SSRG International Journal of Humanities and Social Science
ISSN: 2394-2703/ https://doi.org/10.14445/23942703/IJHSS-V12I6P105

Volume 12 Issue 6, 48-57, Nov-Dec 2025

Original Article

The Perception of Emotional Prosody in English by Chinese L2 Learners

Liu Kaipeng, Wu Ling

Sichuan International Studies University, Chongqing, China.

Corresponding Author : 1284877660@qq.com

Received: 04 October 2025 Revised: 19 November 2025 Accepted: 08 December 2025 Published: 25 December 2025

Abstract *- As Chinese culture globalizes, cross-cultural communication between Chinese students and English speakers has grown increasingly vital. However, the mechanisms underlying emotional encoding and decoding among Chinese L2 learners—particularly college students—in such interactions remain underexplored. Guided by the Auto-segmental Metrical (AM) model, this study investigates how language proficiency, gender, and emotion types influence Chinese English learners' ability to perceive emotional prosody. Behavioral experimentation was conducted using E-Prime to measure perceptual accuracy, SPSS for two-way ANOVA, and Praat to analyze acoustic features of four target emotions. Three core research questions were addressed: 1) Does language proficiency affect perceptual performance? 2) Does gender impact recognition outcomes? 3) Do emotion types influence perceptual accuracy? Key findings reveal that: 1) Learners with higher CET-4 scores demonstrated marginally higher accuracy, though the difference was not statistically significant; 2) Gender exerted no notable effect on perception; 3) Neutral emotion was the easiest to distinguish, followed by sadness, with fear proving the most challenging. Neutrality vs. sadness and fear vs. sadness emerged as the most confusing pairs, with learners showing weak differentiation between these categories—potentially attributed to L1 prosodic transfer. This study contributes to the field of emotional prosody research and offers practical implications for L2 English teaching.*

Keywords *- Chinese L2 learners, Emotional prosody perception, Gender, Language proficiency, Types of emotions.*

1. Introduction

English is essential for Chinese students, yet they often struggle to communicate effectively with native speakers, partly due to insufficient education in emotional prosody. While global research on emotional prosody is extensive, most focus on production rather than perception (Larrouy-Maestri et al., 2024), and few target second language (L2) contexts. Existing work mainly centers on first languages (L1) like English and Chinese, which leaves a gap in understanding how Chinese learners perceive English emotional prosody.

This study addresses this gap by identifying which English emotional prosodies challenge Chinese L2 learners most and exploring links between this ability and L2 proficiency, with findings intended to inform L2 English teaching.

Emotional prosody is critical for emotional communication: it enhances utterance comprehension, accelerates discourse processing, and avoids miscommunication (Li, 2021; Lin & Wang, 2016). For Chinese learners, accurately perceiving English emotional prosody is vital for cross-cultural interaction, yet this skill is rarely taught in traditional English classes. This study uses behavioral experimentation, acoustic analysis, and statistical analysis. It is theoretically grounded in the Auto-segmental Metrical (AM) model— the most authoritative framework for describing intonation across languages, which links prosodic phonology to measurable acoustic parameters (Arvaniti, 2022; Mennen, 2015).

Chinese learners face unique challenges in perceiving English emotional prosody due to cross-linguistic differences: Chinese is a tonal language, where pitch primarily serves to distinguish lexical meaning, whereas English uses pitch dynamically to signal emotion (Ross et al., 1986). This divergence often leads to misperception—for example, interpreting rising English intonation as a question rather than an expression of surprise.

Additionally, L2 proficiency (with proficient learners typically outperforming their peers in prosodic tasks; Zhang & Chen, 2023; Ji et al., 2018) and gender (linked to emotional processing; Neuhaus et al., 2024; Tripp & Munson, 2022) may influence perceptual outcomes, but their specific roles in L2 prosody perception remain underexplored.

To address these gaps, this study compares high- and low-proficiency Chinese L2 learners to answer three research questions: (1) Does L2 proficiency affect the recognition of English emotional prosody? (2) Does gender influence perceptual accuracy? (3) Which emotions are most confusing for Chinese L2 learners?

2. Literature Review

This study draws on two primary theoretical frameworks: the Auto-segmental Metrical model and theories of emotion. Complementary literature on emotional prosody encoding/decoding, L2 proficiency, and gender is also reviewed to contextualize the research.

2.1. The Auto-segmental Metrical Model

The Auto-segmental-Metrical (AM) model of intonational phonology has evolved over four decades, rooted in empirical evidence from speech production and perception (Arvaniti, 2022). Its foundational tenets originate from Janet Pierrehumbert's dissertation, with subsequent refinements based on experimental research across numerous languages (Arvaniti, 2022). A key contribution of the AM model is its distinction between intonation (a phonological construct) and its phonetic realization, granting flexibility to adapt to the descriptive needs of prosodically diverse languages (Arvaniti, 2022).

At its core, the AM model posits that intonation is phonologically represented as a sequence of low (L) and high (H) tones—abstract units defined relative to a speaker's vocal range and other tones in the same contour (Mennen, 2015). Intonation is composed of three pitch events: pitch accents, phrase accents, and boundary tones. Seven pitch accents, like H*, L*+H-, two phrase accents, H-, L-, and two boundary tones, H%, L%, combine to form distinct intonational patterns (Arvaniti, 2022). The "nuclear accent", which is the most prominent pitch accent, carries primary linguistic significance, with local phonetic events driving meaning, rather than pitch curves between events (Ji et al., 2018). This linear tonal structure enables cross-linguistic intonation comparison, making the AM model ideal for analyzing L2 prosody perception (Mennen, 2015).

2.2. Emotional Model

Scholars have proposed two dominant frameworks for describing emotions: discrete (basic) and dimensional models (Silva & Barbosa, 2017). The discrete model, tracing its roots to Darwin and expanded by Ekman (1973), posits a set of evolutionarily derived basic emotions, such as joy, sadness, and fear, that are universally recognized through distinct expressions (Ekman, 1992; Tomkins, 2014). Griffiths (2007) further refined this model by introducing the concept of "emotional programming," suggesting that basic emotions are triggered by specific conditions and manifested through coordinated responses. Ekman's (1992) six basic emotions, joy, sadness, fear, disgust, anger, and surprise, remain the most widely adopted classification in prosody research (Banse & Scherer, 1986).

In contrast, the dimensional model conceptualizes emotions as continuous constructs defined by underlying dimensions, most notably activation, a kind of emotion that is arousal, from calm to agitated, and valence, which is pleasantness, from positive to negative. (Fontaine, 2009; Smith & Ellsworth, 1985). Additional dimensions include dominance, involvement, fairness, and motivation. (Silva & Barbosa, 2017). While dominance was once a focus of research, it has been largely replaced by activation in prosody studies due to inconsistent recognition rates (Silva & Barbosa, 2017). These two models are not mutually exclusive: basic emotions can be mapped onto dimensional scales. For example, joy is equal to positive valence plus high activation, whereas sadness is equal to negative valence plus low activation. (Smith & Ellsworth, 1985).

This study hypothesizes that each basic emotion is associated with distinct perceptual judgments across emotional dimensions, and that ratings for discrete emotions and dimensions correlate with acoustic parameters extracted from speech stimuli.

2.3. The Studies of Emotional Prosody

Research on emotional prosody falls into two broad categories: encoding acoustic features of emotional speech and decoding the perception of emotion from speech. (Larrouy-Maestri et al., 2024). Encoding studies focus on measuring acoustic parameters, such as fundamental frequency (F0), intensity, duration, and spectral characteristics that distinguish emotional expressions (Banse & Scherer, 1996; Williams & Stevens, 1972). For example, Zhang and Chen (2023) found that happiness is characterized by higher pitch, shorter duration, and greater loudness compared to sadness. While some acoustic patterns, like F0 variability for anger, are universal, inconsistencies persist across languages and studies, often due to reliance on sentence-level analyses that overlook lexical interference (Scherer, 1986; Ma & Wang, 2017).

Decoding studies investigate whether listeners can recognize intended emotions from speech samples, with a focus on cross-linguistic universality (Li, 2021). Gussenhoven and Chen (2000) demonstrated partial universality. Chinese, Dutch, and Hungarian listeners all associated higher pitch accents and boundary tones with interrogative sentences. However, Grabe et al. (2003) found that L1 background influences intonation perception, with British, Spanish, and Chinese listeners differing in their classification of rising and falling contours. Ortega-Llebaria and Colantoni (2014) further showed that form-meaning pairing affects perception: Spanish and Chinese learners performed as well as native English speakers in meaning-free tasks but struggled when semantic context was included. For Chinese, a tonal language, emotional prosody perception is complicated by the dual function of pitch (lexical vs. emotional; Ross et al., 1986). Wang and Li (2003) noted that Chinese emotional prosody relies on both F0 changes (e.g., for happiness) and acoustic features (e.g., for sadness and fear), while Li (2006) found that emotional intonation shifts sentence stress to the final position. Liu (2009) added that Chinese emotional intonations are predominantly falling, shaped by pitch range, sentence stress, and syllable length.

2.4 The Studies on the Language Proficiency of Learners

Few studies have explored the link between L2 proficiency and emotional prosody perception, and existing findings are conflicting (Li, 2021). Altro (2013) and Zhu

(2013) reported that learners with more L2 experience outperformed those with less, while Paone and Frontera (2019) found that Russian L2 learners of Italian performed as well as native Italians in emotional prosody perception—supporting the "in-group effect" for geographically distant languages. In contrast, Bhatara et al. (2016) observed that high-proficiency French learners of English had lower accuracy in recognizing positive emotions, suggesting that advanced L2 ability may sometimes interfere with prosodic perception.

Methodological differences may explain these inconsistencies. Ji et al. (2018) grouped learners by CET-4/CET-6 scores and found that Chinese learners' perception of English boundary tones was comparable to native speakers, but they struggled with pitch accents , particularly focus-related distinctions. Zhang and Chen (2023) noted that high-proficiency learners are better at inhibiting irrelevant semantic information during prosodic processing, while low-proficiency learners rely more on L1 prosodic patterns. Chen (2008) and Wang and Liu (2021) further demonstrated that L1 transfer negatively impacts L2 prosodic production, with low-proficiency learners exhibiting greater difficulties in stress and tone perception.

2.5. The Studies on the Gender of Learners

Gender-related research in emotional prosody focuses on two aspects. The first is the gender of the speaker, acoustic differences in emotional speech, and the gender of the listener, which is a perceptual bias. Neuhaus et al. (2024) identified fundamental frequency (F0) and vocal tract length as key determinants of speaker gender perception: higher F0 and shorter vocal tract length increase the likelihood of female classification. Tripp and Munson (2022) expanded this, noting that males typically have lower F0, longer vocal tracts, and narrower vowel space, while females exhibit higher F0, shorter vocal tracts, and more dispersed vowels. Suprasegmental features, such as intonation, speech rate, and pauses, also differ between genders and influence perception (Tripp & Munson, 2022).

For listeners, gender differences in emotional prosody perception are less consistent. Some studies report that females are more accurate in recognizing emotions (Li, 2009), while others find no significant effects (Wang & Ding, 2012). This inconsistency may stem from variations in stimuli, task design, or cultural context—gaps that the current study aims to address.

3. Method

On the basis of the literature review, this part mainly describes the research questions, participants, stimuli, and procedure of the experiments in this paper. It is divided into four parts: the first part is about research questions; the second is about the participants; the third is about the stimuli; and the last is about the procedure.

3.1. Research Questions

This paper tries to answer three questions: first, does language proficiency influence the results of perception? Second, does gender influence the perception? Third, do the types of emotion influence the results of perception? The following text will answer them one by one.

3.2. Participants

Two groups of participants, 60 students in total, participated in the experiment. In order to complete the goal, 60 Chinese L2 learners of English with different level (measured by their Cet-4 scores) were recruited to investigate the influence of language proficiency and gender upon the capability of emotional perception. One hundred students were randomly investigated with their Cet4 scores, and then we chose the last thirty students in the 50 people below the median score (with the highest point being 474 and the lowest 314) as the unskilled group. Also, we chose the first 30 students from the other 50 people (with the highest score of 610 and the lowest 487) as the skilled group. The first group of participants (Group A) consists of 30 college students (fifteen females and fifteen males) and is a less proficient group. The second group of participants (Group B) consists of 30 college students (fifteen males and fifteen females), with proficient English level. They were either undergraduate or graduate students and reported having no hearing impairment. The average age is 21.45, ranging from 19 to 26. and have studied English for more than twelve years. Both groups of participants are from the Shandong region and do not have the experience of going abroad. All subjects provided written informed consent prior to beginning the study.

3.3. Stimuli

The stimuli used in the present study consist of 2 sentence samples read by 4 English speakers, two males and two females, with four emotions: fear, happiness, sadness, and neutrality. The neutral sentence is the baseline. The semantic meaning of the two sentences is controlled and will not confuse subjects in their understanding, and that controls the impact brought by the semantic aspect; they are "Kids are talking by the door" and "Dogs are sitting by the door". Each sentence of the stimuli has a duration between 1 and 3 seconds and is of acceptable quality for performing acoustic analysis. All the speakers of these sentences are from North America, and the quality is validated. In the test, there are another ten sentences read by another two speakers used as a practice block.

3.4. Procedure

All participants of this study signed the informed consent form, and after the experiments, they completed a questionnaire, which inquired about whether there were any difficulties in understanding the stimuli. The results showed that the semantic meaning did not interfere with the perception of the prosody.

All participants taking part in these experiments are in quiet places, like a class without noise to disturb them. Subjects were seated in front of a computer; the program of the experiment was run using E-prime software version 3.0. They entered the experiment and received instructions for an unlimited amount of time. After reading the instructions,

the participants pressed the "X" key to enter the exercise module to deepen their familiarity with the experimental process.

The exercise module consists of ten stimuli different from the formal experimental module. First, the fixation point appears for 500ms. After the fixation point disappears, a stimulation page will appear, and E-Prime will play a stimulus randomly. Participants need to press the "D", "F", "J", and "K" keys to select the emotion that appears in the sentence, and after making a selection, it will automatically jump to the next interface. After that, the subject presses the "X" key, and the next fixation point for the stimulus appears, and so on. This study allows participants to enter the next stimulus according to their own intention, by which they can choose to relax when feeling tired. The stimulation for both the exercise module and the formal experiment is played randomly, and the order of options has also been disrupted. After the exercise module is completed, participants can choose to practice again or enter the formal trial. Press "P" to enter the formal trial, and press "Q" to practice again.

The formal experiment has 32 stimuli, different from those in the practice part, and the experimental procedure is the same as the exercise module. It was not possible to return to the previous page or to proceed to the next one without having marked the response on the scale. After gaining the data, this paper used Excel and SPSS to analyze them and used Praat to extract the acoustic features of the stimulus sentences.

4. Results and Discussion

The results of this study are the response accuracy and reaction time. The average accurate rates and reaction time of each type of emotion of the two groups were logged, but the reaction time was not discussed because there was no big difference between the two groups and the two genders. The results are as follows.

4.1. Results

Using SPSS, a T-test and ANOVA were conducted on the exam data. In the T-test, this paper examined whether there is a significant difference between the skilled group and the unskilled group. In the ANOVA, this paper investigated which factor among language proficiency, types of emotions, and gender has a crucial impact on the perception results.

4.1.1. The Effect of Language Proficiency

According to Levene's test, the variances of each emotion in the two groups are equal to each other (p=0.806). In Table 1, it is easy to see that the accurate rates of each type of emotion in the skilled group are different from that in the unskilled group (Skilled group: happiness 55%, sadness 51%, fear 47%, neutrality 76%. Unskilled group: happiness 49%, sadness 59%, fear 40%, neutrality 72%). The rates in the skilled group are higher than those in the unskilled group, but there is no statistical significance in the T-test (p=0.314, 0.368, 0.941, 0.984) and in ANOVA in Table 2 (F=0.636, p=0.426).

Table 1. The mean correct rates of the skilled and unskilled groups

	Neutrality	Happiness	Sadness	Fear	Total
Skilled group	77%	55%	51%	47%	57%
Unskilled group	72%	49%	59%	41%	55%
Total	74%	52%	55%	44%	56%

Table 2. The results of the test of the between-subjects

Source	Df	Mean square	F	Sig	Partial Eta square
Corrected model	7	0.444	12	<0.001	0.266
Intercept	1	75.8	2049	<0.001	0.898
Group	1	0.024	0.636	0.426	0.003
Emotion	3	0.962	26	<0.001	0.252
Group*emotions	3	0.066	1.79	0.15	0.023

Table 3. The multiple comparisons of different emotions.

Emotions(I)	Emotions(J)	Mean difference	Sig	95% Confidence Interval	
				Lower bound	Higher bound
Fear	Happiness	-.008	0.018	-0.152	-0.141
	Neutrality	-0.3	<0.001	-0.369	-0.230
	Sadness	-0.11	0.001	-0.183	-0.04
Happiness	Fear	0.08	0.018	0.0141	0.152
	Neutrality	-0.21	<0.001	-0.285	-0.147
	Sadness	-0.03	0.374	-0.1	0.037
Neutrality	Fear	0.3	<0.001	0.23	0.369
	Happiness	0.21	<0.001	0.147	0.285
	Sadness	0.18	<0.001	0.116	0.254
Sadness	Fear	0.11	0.001	0.045	0.183
	Happiness	0.031	0.374	-0.037	0.1-
	Neutrality	-0.18	<0.001	-0.254	-0.116

Table 3, which shows multiple comparisons, shows that neutrality and fear have a significant difference in terms of mean, but happiness and sadness are not as significant as the others. Emotions(I) means the type of emotion used to compare, and J means the ones used to be compared. The p-value is less than 0.05, which means the two types of emotions have a significant difference; if not, there is no significant difference.

4.1.2. The Effect of Gender

In Table 4, the performance of males is different from females; the correct rates of males in the skilled group are happiness 53%, sadness 57%, fear 46%, neutrality 72%, while the rates of females are happiness 58%, sadness 46%, fear 48%, neutrality 80%. The test of Levene shows that the variances are equal in each emotion of both males and females. The accuracy rates in unskilled males are happiness 48%, sadness 59%, fear 45%, neutrality 71%, and those of the females are happiness 50%, sadness 59%, fear 37%, neutrality 73%. The result of Leneve's test shows that the variances of four emotions are equal (p=0.375). According to the results of ANOVA, in Table 5, the influence of the factor of gender is not as significant as emotion (F=0.141, p=0.708). In Table 6, the multiple comparisons show the same results as those in the language proficiency. All the correct rates are above the chance rate, which is 25%; thus, it is easy to see that the results have validity.

Table 4. The descriptive statistics of the mean correct rates of emotions between genders

Gender	Emotions	Mean	SD	N
Female	Fear	46%	0.18	30
	Happiness	53%	0.17	30
	Neutrality	75%	0.18	30
	Sadness	52%	0.23	30
	Total	57%	0.22	120
Male	Fear	41%	0.2	30
	Happiness	51%	0.19	30
	Neutrality	72%	0.17	30
	Sadness	58%	0.18	30
	Total	56%	0.21	120
Total	Fear	44%	0.19	60
	Happiness	52%	0.18	60
	Neutrality	74%	0.18	60
	Sadness	55%	0.21	60
	Total	56%	0.22	240

Table 5. The tests of between-subjects effects

Source	Df	Mean square	F	Sig	Partial
squared					
Corrected model	7	0.426	11.367	<0.001	0.255
Intercept	1	75.797	2020.6	<0.001	0.897
Gender	1	0.005	0.141	0.708	0.001
Emotions	3	0.962	25.642	<0.001	0.249

Table 6. Multiple comparisons of emotions between genders

Emotion(I)	Emotion(J)	Mean difference	Sig	95% Confident Interval
Fear	Happiness	-0.08	0.19	-0.153
	Neutrality	-0.3	<0.001	-0.369
	Sadness	-0.11	0.001	-0.184
Happiness	Fear	0.08	0.019	0.013
	Neutrality	-0.21	<0.001	-0.286
	Sadness	-0.03	0.378	-0.1
Neutrality	Fear	0.035	<0.001	0.23
	Happiness	0.216	<0.001	0.147
	Sadness	0.185	<0.001	0.115
Sadness	Fear	0.114	0.001	0.044
	Happiness	0.031	0.378	-0.038
	Neutrality	-0.185	<0.001	-0.255

4.1.3. The effect of Types of Emotions

As for the wrong answers, Table 7 A and B show a similar pattern in both the skilled and unskilled groups, as well as between males and females. When the correct answer is sadness, most of the participants who are wrong choose neutrality. In the skilled group, 75% of the males and 88% of the females, and in the unskilled group, 67% of the males and 71% of the females.

When the correct answer is happiness, participants who select the wrong answer most choose neutrality (60% of the males and 78% of the females in the skilled group, and 68% of the males and 68% of the females in the unskilled group). When the correct answer is fear, most of the people who are wrong choose sadness and neutrality (In the skilled group, 40% of the males choose sadness, and 32% of them choose neutrality, 51% of the females choose sadness, and 37% choose neutrality. In the unskilled group, 57% of the males choose sadness 28% of them choose neutrality, 55% of the females choose sadness, and 33% of them choose neutrality. When the correct answer is neutrality, most people who are wrong choose sadness (68% of the males and 83% of the females in the skilled group, and 63% of the males and 54% of the females in the unskilled group). There is a rule for the wrong answers of the subjects, regardless of whether they are male or female, in the skilled group or in the unskilled group. To sum up, neutrality is the most common emotion students tend to choose, and it is easily confused with sadness and happiness. Fear is the most difficult emotion for subjects to distinguish, and it is confused with sadness and neutrality.

Table 7. The matrix of answer (A) and the most easily confused types of emotion (B)

A

Emotions	Neutrality	Happiness	Sadness	Fear
Neutrality	74%	1%	17%	8%
Happiness	33%	52%	7%	8%
Sadness	35%	3%	55%	7%
Fear	20%	9%	27%	44%

B

Correct answer	Wrong answer	Male in skilled group	Female in skilled group	Male in an unskilled group	Female in an unskilled group
Neutrality	Sadness	68%	83%	54%	63%
Happiness	Neutrality	60%	78%	68%	68%
Sadness	Neutrality	75%	88%	67%	71%
Fear	Sadness	33%	51%	57%	55%
Fear	Neutrality	40%	37%	28%	33%

4.2. Discussion

The results are not consistent with the studies of Zhu (2013), Altro (2013), and Paone and Frontera (2019), but in line with that done by Bhatara (2016). However, in the study done by Bhatara, the higher the language proficiency, the more difficult the recognition of positive emotion, without an impact on the negative ones. But in this study, it has been found that language proficiency will help the recognition of the positive emotion, but has a negative effect on the distinction of the negative emotions, like sadness.

The difference in the results between this paper and other research is due to the difference between Chinese emotional prosody and English. When people hear prosody in other languages, they tend to interpret it based on the prosodic patterns of their mother tongue. First, the difference is embodied in the acoustic parameters exhibited by the emotional prosody of English and Chinese. According to Gong (2022), Chinese emotional prosody shares some common features in pitch and intensity. What is more, Gong has his correct rates of neutrality, happiness, sadness, and fear, which are 41%, 36%, 48% and 30%, which are lower than those of this study. This study compared these features in Chinese emotional prosody with those in our stimuli and found that there is a difference between them. First, using the software of Praat, the acoustic parameters of our stimuli were analyzed. Taking the sentences uttered by male speakers as an example, the mean pitch of neutrality is 112.9Hz, happiness 237Hz, sadness 207.1Hz, fear 143.2Hz (happiness>sadness>fear>neutrality), and the SD of pitch of neutrality is 18.2Hz, happiness 47Hz, sadness 34Hz, and fear 17.7Hz, as shown in Table 8. Secondly, in Table 9, the pitch range of the four emotions is that neutrality is 63.8Hz from 86.5 to 150.3Hz, happiness is 271.6Hz from 77.3 to 348.9Hz, sadness is 229.3Hz from 85.1 to 314.4, and fear is 66.3Hz from 107.1Hz to 183.4Hz (happiness=sadness > fear=neutrality). Thirdly, the duration of the four emotions is that neutrality 1.51s, happiness 1.35s, sadness 1.32s, and fear 1.26s. Fourthly, the intensity of the four emotions is that neutrality 33.8dB, happiness 76dB, sadness 63.4dB, and fear 66dB.

On the other hand, those parameters in Chinese are shown in Table 9. The means of pitch are neutrality 181.94Hz, happiness 246.21Hz, sadness 182.72Hz, and fear 166.97Hz (happiness>sadness = neutrality>fear). The ranges of the pitch are neutrality 168.78Hz from 111.1 Hz to 279.88Hz, happiness 245.03Hz from 145.75Hz to 390.8Hz, sadness 135.29Hz from 124.33Hz to 259.62Hz,

and fear 129.75Hz from 112.9Hz to 242.7Hz (happiness>neutrality > sadness>fear). The intensity of four emotions is neutrality 77.2dB, happiness 77.9dB, sadness 77.12dB, while the pitch of fear is 75.81dB (happiness>neutrality = sadness>fear).

Table 8. The parameters of Chinese and English emotions Acoustic parameters (mean) in English emotions

Emotions	Mean of Pitch (Hz)	Max pitch (Hz)	Min pitch (Hz)	Pitch SD (Hz)	Pitch range (Hz)	Intensity (dB)	Duration (s)
Neutrality	112.9	150.3	86.5	18.2	63.8	33.8	1.51
Happiness	237	348.9	271.6	47	237	76	1.35
Sadness	207.1	314.4	85.1	34	229.3	63.9	1.32
Fear	143.2	183.4	107.1	17.7	66.3	66	1.26

Table 9. Acoustic parameters (mean) in Chinese emotions

Emotions	Mean of Pitch (Hz)	Max pitch (Hz)	Min pitch (Hz)	Pitch SD (Hz)	Pitch range (Hz)	Intensity (dB)	Duration (s)
Neutrality	181.9	279.8	111.1	37.1	168.8	77.2	2.72
Happiness	246.2	390.7	145.7	57.3	245.0	77.9	2.65
Sadness	182.7	259.6	124.3	26.4	135.2	77.1	3.1
Fear	166.9	242.7	112.9	29.3	129.7	75.8	3.1

According to Gong (2022), the correct rates of neutrality are in positive relation with the duration (the standardized coefficient is 0.291), but in negative relation with the mean of pitch, minimum of pitch, the SD of pitch, and the intensity (the standardized coefficients are -0.432, -0.267, -0.03, -0.119). That may be why the correct rate of neutrality in this paper is higher than that in Gong's, as the mean and minimum of pitch and the intensity of our research are lower than those in Gong's study. In terms of the happiness, according to Gong, the rate of correct recognition is in positive relation with the duration (0.049), mean of the pitch (0.025) and the SD of the pitch (0.044), but it is in negative relation with the minimum of the pitch (-0.026), the maximum of the pitch (-0.207). Therefore, the correct rates of happiness are close to each other.

Sadness is in positive relation with the duration (0.292), mean of the pitch (0.702), the maximum of the pitch (0.258), but it is in negative relation with the SD of pitch (-0.463), the minimum of pitch (-0.436), and the intensity (-0.132). Given that the mean maximum of pitch and the intensity in English sentences are higher than the Chinese ones, and the minimum of pitch is lower, the correct rate of this paper in sadness is a little higher than that in Gong's study. The fear is in positive relation with duration (0.123), mean of pitch (0.036), SD of the pitch (0.092), the minimum of pitch (0.029), but in negative relation with the maximum of pitch (-0.202) and intensity (-0.086). Therefore, the correct rate of fear in this study is a little higher than that of Gong's, as the mean and the maximum of pitch and the intensity of English stimuli are lower than those of Gong's.

The acoustic parameters may account for part of the reasons why this paper has found results different from other studies. Another reason is the culture. Bhatara (2016) has spotted that when the listeners are from two cultures that are far from each other, it will be difficult for them to decode the emotional prosody of the opposite culture. Additionally, she also mentioned the "in-group" hypothesis, which argues that people living in the same language community could understand some types of emotion easily, and the mistakes they made in their mother language will be transformed i n to the second language. Thus, the difference in the results may be the consequence of the cultural element.

The reason why neutrality is the most chosen emotion can be attributed to its duration and pitch. The duration of neutrality is the longest among the four emotions, and according to Gong, it may lead to the choice of neutrality. The reason why people do not choose happiness, sadness, and fear but choose neutrality may be that the pitch of neutrality is close to others, and there is no fluctuation in the F0, which makes it easy to recognize it.

Additionally, in a study conducted by Gong, when the correct answer is happiness, many of the people who made the wrong answer chose neutrality (37%) among the other five wrong answers. When the correct answer is sadness, 24% of the subjects chose neutrality among the sixth emotions in total, and when the answer is fear, 58% of the subjects chose sadness. The perception of the English prosody may be influenced by the perception of the Chinese emotional prosody, which is another reason for the results.

According to Kong and Yang (2023), the emotions of neutrality and sadness cannot be separated by the pitch and duration; that is to say, they need other acoustic parameters to be recognized. And the emotion of happiness and neutrality can be perceived by the pitch and duration, but the stimuli in this study may not be so significant, thus participants did not distinguish them completely.

The AM theory can also partially explain the findings in this study. It describes the intonation into two kinds of tonal events, H and L. From Figures 1 and 2, by the bule lines, in the intonation of sadness, the intonation can be described as LHL%, and the intonation of neutrality can be described as HL%. The end part of the intonation of the two emotions is similar, thus they are easily mistaken.

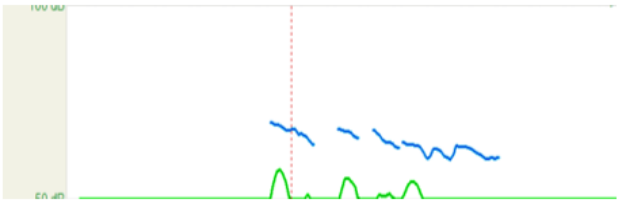

Fig. 1 The curve of the intonation of sadness in English

Fig. 2 The curve of the intonation of neutrality in English

In the same way, in Figure 3, the intonation curve of fear is the same as that of sadness and neutrality. It can be described as L*HL%, because the beginning part of it is a little higher than the pitch accent. But as a drop tonal event, it is still seen as a L.

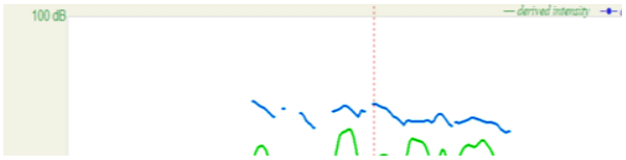

Fig. 3 The curve of the intonation of fear in English

Fig. 4 The curve of the intonation of happiness in English

From Figure 4, it can be seen that the last tone of happiness is H, different from the other three types of emotion; thus, it is the most difficult to confuse with the other three types.

Lastly, according to Chen (2008), this difference may be attributed to the fact that Chinese is a tonal language, whereas English is a stress-timed language. Chinese people depend on prosody to distinguish meaning and are thus sensitive to it. English people do not need to do this; therefore, the emotional prosody is easier to recognize.

5. Conclusion

This paper explores the influence of language proficiency, gender, and types of emotion upon the perception of emotional prosody. And it is affected by these three factors. The higher the language proficiency, the higher the correct rates, though there is no statistically significant effect. Gender will not influence the results of perception. Emotions have the greatest impact on the outcome of perception, beyond language proficiency and gender. In terms of the types of emotion, the emotion of neutrality is the easiest for Chinese students to recognize, sadness is the second easiest emotion, and happiness and fear follow. What is more, neutrality is always confused with sadness and happiness by the Chinese learners, and fear is always confused with sadness and neutrality. That may be accounted for by the influence of the emotional prosody in Chinese, and the tonal events of the stimuli of fear and neutrality are the same as those of sadness and happiness.

The outcome shows little difference between males and females, which means that males and females possess, to some extent, a similar perception system, or the emotional prosody is universal toward both sexes. The reason for the difference in perception between Chinese learners and native English speakers is probably due to the different acoustic parameters they use in communication. In terms of neutrality, the mean of the pitch and the duration may be the principal influential acoustic factor. Mean of the pitch is the most influential to sadness, and duration to the happiness, range of pitch and duration to fear.

This paper cannot be said to be perfect, and it has some drawbacks. The first is that the emotions this paper studied are small. The emotions in the discrete emotion model are several, and some of them are from the same family. On the other hand, the dimensional model is also suitable to describe emotion; its valence and motivation can be adopted to measure the reaction of people towards the various emotional stimuli. What is more, the number of subjects is relatively small, and the difference in language proficiency, which may be the reason why there is no significant result in this study.

References

[1] Rene Altrov, "Aspects of Cultural Communication in Recognizing Emotions," *Trames*, vol. 17, no. 2, pp. 159-174, 2013. [Google Scholar] [Publisher Link]

[2] Amalia Arvaniti, "The Auto-Segmental Metrical Model of Intonational Phonology," *Netherlands: Radboud University Press*, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] Rainer Banse, and Klaus R. Scherer, "Acoustic Profiles in Vocal Emotion Expression," *Journal of Personality and Social Psychology*, vol. 70, no. 3, pp. 614-636, 1996. [CrossRef] [Google Scholar] [Publisher Link]

[4] Anjali Bhatara et al., "Second Language Ability and Emotional Prosody Perception," *PloS One*, vol. 11, no. 6, pp. 1-13, 2016. [Google Scholar]

[5] Paul Ekman, *Darwin and Facial Expression*: *A Century of Research in Review*, New York: Academic Press, 1973. [Google Scholar] [Publisher Link]

[6] Paul Ekman, "An Argument for Basic Emotions," *Cognition & Emotion*, vol. 6, no. 3-4, pp. 169-200, 1992. [CrossRef] [Google Scholar] [Publisher Link]
[7] J. Fontaine, *The Oxford Companion to Emotion and the Affective Sciences*, Oxford: Oxford University Press, 2014. [Google Scholar]
[8] Bingyan Gong et al., "The Mandarin Chinese Auditory Emotions Stimulus Database: A Validated set of Chinese Pseudo-Sentences," *Behavior Research Methods*, vol. 55, pp. 1441-1459, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[9] Esther Grabe et al., "Perception of English Intonation by English, Spanish, and Chinese Listeners," *Language and Speech*, vol. 46, no. 4, pp. 375-401, 2003. [CrossRef] [Google Scholar] [Publisher Link]
[10] Carlos Gussenhoven, and Aoju Chen, "Universal and Language-Specific Effects in the Perception of Question Intonation," *Sixth International Conference on Spoken Language Processing*, ICSLP 2000, Beijing, China, pp. 1-4, 2000. [Google Scholar] [Publisher Link]
[11] Paul E. Griffiths, *What Emotions Really Are: The Problem of Psychological Categories*, Chicago: University of Chicago Press, 2007. [Google Scholar] [Publisher Link]
[12] X.L. Ji, H. Zhang, and A.J. Li, "An Empirical Study of Chinese Learners' Perception of English Intonation," *Foreign Language Teaching and Research*, vol. 50, no. 3, pp. 393-406, 2018. [Online]. Available: https://kns.cnki.net/kcms2/article/abstract?v=zE0--Q1IihQ3vXtw5ez8rPmVG_A0sj3YqN3pXK6PUd-1Y_G9V5iZUhwuri1ajRdV-QaugKkMxjzy1oITEVeqM0eBYW25WI69qfsNvqXv7CVKafoPz0eSwYWqlBzGYa1dccsxHutq1fLhe4jLlzuOBKY4XStqz0nTB76MMXJ0p7gyNLxV7-BiVg==&uniplatform=NZKPT&language=CHS
[13] Pauline Larrouy-Maestri, David Poeppel, and Marc D. Pell, "The Sound of Emotional Prosody: Nearly 3 Decades of Research and Future Directions," *Perspectives on Psychological Science*, vol. 20, no. 4, pp. 623-638, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[14] Petri Laukka, Patrik Juslin, and Roberto Bresin, "A Dimensional Approach to Vocal Expression of Emotion," *Cognition & Emotion*, vol. 19, no. 5, pp. 633-653, 2005. [CrossRef] [Google Scholar] [Publisher Link]
[15] A.J. Li, "Stress Patterns in Emotional Sentences, Song," *Proceedings of the 7th National Conference on Phonetics and International Forum on Frontiers of Phonetics*, Beijing: Foreign Language Teaching and Research Press, pp. 132-140, 2006. [Online]. Available: https://www.zhangqiaokeyan.com/conference-cn-22870/
[16] J.N. Li, and J. Wu, "Chinese Learners' Perception of English Emotional Intonation," *China Foreign Language Studies*, vol. 8, no. 1, pp. 54-62, 2021. [Online]. Available: https://kns.cnki.net/kcms2/article/abstract?v=zE0--Q1IihTXLaGpiIMa1D5SxT7yva7EVc_jkochpfAefMT8O3BSo10UfAWSwpyxWEr3lmIEsAERJyJQGe2SxnoZ1JwaQVL9XLwS8U1aznjzm7sb6RpLGuDgB3gQdK73_n4PmV3vBk6c87M49FGoZtjn-gbje6Dk7hXJocvE76JLm9l_i8N4YQ==&uniplatform=NZKPT&language=CHS
[17] T. Lin, and L.J. Wang, *A Course in Phonetics*, Beijing: Peking University Press, 2016. [Google Scholar]
[18] Ineke Mennen, "Beyond Segments: Towards a L2 Intonation Learning Theory," *Prosody and Language in Contact*/Springer: Berlin Heidelberg, pp. 171-188, 2015. [CrossRef] [Google Scholar] [Publisher Link]
[19] Taylor J. Neuhaus, Ronald C. Scherer, and Jason A. Whitfield, "Gender Perception of Speech: Dependence on Fundamental Frequency, Implied Vocal Tract Length, and Source Spectral Tilt," *Journal of* Voice, pp. 1-17, 2024. [CrossRef] [Google Scholar] [Publisher Link]
[20] Marta Ortega-Llebaria, and Laura Colantoni, "L2 English Intonation: Relations between form Meaning Associations, Access to Meaning and L1 Transfer," *Studies in Second Language Acquisition*,vol. 36, no. 2, pp. 331-353, 2014. [CrossRef] [Google Scholar] [Publisher Link]
[21] Emanuela Paone, and Manuela Frontera, "Emotional Prosody Perception in Italian as a Second Language," *Proceedings of the 10th International Conference of Experimental Linguistics*, *Exling Society*, pp. 161-164, 2019. [Google Scholar] [Publisher Link]
[22] Marc D. Pell, and Vera Skorup, "Implicit Processing of Emotional Prosody in a Foreign Versus Native Language," *Speech Communication*, vol. 50, no. 6, pp. 519-530, 2018. [CrossRef] [Google Scholar] [Publisher Link]
[23] Elliott D. Ross, Jerold A. Edmondson, and G. Burton Seibert, "The Effect of Affect on Various Acoustic Measures of Prosody in Tone and Nontone Languages: A Comparison Based on Computer Analysis of Voice," *Journal of Phonetics*, vol. 14, no. 2, pp. 283-302, 1986. [CrossRef] [Google Scholar] [Publisher Link]
[24] Klaus R. Scherer, "Vocal Affect Expression: A Review and a Model for Future Research," *Psychological Bulletin*, vol. 99, no. 2, pp. 143-165, 1986. [CrossRef] [Google Scholar] [Publisher Link]
[25] K.R. Scherer, "Psychological Models of Emotion," *The Neuropsychology of Emotion*, vol. 137, no. 3, pp. 137-162, 2000. [Google Scholar] [Publisher Link]
[26] Klaus R. Scherer, and Paul Ekman, *Approaches to* Emotion, Psychology Press, 2014. [CrossRef] [Google Scholar] [Publisher Link]
[27] W. da Silva, P.A. Barbosa, "Perception of Emotional Prosody: Investigating the Relation between the Discrete and Dimensional Approaches to Emotions," *Journal of Language Studies*, vol. 25, no. 3, pp. 1075-1103, 2017. [CrossRef] [Google Scholar]
[28] Craig A. Smith, and Phoebe C. Ellsworth, "Patterns of Cognitive Appraisal in Emotion," *Journal of Personality and Social Psychology*, vol. 48, pp. 813-838, 1985. [Google Scholar] [Publisher Link]
[29] Silvan S. Tomkins, *Affect Theory*, East Sussex: Psychology Press, 2014. [Google Scholar] [Publisher Link]
[30] Alayo Tripp, and Benjamin Munson, "Perceiving Gender while Perceiving Language: Integrating Psycholinguistics and Gender Theory," *Wiley Interdisciplinary Reviews: Cognitive Science*, vol. 13, no. 2, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[31] H.B. Wang, and A.J. Li, "Establishment of a Mandarin Emotional Speech Corpus and Perception Experiments," *Proceedings of the 6th National Conference on Modern Phonetics*, Beijing: Tsinghua University Press, pp. 119-124, 2003. [Online]. Available: https://kns.cnki.net/kcms2/article/abstract?v=zE0--Q1IihTuAxQ0yADeyqw3hFZrUEnnhumwfQYW6ZBRzCi88x_nJh5ikAYh_GU1_Z4bWWVT1UebObJADeo-wxndyONRrvaj-IxBg-lJXIECYKjfeS95rcWlXyLJeAUY_xADSPcoVqsA1wdGera1nDJGlF449JcGoiVFM65mbqvwcs6UYYjMX0d4fCAHhq2e&uniplatform=NZKPT&language=CHS

[32] Wang Libin, and Liu Yafang, "An Experimental Study on the Influence of Second Language Proficiency on English Word Stress Production," *Journal of Chongqing University of Science and Technology*, no. 2, pp. 113-118, 2021. [CrossRef] [Publisher Link]

[33] E. Williams, and N. Kenneth, "Emotions and Speech: Some Acoustical Correlates," *The Journal of the Acoustical Society of America*, vol. 52, pp. 1238-1250, 1972. [Google Scholar]

[34] J. Yang, and J.P. Kong, "A Study of Perceptual Categories of Mandarin Emotional Speech Based on Intonation," *Applied Linguistics*, no. 4, pp. 72-81, 2023. [Online]. Available: https://kns.cnki.net/kcms2/article/abstract?v=zE0--Q1IihSdljNLzl7M_3qhznYNmCcOHhn4ZaRftHG9b2seMO018bpxUZH8572B-y69rGQkAPlekDZ5tUFsbj3GhaWFbdS2XAnLSWrmQkADEDYjxqZCM7edJeqwDPn0BNFWm2F6d358i5QUh1I2HN83cJtMPY9oB0qE4zCcCgYXsuPpJcfH_Q==&uniplatform=NZKPT&language=CHS

[35] G.D. Zhang, and F. Chen, "A Study of Semantic-Prosodic Processing of Emotional Words by Chinese English Learners of Different Proficiency Levels," *Foreign Language Teaching and Research*, vol. 55, no. 1, pp. 79-90, 2023. [Online]. Available: https://kns.cnki.net/kcms2/article/abstract?v=zE0--Q1IihReMCkVGYEpPwq5i9cQ1SznLK2zDWX7q8jVicpNTh-lBpXs19g1KJsFWeMzAGnWRl9uFlSeNwRl9lSAWLiP2AiyqEUCp31YmmFd_v1KurHL8LUok_O983KYQf0HtIm64x3aUG9XWgmILvtt-6eGPvJ7YxWun__T1_lcJ4HXejsTYA==&uniplatform=NZKPT&language=CHS

[36] Yinyin Zhu, "Which is the Best Listener Group: Perception of Chinese Emotional Prosody by Chinese Natives, NaïveDutch Listeners and Dutch L2 Learners of Chinese," *Dutch Journal of Applied Linguistics*, vol. 2, no. 2, pp. 170-183, 2013. [CrossRef] [Google Scholar] [Publisher Link]