On English speakers' ability to communicate emotion in Mandarin

Hua-Li Jian

Abstract

The ability of Mandarin learners to express emotion in Mandarin speech has received little attention. This study examines how English L1 users express emotions in Mandarin and how these differ from those of Mandarin L1 users. Scenarios were adopted to elicit joy, anger, sadness, fear, and neutrality. Both groups articulated anger, joy, and fear with a high pitch. Both groups also employed high intensity for anger and joy, while low intensity for sadness and fear. Learners generally employed larger F0 ranges than natives, particularly for anger and fear. Learners articulated level tones with lengthened duration and contour tones with shortened duration, affecting the correct emotional portrayal. Learners used similar intensity range for all emotions, whereas natives tended to vary the intensity with different emotions. The results have implications for teaching Mandarin as a second language with special reference to prosodic naturalness in expressing emotions.

Keywords: emotional speech, English L1, Mandarin L2, fundamental frequency, duration, intensity

L'expression des émotions pour les mandarin-L2 a été peu étudiée. Cette étude examine comment les anglais-L1 expriment en mandarin les émotions et comment cela diffère de celle mandarin-L1. Des scénarios ont été utilisés pour exprimer joie, colère, tristesse, peur, et indifférence. Les deux groupes articulent colère, joie, et peur avec un ton élevé, et ont également employé une ton élevée pour colère et joie, tandis qu'un ton bas pour tristesse et

peur. Les apprenants utilisent une échelle F0 plus large que les mandarin-L1, particulièrement pour colère et peur et combinent des niveaux de ton d'une grande durée avec des tons d'une courte durée, affectant la représentation émotionnelle réelle. Les apprenants utilisent une gamme d'intensité similaires pour toutes les émotions alors que les mandarin-L1 ont tendance à varier l'intensité suivent l'émotion. Ces résultats ont une conséquence pour l'enseignement mandarin-L2 avec une connexion particulière à la naturalité prosodique dans l'expression des émotions.

Vocal expression is important for distinguishing different affective states. Important acoustic parameters include the mean and range of the fundamental frequency (F0), speech rate, and intensity (loudness) for distinguishing emotion categories due to various degrees of arousal of emotions portrayed (Scherer, 1986). Yet, no intonational patterns have been found to be representative of emotion categories (Banziger & Scherer, 2005). Higher F0 levels were found to occur with anger and joy which are high arousal emotions (Scherer, 1986). Anger, joy and fear have also been associated with higher F0 level and range than neutral (Paeschke, Kienast, & Sendlmeier, 1999). Sadness has been associated with low F0 and small spread (Pell, Paulmann, Dara, Alasseri, & Kotz, 2009). Specifically, anger, fear and joy exhibited larger spread in pitch than sadness (Zhang, Ching, & Kong, 2006; Paeschke & Sendlmeier, 2000; Murray & Arnott, 1993; Scherer, 1986).

Placement of stress and speech rate may also communicate emotions. The speech rate is related to the syllable durations where shorter syllables are associated with fast speech rates and vice versa. Different emotions have been found to exhibit different speech rates (Banse & Scherer, 1996). Anger, fear, and joy have been associated with higher speech rates, while sadness has been associated with lower speech rates (Gu & Lee, 2007; Zhang et al., 2006;

Ververidis & Kotropoulos, 2006; Schröder, 2001; Murray & Arnott, 1993; Scherer, 1986), suggesting that duration varied either at sentence or word level. Stressed vowels of sadness had shorter durations compared to anger and happiness, although the consonantal duration in sadness was not shorter than that of other emotions (Pereira & Watson, 1998). Moreover, stress distribution of emotions varies (Paeschke et al., 1999).

Irrespective of emotion, tones and intonation may be affected by the stress patterns in a phrase and stressed words often occur with higher intensity (Tao, Kang, & Li, 2006). Joy and anger showed higher mean intensity than sadness (Murray & Arnott, 1993; Scherer, 1986). The intensity of fear was similar to that of neutral (Murray & Arnott, 1993), but it also exhibited an energy increase in higher frequency bands above 1 kHz (Scherer, 1986). Joy was frequently confused with hot anger and panic fear (Banse & Scherer, 1996). Mean syllable intensity for joy and anger were also found to be similar and higher than fear, sadness and neutrality, with sadness being lowest (Zhang et al., 2006).

Several studies have addressed how different emotional expressions are recognized in first languages versus second languages (Graham, Hamblin, & Feldstein, 2001; Dromey, Silveira, & Sandor, 2005; Pell et al., 2009). A study of affective word processing in a first versus second language has shown that extensive L2 use leads to more rapid activation of the emotional meaning of the words (Degner, Doycheva, & Wentura, 2012). Further, L1 users showed greater arousal in skin conductance levels when responding to negative words compared to neutral and positive words than non-native speakers (Eilola & Havelka, 2011). No difference in other behavioural responses between the speaker groups was observed; interference patterns from negative words appeared identical in L1 and L2. Moreover, L1 was rated as more emotional than other languages and was frequently used for positive and

negative affects (Dewaele, 2010). Interaction of effects such as acquisition age, frequency and context of use may influence affective processing in L1 vs L2 (Pavlenko, 2012).

Less attention is given to second language emotional speech (Dewaele, 2013; Pavlenko, 2012). Language learners need to express speech with adequate affective states. Few studies have focused on L1-L2 pairings in relation to second language learners' production of emotional expressions. Most research has examined emotional speech by L1 speakers or emotionally-charged words by L1 and L2 speakers. Currently, few studies of production exist. This study bridges this gap by investigating Mandarin emotional speech uttered by native English speakers.

Each Mandarin syllable carries one of four tones (HIGH, RISE, LOW and FALL) to denote meaning. Note that the tone denoted LOW is actually a fall-rise tone going from medium low to low and then rises via a turning point. Mandarin learners may have difficulties expressing affective states due to first language effects such as incorrect tonal or sandhi production, where tone sandhi refers to rules that dictate that certain tones change into other tones with certain sequences; for example, the sandhi rule that LOW-LOW syllable pairs change into RISE-LOW. Comparatively, Mandarin is found to exhibit a larger pitch range than English for single word utterances (Keating & Kuo, 2010) and for news broadcasts (Yuan & Liberman, 2010). This is in conflict with the results of Eady (1982) which found no difference in F0 standard deviations between the tone language Mandarin and stress language English (Eady, 1982).

Differences in emotional speech are expected to occur in L1 versus L2 users' realizations. Key questions include:

Pitch: Higher F0 levels and larger ranges tend to occur for the emotions anger, joy, and fear, while sadness exhibits a smaller F0 mean and range (Zhang et al., 2006; Paeschke & Sendlmeier, 2000; Murray & Arnott, 1993; Scherer, 1986). Will L2 users realize these emotions similarly to L1 users? Will L2 utterances exhibit smaller F0 ranges or mean F0 than Mandarin L1 utterances in all, or just some, emotions?

Duration: The vowel durations associated with anger, fear, and joy tend to be shorter than those of sadness (Gu & Lee, 2007; Zhang et al., 2006; Ververidis & Kotropoulos, 2006; Schröder, 2001; Murray & Arnott, 1993; Scherer, 1986). Will L2 users realize these emotions similarly to natives? Will L2 utterances be larger and more varied than L1 utterances in all, or just some, emotions?

Intensity: The intensity tends to be higher for joy and anger, lower for sadness (Zhang et al., 2006; Murray & Arnott, 1993; Scherer, 1986), similar (Murray & Arnott, 1993) or lower (Zhang et al., 2006) for fear, or to increase in high-frequency energy for fear (Scherer, 1986). Will L2 users realize these emotions similarly to natives? Will L2 users show larger intensity range and greater intensity variation than L1 users in all, or just some, emotions?

Research into emotional speech is often based on TV or movie material, real utterances or pseudo-utterances, acted or controlled, scripted, partially scripted or unscripted, and real life situations. Scherer (2003) and Douglas-Cowie, Campbell, Cowie, & Roach (2003) describe methodologies for eliciting emotional speech. The results of a completely spontaneous approach may be hard to assess due to individual variations (Scherer, 2013). As suggested, producing emotions by enacting a script is more likely to render consistent and genuine

emotions (Douglas-Cowie el al., 2003). This study thus used a scenario approach where a situation is simulated to elicit subjects' emotional expressions (Enos & Hirschberg, 2006); semantically-neutral real utterances are constructed to avoid emotionally-colored implications.

This rest of this paper is structured as follows. First the methodological framework employed is discussed. Next follow the results including F0-ranges for level and contour sentences, F0 level, duration and intensity. The results are then scrutinized in the discussion section, immediately followed by a reflection of the limitation of this study. The conclusion section closes the paper.

Method

Participants

Sixteen English-L1 (mean age 28) and sixteen Mandarin-L1 users (mean age 26) participated in the study, with eight males and eight females in each group. All the English-L1 participants had studied Mandarin formally for at least one year at the language centre at National Cheng Kung University. Six Mandarin-L1 participants studied humanities, five studied social sciences and five engineering at the same university.

Emotion categories

This study examined joy, sadness, anger, and fear, with neutral as control. Acted expressions are adopted for this study since spontaneous elicitation of emotions may be unpredictable and not easily observable. When people act out emotions, they find references to their own

experiences. Also, acted expressions are similar to natural expressions (Banse & Scherer, 1996). Participants' acted production can therefore be assumed to be reliable given relevant scenarios.

A scenario approach introduced by Enos and Hirschberg (2006) was employed. The components include the character, the description of the situation, the goal which needs to be accomplished, and the obstacles which evoke the emotions. Both Mandarin and English translation were provided to help the participants understand the scenario. Scenarios embedded with each intended emotion were first written in Mandarin and then translated into English. Quality was ensured by translating the English version back into Mandarin by another bilingual. Adjustments were made to the English translation for all passages where the original and the back-translated Mandarin material differed.

Materials

The material used for the experiment consisted of one compound sentence, i.e., with two related complete sentences: *wo3 shou1 dao4 yi1 feng1 xin4, shi4 lao3 ban3 ji4 lai2 de5* (*I received a letter; it's from the boss*). The sentences are grammatically correct and semantically neutral to avoid random word emphasis or stress placement. RISE is considered the most difficult for L2 users to perceive and produce, and LOW the second most difficult (see for instance Tsai (2008)). The two sentences were designed to include all four tones and the second sentence contains a tone sandhi in the phrase (*lao3 ban3*) (*the boss*). The first sentence *wo3 shou1 dao4 yi1 feng1 xin4,* termed level tone sentence, is comparatively easier without a RISE, while the second sentence *shi4 lao3 ban3 ji4 lai2 de5*, termed contour tone sentence, may be more difficult as it has more tones.

Recording procedure

The participants were allowed to familiarize themselves with the scenarios prior to the recording. They were then asked to memorize the designated sentences. A role play was carried out between the participant and a native speaker of Mandarin to ensure that the participant was able to express convincing emotions. The 160 compound sentences (5 emotions × 32 subjects) were digitally recorded in a quiet room using a high-quality microphone.

Recognizable emotional utterances

The utterances were presented to 16 Mandarin-L1 raters (mean age 26) to ensure that the emotions conveyed by the participants were valid, using a 160-item questionnaire. For each item the raters were asked to select one of anger, joy, sadness, fear, neutral, don't know, or other. The sentence order was randomized. Items with a recognition rate less than 60% were discarded. The limit of 60% was chosen based on Tickle's (2000) recommendation.

F0 range and mean

The F0 ranges were measured at sentence level by eliciting the difference between the maximum and minimum F0. Next, the mean F0 range for each emotion category was calculated. Larger pitch variations were predicted for the contour tone sentence than the level tone sentence since the level tone sentence consists of more level tones. In addition, the mean pitch of each syllable within different emotion categories in the groups was computed.

Duration

Sentence durations were obtained by measuring the difference between their start and end-times. Moreover, the duration of each syllable was measured to verify if lengthening of specific words occurred in each emotion between the two groups.

Intensity

The intensity peak of each syllable in the sentences for each emotion was elicited and compared to that of neutral for both groups.

FIGURE 1 approximately here.

Results

Emotion recognition

Figure 1 shows the recognition rates for the Mandarin-L1 and Mandarin-L2 utterances. The recognition rates for Mandarin-L1 utterances were higher than Mandarin-L2 utterances for all the emotional categories. Joy in Mandarin-L1 showed the highest recognition rate (90.6%). The recognition rates for the other emotions anger, sadness, fear, and neutral were all higher than 80%. As for the Mandarin-L2 utterances, joy showed the lowest recognition rate (49.6%).

The recognition rate of anger and fear in Mandarin-L2 were 63.3% and 60.2% respectively. Sadness (73.0%) and neutral (74.6%) had the highest recognition rates.

Most emotions expressed by L1 speakers were recognized, while emotions expressed by L2 users were harder to recognize, especially joy. The results show that learners had more difficulty expressing high activation emotions than low activation emotions, in particular, joy. This lower recognition rate of joy by L2 users corresponds to the previous work based on meaningless utterances (Scherer, Banse, & Wallbott, 2001; Scherer, Banse, Wallbott, & Goldbeck, 1991; Banse & Scherer, 1996) and a more recent study using meaningless Mandarin sentences (Liu & Pell, 2012). First language effects, inaccurate L2 tones and unfamiliarity with emotional realization in L2 are possible factors. However, the exploration of these factors requires experimental paradigms beyond the scope of this study.

FIGURE 2 approximately here.

F0 range

F0 range is a measure of pitch variation. As expected the patterns of pitch variation for the level and contour tone sentences differed (see Figure 2). First, the F0 ranges of all emotion categories for the level tone sentence, consisting of tones with less pitch variation, were larger than that of neutral. The result corresponds to the prediction that anger, joy, and fear have a larger F0 range, except for sadness, which is assumed to have a smaller F0 range. The large

sadness F0 range by Mandarin-L1 in the contour tone sentence might have been triggered by the high arousal despaired sadness employed (Banziger & Scherer, 2005).

Anger and fear were the only emotions with a significant difference between the two groups for the level sentences ($t = 6.9$; $df = 26$; $p < .05$) and ($t = 9.7$; $df = 25$; $p < .05$), with Mandarin-L2 users exhibiting a larger range than the Mandarin-L1. However, effects from English as L1 was not evident, since their utterances had a larger range than the native speakers (cf. Keating & Kuo, 2010).

Next, the patterns for the contour tone sentence, which is the sentence with more variation in pitch, were slightly different. Only anger and sadness exhibited larger mean F0 ranges in the contour tone sentence than neutral (see Figure 2), unlike the level tone sentence where all emotions exhibited larger F0 ranges than neutral.

The F0 ranges for the different emotions were expected to be more varied compared to neutral as emotional utterances are expressed with more rising and falling pitch to indicate moods. However, the variation was close to neutral, particularly when expressing low activation emotions such as sadness.

Comparing the F0 range across the two speaker groups, only joy was significantly different ($t = 5.4$; $df = 23$; $p < .05$). The F0 range of joy in Mandarin-L2 was larger than that of Mandarin-L1. Although emotions with larger F0 ranges were easier to recognize (Paulmann, Pell, & Kotz, 2008; Bachorowski, 1999; Banse & Scherer, 1996), joy in Mandarin-L2 had lower recognition rates than Mandarin-L1 despite a larger F0 range. The measured F0 range

of contour tone sadness was larger for Mandarin-L1 than Mandarin-L2 although the difference is not significant ($t = 2.8$; $df = 26$; $p > .05$).

The two speaker groups realized HIGH and FALL differently, which are believed to be more difficult to utter for foreign learners (Shen, 1989). L2 users usually employed high pitch as HIGH and utter FALL with additional loudness instead of a falling pitch. Contrastively, when expressing anger and fear, the Mandarin-L2 realized both HIGH and FALL with steep fall, while Mandarin-L1 employed a less steep fall. Overall, the F0 range varied more across emotions for Mandarin-L2 than Mandarin-L1.

Emotions had larger F0 ranges in the contour tone sentence than in the level tone sentence. This could be due to the stronger presence of contour tones in the contour tone sentence. Sentence tonal combinations therefore appear to affect the F0 range of emotional expressions.

FIGURE 3 approximately here.

FIGURE 4 approximately here.

Mean syllable F0

The mean syllable F0 represents the absolute F0 level at syllable level. Figure 3 shows the mean F0 across the emotions for the speaker groups and Figure 4 shows the mean F0 for the syllables over the sentence for the emotions. The mean pitch of joy, anger, and fear was high for Mandarin-L1, corresponding to the prediction that these are high arousal emotions pronounced with high pitch. The mean pitch of sadness was slightly higher than neutral, consistent with the reported English-L1 results (Pell et al., 2009). The difference was comparatively larger for the contour tone sentence than for level tone sentence.

The Mandarin-L2 users uttered joy with the highest mean pitch, similar to the Mandarin-L1. Contrary to expectations, fear showed a higher mean pitch than anger. Neutral and sadness were realized in a similar manner to those by Mandarin-L1. Both neutral and sadness had similar pitch in the level tone sentence and a higher mean pitch in the contour tone sentence.

Effect of sandhi

The pitch means for *feng1, xin4, lao3* and *ji4* uttered as joy are significantly different, namely ($t = 5.31$; $df = 23$; $p < .05$), ($t = 5.17$; $df = 23$; $p < .05$), ($t = 5.53$; $df = 23$; $p < .05$) and ($t = 5.09$; $df = 23$; $p < .05$), respectively. The syllables *feng1* and *xin4* are parts of the tri-syllabic phrase *yi1 feng1 xin4* meaning 'a letter' with HIGH-HIGH-FALL. This sequence was realized as FALL-HIGH-FALL because of the HIGH sandhi effect of *yi1* 'one; a'. The HIGH syllable *yi* 'one; a' remained HIGH when uttered in isolation, but changed into RISE when followed by FALL, and turned into FALL when followed by any other tones. The second syllable *feng1* with HIGH in the phrase was much lower due to the low ending pitch of the previous FALL, which is a common downstep phenomenon. The last FALL syllable of the tri-syllabic phrase *xin4* (FALL) showed an even lower pitch compared to the preceding FALL syllables *dao4*

and *yi4* (now as realized sandhi tone) within the level tone sentence. Also, *xin4* being the last word, showed a natural pitch fall towards the end of the first declarative sentence. The declination effect observed among the Mandarin-L2 was not as strong among the Mandarin-L1. Mandarin-L2 users realized the tri-syllabic phrase *yi1 feng1 xin4* 'a letter' differently from the natives, probably due to unfamiliarity with the tone sandhi of the first syllable. The drop in pitch for the following two syllables *feng1 xin4* was thus significantly smaller.

Lao3 is the first syllable of the disyllabic phrase *lao3 ban3* 'boss'. The sandhi rule triggers, where the first LOW changes into RISE when followed by a LOW. The pitch of Mandarin-L1 *lao3* was high and level with fewer peaks and troughs in joy, while the Mandarin-L2 pitch was more varied. The Mandarin-L2 users uttered RISE-FALL-RISE patterns for the LOW LOW tone sequence more carefully with more details than Mandarin-L1 users.

Ji4, the syllable immediately following the sandhi disyllabic *lao3 ban3*, is the verb of the embedded clause led by *shi4* 'is' of the contour tone sentence. The FALL syllable *ji4* appeared challenging to the Mandarin-L2 resulting in a significant difference in the realization of fall compared to the natives.

The Mandarin-L1 users employed a higher pitch than the Mandarin-L2 when expressing anger, except for the last syllable. Both groups showed an overall lower pitch range than that of joy. The Mandarin-L2 uttered *Feng1*, *xin4*, *lao3* and *ji4* with a lower pitch than the Mandarin-L1. The Mandarin-L2 used similar pitches to express the LOW sequence *lao3 ban3*, while the Mandarin-L1 applied a higher pitch to the first syllable *lao3*, realizing the tone sandhi.

No significant differences were detected between Mandarin-L1 and Mandarin-L2 when expressing sadness. The speakers employed low pitch with small variations to express sadness, as reported in previous studies (Pell et al., 2009; Murray & Arnott, 1993; Burkhardt & Sendlmeier, 2000; Zhang et al., 2006; Yuan et al., 2002). The Mandarin-L2 were comparatively more successful in depicting emotions with low and stable pitch such as sadness than emotions with high pitch and larger pitch variation.

The mean pitch of fear was slightly higher than that of sadness although the difference is not significant. Fear and sadness also exhibited flatter F0 contours than other emotions, consistent with the previous studies (Gu & Lee, 2007; Yuan et al., 2002).

In summary, most emotions were uttered in agreement with the patterns reported in the literature with respect to pitch means. Joy, anger and fear exhibited high pitch, while sadness and neutral had similar and lower pitch. Mandarin-L2 realized fear in higher pitch than anger, while Mandarin-L1 portrayed anger with higher pitch than fear. In addition, Mandarin-L1 used larger pitch ranges to express different emotions than Mandarin-L2. Pitch realizations were more consistent across Mandarin-L1 utterances and more varied across Mandarin-L2 utterances.

FIGURE 5 approximately here.

Duration

The results show that natives act emotions with similar durational attributes. The sentence containing level tones is uttered with shorter duration than the sentence with contour tones (see Figure 5), although these differences are not significant.

Contrary to predictions, the Mandarin-L2 users utter the sentence containing level tones with longer duration than the sentence with contour tones (see Figure 5). This may be explained by L2 users' uttering level tones with longer durations in an attempt to make level tones more perceivable, which are acoustically unmarked in their first language (Shen, 1989). Alternatively, it may be that there are more unvoiced-initial syllables such as *shou1*, *feng1*, and *xin4* in the level tone sentence. These syllables, with unvoiced initials, have longer durations (Shen, 1989). Moreover, L2 users' over-articulation may magnify this effect, contributing to the difference in duration.

Furthermore, semantic salience might have been realized differently by the L2 users who may have lengthened the level tone sentence (*I received a letter*) to denote its greater importance, while the contour tone sentence (*It's from the boss*) is taken as a predicate for further elaboration on the level tone sentence.

FIGURE 6 approximately here.

Figure 6 shows the durations of the emotion categories at syllable level. The syllable durations of joy is different for the two groups except for the final segment *de5* ($t = 1.06$; $df = 23$; $p > .05$). *De5* of Mandarin-L2 is the only syllable that has shorter duration than Mandarin-

L1. The final syllables (*xin4* and *de5*) of both sentences are lengthened by both groups. The phenomenon is stronger for the Mandarin-L2.

The anger syllable durations were significantly different across the two groups except for *ji4* ($t = 0.1$; $df = 26$; $p > .05$). The duration of *ji4* was 0.19 seconds for both groups. Mandarin-L2 employed longer syllable durations in both sentences. The overall speech rate of anger was fast for both groups and the lengthening in the final syllable *de5* in the contour tone sentence was small. This pattern differs from that of joy where the final segments of both sentences were lengthened.

The syllable duration associated with sadness was significantly different for *shou1* ($t = 8.9$; $df = 23$; $p < .05$), *dao4* ($t = 26.2$; $df = 23$; $p < .05$), *yi1* ($t = 5.3$; $df = 23$; $p < .05$), *xin4* ($t = 6.2$; $df = 23$; $p < .05$), *shi4* ($t = 8.7$; $df = 23$; $p < .05$), *lao3* ($t = 11.2$; $df = 23$; $p < .05$) and *ban3* ($t = 11.1$; $df = 23$; $p < .05$). The other syllables *wo3*, *feng1*, *ji4*, *lai2* and *de5* had similar durations. Sadness was realized with a slow speech rate towards the sentence final position, which seemingly posed no difficulties for the L2 users.

For fear the syllables *dao4* ($t = 12.0$; $df = 25$; $p < .05$) and *xin4* ($t = 13.0$; $df = 25$; $p < .05$) were significantly different. The Mandarin-L2 uttered the two falling tone segments with a longer duration than the Mandarin-L1.

TABLE 1 approximately here.

Table 1 summarizes the percentage of syllables with similar durations, that is, durations that are not significantly different. It shows that the Mandarin-L2 portrayed the emotion fear with the most similar duration to that of the natives. The learners tended to lengthen each syllable to express emotions in Mandarin, except for fear, where the vocal duration was similar to that of the natives. In particular, for joy and anger, most segments were lengthened by the Mandarin-L2. The difference in syllable duration uttered by the Mandarin-L2 may be one factor that contributes to the low recognition rate of joy and anger. The learners expressed sadness, an emotion with a slower speech rate and larger syllable duration, more successfully than other emotions. Fear, characterized by its fast speech rate and shorter syllable duration, was realized with a longer duration than neutral by the Mandarin-L1 and with an even longer duration by the Mandarin-L2.

The neutral tone, reported to have shorter duration than other tones (Lee, Tseng, & Ouh-Young, 1989), showed a larger duration than other tones in both groups. *De5*, the final segments in the contour tone sentence underwent reduplication of the rising pitch tail of the preceding rising tone *lai2*. Consequently, the duration became longer. The pitch of the reduplication tail may have been affected by the emotion category uttered. The segment might be lengthened to show a designated emotion. The syllable *ji4* also exhibited similar durations in most emotional expressions except in joy. Emotions with higher arousal such as joy, anger and fear had a high pitch, while the pitch of sadness was falling with longer duration. Emotions were thus realized in both the pitch and duration of the final segment.

FIGURE 7 approximately here.

Intensity

Intensity, or loudness, characterizes the energy a speaker puts into an utterance. Figure 7 illustrates the intensity for each emotion across the two groups. No significant difference was found between the groups for joy, although some FALL syllables (*shi4* and *ji4*) were pronounced with stronger intensity by the Mandarin-L2 than Mandarin-L1 users, consistent with the view that Mandarin users of English L1 regard falling tones as stressed segments (White, 1981). Both groups employed stronger intensity in all segments for joy than neutral.

The intensity of anger is similar to that of joy, although the strengths of *dao4* and *lao3* were significantly different across the two groups. *Dao4*, the accomplishment marker of the verb *shou1*, was articulated as the accented syllable in the phrase *shou1 dao4*. Contrastively, the Mandarin-L2 placed similar weight on both syllables. Both groups took the disyllable phrase *lao3 ban3* as a left-headed foot, where the Mandarin-L1 applied more weight on the first syllable as a way of showing anger while the Mandarin-L2 placed similar weight on each syllable leading to a flatter intensity contour.

The intensity of sadness was similar to that of neutral for both groups, suggesting that they had similar stress placement. Significant differences between the groups were observed for *wo3* ($t = 4.1$; $df = 26$; $p < .05$), *xin4* ($t = 10.0$; $df = 26$; $p < .05$) and *shi4* ($t = 8.7$; $df = 26$; $p < .05$).

Fear was not expected to be comparable to joy and anger in terms of high intensity. Mandarin-L2 exhibited the highest intensity. The intensity of *shou1* ($t = 8.2$; $df = 25$; $p < .05$), *xin4* ($t = 7.1$; $df = 25$; $p < .05$) and *shi4* ($t = 6.0$; $df = 25$; $p < .05$) were significantly different to those of Mandarin-L1. Common to these is the initial voiceless hissing sound, uttered with more weight by the Mandarin-L2 than Mandarin-L1 users.

In summary, the Mandarin-L1 used a wider intensity range to express emotions than the Mandarin-L2. In particular, joy and anger exhibited high intensity, as also found in previous studies (Scherer, 1986; Murray & Arnott, 1993), revealing loudness as means of emphasizing these emotions. Both groups employed similar stress in expressing sadness and neutrality and less stress when expressing fear and sadness, although the Mandarin-L2 employed higher intensity than the natives. The intensity results are consistent with the observations that high pitch is used by the Mandarin-L2 users to show sadness and fear. The L2 users, however, used similar intensity ranges to express different emotions.

Discussion

The L2 users' emotional speech was presented to 16 Mandarin-L1 raters to verify which ones were recognizable prior to acoustic analysis. The outcome revealed that the learners' portrayals were harder to recognize than those of natives, especially joy. This low recognition of joy compared to the other emotions bears similarity with that of previous work employing meaningless sentences produced by natives (Liu & Pell, 2012), where they also observed discrepancies in recognition and acoustic patterns.

F0 range

The F0 range, expressing pitch variations, of the level tone sentence produced by the Mandarin-L2 showed more variation compared to that by the Mandarin-L1 users, but only anger and fear were associated with significantly larger F0-ranges. In the contour tone sentences, joy is the only emotion that was significantly different, where the Mandarin-L2 employed larger ranges than the Mandarin-L1. Mandarin-L2 sadness F0-range was non-significantly smaller than that of Mandarin-L1. Mandarin-L1 employed non-significantly smaller F0-ranges when expressing joy than fear and significantly smaller F0 ranges for joy than neutrality. The fact that the neutral F0-ranges were larger than those of joy in the contour tone sentence may be related to pragmatic interpretations, as discussed in subsequent paragraphs.

Contour tones resulted in larger F0-ranges than level tones across emotions, more so for L1 than L2 users. Learners generally employed larger F0-ranges.

The L2 users uttered level tones with less stable and larger F0 ranges, suggesting that a pitch-sustaining effort to strengthen the impression of 'level' tone may be at work, in particular for the high-arousal emotions anger and fear. It may also be possible that learners' presupposed expectation of emotion for a tone language is such that tone must be "highlighted" or "emphasized" one way or another in order to denote specific emotions, thereby exercising their "highlighting" tonal effect, even on the level tone, and hence resulting in their less stable and larger F0 ranges. The realization of specific tonal patterns in a tone language has certain impact on L2 users' emotional portrayals, observably in a more exaggerated fashion. As shown, the learners articulated contour tones in joy with larger F0 range than the natives, while with non-significant smaller F0 range in sadness than the natives.

In terms of F0 range, the L1 and L2 users' emotions differed more for the level tone sentence than the contour tone sentence. For the level tone sentence, only the F0 range of joy was similar for both groups. The semantic impact of the level tone sentence ("I received a letter") tends towards fact-stating, while the contour tone sentence ("It's from the boss") carries more weight or suspense. This functional cue requires L2 users' attentions, in addition to attending to tonal realizations.

The results reported herein reveal that both L1 and L2 users employed larger F0 ranges in anger than in fear for both sentences within their own emotion categories, implying that these two emotions are probably cross-linguistically or culturally similar in terms of F0 variation. Also, L2 users employed similar F0 ranges for these emotions in the level tone sentence, but with a larger F0 range difference in the same emotions in their own contour sentence. This suggests that contour tones contribute to larger F0 variation within L2 users' emotional expressions than level tones. Hence, the number and type of tones in the speech material have an effect on L2 users' emotional realization. Note that L2 users' anger had very similar range in both sentences, while L2 users' fear exhibited a larger difference in range across the two sentences. That is, fear had much larger F0 range in the level sentence than in the contour sentence. This larger variation in F0 by L2 users may have been induced by their tonal realizations in the contour sentence, suggesting that Mandarin tones, particularly HIGH, FALL, and LOW, do have an effect on expressing emotions among the L2 users. For the contour sentence, the natives revealed a smaller F0 range for joy than fear, followed by neutral. This pattern was reversed for the natives' level sentence. Natives demonstrated a larger variation in range across the two sentences than L2 users. In other words, L1 level

sentences exhibited a smaller range than L1 contour sentences, while L2 users' F0 ranges remained large for both sentences.

The F0 range was smaller for joy than anger for both speaker groups in the contour tone sentence, while only L2 users showed the same characteristics in the level tone sentence. The natives employed a larger F0 range to express joy than anger in the level tone sentence. Similar patterns have also been reported for meaningless native speech (Lin & Pell, 2012). One explanation could be that native speakers strengthen or dramatize meaningless expressions that contain emotions. If expressions have no meaning, the focus may be shifted to "properly" produce the emotion intended, and hence unintentionally strengthening the emotions. However, when semantic information is embedded into expressions that also contain emotions, the boundary between joy and anger F0 range becomes blurred. For L2 users, the semantic burden tends to be heavier than for L1 users. This could also explain why L2 users generally used larger F0 ranges in both sentences compared to natives. In terms of mean F0, natives employed a higher pitch (240 Hz) for anger than fear (231 Hz). L2 users uttered fear with a slightly higher pitch (222 Hz) than anger (207 Hz). Comparatively, anger has been reported to show a higher mean F0 than fear among native meaningless utterances (Lin & Pell, 2012). Revealingly, Mandarin linguistic tone has an impact and also imposes a semantic burden for L2 users. Hence, L2 users' emotional speech can be expected to be more ambiguous, especially where there is overlapping use of F0 range and mean in certain emotion categories such as fear and anger.

The acoustic differences observed for anger, fear and joy between L1 and L2 users' Mandarin speech agree with previous literature (Jaywant & Pell, 2012; Liu & Pell, 2012; Pell et al., 2009; Juslin & Laukka, 2001), where these acoustic differences are larger than for other

emotions such as sadness, surprise and disgust. Moreover, there may be potential culture differences associated with expressing anger, fear and joy, contributing to a possible L1 effect.

The L2 users employed larger F0 ranges in the level tone sentence than the L1 users, except for similar joy F0-ranges. In the contour tone sentence, the L2 users' F0 ranges were larger for joy and smaller for sadness compared to natives, suggesting over expressing joy and sadness. The three other emotions were similar though slightly larger than those of natives. Thus, the acoustic F0 variation may seem larger for the L2 than for L1 users, whereas the recognition rate was lower for L2 than L1 speech. A distinct F0 range is more likely to be recognized. This inconsistency could occur if Mandarin-L2 users over-express the emotions for any reasons, such as possible involuntary gestures, since this is a foreign language where they concentrate to express themselves. Imprecise knowledge about the culture of the target language could also be a cause. Further, interactional effects may be caused by the participants' mental or physical state during the enactment or interferences from L2 users' own culture. Also, the brief exposure to the English translation of the scenario may have unintentionally trigged the L2 user's own culture during recording.

The fact that L2 users generally used larger F0 ranges than L1 users suggests larger tonal variation in L2 users' realizations. Such vocal expression may be less stable and confuse L1 listeners, thus leading to a lower recognition rate.

Certain phonological realizations involving tonal changes were executed differently. Mandarin-L2 realized joy with the tri-syllabic phrase *yi1 feng1 xin4* differently from Mandarin-L1. Level and falling tones and sandhi rules were problematic for Mandarin-L2 users. Moreover, the pitch step-down in the sentence-final element was smaller among the L2

users. For the disyllabic phrase involving contour tones, *lao3 ban3,* the sandhi tone change was articulated with more elaborate and detailed pitch fall and rise by the L2 users. Tonal characteristics realized in sandhi sequence by L2 users probably impacted their emotional portrayals.

Mean F0

Measurements of mean F0 for both groups are consistent with the literature where high pitch correlates with high activations emotions such as joy, anger, and fear. Sadness and neutral were low in mean pitch, with the former slightly higher than the latter for both groups. The natives exhibited a higher mean F0 when portraying anger than fear, while the L2 users exhibited a higher mean F0 when expressing fear than anger. The relative different use of mean pitch in portraying fear and anger by the two speaker groups need further investigation. One possible explanation could be the subtle distinction between panic fear and hot anger (cf. Banziger & Scherer, 2005), where panic fear may exhibit a higher pitch than hot anger. Emotional expressions involving contextual-pragmatic interpretations may also play a part.

Both groups expressed anger with a lower mean pitch than joy. In particular, the L2 users employed significantly lower pitch in *feng1, xin4, lao3,* and *ji4* than the natives for both joy and anger. Both speaker groups also expressed sadness with lower and more stable pitch than joy, anger, and fear. The results are different to those reported for meaningless utterances in the literature, where anger resulted in a high mean pitch and happiness medium mean pitch, with sadness and fear low pitch (Liu & Pell, 2012). A possible explanation is that realizing meaningless utterances requires less mental effort than realizing meaningful utterances as used in this study.

Concerning the first research question addressing pitch, L2 users' emotional utterances revealed larger F0 ranges than natives except for sadness in the contour tone sentence and slightly smaller yet close range for joy in the level sentence, suggesting that L2 users tend towards less stable or more dramatic tonal use during emotional expressions. Both groups exhibited a pattern where anger, joy, and fear have higher mean F0 than sadness (Zhang et al., 2006; Yuan et al, 2002; Murray & Arnott, 1993; Scherer, 1986), than neutral (Paeschke et al., 1999) or than sadness/neutral (Gu & Lee, 2007). Both groups exhibited a higher mean F0 for joy than anger, and the opposite pattern for anger versus fear. Comparatively, L1 and L2 use of mean F0 revealed a common trend in expressing joy and anger, perhaps implying a cross-linguistic or cultural similarity for joy and anger, and a more subtle display of anger versus fear.

Pitch effect on emotional speech

The results suggest that pitch height and contour changes for Mandarin tonal realizations may interfere with pitch use for L2 emotional expressions. English speakers employed mainly pitch contours to characterize their affective speech (Ross, Edmondson, Seibert, & Homan, 1987), with intensity and duration as contributory factors (Williams & Stevens, 1972). For Mandarin learners of English L1, lexical tones posed difficulty due to their unfamiliarity with the associations between F0 and segmental features (Shen, 1989) and their learning was affected by the knowledge of pitch use for English stress and intonation systems (White, 1981; Broselow, Hurtig, & Ringen, 1987). A recent study comparing tone language L1 vs non-tone L1 speakers learning Mandarin suggested that contour tonal contrast is difficult for L2 learners, irrespective of their L1 (Hao, 2012) and that other factors than L1 may affect their

tonal acquisition. This study confirms the trend of difficulty associated with pitch and contour realizations in the context of emotional speech.

Further, it was found that English preschoolers had difficulty interpreting pitch contours portraying emotions such as happiness and sadness and that the connections between stereotypical intonations and their intended emotions may need to be learned through experience (Quam & Swingley, 2012). For foreign or second language learners to successfully acquire and communicate their Mandarin L2 emotional expressions, including English L1 speakers, this may also be the case. Additionally, a theory of language embodiment proposed by (Pavlenko, 2005) suggests an integrated model of children's emotion socialization where learning of linguistic/phonological elements is mingled with other stimuli such as visual, auditory, and visceral sensations, autobiographical memories and affect. These studies suggest that language acquisition requires a process of maturation of linguistic input integrated with other dimensions. In the context of this study relevant dimensions include autobiographical memories, visceral mode, and emotion.

One may question whether emotion expression in nontonal-nontonal L1-L2 is easier. Further research is needed, since other factors including culture contribute to emotional utterances.

Duration

L2 users' sentences were uttered with longer durations than native articulations. The lengthening of L2 users' tones suggests that they emphasized their "levelness". The L2 users expressed joy with larger syllabic durations than the natives with the exception of utterance-finals. Both groups expressed anger with faster speech rates than joy and the L2 users'

durations were significantly longer. Sadness was associated with the most significant differences between the groups at syllable level, in particular for HIGH and FALL, and a sandhi effect for HIGH and LOW. Fear was the most similar emotion in terms of duration. Unexpectedly, the neutral tone sentence-final *de5* exhibited longer durations than other syllables for both groups. It was uttered using a high pitch with joy, anger and fear and using a FALL with sadness.

Concerning the second research question addressing duration, the results for both groups agree with the literature where utterances related to anger, fear and joy are shorter than those of sadness (Gu & Lee, 2007; Zhang et al., 2006; Ververidis & Kotropoulos, 2006; Schröder, 2001; Murray & Arnott, 1993; Scherer, 1986). Of the three high-arousal emotions, anger was shorter than joy. L2 users' utterances were the longest. The duration of fear was the most similar across the speaker groups, while that of sadness was the most dissimilar. L2 users' anger was associated with the second longest duration, while fear was second longest for natives. L2 users employed longer durations than natives.

Intensity

L2 users exhibited lower and flatter intensity contours for joy and anger than natives, and sadness and fear with higher and flatter intensity contours compared to natives. Both groups applied less stress when communicating fear and sadness compared to joy and anger, agreeing with previous work (Zhang et al., 2006). L2 users articulated FALL with higher intensity than natives irrespective of emotion. Natives placed more weight on the first sandhi syllable in LOW sandhi sequences. The two speaker groups thus realized intensity strength and variation differently.

Of the three high activation emotions, natives expressed joy and anger with similar intensity, consistent with previous studies (Zhang et al., 2006; Murray & Arnott, 1993; Scherer, 1986), and fear with lower intensity. Sadness exhibited the lowest intensity, also agreeing with previous studies (Zhang et al., 2006; Murray & Arnott, 1993; Scherer, 1986). L2 users uttered the four emotions with more similar intensity signatures than natives. Observable differences include L2 users' portrayal of fear with higher intensity than natives. Sadness was portrayed with lower intensity compared to anger, joy and fear.

Concerning the third research question addressing intensity, the results agree with the literature reporting high mean intensity for joy and anger and low for sadness (Zhang et al., 2006; Murray & Arnott, 1993; Scherer, 1986). There was no noticeable difference in intensity variations across the two groups and there is hence no obvious L1 effect.

Limitations of this study

The scenarios employed consisted of a set of contexts; both English and Mandarin texts were presented to assist participants' understanding. It may be possible that the brief exposure to the English translations of the scenarios had an effect, in which the first language and the first socio-culture interfered. The recording material comprised a compound sentence, with two complete and related sentences. One may thus argue that the data gathered was too limited.

Also, it is challenging to find suitable L2 participants at the same level of Mandarin proficiency, further complicated by the fact that emotion is generally not taught in Mandarin language programs. Next, the two sentence types, the level tone sentence and contour tone

sentence, are not completely pure in terms of level and contour tones as it is nearly impossible to construct meaningful sentences consisting of only level tones and contour tones, respectively. Another source of error could be that tonal characteristics realized in sandhi sequence by L2 users probably impacted their emotional portrayals.

Summary and conclusion

The F0 ranges of anger and fear of the level tone sentence by L2 users were significantly larger than for natives, and F0 ranges of the contour tone sentence by L2 users were larger than by natives except sadness. Both groups employed higher F0 when expressing joy than anger. The groups diverged when expressing anger and fear—natives used higher F0 for anger than for fear, while L2 users employed higher F0 for fear than for anger. Sandi rules complicated realizations by L2 users. L2 users generally articulated more slowly, with fear showing the most similar duration in both groups while sadness the most dissimilar. Generally L2 users did not vary intensity when expressing emotions. F0 appears to be the most salient feature contributing to a foreign accent of emotional Mandarin speech. Future pedagogical applications should place primary focus in tonal training in combination with emotional traits. Articulation drills could be developed to ensure that high level tones are articulated as level and not falling when expressing anger and fear and that falling tones are articulated with falling pitch and not increased loudness. Moreover, L2 users should be drilled in articulating shorter level tones and longer contour tones. In terms of intensity, L2 users should be drilled in utilizing the intensity repertoire when expressing various emotions. Further, linking emotions in Mandarin speech with tonal practice may benefit non-tone language learners in both vocal and emotional communication. Examples of language and emotional use in socio-cultural context should be considered for practical training in varied forms both as in class

activities and extracurricular project work. Cultural issues in emotional portrayal are not specifically examined in this study; future research should address issues involving prosodic impact on L2 emotional speech or vice versa.

Acknowledgements

The author thanks Jessica Hong for assistance with collecting and processing the data and the anonymous reviewers for their invaluable input.

References

Bachorowski, J. (1999). Vocal expression and perception of emotion. *Current Directions in Psychological Science, 8*(2), 53-57.

Banse, R., & Scherer, K. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*, 614-636.

Banziger, T., & Scherer, K. (2005). The role of intonation in emotional expressions. *Speech Communication*, *46*(3-4), 252-267.

Broselow, E., Hurtig, R.R., & Ringen, C. (1987). The perception of second language prosody. In G. Ioup & S.H. Weinberger (Eds.), *Inter-language Phonology: The Acquisition of Second Language Sound System* (pp. 350-361). Cambridge: Newbury House Publishers.

Burkhardt, F., & Sendlmeier, W. (2000). Verification of acoustical correlates of emotional speech using formant-synthesis. *Proceedings of ITRW on Speech and Emotion*, 151-156.

Degner, J., Doycheva, C., & Wentura, D. (2012). It matters how much you talk: On the automaticity of affective connotations of first and second language words. *Bilingualism: Language and Cognition, 15*, 181–189.

Dewaele, J.-M. (2010). *Emotions in Multiple Languages*. Basingstoke: Palgrave Macmillan.

Dewaele, J.-M. (2013). *Emotions in Multiple Languages* (2nd ed.). Basingstoke: Palgrave Macmillan.

Douglas-Cowie, E., Campbell, N., Cowie, R., & Roach, P. (2003). Emotional speech: towards a new generation of databases. *Speech Communication, 40*, 33-60.

Dromey, C., Silveira, J., & Sandor, P. (2005). Recognition of affective prosody by speakers of English as a first or foreign language. *Speech Communication, 47*(3), 351-359.

Eady S. J. (1982). Differences in the F0 patterns of speech: Tone language versus stress language. *Language and speech, 25*, 29-42.

Eilola, T. M., & Havelka, J. (2011). Behavioural and physiological responses to the emotional and taboo Stroop tasks in native and non-native speakers of English. *International Journal of Bilingualism, 15*, 353-369.

Enos, F., & Hirschberg, J. (2006). *A framework for eliciting emotional speech: capitalizing on the actor's process*. In *Workshop on Corpora for Research on Emotion and Affect*. *5th International Conference on Language Resources and Evaluation* (*LREC2006)*, 6-10.

Graham, C., Hamblin, A., & Feldstein, S. (2001). Recognition of emotion in English voices by speakers of Japanese, Spanish and English. *IRAL-International Review of Applied Linguistics in Language Teaching, 39*(1), 19-37.

Gu, W. & Lee, T. (2007). Quantitative analysis of F0 contours of emotional speech of Mandarin. *Proceedings of 6th ISCA Workshop on Speech Synthesis*, 228-233.

Hao, Y-C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics, 40*, 269-279.

Jaywant, A., & Pell, M. D. (2012). Categorical processing of negative emotions from speech prosody. *Speech Communication, 54*, 1-10.

Juslin, P. N., & Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion, 1*, 381-412.

Keating, P. & Kuo, G. (2010). Comparison of speaking fundamental frequency in English and Mandarin. *UCLA Working Papers in Phonetics, 108*, 164-187.

Lee, L., Tseng, C., & Ouh-Young, M. (1989). The synthesis rules in a Chinese text-to-speech system. *IEEE Transactions on Acoustics, Speech and Signal Processing, 37*(9), 1309-1320.

Liu, P., & Pell, M.D. (2012). Recognizing vocal emotions in Mandarin Chinese: A validated database of Chinese vocal stimuli. *Behavior Research Methods, 44*, 1042-1051.

Murray, I., & Arnott, J. (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*, 1097-1108.

Paeschke, A., Kienast, M., & Sendlmeier, W. (1999). F0-contours in emotional speech. *Proceedings of the 14th International Congress of Phonetic Sciences,* 929-932.

Paeschke, A. & Sendlmeier, W. F. (2000). Prosodic characteristics of emotional speech:

Measurements of fundamental frequency movements. *Proceedings of ITRW on Speech and Emotion*, 75-80.

Paulmann, S., Pell, M., & Kotz, S. (2008). How aging affects the recognition of emotional speech. *Brain and language, 104*(3), 262-269.

Pavlenko, A. (2005). *Emotions and Multilingualism*. Cambridge, MA: Cambridge University Press.

Pavlenko, A. (2012). Affective processing in bilingual speakers: Disembodied cognition? *International Journal of Psychology, 47*(6), 405-428.

Pell, M., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S., (2009). Factors in the recognition of vocally expressed emotions: a comparison of four languages. *Journal of Phonetics, 37*, 417-435.

Pereira, C., & Watson, C. (1998). Some acoustic characteristics of emotion. *Proceedings of 5th International Conference on Spoken Language Processing*, paper 0684.

Quam, C., & Swingley, D. (2012). Development in children's interpretation of pitch cues to emotions. *Child Development, 83*, 236-250.

Ross, E. D., Edmondson, J. A., Seibert, G. B., & Homan, R. W. 1987. Acoustic analysis of affective prosody during right-sided Wada test: A within-subjects verification of the right hemisphere's role in language. *Brain and Language, 33*, 128-145.

Scherer, K. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99*(2), 143-165.

Scherer, K. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40*, 227-256.

Scherer, K. (2013). Vocal markers of emotion: Comparing induction and acting elicitation. *Computer Speech & Language, 27*(1), 40-58.

Scherer, K., Banse, R., & Wallbott, H. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32*(1), 76-92.

Scherer, K., Banse, R., Wallbott, H., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion, 15*(2), 123-148.

Schröder, M. (2001). Emotional Speech Synthesis - A Review. *Proceedings of Eurospeech, 1*, 561-564.

Shen, X. (1989). Interplay of the four citation tones and intonation in Mandarin Chinese. *Journal of Chinese Linguistics, 17*(1), 61-74.

Tao, J., Kang, Y., & Li, A. (2006). Prosody conversion from neutral speech to emotional speech. *IEEE Transactions on Audio, Speech, and Language Processing, 14*(4), 1145-1154.

Tickle, A. (2000). English and Japanese speaker's emotion vocalizations and recognition: a comparison highlighting vowel quality. *Proceedings of ISCA Workshop on Speech and Emotion,* 104-109.

Tsai, P. L. (2008). *Perception and Production of Chinese Lexical Tone by Adult English Speaking Learners* (Unpublished master's thesis).

Ververidis, D., & Kotropoulos, C. (2006). Emotional speech recognition: Resources, features, and methods. *Speech Communication, 48*(9), 1162-1181.

White, C.M. (1981). Tonal perception errors and interference from English intonation. *Journal of Chinese Language Teachers Association, 16*, 27-56.

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *Journal of the Acoustical Society of America, 52,* 1238-1250.

Yuan, J., & Liberman, M. (2010). F0 declination in English and Mandarin broadcast news speech. *Proceedings of Interspeech,* 134-137.

Yuan, J., Shen, L., & Chen, F. (2002). The acoustic realization of anger, fear, joy and sadness in Chinese. *Proceedings of ICSLP*, 2025-2028.

Zhang, S., Ching, P.C., and Kong, F. (2006). Acoustic analysis of emotional speech in Mandarin Chinese. *Proceedings of ISCSLP*, 57-66.

Figure captions

FIGURE 1 Emotion recognition for Mandarin-L1 and Mandarin-L2 utterances

FIGURE 2 F0 range. Error bars indicate standard deviation.

FIGURE 3 Mean F0 across emotions for Mandarin-L1 and Mandarin-L2. Error bars show standard deviation.

FIGURE 4 Pitch means for the syllables over the sentences for the emotions by Mandarin-L1 and Mandarin-L2 users

FIGURE 5 Mean sentence durations

FIGURE 6 Syllable durations for joy, anger, sadness, and fear

FIGURE 7 Intensities associated with emotions

Table titles

TABLE 1 Percentage of similarity in syllable duration of different emotions between the two groups

Figures

FIGURE 1

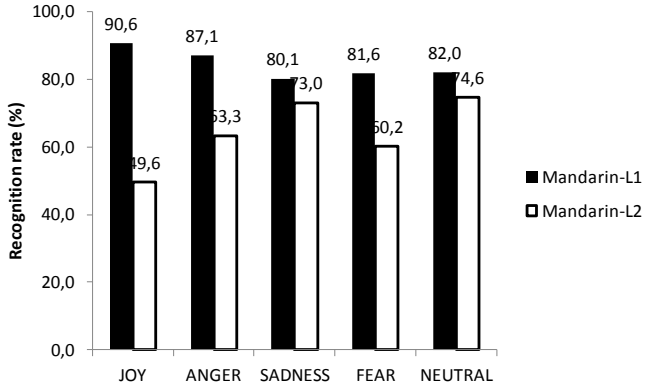
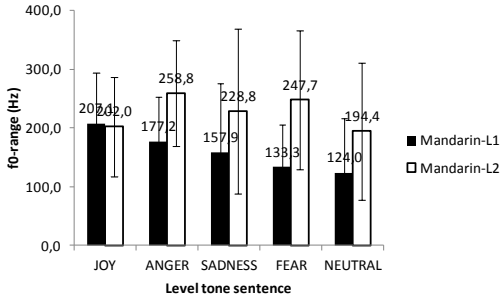


FIGURE 2


400,0
300,0
200,0
100,0
0,0
f0-range (Hz)
207,4
242,0
177,2
258,8
157,9
228,8
133,3
247,7
124,0
194,4
JOY
ANGER
SADNESS
FEAR
NEUTRAL
Level tone sentence
Mandarin-L1
Mandarin-L2

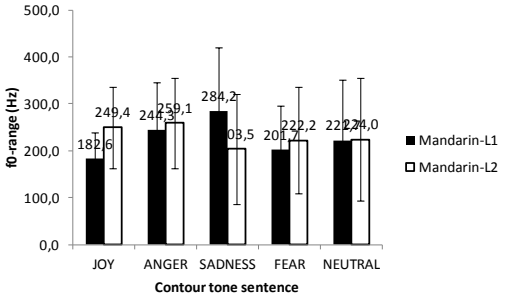

500,0
400,0
300,0
200,0
100,0
0,0
f0-range (Hz)
182,6
249,4
244,3
259,1
284,2
203,5
201,7
222,2
222,7
224,0
JOY
ANGER
SADNESS
FEAR
NEUTRAL
Contour tone sentence
Mandarin-L1
Mandarin-L2

FIGURE 3

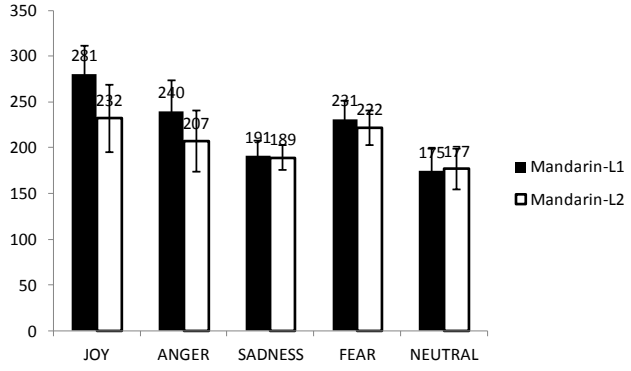

350
300
250
200
150
100
50
0
281
232
240
207
191
189
231
222
175
177
JOY
ANGER
SADNESS
FEAR
NEUTRAL
Mandarin-L1
Mandarin-L2

FIGURE 4

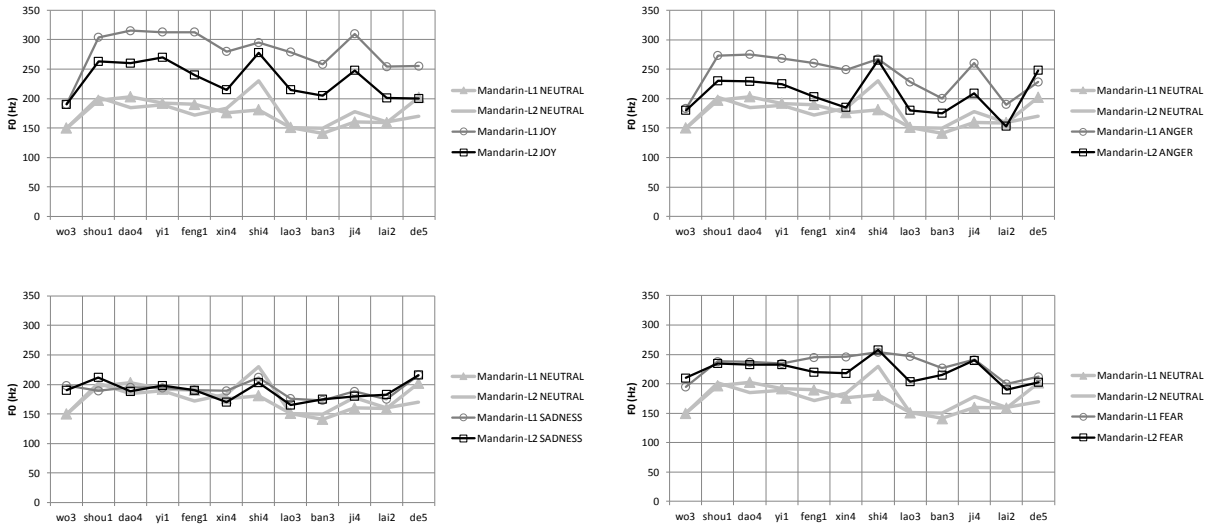

F0 (Hz)
350
300
250
200
150
100
50
0
wo3 shou1 dao4 yi1 feng1 xin4 shi4 lao3 ban3 ji4 lai2 de5
Mandarin-L1 NEUTRAL
Mandarin-L2 NEUTRAL
Mandarin-L1 JOY
Mandarin-L2 JOY
Mandarin-L1 ANGER
Mandarin-L2 ANGER
Mandarin-L1 SADNESS
Mandarin-L2 SADNESS
Mandarin-L1 FEAR
Mandarin-L2 FEAR

FIGURE 5

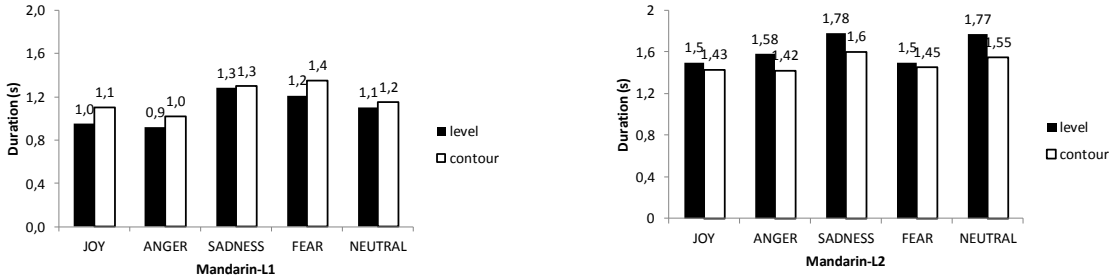

Duration (s)
2,0
1,6
1,2
0,8
0,4
0,0
1,0
1,1
0,9
1,0
1,3
1,3
1,2
1,4
1,1
1,2
JOY
ANGER
SADNESS
FEAR
NEUTRAL
Mandarin-L1
level
contour
Duration (s)
2
1,6
1,2
0,8
0,4
0
1,5
1,43
1,58
1,42
1,78
1,6
1,5
1,45
1,77
1,55
JOY
ANGER
SADNESS
FEAR
NEUTRAL
Mandarin-L2
level
contour

FIGURE 6

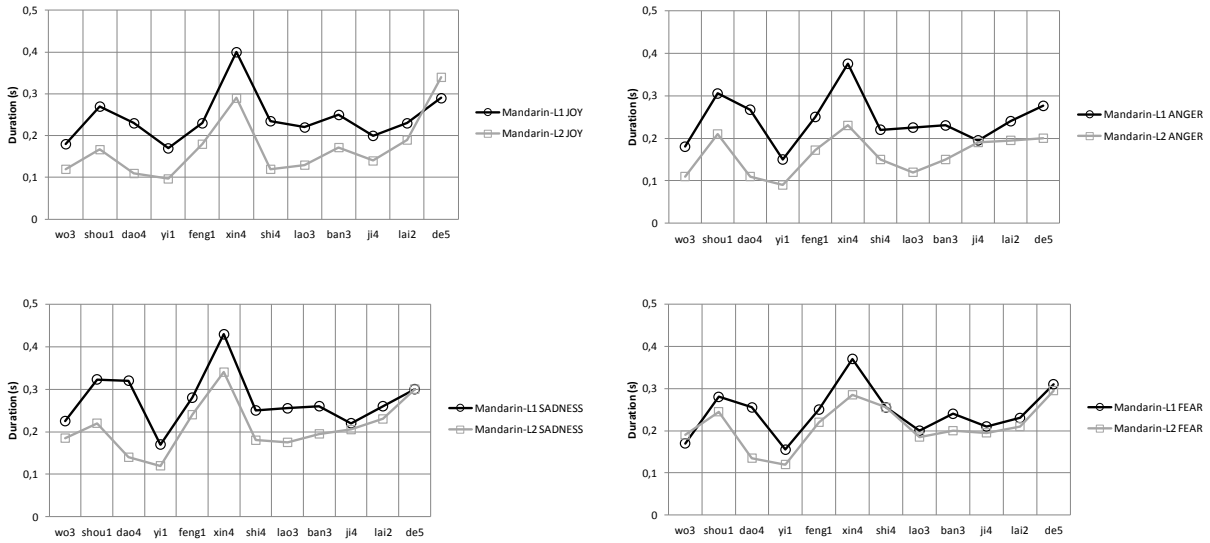
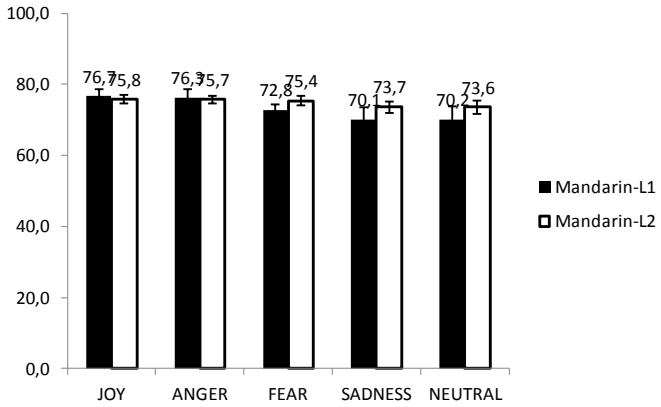

Duration (s)
0,5
0,4
0,3
0,2
0,1
0
wo3 shou1 dao4 yi1 feng1 xin4 shi4 lao3 ban3 ji4 lai2 de5
Mandarin-L1 JOY
Mandarin-L2 JOY
Mandarin-L1 ANGER
Mandarin-L2 ANGER
Mandarin-L1 SADNESS
Mandarin-L2 SADNESS
Mandarin-L1 FEAR
Mandarin-L2 FEAR

FIGURE 7


100,0
80,0
60,0
40,0
20,0
0,0
76,7
75,8
76,3
75,7
72,8
75,4
70,1
73,7
70,2
73,6
JOY
ANGER
FEAR
SADNESS
NEUTRAL
Mandarin-L1
Mandarin-L2

Tables

TABLE 1

Emotion	Matches
Joy	0 %
Anger	14 %
Sadness	28 %
Fear	72 %
Neutral	0 %