THE JOURNAL OF ASIA TEFL
Vol. 19, No. 1, Spring, 2022, 125-140
http://dx.doi.org/10.18823/asiatefl.2022.19.1.8.125

The Journal of Asia TEFL

http://journal.asiatefl.org/

e-ISSN 2466-1511

No Prosody, No Emotion: Affective Prosody by Chinese EFL Learners

Zepeng Wang
Harbin Engineering University, China

Yansheng Mao
Harbin Engineering University, China

The current paper intended to investigate the performance of affective prosody by Chinese English as a foreign language (EFL) learners. More specifically, the current study examined and compared aspects contributing to perceived vocal affect in the speech of Chinese EFL learners and native English speakers (hereafter NSs). The data were collected from 20 university students who took part in two recording sessions designed by the authors. In comparison to previous studies on L2 acquisition of English prosody, this research offers a detailed description of the performance of affective prosody in the speech of Chinese EFL learners. The results show that, for the majority of Chinese EFL learners, their pitch range and vowel duration are narrower and shorter than for NSs, all of which contribute to their low scores for affective involvement. In addition, a number of pedagogical suggestions have been proposed by this study, which will help Chinese EFL learners to deal with the prosodic problems they encounter when speaking English.

Keywords: Chinese EFL learners, acquisition, affective prosody, pitch range, vowel duration

Introduction

This study reports on an investigation into the performance of affective prosody by Chinese EFL learners. Successful second language (L2) pronunciation not only involves the authentic production of all the individual sounds, but also the acquisition of the L2's unique prosody (Mennen & de Leeuw, 2014). However, given the complexity of prosody, acquiring the prosody of an L2 is generally considered to be extremely challenging for most language learners (Bongaerts et al., 1997; Cruz-Ferreira, 1989). For example, the English prosody is often seen as "the final hurdle, which a vast majority of speakers of English as a foreign language never manage to cross" (Banjo, 1979, p. 12).

According to its different functions, prosody can be divided into stress prosody which entails decisions about semantic meaning, and affective prosody which conveys information about the interlocutors' emotional state (Bozikas et al., 2006; Edwards et al., 2002). Recent years have witnessed a growing academic interest in the prosodic aspects in the field of second language acquisition (SLA) (Hua & Li, 2016; Kashiwagi & Snyder, 2008; Mennen & de Leeuw, 2014; Mennen et al., 2014; Yang, 2016), and many important findings emerged. The most pertinent findings to the present study are that some English as a second language (ESL) learners may use narrower or less variable pitch range than native speakers (Backman, 1979; Busà & Urbani, 2011; Pickering, 2004), and ESL learners are likely to produce longer utterances compared to native speakers (Aoyama & Guion, 2007; Guion et al., 2000; Munro & Derwing, 1995). However, these studies are carried out from the perspective of stress prosody, very little is

currently known about the performance of affective prosody by ESL learners, even though affects are also conveyed by prosody and play an important role in interpersonal communication (Couper-Kuhlen, 1986; Hirschberg, 2004).

Against this backdrop, it is of enormous value to examine to what extent can ESL learners flexibly use prosody to convey affects in the target language. Accordingly, follow-up research focusing on this issue may help not only test the universality of previous findings in the studies that are conducted from the perspective of stress prosody but also map out the status quo of the acquisition and application of affective prosody by ESL learners, which is bound to be valuable to ESL teaching and learning. Thus, the current study set out to investigate the performance of affective prosody among Chinese EFL learners by comparing their pitch range and vowel duration with those of NSs, in both their spontaneous speech (hereafter SS) and read speech (hereafter RS). It is hoped that this research will timely contribute to a more comprehensive understanding of the prosodic acquisition by Chinese EFL learners.

Literature Review

Studies on Affective Prosody

In addition to its unique role in distinguishing semantic and syntactic information, prosody also conveys the affective state of an interlocutor, which is termed as affective prosody (Bozikas et al., 2006; Edwards et al., 2002). To date, empirical studies on affective prosody can be divided into three main categories: (1) those focusing on the relationship between affect and acoustic parameters, (2) those concerning the performance of affective prosody among particular groups, and (3) those considering the relationship between affective prosody and left/right hemisphere. In the following section, we will review categories (1) and (2) because they are the most relevant to the current study.

There is a large volume of published studies describing the acoustic parameters of human vocal affects (Banse & Scherer, 1996; Tartter, 1980; Williams & Stevens, 1972, to name only a few), among which fundamental frequency (F0; associated with perceived voice pitch) and duration (associated with perceived length) are the most frequently investigated parameters. One well-known study that is often cited in this field is that of Williams and Stevens (1972), who found that anger produced a higher pitch and wider pitch range than fear, sorrow, or being neutral. They also noted that the duration of utterances spoken in anger was generally longer. Later, to contribute to the understanding of the acoustic properties of affects of which only negative emotions had been dealt with, Tartter (1980) investigated the acoustic effects of smiling on speech. The results showed that smiling raised pitch in all participants, as well as the duration of speech in some participants. Along the same lines, many other experiments have been conducted. For instance, a detailed summary of the research evidence has been provided in the work of Banse and Scherer (1996; see also Murray & Arnott, 1993). To be specific, anger was generally characterized by an increase in mean pitch; fear was characterized by an increase in mean pitch and pitch range. Taken together, these studies indicate that the prosodic patterns including pitch and duration are not consistent and affective speech demonstrates a wider pitch range and longer duration than being neutral.

Similarly, pitch and duration also are the most frequently examined parameters in studies concerning the performance of affective prosody among particular groups (Reilly et al., 1990; Setter et al., 2007; Wang & Liang, 2013, to name only a few), which have become available during the past 30 years. The first systematic study in this field was reported by Reilly et al. (1990). In their study, they designed a storytelling task to investigate the use of affective vocal prosody among a small group of adolescents with Williams Syndrome (WS). The results reported that children with WS made extensive use of affective prosody in comparison to adolescents with Down's Syndrome. However, one deficiency of this study is that the researchers only used impressionistic data. To address the lack of acoustic measurements in Reilly et al.'s (1990) study, Setter et al. (2007) measured the pitch range and vowel duration in children with WS and compared these parameters with those of typically developing children. Their results

showed that compared to typically developing children, children with WS had a much greater average pitch range and tended to use longer vowels, which were consistent with Reilly et al.'s (1990) claim that the use of affective prosody by individuals with WS was somewhat aberrant. The evidence indicated a strong relationship between extensive use of affective prosody and children with WS. Nevertheless, these studies were based on children who spoke English as their native language, the universality of this finding needs to be verified by further research which focuses on children whose native language is not English. As such, to further examine the universality of the previous findings, Wang and Liang (2013) designed a storytelling task to explore the affective prosody of mentally retarded Chinese children. The results showed that these children tended to have a greater pitch range and used longer vowels than their typically developing peers, a finding which was in line with Setter et al.'s (2007) study on children from the UK.

Overall, these studies provide reasonably consistent evidence of an association between the extensive use of affective prosody (wider pitch range and longer duration) and atypically developing populations, irrespective of their native language. However, far too little attention has been given to the performance of affective prosody by L2 learners, an issue that is worth more focused studies in SLA. In this connection, we argue that a timely investigation into the performance of affective prosody by L2 learners can serve as the departure for finding out the status quo of the affective expression by L2 learners in interpersonal interaction and instructing them to appropriately use prosody to convey affects since inappropriate prosodic encoding of affects can wreak havoc and cause significant damage to interpersonal relationships.

Before we introduce this study, the following section will review existing studies on ESL learners' prosodic performance, especially of pitch range and duration, as they are the most relevant prosodic parameters to the current study.

Studies on the Prosodic Performance of ESL Learners

To date, many studies have examined the pitch range and duration of ESL learners. However, most of the experiments were designed to investigate the performance of prosody at the lexical, utterance, or discourse level, while the affective prosody of ESL learners is seldom studied and it is unclear to what extent they can adjust pitch range and duration to appropriately convey affects in the target language. Thus, the following review focuses on the investigation of the pitch range and duration of ESL learners from the perspective of stress prosody.

There is evidence that some ESL learners may use narrower or less variable pitch range than native speakers (Backman, 1979 for Spanish ESL learners; Busà & Urbani, 2011 for Italian ESL learners; Pickering, 2004 for Chinese EFL learners), which made the identification of prosodic units difficult (Pickering, 2004; Wennerstrom, 1998). For example, Backman (1979) examined the intonation errors in the speech of eight native Venezuelan Spanish ESL learners. The results of analyses indicated that these learners' pitch ranges were narrower than those of native American English speakers. Busà and Urbani (2011) made a similar point in their study testing the hypothesis that Italian ESL learners have a narrower pitch range and less pitch variation than native English speakers. Consequently, the pitch ranges produced by 18 native American English speakers and 18 native Italian speakers were compared, and the results showed that the overall pitch ranges in Italian speakers' English utterances were narrower than those in native speakers. In addition, Italian speakers' English speech showed less pitch variation than native speakers.

Similar results can also be found in the speech by Chinese EFL learners. In a study investigating the systematic use of pitch and pause cues to create intonational paragraphs in teaching discourse, Pickering (2004) found that Chinese speakers were unable to consistently manipulate key and tone choices to create intonational paragraphs, and one contributor was their overall narrower pitch range. The Chinese participants who were teaching assistants (TA) were speaking more monotonously than their native-speaking colleagues, "Analysis of prosodic units defined primarily by pitch level was hampered at the outset by a compression of overall pitch range in the ITA teaching presentations as compared to the pitch ranges found in the NS TA data set.", as Pickering argued (2004, p. 31). Further evidence that Chinese EFL learners have a narrower pitch range can also be found in the study of Wennerstrom (1998). To be

specific, by conducting a quantitative study on the intonation of 18 Mandarin Chinese speakers giving academic lectures in English, Wennerstrom (1998) found that Chinese speakers generally made a smaller distinction between contrast words and given words than native English speakers, although some individual subjects performed better. She argued that "transfer could lead to less exploitation of the pitch range in contrasts" (Wennerstrom, 1998, p. 21), by drawing on the work of Shen (1990) who suggested that contrastive stress in Mandarin Chinese is achieved more through duration than pitch. In spite of the differences in the backgrounds of participants and stimuli materials, a lot of similar results may indicate that the narrower pitch range is a universal problem encountered by many ESL learners.

Previous research also has shown that ESL learners are likely to produce longer utterances than native speakers (Aoyama & Guion, 2007 for Japanese ESL learners; Guion et al., 2000 for both Italian and Korean ESL learners; Munro & Derwing, 1995 for Chinese EFL learners). In other words, ESL learners' utterances in English tend to be longer than those of native speakers. For example, Aoyama and Guion (2007) noted that the absolute durations of syllables and utterances tended to be longer in the native Japanese speakers' utterances than in those of native English speakers' speech. The authors also found that "the durations of function words were proportionately longer in the NJ speakers' utterances than in the NE speakers" (Aoyama & Guion, 2007, p. 294). It is suggested that function words were not as reduced in non-native speech as in native speech.

The relatively longer duration was also demonstrated in the English speech by both Italian and Korean speakers. In a study examining the correlation between the native Italian speakers' age of arrival in an English-speaking country and their duration of English sentences, Guion et al. (2000) found that the later the native Italian speakers arrived in Canada, the longer their sentences were. They also noted that the English sentences' durations produced by the latest-arriving groups of native Italian speakers were 15% longer than those of the native control group. To check whether the correlation is generalizable, the authors conducted another experiment on the sentence durations of native Korean speakers. The results showed that "the later in life the native Korean subjects began to learn English, the longer in duration their English sentences were" (Guion et al., 2000, p. 216) and the durations produced by the latest-arriving groups of native Korean speakers were 27% longer than those of the native control group. The comparatively longer duration can also be observed in the English speech by Chinese EFL learners. In a study examining the effect of a foreign accent on sentence processing time, Munro and Derwing (1995) had 10 adult native Mandarin Chinese speakers who had lived in Canada for 4 years on average and 10 native English speakers read 40 English sentences. The analysis of the duration data indicated that "the Mandarin talkers generally spoke more slowly than the native English speakers" (Munro & Derwing, 1995, p. 301).

It is thus clear from the above review that ESL learners tend to have a narrower pitch range and longer duration than native speakers. However, these studies were designed from the perspective of stress prosody, while the affective aspect of prosody remains largely unexamined. Against this backdrop, many questions about the prosodic acquisition of ESL learners remain unsolved and need to be answered, such as to what degree can ESL learners appropriately use prosody to convey specific affects. As such, the current study aims to examine the production of affective prosody by Chinese EFL learners, on the basis of pitch range and vowel duration (Cutler et al., 1997; Pell, 2001). In this sense, the paper may, from the perspective of affective prosody, add to the existing literature about the prosodic acquisition by Chinese EFL learners to shed new light on the opinions concluded from the perspective of stress prosody.

Methodology

Research Questions

As pointed out in the previous chapters, there exists a huge research gap in terms of prosodic acquisition by L2 learners. Accordingly, many crucial questions in this field remain unexplored. For

example, whether the previous findings from the perspective of stress prosody can be applicable to the context of affective prosody by L2 learners. Aiming at contributing to the existing literature on the acquisition of prosody by L2 learners, the current study intends to compare the pitch range and vowel duration of Chinese EFL learners with those of native speakers in both their SS and RS to answer the following research questions:

(1) Do Chinese EFL learners have difficulty in using prosody to convey affects? If yes, in what ways?
(2) Are there any differences in affective prosody within the group of Chinese EFL learners? If yes, what are these differences?

Participants

20 voluntary university students, 10 males, and 10 females, aged between 17 and 26, with a mean age of 21, participated in the study. The length of time the participants had been learning English varied from 9 to 17 years, with a mean of 13 years. All participants had passed the Chinese College English Test Band 4 examination and therefore had acquired sufficient English to be able to read and generate simple texts for analysis. As a comparison group, one male native American English speaker (in his 60s) who taught English in China was recruited for both SS and RS. Recordings of material read by a another two NSs were also used for analysis. All participants indicated that they did not suffer from either a hearing or speaking impairment and gave their consent for the data to be used in this research.

Materials

In line with previous studies on the production of affective prosody, the current paper used a picture storytelling task, which was introduced by Bamberg (1987) to investigate the narrative function and later was widely adopted in the field of language development. The present study chose the wordless picture book *Frog, where are you?* (Mayer, 1969) as the material, which was also used in previous studies and turned out to be very effective in eliciting the subjects' affects (see e.g., Reilly et al., 1990; Setter et al., 2007; Wang & Liang, 2013). As storytelling is "the art of using language, vocalization, and/or physical movement and gesture to reveal the elements and images of a story to a specific, live audience" (National Storytelling Association, n.d.), thus affective involvement is more often required in the storytelling context than in other narratives. As argued by Theune et al. (2006) and Verma et al. (2015), a greater variation in pitch, tempo, and duration is observed in storytelling than in other narratives.

Procedures

The current study is made up of two recording experiments. The procedure of the first recording experiment is as follows. At first, the first author chatted to each participant to make him/her feel more at ease. The author then introduced the book to each subject with the introductory words "This is a very interesting picture book, it's about a little boy, a little dog, and a little frog. You can have a quick look at this book and then I hope you can tell me the story as you progress page by page through the book". Each participant was given enough time to prepare their storytelling. During the recording, the author also used certain phrases to encourage each subject to continue telling the story, such as "oh, what happened?".

Given the requisite skills for producing a good narrative involve complex linguistic, cognitive, and affective abilities (Reilly et al., 2004), the performance of affective prosody may be hindered by participants' linguistic ability, although they were given enough time to prepare in the first recording. Thus, to eliminate the effect of linguistic knowledge such as lexicon and grammar, on the production of affective prosody, we conducted another recording experiment. The procedure was the same as the one in the first recording experiment with the exception that the participants were provided with a prepared

script of the Frog story when telling the story. Similarly, participants were given enough time to read the script until they were familiar with the content.

All utterances were recorded in a quiet room by using Praat (Boersma & Weenink, 1992-2022) on a Macbook Air using a microphone at a sample rate of 44.1kHz and 16-bit quantization. The microphone was placed approximately 1-1.5 inches away from the speaker's mouth. Considering that affective expression becomes more involved during the storytelling climax (Wang & Liang, 2013), we extracted 40 WAV format sound files for Chinese EFL learners and 4 WAV format sound files for native speakers at the climax part of participant's storytelling speech.

Analysis

The recorded utterances were analyzed using Praat (Boersma & Weenink, 1992-2022). The mean, minimum and maximum pitch, and the pitch range were extracted. The measurement of vowel duration for each target syllable was based on sound waveforms and spectrograms. Due to the difference in the number of syllables in the participants' SS, the overall duration of their vowels was divided by the total number of syllables to give the mean vowel duration. Statistical Product and Service Solutions (abbr. SPSS) 26.0 software was used to statistically analyze the data.

In this study, a five-point Likert scale was used. An American English teacher and a Chinese English teacher were invited to rate the affective richness of the 40 sound files. A score of 1 means a very low affective involvement, and a score of 5 means a very high affective involvement. After that, we performed a Kendall correlation test on the scores provided by the two raters and found that the scores were very consistent ($r = 0.508$, $p < 0.001$).

Results

Information relating to the pitch ranges and vowel durations of the participants was presented in Table 1 shown below. In terms of affective involvement, Chinese EFL learners received only 2.1 out of 5 for their SS, and 2.5 out of 5 for their RS, which means that they have considerable difficulty in using prosody to convey affects. Their problems are fully reflected in the following: (i) in terms of pitch, the pitch ranges of Chinese EFL learners are narrower than those of NSs in both SS and RS; in addition, the pitch standard deviation (hereafter pitch SD) of Chinese EFL learners is smaller than the corresponding value for NSs, which leads to a more level intonation as perceived by the listener; (ii) with regard to vowel duration, although the majority of Chinese EFL learners are able to differentiate content words from function words, their overall vowel duration is significantly shorter than the corresponding value for NSs. All these factors contribute to their low affective involvement scores. The findings of the current study will be presented in the following sections, beginning with the pitch performance of Chinese EFL learners.

TABLE 1
Information Relating to the Prosodic Correlates of the Participants

	Participants		Pitch Range		Pitch SD		Vowel Duration
			(Hz)	(Semitone)	(Hz)	(Semitone)	(ms)
SS	20 Learners	Mean	76.7	7.9	16.7	1.7	161.1
	1 NS	Absolute	132.5	12.99	44.61	4.4	216.5
RS	20 Learners	Mean	96.1	9.2	21.6	2.1	166.3
	3 NSs[1]	Absolute	149.8	12.5	39.2	3.4	202.5

[1] Due to the COVID-19 pandemic, the two additional sound files were obtained from the internet, as mentioned in *Participants*.

Pitch Range

Pitch conveys a substantial amount of information about affects (Murray & Arnott, 1993; Williams & Stevens, 1972). Unlike neutral speech, a larger pitch value and greater pitch variation are observed in storytelling speech (Verma et al., 2015). The section of recording that we chose to analyze involves a little boy finally finding his pet frog after enduring hardships, which is obviously a happy context. Therefore, a comparatively wider pitch range is expected in the storytelling task undertaken by our participants. Table 2 lists the pitch information for the climax of the participants' SS. It is evident that the pitch range and pitch SD of Chinese EFL learners are significantly narrower and smaller than those of the native English speaker.

TABLE 2
Information Relating to Pitch in the Participants' SS

	Participants					
	20 Learners					1 NS
	Mean	SD	df	*t*	*p*	
Pitch Range	76.7 Hz	45.945 Hz	19	-6.349	< 0.05	132.5 Hz
Pitch Range	7.9 st	3.534 st	19	-7.956	< 0.05	12.99 st
Pitch SD	16.7 Hz	12.326 Hz	19	-13.053	< 0.05	44.61 Hz
Pitch SD	1.7 st	1.148 st	19	-13.5884	< 0.05	4.4 st

By conducting a one-sample *t*-test, we found a significant difference in pitch range and pitch SD between Chinese EFL learners and the native English speaker in their SS ($p < 0.05$). During the climax to the story, the mean pitch range of the learners is only 7.9 st, which is far lower than the value of 12.99 st obtained for the native speaker; in addition, the pitch SD of the learners is just 1.7 st, while the equivalent value for the native speaker is 4.4 st. The narrower pitch range and smaller pitch SD might indicate that Chinese EFL learners have encountered problems in using pitch to convey affects appropriately. Similarly, in their RS, Chinese EFL learners again used a narrower pitch range and smaller pitch SD than NSs. Table 3 lists the pitch information for the climax of the participants' RS. It can be seen that the mean pitch range of the learners is 9.2 st, a higher value than that obtained in their SS (7.9 st), but is lower than the value of 12.5 st obtained for the NSs; in addition, the pitch SD of the learners is 2.1 st, which again is higher than that obtained for their SS (1.7 st), but is still lower than the value of 3.4 st obtained for NSs. Although the climax of the story that we chose to analyze is obviously a happy and exciting context, the pitch characteristics of the Chinese EFL learners' speech do not show the same level of variation as the speech of NSs does. These results might indicate that Chinese EFL learners have difficulty in using prosody to convey affects flexibly.

TABLE 3
Information Relating to Pitch in the Participants' RS

	Participants					
	20 Learners					3 NSs
	Mean	SD	df	*t*	*p*	
Pitch Range	96.1 Hz	51.121 Hz	19	-4.7	< 0.05	149.8 Hz
Pitch Range	9.2 st	3.688 st	19	-4.049	< 0.05	12.5 st
Pitch SD	21.6 Hz	12.277 Hz	19	-6.403	< 0.05	39.2 Hz
Pitch SD	2.1 st	1.158 st	19	-4.898	< 0.05	3.4 st

To check whether Chinese EFL learners' pitch performance was better when they were provided with a prepared script during the storytelling, a paired-sample *t*-test was conducted. The results showed no significant difference in the pitch range between the SS and RS of Chinese EFL learners. In detail, learners' pitch ranges (st) in SS were narrower than in RS, but without a statistical difference ($t = -1.337$, df = 19, $p = 0.197$, $p > 0.05$, MD = -1.235). Similar results were also obtained for their pitch SD (st) ($t = -1.256$, df = 19, $p = 0.224$, $p > 0.05$, MD = -0.395) in SS and RS. Thus, it could be concluded that many Chinese EFL

learners have an ingrained problem with the adjustment of pitch to convey affects.

To further illustrate the difference in pitch performance between Chinese EFL learners and NSs, we drew pitch contours for the climax of the story in both SS and RS for one learner #5 (L#5) and one native speaker, as shown in Figures 1 to 4.

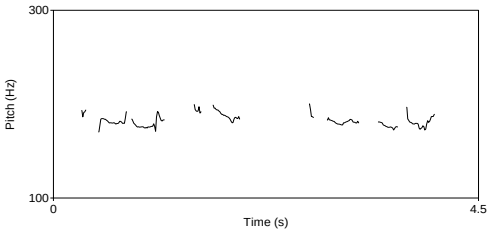

Figure 1. Pitch contour of the SS climax of L#5.

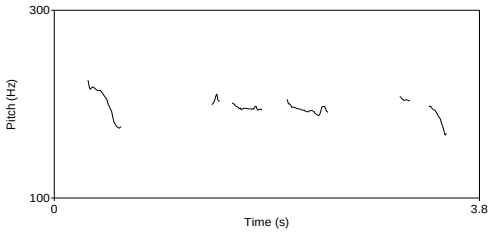

Figure 2. Pitch contour of the RS climax of L#5.

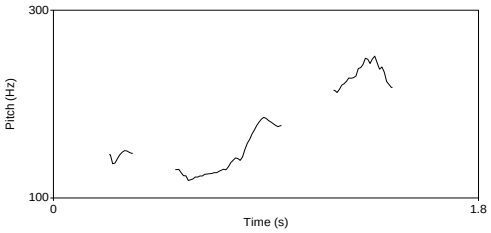

Figure 3. Pitch contour of the SS climax of the NS.

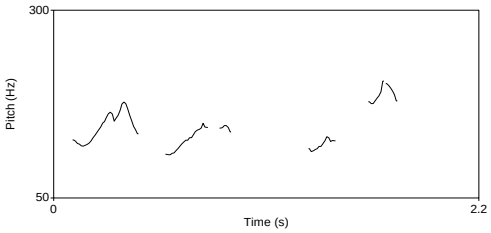

Figure 4. Pitch contour of the RS climax of the NS.

Figure 1 is the pitch contour for learner #5's speech "he found two frogs behind the branch" in the spontaneous recording, while Figure 2 shows the pitch contour for her read speech "there they found the boy's pet frog". Figure 3 presents the pitch contour for the native speaker's SS "he found his frog", and Figure 4 is his pitch contour for "there they found the boy's pet frog". The x-axis represents time (s) and the y-axis represents pitch (Hz). Obviously, the fluctuation degree of the pitch contour in Figure 1 (Pitch Range: 2.70 st, Pitch SD: 0.57 st) and in Figure 2 (Pitch Range: 5.15 st, Pitch SD: 0.95 st) is less marked than that in Figure 3 (Pitch Range: 12.99 st, Pitch SD: 4.4 st) and Figure 4 (Pitch Range: 12.8 st, Pitch SD: 3.55 st). This learner's speech demonstrates a comparatively narrower pitch range and smaller pitch SD than the native speaker and is a typical case of deficient affective involvement, which is also reflected in the average scores she received from the raters: 1 out of 5 for her SS and 1.5 out of 5 for her RS.

Vowel Duration

To further investigate the affective prosody of Chinese EFL learners, we analyzed the vowel duration of the participants. Table 4 gives information relating to the vowel duration of the participants in both SS and RS. In general, the vowel duration of Chinese EFL learners is shorter than that of NSs.

TABLE 4
Information Relating to the Participants' Vowel Duration

		Mean (ms)	Standard Deviation (ms)	df	*t*	*p*
SS	20 learners	161.1	34.465	19	-7.191	< 0.05
	1NS	216.5	-			
RS	20 learners	166.3	31.534	19	-5.136	< 0.05
	3 NSs	202.5	37.877			

The results of a one-sample *t*-test indicate that there is a significant difference in the vowel duration between Chinese EFL learners and NSs ($p < 0.05$). The average vowel duration in the SS of the learners is 161.1 ms, which is much shorter than the equivalent value of 216.5 ms for the native speaker; in RS, the average vowel duration of the learners is 166.3 ms, which again is much shorter than the value of 202.5 ms for the NSs. In order to have a comprehensive understanding of the difference in vowel performance between Chinese EFL learners and NSs, we measured their vowel duration of /aʊ/ in *found*, /ɔɪ/ in *boy's*, /ɛ/ in *pet*, and /ɑ/ in *frog* in the sentence "there they found the boy's pet frog" produced in the second recording experiment. The results of a one-sample *t*-test demonstrated that the duration of /ɛ/ in *pet* is significantly shorter for the learners (119.9 ms) than for the NSs (149 ms) ($t = -2.964$, df = 19, $p = 0.008$, $p < 0.05$, MD = -29.1). A similar result is obtained for the mean length of /ɑ/ in *frog*, with a duration of 186.3 ms for the learners and a duration of 354.3 ms for the NSs. To further illustrate the vowel performance of Chinese EFL learners during storytelling, we drew spectrograms for the vowel /ɑ/ in *frog* in the climax of the SS and RS of one learner #1 and one native English speaker, as shown in Figures 5 to 8.

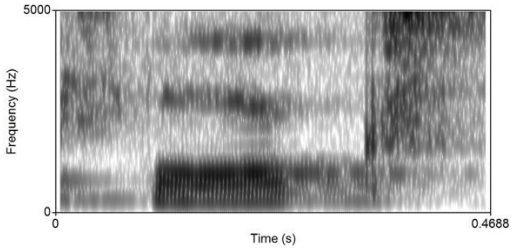

Figure 5. Spectrogram of /ɑ/in the learner's SS

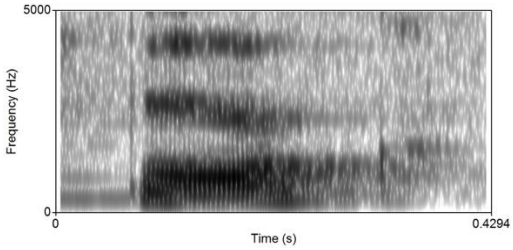

Figure 6. Spectrogram of /ɑ/ in the learner's RS

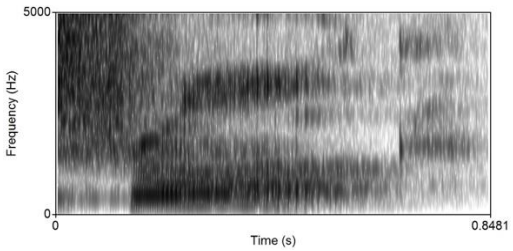

Figure 7. Spectrogram of /ɑ/ in the NS's SS

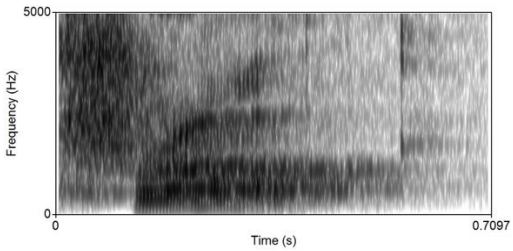

Figure 8. Spectrogram of /ɑ/ in the NS's RS

Figure 5 is the spectrogram obtained for the learner's production of /ɑ/ in *frog* in SS, which lasts approximately 136 ms. Figure 6 illustrates the spectrogram of this learner's production of /ɑ/ in *frog* in RS, which lasts about 152 ms. Figure 7 presents the spectrogram of the native speaker's production of /ɑ/ in *frog* in SS, which lasts approximately 398 ms. Figure 8 provides the spectrogram of this native speaker's production of /ɑ/ in *frog* in RS, which lasts around 282 ms. It is worth noting that the duration of the vowel /ɑ/ produced by the native speaker is almost twice the duration of the same vowel produced by the Chinese EFL learner in the storytelling.

To determine whether the vowel duration of the Chinese EFL learners was longer in their RS than SS, we conducted a paired-sample *t*-test. The results obtained showed no significant difference in the vowel duration between their SS (161.1ms) and RS (166.3ms) ($t = -0.632$, df = 19, $p = 0.535$, $p > 0.05$, MD = -5.2), which suggested that even they were permitted to tell the story with the script, Chinese EFL learners did not perform better in terms of vowel duration. To further investigate the acquisition of vowel duration by Chinese EFL learners, we calculated their vowel duration of content words and function words in both their SS and RS. The results showed that Chinese EFL learners were able to distinguish content words from function words. Specifically, regarding SS, learners' production of vowels in content words lasted

approximately 186.9 ms, which was much longer than that of vowels in function words (MD = 36.747, df = 18[2], p = 0.033, $p < 0.05$); concerning RS, the average vowel duration of content words is 182.1 ms, which is also significantly longer than the average vowel duration of function words (140.1 ms) (MD = 42, df = 19, p = 0.012, $p < 0.05$). This is in agreement with Chen et al.'s (2001) finding that Mandarin speakers produce stressed words with significantly longer vowel durations than those of unstressed words.

Nevertheless, in the case of storytelling, Chinese EFL learners' vowel duration of both content words and function words is shorter than the equivalent duration for NSs, another factor contributing to their low affective involvement scores.

Discussion

It is undeniable that English is fast becoming a lingua franca (ELF) through which people from other cultures or who speak languages other than English as their first language can communicate. As approximately one in four users of English across the world is a native speaker of the language (Crystal, 2003), most ELF interactions take place between non-native speakers of English. Against this backdrop, the acquisition of a native-like accent is no longer the ultimate objective of the majority of global English learners (Jenkins, 1998). Many scholars even have attempted to establish some form of simplified, universal pronunciation which is intelligible and acceptable to both native and non-native speakers of English. For example, Gimson (1978) reduced 24 consonant sounds and 20 vowel sounds into 14 and 15 sounds respectively. However, linguistic information (word intelligibility) is not the only thing conveyed in communication, affects are also important, particularly in narratives in which affects are strongly needed, such as the storytelling speech. Therefore, L2 learners not only have to grasp the usage of prosody in conveying grammatical meaning at the lexical, utterance, or discourse level, but also have to learn and use prosody to convey their affects appropriately. However, the results of the current study showed that the majority of Chines EFL learners had problems in the performance of affective prosody.

The overall narrower pitch range of Chinese EFL learners than native speakers was not only shown in their SS but also in their RS in which the script of the Frog story was provided to them to eliminate the influence of linguistic knowledge on the production of affective prosody. However, no significant difference in the pitch range was found between their SS and RS. Thus, it could be concluded that many Chinese EFL learners have an ingrained problem with the acquisition of pitch to convey affects. The reduced variation in pitch observed in the learners' speech might make their speech monotonous, which is likely to, as one rater commented, 'drive people to sleep'. This finding, together with previous findings that ESL learners tended to use a narrower pitch range than native speakers, may indicate that ESL learners have difficulty in using pitch flexibly as native speakers do, although most of the related studies were carried out from the perspective of stress prosody.

Contrary to the findings of most studies dealing with the pitch range exhibited by ESL learners, Aoyama and Guion (2007) found a wider pitch range in the English speech by native Japanese speakers than native speakers. The authors argued that the overall greater pitch range might be attributed to the phonological difference between Japanese and English since the former is considered as a pitch accent language in which pitch has a phonemic function (Yamazawa & Hollien, 1992) and is the only consistent cue for accented syllables (Beckman, 1986). It is suggested that the native language transfer happened during Japanese learners' acquisition process of English, thus, they were liable to use a relatively wider pitch range when speaking English. The potential transfer from ESL learners' native language into their use of target language may be a plausible explanation for their marked difference in pitch range from native speakers, given the overall pitch range of their native language is narrower or wider than that of English. However, Mandarin Chinese was found to have a wider pitch span compared to English (Keating & Kuo, 2012), which suggests that the narrower pitch range in Chinese EFL learners' storytelling speech

[2] The degree of freedom here is 18, instead of 19, the reason for this being that there was no function word in one participant's SS, "find frog family".

might be attributed to other factors.

Individuals' personality is one possible factor influencing the learners' production of affective prosody. Many researchers have found a positive correlation between extroversion and L2 learning (Chastain, 1975; Pritchard, 1952), while introverted traits such as shyness are a disadvantage for L2 learning, particularly for L2 pronunciation. As Kainada and Lengeris (2014) argued, the use of a wider pitch range can make English speakers sound overexcited or aggressive to German speakers, whose native language has been found to employ lower pitch levels and narrower pitch ranges than English (Mennen, 2007). Besides, lack of confidence or anxiety may also contribute to the monotonous speech by L2 learners' narrower pitch range (Mennen, 1998; Ordin & Mennen, 2017; Zimmerer et al., 2014). Chinese EFL learners may be too shy to use the variation in pitch demanded by English in their speech and therefore tend to use a more monotonous tone. Another contributor to the relatively narrower pitch range by Chinese EFL learners is the lack of storytelling practice in academic curricula (McKay & McKay, 2002), which, in part, can hinder learners' ability to differentiate between storytelling speech and other speaking styles.

Previous studies have shown that learners' pitch performance can be improved through particular training (Celce-Murcia et al., 1996; Nagamine, 2002). For example, Nagamine (2002) examined the validity of 'hyper-pronunciation', a pedagogical method aiming at teaching English intonation proposed by Todaka (1990), the results showed that after receiving a thirteen-week training of 'hyper-pronunciation' which includes exaggerating pitch spans, the pitch ranges exhibited by Japanese college students were much broader and closer to those of NSs. As such, to improve Chinese EFL learners' ability to convey affects through their use of prosody, we would recommend that not only phonological knowledge and training are needed to improve learner's pitch performance, but, more importantly, psychological encouragement as well, so that more reticent students will come out of their shells.

There is evidence that the duration of some ESL learners is longer than native speakers of English (Aoyama & Guion, 2007 for Japanese L2 learners of English; Guion et al., 2000 for both Italian and Korean L2 learners of English; Munro & Derwing, 1995 for Mandarin L2 learners of English). However, the results of the current study showed that the overall duration of Chinese EFL learners was significantly shorter than the corresponding value for NSs, which contradicted Munro and Derwing's (1995, p. 301) finding that "the duration data reported here indicate that the Mandarin talkers generally spoke more slowly than the native English speakers". We propose that the opposite results between these studies are at least partly due to the different genres of stimuli, that is, the speech examined in the present study was elicited in the context of storytelling; while the speech recorded in Munro and Derwing's study was based on 40 short sentences which were obviously true or false, thus, those sentences are much more information oriented. For example, in storytelling speech, an unusual lengthening of vowels (a slow tempo) is observed and used to attract the listeners' attention and keep them engaged with the story (Theune et al., 2006; Verma et al., 2015).

One possible reason for the marked storytelling performance of the Chinese EFL learners might be the lack of storytelling practice in academic curricula (McKay & McKay, 2002), which, in part, can hinder the learners' ability to realize and differentiate the speaking styles of different narrative genres. For example, the storytelling speech is slower than the speech of the newsreaders which is more neutral (Fackrell et al., 2000; Theune et al., 2006). Personality factors might also have contributed to the constant vowel duration of Chinese EFL learners. Some learners feel uncomfortable with the target language's speech rhythm and melody patterns, while some of them even feel stupid pronouncing the weird sounds (Gilakjani & Ahmadi, 2011). Shyness or lack of confidence could hinder the learners to express themselves in English (Tong, 2004), which may result in their inflexible duration pattern in speech no matter the narrative genre.

As indicated, simply learning the difference between stressed and unstressed syllables is not sufficient, particularly when attitudinal expressions and different discourse genres are taken into consideration. In order that Chinese EFL learners' English speech sounds more expressive and engaging, it is recommended that activities that help learners differentiate the prosodic patterns of various narratives should be included in current academic curricula; furthermore, learners should try to imitate the prosodic way native speakers adopted to tell stories.

Conclusion

This study has found that Chinese EFL learners experience difficulties in mastering and using affective prosody when speaking English. More specifically, they tend to use a narrower pitch range, with less salient pitch variations than NSs; in addition, their vowel duration is much shorter than that of NSs.

Concerning pitch, the acoustic analyses suggest that the pitch range is significantly narrower for Chinese EFL learners than for NSs in the climax of both their SS and RS. Furthermore, for learners, their pitch standard deviations are considerably smaller than those of NSs, which indicates that Chinese EFL learners have used a relatively level intonation compared to NSs. From the perspective of a native English speaker, these prosodic characteristics tend to create the impression that learners 'really drive people to sleep' as the tone of their speech is rather monotonous. Another interesting finding obtained from this study is that the pitch range used by the learners in their SS and RS does not show any significant differences, which is contrary to our hypothesis that the pitch range of learners tends to be wider if they are given a prepared script. Nevertheless, this supports the argument that many learners have difficulty in using pitch to convey affects appropriately.

In the case of Chinese EFL learners, although we have found that the duration of the vowels in content words is significantly longer than the duration of vowels in function words, their general vowel duration, in both SS and RS, is significantly shorter than that of NSs. In storytelling speech, NSs tend to reduce their speech rate to lengthen certain vowels, thereby increasing their affective involvement. However, in contrast, many Chinese EFL learners are not aware of the different speech rates associated with different speaking styles. Instead, they talk in their usual conversational manner.

The analysis of affective prosody among Chinese EFL learners provides a number of suggestions that learners can use to improve their affective involvement and to sound more expressive and attractive when speaking English. We admit that this paper does not fully consider the Chinese EFL learners' perception of affective prosody, and other in-depth studies on this topic could contribute to our understanding of the reasons for their prosodic performance. However, the current study could serve as the point of departure for subsequent studies, which could ultimately shed further light on the prosodic aspect in SLA.

Acknowledgements

We would like to express our special thanks to all the participants from Harbin Engineering University in our study.

The Authors

Zepeng Wang (first author) received his B.A. in English from Harbin Engineering University and is an M.A. candidate in the School of Foreign Studies at Harbin Engineering University. His current research interests include the interface study of prosody and pragmatics and second language acquisition.

School of Foreign Studies
Harbin Engineering University
Harbin, 150001, China
Tel: +86 15765598240
Email: leowang2u@foxmail.com

Yansheng Mao (corresponding author) is professor in the School of Foreign Studies at Harbin Engineering University. His current research interests cover pragmatics and language assessment. His recent publications include articles in *Pragmatics & Society*, *Discourse, Context & Media*, *Language & Communication*, *Discourse Studies*, *Health Communication*, and *System*, etc.

School of Foreign Studies
Harbin Engineering University
Harbin, 150001, China
Tel: +86 13936583108
Email: wiltonmao@163.com

References

Aoyama, K., & Guion, S. G. (2007). Prosody in second language acquisition. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 281-297). John Benjamins.

Backman, N. (1979). Intonation errors in second-language pronunciation of eight Spanish-speaking adults learning English. *Interlanguage Studies Bulletin*, *4*(2), 239-265.

Bamberg, M. (1987). *The acquisition of narratives*. Mouton de Gruyter.

Banjo, A. (1979). Beyond intelligibility in Nigerian English. In E. Ubahakwe (Ed.), *Varieties and functions of English in Nigeria* (pp. 7-13). African Universities Press.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614-636. https://doi.org/10.1037/0022-3514.70.3.614

Beckman, M. E. (1986). *Stress and non-stress accent*. Foris Publications.

Boersma, P., & Weenink, D. (1992-2022). *Praat: Doing phonetics by computer* (Version 6.1.11). http://www.praat.org/

Bongaerts, T., van Summeren, C., Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition*, *19*(4), 447-465. http://doi.org/10.1017/S0272263197004026

Bozikas, V. P., Kosmidis, M. H., Anezoulaki, D., Giannakou, M., Andreou, C., & Karavatos, A. (2006). Impaired perception of affective prosody in schizophrenia. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *18*(1), 81-85. https://doi.org/10.1176/jnp.18.1.81

Busà, M. G., & Urbani, M. (2011). A cross linguistic analysis of pitch range in English L1 and L2. In *Proceedings of the 17th ICPhS* (pp. 380-383). International Congress of Phonetic Sciences (ICPhS).

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge University Press.

Chastain, K. (1975). Affective and ability factors in second-language acquisition. *Language Learning*, *25*(1), 153-161. https://doi.org/10.1111/j.1467-1770.1975.tb00115.x

Chen, Y., Robb, M. P., Gilbert, H. R., & Lerman, J. W. (2001). A study of sentence stress production in Mandarin speakers of American English. *The Journal of the Acoustical Society of America*, *109*(4), 1681-1690. https://doi.org/10.1121/1.1356023

Couper-Kuhlen, E. (1986). *An introduction to English prosody*. Niemeyer.

Cruz-Ferreira, M. (1989). A test for non-native comprehension of intonation in English. *International Review of Applied Linguistics in Language Teaching*, *27*(1), 23-39. https://doi.org/10.1515/iral.1989.27.1.23

Crystal, D. (2003). *English as a global language* (2nd ed.). Cambridge University Press.

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

Edwards, J., Jackson, H. J., & Pattison, P. E. (2002). Emotion recognition via facial expression and affective prosody in schizophrenia: A methodological review. *Clinical Psychology Review*, *22*(6), 789-832. https://doi.org/10.1016/s0272-7358(02)00130-7

Fackrell, J., Vereecken, H., Buhmann, J., Martens, J. P., & Coile, B. V. (2000). Prosodic variation with text type. In *Sixth International Conference on Spoken Language Processing* (pp. 231-234). International Speech Communication Association.

Gilakjani, A. P., & Ahmadi, M. R. (2011). Why is pronunciation so difficult to learn?. *English Language Teaching*, *4*(3), 74-83. https://doi.org/10.5539/elt.v4n3p74

Gimson, A. C. (1978). Towards an international pronunciation of English. In P. Strevens (Ed.), *In Honour of A.S. Hornby* (pp. 45-53). Oxford University Press.

Guion, S. G., Flege, J. E., Liu, S. H., & Yeni-Komshian, G. H. (2000). Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics*, *21*(2), 205-228. http://doi.org/10.1017/S0142716400002034

Hirschberg, J. (2004). Pragmatics and intonation. In L. R. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 515-537). Blackwell.

Hua, C., & Li, B. (2016). Nuclear stress patterns in reading by adult Chinese EFL learners: Explicit training or implicit learning?. *The Journal of Asia TEFL*, *13*(1), 1-15. http://dx.doi.org/10.18823/asiatefl.2016.13.1.1.1

Jenkins, J. (1998). Which pronunciation norms and models for English as an international language? *ELT Journal*, *52*(2), 119-126. https://doi.org/10.1093/elt/52.2.119

Kainada, E., & Lengeris, A. (2014). The acquisition of English intonation by native Greek speakers. In N. Lavidas, T. Alexiou, & A. M. Sougari (Eds.), *Major trends in theoretical and applied linguistics 1* (pp. 141-156). Versita.

Kashiwagi, A., & Snyder, M. (2008). American and Japanese listener assessment of Japanese EFL speech: Pronunciation features affecting intelligibility. *The Journal of Asia TEFL*, *5*(4), 27-47.

Keating, P., & Kuo, G. (2012). Comparison of speaking fundamental frequency in English and Mandarin. *The Journal of the Acoustical Society of America*, *132*(2), 1050-1060. https://doi.org/10.1121/1.4730893

Mayer, M. (1969). *Frog, where are you?* Dial Press.

McKay, S. L., & McKay, S. (2002). *Teaching English as an international language: An introduction to the role of English as an international language and its implications for language teaching*. Oxford University Press.

Mennen, I. (1998). Can language learners ever acquire the intonation of a second language?. In *Proceedings of STiLL-Speech Technology in Language Learning* (pp. 17-20). International Speech Communication Association.

Mennen, I. (2007). Phonological and phonetic influences in non-native intonation. In J. Trouvain & U. Gut (Eds.), *Non-native prosody: Phonetic description and teaching practice* (pp. 53-76). Mouton de Gruyter.

Mennen, I., & de Leeuw, E. (2014). Beyond segments: Prosody in SLA. *Studies in Second Language Acquisition*, *36*(2), 183-194. http://doi.org/10.1017/S0272263114000138

Mennen, I., Schaeffler, F., & Dickie, C. (2014). Second language acquisition of pitch range in German learners of English. *Studies in Second Language Acquisition*, *36*(2), 303-329. http://doi.org/10.1017/ S0272263114000023

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, *38*(3), 289-306. https://doi.org/10.1177/002383099503800305

Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America*, *93*(2), 1097-1108. https://doi.org/10.1121/1.405558

Nagamine, T. (2002). An experimental study on the teachability and learnability of English intonational aspect: Acoustic analysis on F0 and native-speaker judgment task. *Journal of Language and Linguistics*, *1*(4), 362-399.

National Storytelling Association. (n.d.). *What is Storytelling?*. National Storytelling Association. https://www.eldrbarry.net/roos/st_defn.htm

Ordin, M., & Mennen, I. (2017). Cross-linguistic differences in bilinguals' fundamental frequency ranges. *Journal of Speech, Language, and Hearing Research*, *60*(6), 1493-1506. https://doi.org/10.1044/2016_JSLHR-S-16-0315

Pell, M. D. (2001). Influence of emotion and focus location on prosody in matched statements and questions. *The Journal of the Acoustical Society of America*, *109*(4), 1668-1680. https://doi.org/10.1121/1.1352088

Pickering, L. (2004). The structure and function of intonational paragraphs in native and nonnative speaker instructional discourse. *English for Specific Purposes*, *23*(1), 19-43. https://doi.org/10.1016/S0889-4906(03)00020-6

Pritchard, D. F. L. (1952). An investigation into the relationship between personality traits and ability in modern languages. *British Journal of Educational Psychology*, *22*(2), 147-148. https://doi.org/10.1111/j.2044-8279.1952.tb02817.x

Reilly, J., Klima, E. S., & Bellugi, U. (1990). Once more with feeling: Affect and language in atypical populations. *Development and Psychopathology*, *2*(4), 367-391. http://doi.org/10.1017/S0954579400005782

Reilly, J., Losh, M., Bellugi, U., & Wulfeck, B. (2004). "Frog, where are you?" Narratives in children with specific language impairment, early focal brain injury, and Williams syndrome. *Brain and Language*, *88*(2), 229-247. https://doi.org/10.1016/S0093-934X(03)00101-9

Setter, J., Stojanovik, V., Van Ewijk, L., & Moreland, M. (2007). Affective prosody in children with Williams syndrome. *Clinical Linguistics & Phonetics*, *21*(9), 659-672. https://doi.org/10.1080/02699200701539056

Shen, X. S. (1990). *The prosody of Mandarin Chinese*. University of California Press.

Tartter, V. C. (1980). Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception & Psychophysics*, *27*(1), 24-27. https://doi.org/10.3758/BF03199901

Theune, M., Meijs, K., Heylen, D., & Ordelman, R. (2006). Generating expressive speech for storytelling applications. *IEEE Transactions on Audio, Speech, and Language Processing*, *14*(4), 1137-1144.

Todaka, Y. (1990). An error analysis of Japanese students' intonation and its prosodic analysis. [Unpublished master's thesis]. University of California, Los Angeles.

Tong, J. (2004). Little voice: Students' confidence and their responses in English lessons. *The Journal of Asia TEFL*, *1*(1), 197-212.

Verma, R., Sarkar, P., & Rao, K. S. (2015). Conversion of neutral speech to storytelling style speech. In *2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR)* (pp. 1-6). International Conference on Advances in Pattern Recognition (ICAPR).

Wang, F., & Liang, D. (2013). Affective prosody in children suffering from mental retardation in spontaneous speech. *Journal of Psychological Science*, *36*(2), 401-405.

Wennerstrom, A. (1998). Intonation as cohesion in academic discourse: A study of Chinese speakers of English. *Studies in Second Language Acquisition*, *20*(1), 1-25. http://doi.org/10.1017/S0272263198001016

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *The Journal of the Acoustical Society of America*, *52*(4B), 1238-1250. https://doi.org/10.1121/1.1913238

Yamazawa, H., & Hollien, H. (1992). Speaking fundamental frequency patterns of Japanese women. *Phonetica*, *49*(2), 128-140. https://doi.org/10.1159/000261907

Yang, C. (2016). *The acquisition of L2 Mandarin prosody: From experimental studies to pedagogical practice*. John Benjamins.

Zimmerer, F., Jügler, J., Andreeva, B., Möbius, B., & Trouvain, J. (2014). Too cautious to vary more? A comparison of pitch variation in native and non-native productions of French and German speakers. In N. Campbell, D. Gibbon, & D. Hirst (Eds.) *Proceedings to the 7th Speech Prosody Conference* (pp. 1037-1041). International Speech Communication Association.

(Received November 6, 2021; Revised February 20, 2022; Accepted March 18, 2022)

