an interaction between emotion type and syllable length on the perception of emotional prosody: native Chinese speakers exhibited the lowest accuracy in identifying positive emotional prosody such as 'joy' in monosyllabic stimuli, whereas native English speakers, including L2 and non-native groups, both had the lowest accuracy in recognizing 'neutral' prosody in monosyllables. In summary, this study sheds light on the complex nature of emotional prosody perception in tonal languages and highlights the effects and interactions of speaker group, emotion type, and syllable length on emotional prosody perception. Future research should consider both cultural background and linguistic experience when studying emotional prosody perception in the context of second language acquisition.

Acknowledgements

Declaration of conflicting interests

Funding

Data availability

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

ORCID iD

Cheng Xiao 🆔 https://orcid.org/0000-0001-5057-9143

Notes

1. Additional regression models for gender and age were run separately, indicating that neither gender nor age were significant predictors across groups for emotional prosody perception.
2. Two L2 learners just finished their first-semester Chinese course at the time of their participation. The reported L2 Chinese learning experience comprises only the coursework.
3. Some words (e.g., *Sunday*) may carry positive semantic valence implicitly. After excluding these words, additional statistical analyses confirmed that our findings were not driven by the emotion-laden words.

References

Alm CO and Llorà X (2006) Evolving emotional prosody. In: *Interspeech 2006*. Available at: https://doi.org/10.21437/interspeech.2006-504 (accessed September 2024).

Altrov R (2013) Aspects of cultural communication in recognizing emotions. *Trames* 17: 159–74.

Bachorowski JA and Owren MJ (1995) Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context. *Psychological Science* 6: 219–24.

Barr DJ, Levy R, Scheepers C, and Tily HJ (2013) Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68: 255–78.

Beier E and Zautra AJ (1972) The identification of vocal expressions of emotion across cultures. *Journal of Consulting and Clinical Psychology* 40: 560.

Ben-David BM, Multani N, Shakuf V, Rudzicz F, and van Lieshout PH (2016) Prosody and semantics are separate but not separable channels in the perception of emotional speech: Test for rating of emotions in speech. *Journal of Speech, Language, and Hearing Research* 59: 72–89.

Bhatara A, Laukka P, Boll-Avetisyan N et al. (2016) Second language ability and emotional prosody perception. *PLoS One* 11: e0156855.

Blicher DL, Diehl RL, and Cohen LB (1990) Effects of syllable duration on the perception of the Mandarin Tone 2 / Tone 3 distinction: Evidence of auditory enhancement. *Journal of Phonetics* 18: 37–49.

Boersma P and Weenink D (2023) *Praat: Doing phonetics by computer: Version 6.3.10* [computer program]. Available at: http://www.praat.org (accessed September 2024).

Brooks JA, Chikazoe J, Sadato N, and Freeman JB (2019) The neural representation of facial-emotion categories reflects conceptual structure. *Proceedings of the National Academy of Sciences* 116: 15861–70.

Chang HS, Lee CY, Wang X et al. (2023) Emotional tones of voice affect the acoustics and perception of Mandarin tones. *PLoS One* 18: e0283635.

Cho CM and Dewaele JM (2021) A crosslinguistic study of the perception of emotional intonation. Influence of the pitch modulations. *Studies in Second Language Acquisition* 43: 870–95.

Chronaki G, Wigelsworth M, Pell MD, and Kotz SA (2018) The development of cross-cultural recognition of vocal emotion during childhood and adolescence. *Scientific Reports* 8: 1–17.

Cowen AS, Laukka P, Elfenbein HA, Liu R, and Keltner D (2019) The primacy of categories in the recognition of 12 emotions in speech prosody across two cultures. *Nature Human Behaviour* 3: 369–82.

Cunnings I (2012) An overview of mixed-effects statistical models for second language researchers. *Second Language Research* 28: 369–82.

Cutler A, Dahan D, and Van Donselaar W (1997) Prosody in the comprehension of spoken language: A literature review. *Language and Speech* 40: 141–201.

Dewaele JM (2005) Investigating the psychological and emotional dimensions in instructed language learning: Obstacles and possibilities. *The Modern Language Journal* 89: 367–80.

Dromey C, Silveira J, and Sandor P (2005) Recognition of affective prosody by speakers of English as a first or foreign language. *Speech Communication* 47: 351–59.

Duanmu S (2007) *The phonology of Standard Chinese*. Oxford: Oxford University Press.

Ekman P and Friesen WV (1986) A new pan-cultural facial expression of emotion. *Motivation and Emotion* 10: 159–68.

Ekman P, Sorenson ER, and Friesen WV (1969) Pan-cultural elements in facial displays of emotion. *Science* 164: 86–88.

Elfenbein HA (2013) Nonverbal dialects and accents in facial expressions of emotion. *Emotion Review* 5: 90–96.

Elfenbein HA and Ambady N (2002a) On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin* 128: 203–35.

Elfenbein HA and Ambady N (2002b) Is there an in-group advantage in emotion recognition? *Psychological Bulletin* 128: 243–49.

Elfenbein HA and Ambady N (2003) Universals and cultural differences in recognizing emotions. *Current Directions in Psychological Science* 12: 159–64.

Gendron M, Crivelli C, and Barrett LF (2018) Universality reconsidered: Diversity in making meaning of facial expressions. *Current Directions in Psychological Science* 27: 211–19.

Graham CR, Hamblin AW, and Feldstein S (2001) Recognition of emotion in English voices by speakers of Japanese, Spanish and English. *International Review of Applied Linguistics in Language Teaching* 39: 19–37.

Hunter EM, Phillips LH, and MacPherson SE (2010) Effects of age on cross-modal emotion perception. *Psychology and Aging* 25: 779–787.

Ip MHK and Cutler A (2020) Universals of listening: Equivalent prosodic entrainment in tone and non-tone languages. *Cognition* 202: 104311.

Jack RE, Garrod OG, Yu H, Caldara R, and Schyns PG (2012) Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences* 109: 7241–44.

Jaeger TF (2008) Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of memory and language* 59: 434–46.

Juslin PN and Laukka P (2003) Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin* 129: 770.

Kaneko I and Yamane N (2022) Emotional prosody of love and sorrow: L1 English, TTS and EFL learners. In: Levis J and Guskaroska A (eds) *Proceedings of the Twelfth Pronunciation in Second Language Learning and Teaching Conference*. Ames, IA: Iowa State University Digital Press.

Kaushanskaya M, Yoo J, and Van Hecke S (2013) Word learning in adults with second-language experience: Effects of phonological and referent familiarity. *Journal of Speech, Language, and Hearing Research* 56: 667–78.

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

Koolagudi SG and Krothapalli RS (2011) Two stage emotion recognition based on speaking rate. *International Journal of Speech Technology* 14: 35–48.

Kragel PA, Hariri AR, and LaBar KS (2022) The temporal dynamics of spontaneous emotional brain states and their implications for mental health. *Journal of Cognitive Neuroscience* 34: 715–28.

Laukka P and Elfenbein HA (2021) Cross-cultural emotion recognition and in-group advantage in vocal expression: A meta-analysis. *Emotion Review* 13: 3–11.

Laukka P, Elfenbein HA, Thingujam NS et al. (2016) The expression and recognition of emotions in the voice across five nations: A lens model analysis based on acoustic features. *Journal of Personality and Social Psychology* 111: 686.

Lengeris A (2012) Prosody and second language teaching: Lessons from L2 speech perception and production research. In: Romero-Trillo J (ed.) *Pragmatics and prosody in English language teaching*. Dordrecht: Springer, pp. 25–40.

Lenth R, Singmann H, Love J, Buerkner P, and Herve M (2020) *emmeans: Estimated marginal means, aka least-squares means: R package: Version 1.5.3* [computer software]. Available at: https://cran.r-project.org/web/packages/emmeans/index.html (accessed September 2024).

Lin Y, Ding H, and 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: 896–912.

Lin Y, Ding H, and Zhang Y (2021a) Unisensory and multisensory Stroop effects modulate gender differences in verbal and nonverbal emotion perception. *Journal of Speech, Language, and Hearing Research* 64: 4439–57.

Lin Y, Ding H, and Zhang Y (2021b) Gender differences in identifying facial, prosodic, and semantic emotions show category-and channel-specific effects mediated by encoder's gender. *Journal of Speech, Language, and Hearing Research* 64: 2941–55.

Liu J and Wiener S (2020) Homophones facilitate lexical development in a second language. *System* 91: 102249.

Liu P and Pell MD (2012) Recognizing vocal emotions in Mandarin Chinese: A validated database of Chinese vocal emotional stimuli. *Behavior Research Methods* 44: 1042–51.

Liu P, Rigoulot S, Jiang X, Zhang S, and Pell MD (2021) Unattended emotional prosody affects visual processing of Facial expressions in Mandarin-speaking Chinese: A comparison with English-speaking Canadians. *Journal of Cross-Cultural Psychology* 52: 275–94.

Liu P, Rigoulot S, and Pell MD (2015) Cultural differences in on-line sensitivity to emotional voices: Comparing East and West. *Frontiers in Human Neuroscience* 9: 311.

Matsumoto Y (1988) Reexamination of the universality of face: Politeness phenomena in Japanese. *Journal of Pragmatics* 12: 403–426.

Min CS and Schirmer A (2011) Perceiving verbal and vocal emotions in a second language. *Cognition and Emotion* 25: 1376–92.

Ouyang IC and Kaiser E (2015) Prosody and information structure in a tone language: An investigation of Mandarin Chinese. *Language, Cognition and Neuroscience* 30: 57–72.

Paone E and Frontera M (2019) Emotional prosody perception in Italian as a second language. In: *ExLing 2019: Proceedings of Tenth International Conference of Experimental Linguistics*. Athens: International Society of Experimental Linguistics, pp. 161–64.

Paulmann S and Uskul AK (2014) Cross-cultural emotional prosody recognition: Evidence from Chinese and British listeners. *Cognition and Emotion* 28: 230–44.

Pell MD and Kotz SA (2011) On the time course of vocal emotion recognition. *PLoS One* 6: e27256.

Pell MD, Monetta L, Paulmann S, and Kotz SA (2009) Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior* 33: 107–20.

R Core Team (2022) *R: A language and environment for statistical computing* [computer program]. Vienna: R Foundation for Statistical Computing. Available at: https://www.R-project.org (accessed September 2024).

Ross ED, Edmondson JA, and Seibert GB (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: 283–302.

Russell JA (1994) Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin* 115: 102.

Russell JA and Barrett LF (1999) Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology* 76: 805.

Sauter DA, Eisner F, Ekman P, and Scott SK (2010) Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences* 107: 2408–12.

Scherer KR (1986) Vocal affect expression: A review and a model for future research. *Psychological Bulletin* 99: 143–165.

Scherer KR, Banse R, and Wallbott HG (2001) Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-cultural Psychology* 32: 76–92.

Sen A, Isaacowitz D, and Schirmer A (2018) Age differences in vocal emotion perception: On the role of speaker age and listener sex. *Cognition and Emotion* 32: 1189–204.

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

Shochi T, Brousse A, Guerry M, Erickson D, and Rilliard A (2016) Learning effect of social affective prosody in Japanese by French learners. In: *Speech Prosody 2016*. Available at: https://doi.org/10.21437/speechprosody.2016-199 (accessed September 2024).

Van Bezooijen R, Otto SA, and Heenan TA (1983) Recognition of vocal expressions of emotion: A three-nation study to identify universal characteristics. *Journal of Cross-Cultural Psychology* 14: 387–406.

Wei H, He Y, Kauschke C, Scharinger M, and Domahs U (2022) An EEG-study on L2 categorization of emotional prosody in German. In: *Speech Prosody 2022*. Available at https://doi.org/10.21437/SpeechProsody.2022-128 (accessed September 2024).

Wilson D and Wharton T (2006) Relevance and prosody. *Journal of Pragmatics* 38: 1559–79.

Xiao C and Liu J (2024) Semantic effects on the perception of emotional prosody in native and non-native Chinese speakers. *Cognition and Emotion*. Advance online publication. https://doi.org/10.1080/02699931.2024.2371088

Xu Y (2005) Speech melody as articulatorily implemented communicative functions. *Speech Communication* 46: 220–51.

Yan Y, Li S, and Chen Y (2022) In-group advantage for Chinese and English emotional prosody in quiet and noise conditions. In: *Thirteenth International Symposium on Chinese Spoken Language Processing (ISCSLP)*. New York: IEEE, pp. 305–09.

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

Yu L and Zhang Y (2018) Testing native language neural commitment at the brainstem level: A cross-linguistic investigation of the association between frequency-following response and speech perception. *Neuropsychologia* 109: 140–48.

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: 170–83.