Research Article **Open Access**

Donna Erickson, Chunyue Zhu, Shigeto Kawahara*, Atsuo Suemitsu

Articulation, Acoustics and Perception of Mandarin Chinese Emotional Speech

DOI 10.1515/opli-2016-0034
Received June 22, 2016; accepted December 8, 2016

Abstract: This paper studies articulatory, acoustic and perceptual characteristics of Mandarin Chinese emotional utterances as produced by two speakers, expressing NEUTRAL, ANGRY, SAD and HAPPY emotions. Articulatory patterns were recorded using ElectroMagnetic Articulography (EMA), together with acoustic recordings. The acoustic and articulatory analysis revealed that HAPPY and ANGRY were generally higher-pitched, louder, and produced with a more open mouth than NEUTRAL or SAD. SAD is produced with low back tongue dorsum position and HAPPY, with a forward position, and for one speaker, duration was longer for ANGRY and SAD. Moreover, F1 and F2 are more dispersed (i.e., hyperarticulated) in emotional speech than NEUTRAL speech. Perception tests conducted with 18 native listeners suggest that listeners were able to perceive the expressed emotions far above chance level. The louder and higher pitched the utterance, the more emotional the speech tends to be perceived. We also explore specific articulatory and acoustic correlates of each type of emotional speech, and how they impact perception.

Keywords: Mandarin Chinese, emotion, articulation, acoustics, perception, jaw displacement, tongue dorsum, F0, F1, F2, intensity, duration

1 Introduction

1.1 General aims

Emotion is part of our human nature, and it is a subject that is interesting to study from a linguistic perspective. Our everyday experience tells us that speakers can change their speech when they are emotional, which inevitably affects its acoustic realizations. Also, listeners can often—perhaps not always—accurately perceive the emotional state of the speaker. Studying this observation in a scientific manner is one task of modern phonetics.

There is an increasing body of literature on the acoustic characteristics of emotional speech, including acted emotion by professionals and non-professionals and also spontaneous, non-scripted emotional speech (see tables comparing different studies in detail in Erickson 2005). A multitude of different languages have been examined in this research tradition, including Mandarin Chinese (e.g., ChangLiao 2004; Gu & Lee 2007; Lin & Fon 2012; Liu & Pell 2012; Yang et al. 2007; Yuan et al. 2002; Wang et al. 2005; Wen et al. 2011; Zhang et al. 2006). However, how speakers articulate emotional speech is less well-studied, with a few exceptions (Erickson, Huang et al. 2008; Li et al. 2010; Nguyen et al. 2008). The perception of emotional speech is also less well studied than its acoustics. To fill these gaps, this paper examines the articulation, acoustics, and perception of emotional speech in Mandarin Chinese, in an attempt to further our understanding of emotional speech utterances.

***Corresponding author: Shigeto Kawahara,** Keio University, Tokyo, Japan, E-mail: kawahara.research@gmail.com
Donna Erickson, Kanazawa Medical University
Chunyue Zhu, Kobe University
Atsuo Suemitsu, Sapporo University of Health Sciences

1.2 Previous studies on emotional speech

We start this paper by reviewing some previous phonetic studies on emotional speech, which served as the basis of the current study (see Erickson 2005 a more extended summary of other studies). Erickson et al. (2000) used ElectroMagnetic Articulography (EMA) to examine properties of acted emotional utterances of two American English speakers. They showed that jaw and tongue dorsum position change as a function of the particular emotion, and specifically, the emotion ANGER may involve increased jaw lowering (throughout this paper, for the sake of exposition, we use SMALL CAPITAL to express emotion types). They also found that emotional speech showed particular acoustic characteristics, realized in terms of F0 and formant structures.

An acoustic and articulatory study of spontaneous Mandarin Chinese by Erickson, Huang et al. (2008) examined a female native speaker, as she was speaking to her friend over a telephone-type connection set up in the lab, recalling past emotional events in her life, including the very SAD story of how her husband was murdered. The acoustic analysis showed that HAPPY syllables were significantly louder, higher in pitch, and shorter in duration than SAD syllables. Also for SAD, a breathy voice quality was found, as well as lowered lip and jaw, and more tongue tip protrusion compared to HAPPY.

Li et al. (2010) reported on acoustic and EMA recordings of Mandarin Chinese emotional speech (ANGRY, SAD, HAPPY and NEUTRAL, 9 vowels with 111 sentences) for a single male speaker. Among their findings was that HAPPY has the highest F0 maximum, followed by ANGRY, and then NEUTRAL/SAD. Articulation also changed as a function of the emotion, such that ANGRY has the highest tongue body position. They also reported increased lip protrusion for SAD as well as for ANGRY. In addition, the study found intonation differences in final boundary tones, with HAPPY having a high rising tone, ANGRY, a high falling tone, and SAD and NEUTRAL, low tones.

Wang et al. (2005) reported that in contrast to neutral speech, the pitch register of HAPPY speech is higher, and the slope of F0 contour of the final syllable of each prosodic word is steeper, especially for the syllable at the end of the sentence.

1.3 The current study

Compared to acoustic examinations of emotional speech, the number of articulatory studies on emotional speech is limited; more case studies are thus warranted to advance our understanding of emotional speech in terms of articulation. The paucity of articulatory data of emotional speech is most likely due to the challenges of making articulatory recordings such as EMA, which is not available in every phonetics laboratory, along with the relative newness of such techniques, combined with the difficulty of recording emotional utterances in a lab setting.

The motivation for our study is to add more data to the literature on emotional speech in Mandarin Chinese—and we hope, more generally in natural languages—by reporting on two additional Mandarin Chinese (male and female) speakers' articulation of emotion; moreover, in contrast to some of the student participants in the pilot tests, the two speakers in our study were middle-aged with considerable ability to comfortably express emotions, as will be described later.

Previous studies on Mandarin Chinese tended to treat all those utterances intended by the speaker as emotional to also convey that intended emotion to listeners. However, even with spontaneous emotional expressions (Erickson et al. 2006; Spring et al. 1992) or professionally-acted emotions (e.g., Dang et al. 2010), not all utterances are perceived as the emotion intended by the speaker. A related theoretical question along these lines is whether emotions are expressed to be communicated to others (Ohala 1994), which predicts that particular emotions should have particular acoustic targets, or are they part of a human's cathartic system, produced to bring about relief to one's intense emotional experiencing of pain, pleasure, fear, sorrow, etc. and as such "targetless" in terms of acoustics (see e.g., Mazo et al. 1994). The answer may actually depend on the speaker and the specific situation (see e.g., Erickson et al. 2012) and will be discussed later in the interpretation of our acoustic and articulatory results.

In this study, we tested what acoustic characteristics allowed listeners to perceive the "right" emotional state. Since only a small number of speakers' articulatory data has been reported on in the literature, the additional two speakers analyzed in this study will both substantiate previous findings, as well as uncover new ones, as part of the ongoing process of understanding the articulation, acoustics and perception of emotional speech in Mandarin Chinese, and again, natural languages in general.

2 Method

Articulatory and acoustic recordings were made of two Mandarin Chinese speakers producing NEUTRAL, ANGRY, SAD and HAPPY utterances. Perception tests were conducted to see how well the emotions were perceived by listeners.

2.1 Data recording

Articulatory recordings were done using the 3D EMA (Carstens AG500) at the Japan Advanced Institute of Science and Technology (JAIST, Kanazawa Japan). Acoustic signals were simultaneously recorded. Based on the findings reviewed in section 1.2., especially Erickson et al. (2000), this paper reports on the recordings of two EMA sensors: one glued to the lower medial incisors to track jaw motion and another glued to the tongue dorsum (See Figure 1).

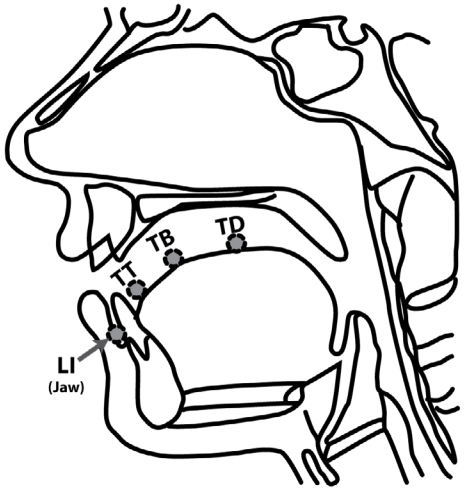

Fig. 1. Placement of sensors on Jaw (Lower Incisor) and Tongue for the EMA experiment. This paper focuses on Jaw and TD (Tongue Dorsum).

Head movement was corrected using four sensors attached to the upper incisors, bridge of the nose, left and right mastoid processes behind the ears. The articulatory and acoustic data were digitized at sampling rates of 200 Hz and 16 kHz, respectively. The occlusal plane was estimated using a bite plane with three additional sensors. In post processing, the articulatory data were rotated to the occlusal plane and corrected for head movement using the reference sensors after low-pass filtering at 20 Hz. The lowest vertical position (maximum displacement) of the jaw with respect to the bite plane was located for each target syllable of the utterance using the MATLAB-based software mview (Haskins Laboratories); this measure indicates the amount of the articulator's displacements from the occlusal plane.

Tongue dorsum measurements were made at the time of the lowest vertical jaw position. The horizontal position of the tongue dorsum relative to the upper incisor is indicated as TDx, and high values indicate forward position of tongue; the vertical position of the tongue dorsum relative to the occlusal plane is indicated as TDz, and large values indicate low tongue dorsum position.

Acoustic measurements of the vowels were obtained using Praat (Boersma & Weenink 2015). Duration, maximum F0, maximum intensity, F1 and F2 were measured, by creating a 10 ms analysis window centered in the middle of the vowel. Approximately 6 repetitions of each utterance-type were successfully recorded and analyzed.

2.2 Stimuli and analyses

For this paper, we report on 3 Mandarin Chinese phrases: 巴拿马 [ba1 na2 ma3] ('Panama'), 妈妈骂马 [ma1 ma ma4 ma3] ('Mother curses the horse'), and大把大把 [da4 ba3 da4 ba3] ('a lot of'). The vowels in the syllables are all the same, /a/, since jaw displacement varies according to vowel quality. For example, in English, high vowels and mid vowels can differ by approximately 2 mm, and so do mid vowels and low vowels (Menezes & Erickson 2013; Williams et al. 2013); similar findings have been reported for Japanese (Erickson & Kawahara 2016; Kawahara et al. 2014, submitted). Although no studies that we know of have addressed this issue in Mandarin Chinese, we assume that there is a similar relationship between vowel height and jaw displacement. This study thus controlled for the vowel quality in the stimuli. We examined the utterances with the low /a/ vowel, because jaw displacement patterns tend to be most clearly articulated with low vowels, but ongoing work is also looking at utterances with high and mid vowels.

For the current acoustic and articulatory analyses, we focus specifically on the word/phrase final syllable for the following reason. Recent research suggests that utterance-final syllables in Mandarin Chinese are *prominent*, which show increased jaw displacement and increased vowel duration (Erickson et al. 2016; Iwata et al. 2015). For the speakers in this study, increased jaw displacement for the final syllable is also seen, regardless of the emotion expressed (see Figures 2 and 3 below). Additional motivation for focusing on the last syllable comes from reports by Wang et al. (2005) that the greatest excursion of F0 occurs on the final syllable of each prosodic word and especially for the sentence final syllables.

2.3 Speakers and elicitation

Two middle-aged native speakers of Mandarin Chinese served as the speakers for the current experiment, one male (C02) and one female (C03), both born and raised in Beijing.

The utterances were spoken with different emotional expressions: Neutral, Angry, Sad and Happy. The speakers were asked to first speak the Neutral sentences, six randomized repetitions as part of a larger data set; then they were instructed to change their emotional attitude by remembering a situation where they felt very Angry, and to speak 6 repetitions of a set of words presented in a randomized order. Then, they were asked to put themselves in a "sad emotional situation" and to speak the sentences, remembering a situation where they were very Sad, and then the same for Happy. Between each set of emotions, the speakers were given time to set their moods.

C02 had no experience of acting, but had participated in amateur comedy talk shows in his college days. C02 describes himself as a "cry-baby" i.e., he cries easily when seeing a movie, reading a book, listening to music. So he was able to cry with tears even during the experiment. During the experiment, he imagined himself in the pain of losing his loved ones for Sad, or being angry with someone for something they had said to him for Anger; for Happy, he thought about his son and daughter.

C03 had no experience of acting, although her aunt was a professional actress, and she also was very convincing in her expressions of emotion. When she was doing the experiment, she imagined that she was very mad at someone who did something wrong (Anger), or felt very relaxed and excited about trip she was going to take (HAPPY), or imagined that she would fail to do anything, so what an unfortunate woman she was (Sad).

Each speaker afterwards reported that they experienced physiological changes for each of the emotion sets—especially C02, who was speaking with tears in his eyes for Sad and with his hands shaking and heart beat rising for Angry. For this speaker, it took about ten minutes to recover from each emotion, before he could do the next one, and music was played to calm him down. C03 also reported experiencing physiological changes, but not to the same degree.

2.4 Perceptual evaluations

The recorded utterances were randomized to make two perception tests, one for C02 and one for C03, for a total of 72 utterances and 73 utterances, respectively. The stimuli were presented to 18 university Mandarin Chinese students at a Japanese university in the Kansai area.[1] Each listener participated in the two perception tests. The listeners were asked to (a) rate how emotional the utterances were (1 to 5, with "5" as "extremely emotional", "3" as "emotional", and "1" as "not emotional") and (b) identify what emotion they heard—ANGRY, HAPPY, SAD, NEUTRAL, or OTHER EMOTION. Each test was preceded by a practice test of 5 utterances. The tests were presented on a computer interface through headphones in a quiet room.

3 Results

3.1 Perception results

The results of the perception tests are the basis for the subsequent acoustic and articulatory analyses. We analyzed those utterances that were (a) rated by listeners as being "emotional" given a rating of 3, 4 and 5, in answer to the first question in the perception test and (b) judged to be the intended emotion (answer to second question in the perception test). In this way, the paper focuses on *perceived* emotion, not *intended* or *produced* emotion, i.e., we examined the acoustic and articulatory characteristics of emotional speech that was perceived correctly.

The overall perception test results for speakers C02 and C03 are shown in Tables 1, 2 and 3. Table 1 shows, on average, how emotional each utterance was judged by the listeners. Table 2 shows the total number of NEUTRAL, ANGRY, HAPPY and SAD utterances, and how many of these were perceived accurately. Table 3 is a confusion matrix.

Table 1 shows that listeners generally judged emotional utterance as more emotional than the NEUTRAL utterances. All the emotional utterances, except for HAPPY of C02, were rated by listeners with "3" or above ("3" was "emotional" on the given scale). The highest rated was C02's expressions of SAD, which were rated as "very emotional". NEUTRAL was rated as not emotional for both speakers; speaker C02's HAPPY was rated as not very emotional.

Table 1. Average judged emotional degree (5=very emotional, 3 =emotional, and 1=not emotional).

Speaker	ANGRY	HAPPY	NEUTRAL	SAD
C02	3.3	2.0	1.2	4.2
C03	3.6	3.2	1.3	3.3

Table 2 shows that SAD utterances were best perceived as the speaker intended and HAPPY, the least well perceived.

Table 2. Total number of NEUTRAL, ANGRY, HAPPY and SAD utterances, and their correct and incorrect identification responses.

Intended Emotion	Correct	Incorrect	Total	% Correct
NEUTRAL	479	133	612	78%
ANGRY	491	157	648	76%
HAPPY	423	279	702	60%
SAD	629	19	648	97%

1 An anonymous reviewer suggested it would be interesting to ask the speakers themselves to do a self-evaluation; while we agree that this is an interesting possibility, we asked other listeners to make evaluations as a way to validate the emotional stimuli for the perception tests. Our analysis is thus based on *perceived* emotion, rather than *intended* emotion.

Table 3. Confusion matrix of the perception test.

		Perceived Emotion			
Intended Emotion	**C02**				
	Angry	Happy	Neutral	Sad	Other
Angry	**68%**	12%	2%	0%	18%
Happy	17%	**47%**	23%	2%	11%
Neutral	2%	1%	**81%**	10%	6%
Sad	0%	0%	0%	**99%**	0%
Intended Emotion	**C03**				
	Angry	Happy	Neutral	Sad	Other
Angry	**83%**	3%	2%	3%	8%
Happy	4%	**72%**	2%	7%	16%
Neutral	6%	1%	**75%**	10%	8%
Sad	2%	2%	1%	**95%**	1%

In Table 3, the accurately-perceived answers are shown in bold. The results show that listeners perceived all expressed emotions far better than chance level (1/5 = 20% is chance level). Listeners' correct identification rate was highest for Sad for both C02 and C03 (99% and 95%, respectively) and lowest for Happy (47% and 72%, respectively). Happy for C02 was confused with Neutral (23%) or Angry (17%) and for C03, with "other" (16%). Recall also that C02's Happy speech was not judged to be very emotional (Table 1).

3.2 Acoustic and articulatory results

3.2.1 Correlation analyses between phonetic characteristics and listener ratings of emotional degree ("emotional-ness")

In order to understand what phonetic characteristics contribute to an utterance being perceived as emotional, a Pearson-correlation analysis was run between the phonetic characteristics of each utterance and the emotional degree, i.e., how emotional the speech was rated by the listeners. The results are shown in Table 4.

Table 4. Results of the correlation analyses between articulatory/acoustic measures and the judged degrees of "emotional-ness".

	Jaw	TDx	TDz	F0 MAX	F1	Intensity MAX	Duration
C02							
r	0.823	0.038	0.055	0.596	0.607	0.870	-0.091
p-value	*p<.001*	*n.s.*	*n.s.*	*p<.001*	*p<.001*	*p<.001*	*n.s.*
C03							
r	-0.222	-0.753	-0.489	0.685	0.428	0.849	0.786
p-value	*n.s.*	*p<.001*	*p<.05*	*p<.001*	*n.s.*	*p<.001*	*p<.001*

The results show that for both speakers, the higher pitched (F0 Max) and louder (Intensity Max), the more emotional the speech was judged to be. For C03, in addition, the longer the syllable, the more emotional, whereas for C02, the more open the jaw/the higher the F1, the more emotional. For C03, the amount of tongue dorsum (TDx) fronting may function as a cue for emotion: the more fronted the tongue dorsum, the more emotional it was perceived to be.

3.2.2 Analyses of specific measures

Next, let us move on to the articulatory characteristics of different types of emotional speech. We analyze here those utterances that were perceived above chance by listeners as the emotion intended by the speaker. The overall pattern of mean jaw displacement for each of the emotional expressions is shown in Figures 2 and 3; in which the larger size bar indicates a greater jaw displacement/mouth opening, and the x-axis indicates syllables. One general pattern we observe is that jaw opens the most in the sentence-final syllables, regardless of the emotional type. Another observation is that on the utterance-final syllables, C02 has large jaw opening for ANGRY voice, whereas C03 shows largest jaw opening for HAPPY speech. C02 also has larger jaw opening for HAPPY compared to NEUTRAL and SAD.[2]

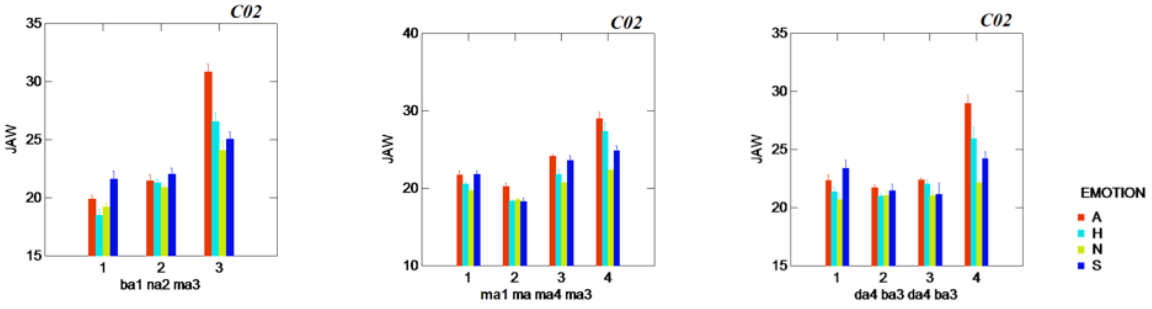

Fig. 2. Amount of jaw displacement by syllable for all utterances that were well-perceived as the emotion intended by speaker C02. (All analyses in this paper concern the emotions that were well-perceived, i.e., perceived above chance, as the emotion intended by speaker.). The y-axis indicates the amount of jaw displacement (mm), so that the larger value indicates a larger mouth opening. The x-axis indicates the syllable. Error bars show standard error. The colors indicate the different well-perceived emotion.

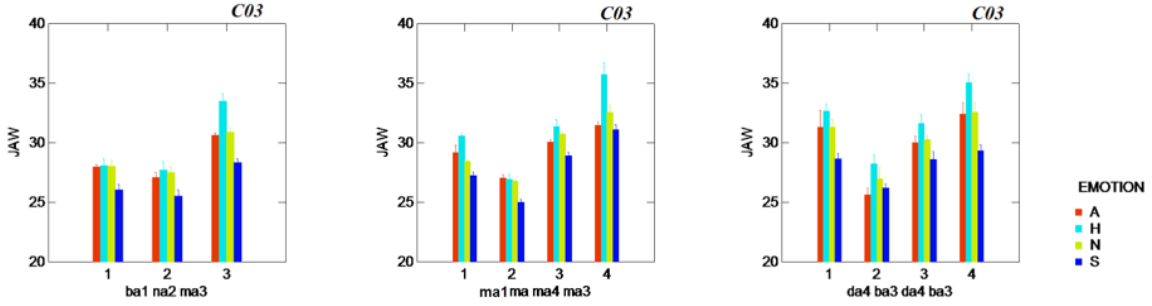

Fig. 3. Bar graph showing amount of jaw displacement (mm) by syllable in the word for all utterances that were well-perceived as the emotion intended by C03. The x-axis indicates the syllable. The colors indicate the different well-perceived emotion.

In order to better document and understand the salient articulatory and acoustic characteristics of each well-perceived (i.e., perceived above chance) emotional expression (NEUTRAL, ANGRY, HAPPY, SAD), Tables 5 and 6 show the mean values (the standard deviations) of the phonetic characteristics of the final syllable: maximum jaw displacement, tongue dorsum; also maximum F0, maximum intensity, duration and F1 and F2. The jaw and tongue dorsum (TDz) measurements are in terms of mm from the occlusal plane; the tongue dorsum-x (TDx) measurements are such that larger negative numbers indicate tongue position is further back in the mouth.

2 An anonymous reviewer asked whether it is a relative difference between the NEUTRAL "baseline" and each particular emotion or it is absolute characteristics of each emotion type that matters for the perception of each type of emotion. It is not surprising if listeners use the NEUTRAL speech as their perceptual baseline for some kind of normalization, but we note that it is not impossible to tell a speaker's emotional state, even when we do not know that person's speech beforehand. This question, therefore, requires further thinking and experimentation.

Table 5. Mean values and standard deviations of phonetic measurements for the final syllables for each emotional category perceived above chance level (20%) for C02. Number of utterances (N): for SAD=15~17; ANGRY=17~18; HAPPY = 16~17; NEUTRAL=19. The jaw and tongue dorsum (TDz) measurements are in terms of mm from the occlusal plane; the tongue dorsum-x (TDx) measurements are such that larger negative numbers indicate tongue position is further back in the mouth.

Emotion	Jaw	TDx	TDz	F0 MAX	F1	F2	Intensity MAX	Vowel Duration
SAD	24.7	-67.7	22.4	192.7	869.9	1532.0	71.5	0.3
	(1.3)	(1.4)	(2.0)	(32.6)	(160.7)	(313.8)	(5.5)	(0.1)
ANGRY	29.5	-62.4	19.8	204.6	907.7	1230.0	81.9	0.3
	(1.9)	(2.0)	(2.2)	(36.2)	(71.5)	(59.9)	(4.5)	(0.1)
HAPPY	26.7	-60.1	19.8	155.2	911.8	1354.3	73.0	0.3
	(2.0)	(1.9)	(1.6)	(15.8)	(65.6)	(42.1)	(2.2)	(0.0)
NEUTRAL	22.9	-62.4	19.4	145.6	749.2	1235.5	68.0	0.3
	(1.6)	(1.3)	(1.2)	(38.6)	(106.0)	(57.9)	(4.0)	(0.1)

Table 6. Mean values and standard deviations of phonetic measurements for the final syllables for each emotional category perceived above chance level (20%) for C03. Number of utterances (N): for SAD=17; ANGRY=13~14; HAPPY = 21, except for TDx and TDz, where N=0 due to sensor attachment problems; NEUTRAL=15~16, except for F2, where N=13.

Emotion	Jaw	TDx	TDz	F0 MAX	F1	F2	Intensity MAX	Vowel Duration
SAD	29.5	-55.4	7.8	221.0	775.6	1333.6	81.7	0.34
	(1.4)	(1.1)	(1.0)	(26.8)	(64.4)	(25.0)	(1.8)	(0.0)
ANGRY	31.3	-52.5	5.5	252.4	1049.2	1473.3	88.7	0.37
	(0.9)	(1.1)	(0.9)	(43.5)	(115.3)	(44.0)	(2.1)	(0.0)
HAPPY	34.8	--	--	275.8	1051.3	1458.7	84.3	0.29
	(2.1)			(63.9)	(128.7)	(33.4)	(2.3)	(0.0)
NEUTRAL	32.0	-47.9	4.9	162.0	815.4	1413.3	75.0	0.28
	(1.5)	(1.2)	(1.0)	(38.0)	(172.9)	(33.1)	(1.9)	(0.1)

In order to examine whether the emotions differ significantly in terms of their phonetic characteristics, ANOVAs were run with item as the random factor, EMOTION as the independent factor and each of the 8 measured phonetic values as the dependents variables. The results, shown in Tables 7 and 8, demonstrate that for both speakers, each of the phonetic values changes significantly, depending on the emotional expression; the one exception is duration for C02 for whom it does not change as a function of the emotion.

Table 7. Results of ANOVA with EMOTION as the factor and each of the 8 measured phonetic values as the dependents for C02

Measured value	F	p-value
Jaw	49.349	*p<.001*
TDx	64.005	*p<.001*
TDz	9.862	*p<.001*
F1	22.602	*p<.001*
F2	29.354	*p<.001*
F0 Max	13.260	*p<.001*
Intensity Max	35.090	*p<.001*
Vowel Duration	0.254	*p*=0.858

Table 8. Results of ANOVA for C03.

Measured value	F	p-value
Jaw	34.435	*p<.001*
TDx	347.000	*p<.001*
TDz	29.426	*p<.001*
F1	23.518	*p<.001*
F2	30.886	*p<.001*
F0 Max	18.521	*p<.001*
Intensity Max	118.107	*p<.001*
Vowel Duration	8.716	*p<.001*

A post-hoc pairwise comparison analysis was done, whose results are shown in Tables 9 and 10.[3] For both of the speakers, jaw displacement is significantly different for each emotion for all pairwise Bonferroni comparisons, except for ANGRY vs. NEUTRAL for C03. As for TDx, the tongue horizontal (forward-backward) position is significantly different for each of the emotions, except for ANGRY vs. NEUTRAL for C02. As for TDz, for both speakers, the tongue vertical (up-down) position is significantly different for SAD compared to the other emotions, except that for C03, there is no significant difference for HAPPY vs. SAD. F1 for C02 is significantly different for NEUTRAL vs. ANGRY/HAPPY but not ANGRY vs. HAPPY; for C03, F1 is significantly different for all emotions, except that HAPPY is not different from ANGRY, and that SAD is not different from NEUTRAL. F2 is not significantly different for NEUTRAL vs. ANGRY for C02, and not for ANGRY vs. HAPPY for C03. F0 Max is significantly different for all the comparisons for C02 except for HAPPY vs. NEUTRAL and ANGRY vs. SAD, and for C03, except for ANGRY vs. HAPPY. Maximum intensity is significantly different for all emotions for C03, but not for SAD vs. HAPPY for C02. As for duration, C02 shows no significant changes in duration as a function of emotion, whereas C03 does, except for HAPPY vs. NEUTRAL/ANGRY.

As an additional way of investigating the characteristics of well-perceived HAPPY, SAD, ANGRY and NEUTRAL speech, we present a set of 4 scatter plots of the measured data. Figure 4 shows the scatter plots for F1 and F2 for both speakers.

Table 9. The *p*-values for results of pairwise comparison for C02. SAD formant frequencies could not be measured, because of its heavy creakiness.

	ANGRY-NEUTRAL	SAD-NEUTRAL	HAPPY-NEUTRAL	HAPPY-SAD	ANGRY-SAD	ANGRY-HAPPY
Jaw	*p<.001*	*p*<.01	*p<.001*	*p<.05*	*p<.001*	*p*<.01
TDx	*n.s.*	*p<.001*	*p*<.01	*p<.001*	*p<.001*	*p<.05*
TDz	*n.s.*	*p<.001*	*n.s.*	*p*<.01	*p*<.01	*n.s.*
F1	*p<.001*	--	*n.s.*	--	--	*n.s.*
F2	*n.s.*	--	*n.s.*	--	--	*p<.001*
F0 MAX	*p<.001*	*p*<.01	*n.s.*	*p*<.01	*n.s.*	*p<.001*
Intensity	*p<.001*	*p<.05*	*p<.001*	*n.s.*	*p<.001*	*p<.001*
Duration	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*	*n.s.*

3 Since this analysis involves multiple comparisons, the *p*-values should be interpreted with caution, as it may inflate the Type-1 error.

Table 10. The *p*-values of pairwise comparison for C03. TDx and TDz could not be measured for HAPPY due to sensor attachment problems.

	ANGRY-NEUTRAL	SAD-NEUTRAL	HAPPY-NEUTRAL	HAPPY-SAD	ANGRY-SAD	ANGRY-HAPPY
Jaw	*n.s.*	*p<.001*	*p<.05*	*p<.001*	*p*<.01	*p<.001*
TDx	*p<.001*	*p<.001*	--	--	*p<.001*	--
TDz	*n.s.*	*n.s.*	--	--	*p<.001*	--
F1	*p<.001*	*n.s.*	*p*<.01	*p<.001*	*p<.001*	*n.s.*
F2	*p<.05*	*p<.001*	*p*<.01	*p<.001*	*p<.001*	*n.s.*
F0 MAX	*p<.001*	*p<.001*	*p<.001*	*p<.05*	*p<.05*	*n.s.*
Intensity	*p<.001*	*p<.001*	*p<.001*	*p*<.01	*p<.001*	*p<.001*
Duration	*p<.001*	*p*<.01	*n.s.*	*p<.001*	*n.s.*	*p*<.01

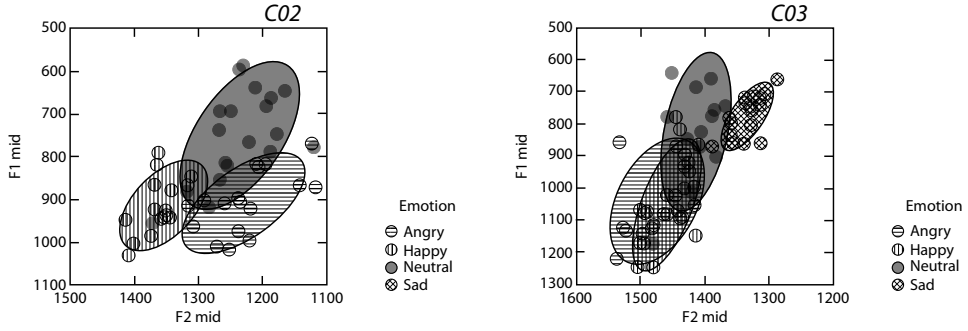

Fig. 4. F1 vs. F2. The graphs show the formant values in terms of the standard phonetic vowel chart (e.g., Ladefoged 2005), such that a high front vowel is at the top left of the graph. The left panel shows values for C02, the right panels, for C03. The formant frequencies for SAD for C02 were difficult to measure, since heavy creaky voice occurred frequently on the final syllable, and are not shown here. Ellipses show ellipse sample (ELL, 0.683).

Figure 4 shows that vowel quality changes as a function of the emotion. For both speakers, the /a/ vowel for HAPPY and ANGRY have higher F1 (i.e., the mouth is more open); in terms of F2, HAPPY /a/ for C02 has higher F2 (i.e., the tongue is more fronted) than ANGRY, but about the same degree of frontness for C03. For both speakers, the vowels in NEUTRAL speech tend to be more central, which indicates that emotional speech may involve some sort of hyperarticulation. For C03, SAD is more back and higher.

As for F2, for C02 it is highest for HAPPY and significantly higher than NEUTRAL or ANGRY. For C03, in contrast, F2 is significantly higher for ANGRY/HAPPY (no significant difference between ANGRY and HAPPY) vs. NEUTRAL vs. SAD.

F0 and intensity also change as a function of the emotion. Figure 5 shows scatter plots for C02 and C03, respectively.

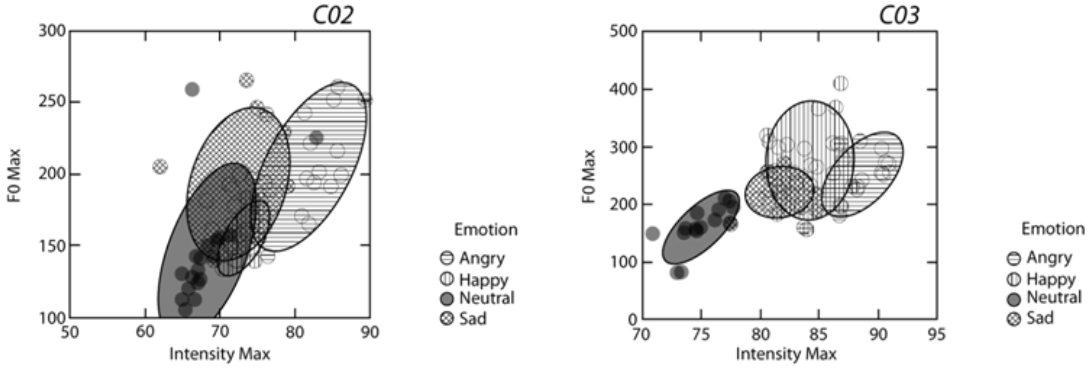

Fig. 5. F0 max in Hz (y-axis) and maximum intensity in dB (x-axis).

For F0 Max of C02, ANGRY is higher than HAPPY, and SAD is higher than HAPPY/NEUTRAL (no significant difference between HAPPY vs. NEUTRAL). For C03, F0 Max is significantly different for all the emotions, except ANGRY vs. SAD; specifically, HAPPY is higher than ANGRY, then SAD, and then NEUTRAL. As for maximum intensity, for C02, it is significantly different for all the emotions, except HAPPY vs. SAD; specifically, ANGRY is louder than HAPPY/SAD, and then NEUTRAL. For C03, maximum intensity is significantly different for all the emotions; specifically, ANGRY is louder than HAPPY, then SAD, then NEUTRAL. Overall, in terms of maximum F0 and intensity, the emotions tend to be better separated for C03 than for C02.

Figure 6 plots jaw opening against duration.

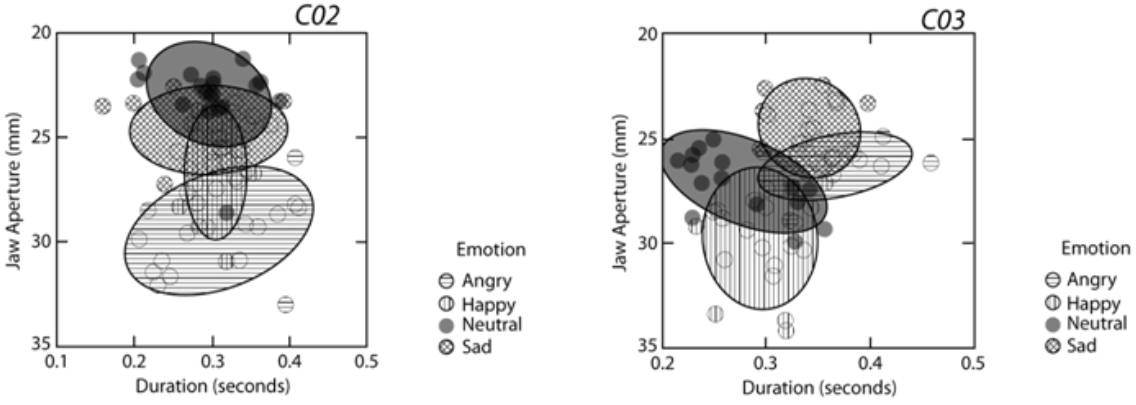

Fig. 6. Jaw displacement in mm with most open jaw at bottom of graph (y-axis) and duration of vowel in ms (x-axis).

For C02, the amount of jaw displacement is significantly different for all the emotions, with significantly larger jaw displacement for ANGRY, then HAPPY, then SAD, and then NEUTRAL; however, duration does not change significantly. For C03, the jaw is significantly different for all the emotions, except for ANGRY vs. NEUTRAL. HAPPY syllables (not ANGRY), are associated with the most open mouth, then NEUTRAL, and then ANGRY and SAD. With regard to duration, for C03, it is significantly longer for ANGRY/SAD than HAPPY/NEUTRAL. In addition, for C03, there is an interesting relationship between jaw opening and duration as a function of the emotion: for ANGRY, it is mostly vowel duration that tends to increase, while for HAPPY and SAD, it is mostly jaw opening that increases. For NEUTRAL, we see that as the jaw opening increases, duration becomes longer (a similar finding for NEUTRAL was also reported by Iwata et al. 2015 and Erickson et al. 2016).

Finally, looking at Figure 7 for C02, we see that SAD has the lowest, most back TD position; for both C02 and C03, the TDx position is significantly different for all the emotions, except for ANGRY vs. NEUTRAL for C02. As for the TDz for both speakers, it is significantly lower for SAD than the other emotions.

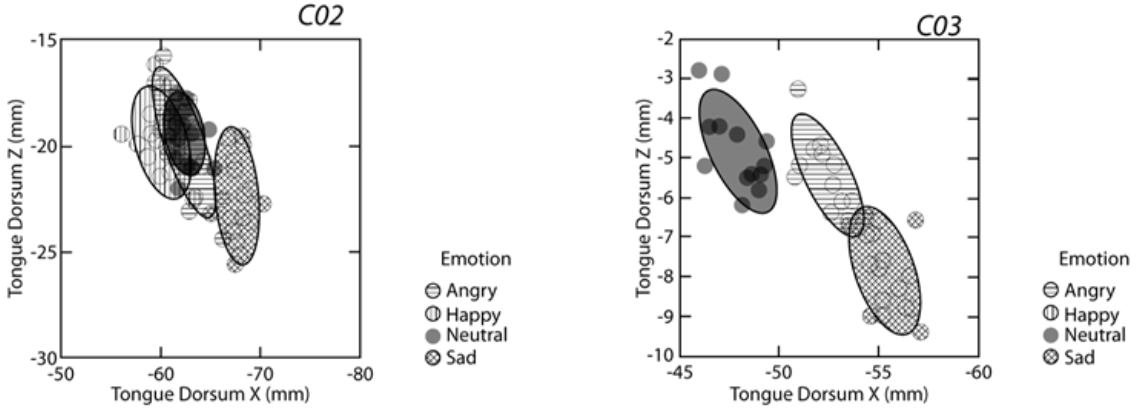

Fig. 7. TDz in mm (y-axis) and TDx in mm (x-axis). The axes are shown such that the speaker is facing left, with the more open mouth and more front tongue at the bottom left of the graph.

A summary of the acoustic and articulatory findings of emotional speech compared with neutral speech is shown in Table 11.

Table 11. Summary of acoustic and articulatory findings of well-perceived emotional speech compared with well-perceived NEUTRAL speech. Significance levels are shown with the number of asterisks: ***= *p<.001*; **= *p<.01*; * = *p<.05*. Blank cells indicate no significant differences.

C02								
	Jaw	**TDx**	**TDz**	**F0 Max**	**F1**	**F2**	**Intensity Max**	**Vowel Duration**
SAD	More open**	More back***	Lower***	Higher**	*Higher**		*Higher**	
ANGRY	More open***		*Lower**	Higher***	Higher***	*Higher**	Higher***	
HAPPY	More open***	More front**	*Lower**	*Higher**	*Higher**	*Higher**	Higher***	
C03								
	Jaw	**TDx**	**TDz**	**F0 Max**	**F1**	**F2**	**Intensity Max**	**Vowel Duration**
SAD	Less open***	More back***	*Higher**	Higher***	*Lower**	Lower***	Higher***	Longer**
ANGRY	*Less open**	More back***	*Higher**	Higher***	Higher***	*Higher**	Higher***	Longer***
HAPPY	*More open**	*NA*	*NA*	Higher***	Higher**	Higher**	Higher***	*Longer**

4 Discussion

First let us summarize the phonetic characteristics that cue an utterance being heard by listeners as emotional, as reported in Table 4. The results suggest that louder and higher pitched utterances are heard by listeners as "emotional", a finding also reported by other studies on emotional speech (see e.g., Erickson 2005). Additionally, increased F1 served as a cue to emotion, similarly reported for happy laughing speech by Szameitat et al. (2011). For one speaker (C03), increased duration is a cue to an utterance being heard as emotional, also as reported by e.g., Erickson (2005). In terms of articulation, for one speaker (C02), the more open the jaw, the more listeners heard the utterance as emotional (see also Erickson et al. 2000); also for this speaker, the lower the tongue dorsum, the greater the perceived emotion, especially for SAD.

The second topic concerns whether listeners could perceive the specific emotions intended by the speakers. This study shows that listeners perceived all emotions better than chance. SAD was best perceived, especially for C02, and HAPPY least well. That C02's SAD was well-perceived as emotional is not surprising, considering that this speaker was actually crying with tears while speaking. Our current analysis focused on a small subset of possible acoustic/articulatory characteristics of emotional expressions; oral, nasal, pharyngeal resonances, which may have arisen due to actual crying, among many other things, are not examined here, but probably contribute to the listeners' perception of SAD. That HAPPY in general is not as well perceived as the other emotional expressions is not surprising either, since similar findings have also been reported in the literature (e.g., Erickson 2005; Scherer et al. 2001). More work needs to be done to address why—however, our conjecture based on work with social affective expressions (e.g., Shochi et al. 2009) is that HAPPY is a socially positive expression and as such, is less well marked. ANGRY and SAD are more marked in that they convey information related to survival and self-protection. Another reason that C02's productions were especially difficult for listeners to recognize as HAPPY may be because the maximum F0 for his HAPPY was rather low, whereas generally HAPPY has a high F0 (e.g., Li et al. 2010). (See Tables 5 and 6 and Figure 5 which show clear contrasts between C02's not well-perceived HAPPY and C03's well-perceived HAPPY).

It is also surprising that C02's ANGRY expressions were not better perceived as ANGRY, especially since C02 was actually physically shaking with anger during the recordings. This may have to do with the fact that the nine ANGRY utterances that ended without rising intonation were those that were perceived as ANGRY, while those nine ANGRY utterances that were perceived as HAPPY ended with rising intonation. A similar finding about boundary tones and emotions was reported by Li et al. (2010), i.e., ANGRY utterances tend not to have boundary tones, while HAPPY ones had rising ones.

A third topic concerns our findings about the acoustic and articulatory characteristics on the word/phrase final syllables of well-perceived emotions (HAPPY, ANGRY, NEUTRAL, SAD). In terms of F0, HAPPY had the highest F0 Max for one of the speakers (C03), similar to that reported by Li et al. (2010) and Wang (2005); however, for the other speaker, C02, ANGRY had the highest F0 Max, and perhaps this is one of the reasons that his ANGRY was sometimes confused with HAPPY, as discussed above. However, for both speakers, ANGRY was the loudest.

In terms of jaw displacement, the largest jaw opening is for ANGRY for C02 (see Erickson et al. 2000 who reported a similar finding), while it is for HAPPY for C03. As for vowel duration, we see for C03 that the negative emotions (SAD and ANGER) have longer durations, as was also reported by Lin and Fon (2012); also by Zhang et al. (2006) for SAD. For C02, however, we do not see this. Concomitant increases in jaw opening and duration have been reported for final (NEUTRAL) syllables by Iwata et al. (2015). It is interesting that for emotional speech, the jaw and duration may, at least for some speakers, work independently. This interplay between emotion, jaw displacement and duration needs to be investigated further.

In terms of tongue dorsum positions, for both speakers SAD has a low back tongue dorsum position and HAPPY the most forward tongue dorsum for Speaker C02. It is interesting that tongue positions (Figure 7) do not seem to match the F1-F2 patterns (Figure 4). Prior work with English comparing formant frequencies with TDx-z positions of emphasized vs. non-emphasized syllables shows a strong match with tongue articulation and formants: for emphasized /a/-vowels, the jaw opens more with the tongue more back and low, resulting in higher F1 and lower F2 (Erickson 2002). However, for emotional speech, at least for these two speakers, we do not clearly see this pattern. We do not have an explanation for this discrepancy at present, other than the conjecture that different types of emotional expressions are produced with different types of tongue (as well as lip, and jaw) articulations. It is possible that these articulatory changes lead to changes in voice qualities. This is a topic of future research.

In actuality, emotions are complex and often there is no single emotion expressed in an utterance (see e.g., Dang et al. 2010). Moreover, the emotional labels used can lead to confusing results. For instance, in this paper, we have analyzed what is often referred to as "Hot Anger", in contrast to "Cold Anger". The characteristics of these two types of angers are different; for instance, the former tends to be high-pitched, loud (e.g., Scherer 1989), and also increased jaw displacement (Erickson et al. 2000), whereas the latter is low-pitched, soft (e.g., Scherer 1989) and decreased jaw displacement (Kim et al. 2014). Also, the term, "Sad"—frequently it refers to a soft, low pitched expression (e.g., Scherer 1989) which is what C03 performed; this contrasts with active grieving sadness (e.g., Erickson et al. 2006; Scherer 1989), which is what C02 performed.

One interesting remaining question is, when experiencing emotion, does the speaker have an acoustic target—i.e., wants to produce a sound such that listeners perceive the emotion? Or, does the emotional experience cause changes in the articulation that result in the acoustic changes? Differently put, is emotional speech listener-oriented or speaker-oriented? Along these lines, according to Nguyen et al. (2008), spontaneous emotional speech has different glottal characteristics than acted emotional speech. Erickson et al. (2006) found that acted emotions are better perceived than real spontaneous emotion. Erickson et al. (2009) reported that smiling speech was sometimes judged as SAD, and they discussed possible reasons for discontinuities between how a speaker produces the sound, and how it is perceived by listeners. Erickson et al. (2006) reported that there may be a difference in the phonetic characteristics of emotional speech as produced by a speaker in a highly intense emotional situation vs. those characteristics evaluated by listeners as being emotional. These findings may suggest that a highly emotional person may be "inside his/her own world" where the expression of emotion is a personal, cathartic activity. This contrasts with a more

"acted" style of emotion, in which the speaker is aware of both him/herself and the other outside person, with the emotional expression more a communicatory activity. With regard to the current two speakers, we might say that the expressions of C03 were well-expressed, well-perceived acted emotions, whereas those of C02, especially for SAD and ANGRY were spontaneous, heart-felt, experienced emotions. The type of spontaneous expressions of emotion by C02 may have no acoustic target *per se*, whereas those of C03 may indeed have acoustic targets, along the lines advocated by Ohala (1994). Much more research is needed on the challenging and complex topics of acted vs. experienced emotions, and the larger topic of the presence of articulatory/acoustic targets.

5 Summary

This study examined the acoustic and articulatory changes associated with emotions as produced by two speakers of Mandarin Chinese. The results confirm that a speaker's voice changes in terms of acoustics and articulation when he or she is expressing different emotions. Moreover, listeners are able to make judgments about how emotional the speech is, and what specific emotion the speakers is expressing, although not all intended-emotions were perceived as that particular emotion by listeners. Another new aspect of this study is that we report on articulatory and acoustic characteristics of "accurately" perceived emotional expressions. In general, emotional speech, and especially ANGRY or HAPPY, tend to be louder and higher pitched; moreover, F1 and F2 (for the vowel /a/) are more dispersed (i.e., hyperarticulated) in emotional speech than non-emotional speech, a finding also reported by e.g., Li et al. (2010).

Other new findings are that HAPPY tends to be produced with a more fronted vowel, and SAD, with a more backed vowel; jaw and tongue dorsum position also tend to be lower for emotional speech, with the tongue more forward for HAPPY speech and more back (and low) for SAD speech. Duration, at least for one speaker, however, does not change as a function of the emotion, even though jaw displacement does. This suggests that jaw displacement and duration can be independent,[4] especially during emotional speech. Numerous other interspeaker differences also are reported, and more work is needed, especially data analysis of more speakers.

This study is but a tip of the iceberg report on various observations about some of the acoustic and articulatory characteristics of Mandarin Chinese emotional speech, specifically for phrase/word final syllables on Tone 3. Future work needs to examine the effect of different tones on emotional expressions, ideally with different sets of vowels. Although this paper is based on a small set of data, it is offered as a stepping stone toward a better understanding of the acoustic and articulatory characteristics of emotional speech expressions in Mandarin Chinese.

Acknowledgements: This work was supported by the Japan Society for the Promotion of Science, Grants-in-Aid for Scientific Research (C) #25370444 and (A) #25240026. A special acknowledgement is made to Mark Tiede for help with 'mview', to Jianwu Dang for the use of the EMA Lab at JAIST, and to Jeff Moore for help improving some of the figures. We are also grateful to two anonymous reviewers, who offered critical yet constructive comments, which improved the content and exposition of the paper.

References

Boersma, Paul, David Weenink. 2015. Praat. Available Jul. 2015 from www.praat.org

ChangLiao, I. 2004. *A study of the influences of emotion on Mandarin Tones*. Master Thesis. National Taiwan University, Taiwan.

Dang, Jianwu, Aijun Li, Donna Erickson, Atsuo Suemitsu, Masato Akagi, Kyoko Sakuraba, Nobuaki Minematsu, Keikichi Hirose. 2010. Comparison of emotion perception among different cultures. *Acoustical Science and Technology* 31, (6), pp. 394-402.

4 Recent work reports either little to no correlation between jaw displacement and acoustic duration for English (Erickson and Kawahara, 2016) and a weak, negative correlation for Japanese (Kawahara et al. 2015).

Erickson, Donna, Arthur Abramson, Kikuo Maekawa, Tokihiko Kaburagi. 2000. Articulatory characteristics of emotional utterances in spoken English. *Proceedings of the International Conference of Spoken Language Processing* 2, pp. 365-368.

Erickson, Donna. 2002. Articulation of extreme formant patterns for emphasized vowels. *Phonetica* 59, pp. 134-149.

Erickson, Donna. 2005. Expressive speech: Production, perception and application to speech synthesis. *Acoustical Science and Technology* 26 (4), pp. 317-325.

Erickson, Donna, Kenji Yoshida, Caroline Menezes, Akinori Fujino, Takemi Mochida, Yoshiho Shibuya. 2006. Exploratory study of some acoustic and articulatory characteristics of *sad* speech. *Phonetica* 63, pp. 1-25.

Erickson, Donna, Chun-Fang Huang, Takaaki Shochi, Albert Rilliard, Jianwu Dang, Ray Iwata, Xugang Lu. 2008. Acoustic and arituclatory cues for Taiwanese, Japanese and American listeners' perception of Chinese *happy* and *sad* speech. *Proceedings of ASJ '2008 Fall Meeting,* 1-Q-14.

Erickson, Donna, Caroline Menezes, Ken-ichi Sakakibara. 2009. Are you laughing, smiling or crying? *APSIPA, Hokkaido, October 2009*, pp.531-537.

Erickson, Donna, Shigeto Kawahara. 2016. Articulatory correlates of metrical structure: Studying jaw displacement patterns. *Linguistic Vanguard* 2, pp. 103-118.

Erickson, Donna, Ray Iwata, Atsuo Suemitsu. 2016. Jaw displacement and phrasal stress in Mandarin Chinese, *TAL 2016*.

Gu, Wentao, Tan Lee. 2007. Quantitative analysis of F0 contours of emotional speech of Mandarin, *Proceedings of 6th ISCA Speech Synthesis Workshop*, pp. 228-233.

Iwata, Ray, Donna Erickson, Yoshiho Shibuya, Atsuo Suemitsu. 2015. Articulation of phrasal stress in Mandarin Chinese. *Acoustical Society of Japan, Fall Meeting*, 207-213.

Kawahara, Shigeto, Donna Erickson, Atsuo Suemitsu. In press. A quantitative study of jaw opening: An EMA study of Japanese vowels. *Acoustical Science and Technology* 38.

Kawahara, Shigeto, Donna Erickson, Atsuo Suemitsu. 2015. Edge prominence and declination in Japanese jaw displacement patterns: A view from the C/D model. *Journal of Phonetic Society of Japan* 19, pp. 33–43.

Kawahara, Shigeto, Hinako Masuda, Donna Erickson, Jeff Moore, Atsuo Suemitsu, Yoshiho Shibuya. 2014. Quantifying the effects of vowel quality and preceding consonants on jaw displacement: Japanese data. *Journal of the Phonetic Society of Japan* 18(2), pp. 54-62.

Kim, Jangwon, Donna Erickson, Sungbok Lee, Shrikanth Narayanan. 2014. A study of invariant properties and variation patterns in the converter/distributor model for emotional speech. *Interspeech 2014*, pp. 413-417.

Ladefoged, Peter. 2005. *A Course in Phonetics*, 5th Ed. Belmont, U.S.A.: Thomson/Wadsworth Publishers.

Li, Aijun, Qiang Fang, Fang Hu, Lu Zheng, Hong Wang, Jianwu Dang. 2010. Acoustic and articulatory analysis on Mandarin Chinese vowels in emotional speech. *Proceedings of Institute of Electrical and Electronics Engineers,* pp. 38-43.

Lin, Hsin-Yi, Janice Fon. 2012. Prosodic and acoustic features of emotional speech in Taiwan Mandarin. *Proceedings of 6th International Conference on Speech Prosody*, pp. 450-453.

Liu, Pan, Marc D. Pell. 2012. Recognizing vocal emotions in Mandarin Chinese: a validated database of Chinese vocal emotional stimuli. *Behav. Res. Methods*, 4, pp. 1042-51. doi: 10.3758/s13428-012-0203-3.

Mazo, Margarita, Donna Erickson, Todd Harvey. 1995. Emotion and expression: Temporal data on voice quality in Russian lament. *The Eighth Vocal Fold Physiology Conference, Kurume, Japan*, pp. 173-187.

Menezes, Caroline, Donna Erickson. 2013. Intrinsic variations in jaw deviation in English vowels. *Proceedings of International Congress of Acoustics*. *Proceedings of Meetings on Acoustics* 19, 060253.

Nguyen, Binh Phu, Isao Tokuda, Donna Erickson. 2008. Analysis of the roles of glottal features for emotion classification in spontaneous and acted emotional signals. *Proc. ASJ '2008 Fall Meeting*, 1-Q-20.

Ohala, John J. 1994. The frequency codes underlies the sound symbolic use of voice pitch. In: Leanne Hinton, Johanna Nichols, John J. Ohala (eds.), *Sound symbolism*. Cambridge: Cambridge University Press, pp. 325-347.

Scherer, Klaus. 1989. Vocal correlates of emotional arousal and affective disturbance. In: Wanger, Hugh L., Antony S. R. Manstead (eds.), *Handbook of Social Psychophysiology*, Hoboken, New Jersey: John Wiley & Sons, Ltd., pp. 165-197.

Scherer, Klaus, Rainer Banse, Harald G. Wallbott. 2001. Emotion inferences from vocal expression correlate across languages and cultures*, J. Cross-Cultural Psychol*. 32, pp. 76-92.

Shochi, Takaaki, Albert Rilliard, Véronique Aubergé, Donna Erickson. 2009. Intercultural perception of English, French and Japanese Social Affective Prosody. In: Hancil, Sylvie (ed.), *The Role of prosody in Affective Speech*, *Linguistic Insights Series. Studies in Languages and Communication* (97), New York, Bern, Berlin, Bruxelles, Frankfurt am Main, Oxford, Wien: Peter Lang Publishing Group, pp. 31-60.

Spring, Cari, Donna Erickson, Thomas Call. 1992. Emotional modalities and intonation in spoken language. *Proceedings of the International Conference on Spoken Language Processing*, pp. 679-682.

Szameitat, Diana P., Chris J. Darwin, André J. Szameitat, Dirk Wildgruber, Kai Alter. 2011. Formant characteristics of human laughter. *Journal of Voice* 25 (1), pp. 32-37.

Wang, Haibo, Aijun Li, Qiang Fang. 2005. F0 contour of prosodic word in happy speech of Mandarin, *Affective Computing and Intelligent Interaction, Lecture Notes in Computer Science* 3784, pp. 433-440.

Wen, Miaomiao, Miaomiao Wang, Keikichi Hirose, Nobuaki Minematsu. 2011. Prosody conversion for emotional Mandarin speech synthesis using the tone nucleus model. *IPSJ SIG Technical Report* vol. 2011-SLP-87 No. 2 2011/7/21.

Williams, J.C., Donna Erickson, Yousuke Ozaki, Atsuo Suemitsu, Nobuaki Minematsu, Osamu Fujimura. 2013. Neutralizing differences in jaw displacement for English vowels, *Proceedings of International Congress of Acoustics*. POMA 19, 060268.

Yang, Yingchun, Zhaohui Wu, Tan Wu, Dongdong Li. 2007. *Mandarin Affective Speech*. LDC2007S09, ISBN: 1-58563-442-5.

Yuan, Jiahong, Liqin Shen, Fangxin Chen. 2002. The acoustic analysis of anger, fear, joy and sadness in Chinese, *Proc. 7th International Conference on Spoken Language Processing,* pp. 2025-2028.

Zhang, Sheng, P. C. Ching, Fanrang Kong. 2006. Acoustic analysis of emotional speech in Mandarin Chinese. *International Symposium on Chinese Spoken Language Processing (ISCSLP 2006).*