Lexicalized emotion?
Tonal patterns of emotion words in Mandarin Chinese

Yao Yao[1], Jingxia Lin[2] and Chu-Ren Huang[1]
[1]*The Hong Kong Polytechnic University* [2]*Nanyang Technological University*

In this study, we compared the tonal patterns of emotion words in Mandarin Chinese to the prosodic patterns of emotional speech. We used statistical methods to model the variation in tonal height and slope of tonal contour of Mandarin emotion words. Our results showed that there was indeed some similarity between the tones of Mandarin emotion words and emotional intonation, suggesting that emotional cues may be lexicalized in lexical tones.

0. Introduction

Few studies have examined the relationship between pronunciation and word meaning, mainly because the form-meaning relationship is widely believed to be arbitrary and language-specific to a large extent. For example, there is no obvious reason why a cat has to be called *cat* [ˈkʰæt] in English but 猫 *māo* [mau55] in Mandarin Chinese. Exceptions certainly exist if we take a closer look at this issue, as several form-meaning corresponding patterns do seem to be either universal or predictable. The best-known counterexamples are found in onomatopoeic words. While the forms meaning "cat" might be arbitrary and different across languages, the words that describe the sound cats make are highly similar in most languages (e.g. *meow* [miˈau] in English, 喵 *miāo* [miau55] in Mandarin), because cats worldwide sing alike. Consider another counterexample: if an English speaker is asked to say the words *up* and *down* without context, it is almost certain that the word *up* will be produced with an overall higher pitch than the word *down* – a fact that is most readily explained by attributing to the meanings of *up* and *down*. Along the same line, according to personal experience and anecdotal evidence, swear words in English are almost exclusively produced with falling intonations, which might be linked with the definitive tone of the embedding message that is in turn attributable to the emotional state of the speaker that would have promoted the use of swear words.

It is the latter type of exceptions discussed above that makes us wonder whether there exists some general mechanism for language to encode word meaning in intonation. Is it possible to find a larger category of words for which the prosody in production systematically reflects some semantic features? To explore this idea, we turned our focus from English to a tonal language, Mandarin Chinese, because tones are lexicalized

prosodic patterns that are undoubtedly an important part of the lexical form. In specific, we chose to study Mandarin words expressing emotions (e.g. 高兴 gāo xìng [kau55 ɕiŋ51] "adj. happy"; 难过 nán guò [næn35 kwɔ51] "adj. sad"), because the lexical meaning of these words, i.e. human emotions, seemed to have a natural linkage with intonation, as further illustrated in Section 1 in the following.

Thus, the overarching goal of this research is to broaden our knowledge about the form-meaning relationship in words. The more specific research question is whether emotional meaning can be lexicalized in tones. To address these questions, we conducted a statistical analysis of the tonal patterns of Mandarin emotion words. Detail of the study is presented below.

1. Background literature

1.1 Emotion studies

Emotion is a fundamental aspect of humanity shared by people living in different times, locations and cultures. It can be defined as a "mental and physiological state associated with feelings, thoughts and behaviour" (Lee, 2010:1; see also Descarte, 1649; James, 1884; Harkins & Wierbicka, 2001; Plutchik, 1962; etc.). Studies on human emotion generally agree on the existence of some basic categories of emotion, while other, more complicated emotions can be construed as the combination of two or more basic emotions (see Lee, 2010 among others). The least contended basic emotion categories include *joy*, *anger*, *sadness*, *fear* and *surprise*.

Human emotion has been the topic of various lines of scholarly inquiry, ranging from linguistic analysis of emotional expressions (Kövecses, 2000; Solomon, 1976; Wierzbicka, 1972; etc.) to neuropsychological studies of the cognition of emotion (Desmet, 2002; Plutchik, 1962, 1980, 1994; Russell, 1980; Turner, 1996, 2000; etc.) to automatic identification of emotion in computational applications (Ahmad, 2008; Chuang & Wu, 2002; Ortony *et al*., 1988; Strapparava & Mihalcea, 2008; Subasic & Huettner, 2001; etc.). Needless to say, the first line of research bears the highest relevance to the current study, therefore we devote the next section to a more detailed literature review along this line, with a focus on the acoustic-phonetic properties of emotional speech.

1.2 Intonation and emotional speech

Emotional speech refers to the speech produced with emotion. It is a well-known fact that utterances produced under different emotional states may sound different, even with the same linguistic content. In fact, studies have shown that normal-developing human infants can sense the emotion in speech (e.g. anger or joy) from as early as 3 months of age – well before they start to acquire words (Blasi *et al*., 2011). Despite the prevalence of emotional speech, the exact acoustic cues of emotional speech are still under debate. A number of acoustic correlates – at both segmental and suprasegmental levels – have been examined in different languages, but the findings are rather mixed.

If we focus on intonation, the existing literature presents two differing patterns of

prosodic variation in emotional speech. On one hand, it has been proposed that pitch height is associated with emotion category (see Juslin & Laukka, 2003 and Scherer, 2003 for recent reviews; see also Li *et al*., 2011), as utterances produced with *joy*, *anger*, and *fear* tend to have higher pitch than those produced with *sadness* and *depression*. On the other hand, it has also been contended that overall pitch height reflects more of emotion intensity (i.e. level of arousal) rather than emotion category (Bänziger & Scherer, 2005; Pakosz, 1983). Specifically, Bänziger and Scherer (2005) found that utterances produced with high-intensity emotions tended to have higher pitch levels than utterances produced with emotions of the same category but with lower intensity. Furthermore, Bänziger and Scherer's results suggested that different emotion categories were distinguished by the shape of pitch contours, as expressions produced with *joy* and *anger* showed greater and faster F_0 changes than those produced with *sadness* and *fear*.

It is not the purpose of the current study to discern which prosodic pattern is more reliably observed in emotional speech; instead, we take these findings as a starting point for predicting what might be observed in the tones of emotion words if there is indeed a parallel between emotional prosody and tones of emotion words.

1.3 Current study

The general hypothesis of the current study is that the lexical tones of Mandarin emotion words would show similar patterns of variation with regard to the emotional meaning (in terms of category, intensity, etc.) as the patterns of prosodic variation observed in emotional speech. Specifically, we formed three predictions of emotional tonal variation based on the findings of emotional prosody:

> **Prediction 1**: The overall tone height of words expressing *joy*, *anger* and *fear* should be higher than the tone height of words expressing *sadness* and *depression.*
> **Prediction 2**: The overall tone height of words expressing high-intensity emotions should be higher than the tone height of words expressing low-intensity emotions
> **Prediction 3:** The tonal contour of words expressing *joy* and *anger* should have greater slopes (i.e. faster pitch change) than the tonal contour of words expressing *sadness* and *fear*.

It should be noted that given the mixed nature of the findings of emotional prosody, it is not realistic to expect that all the predicted patterns will be observed in the tones of emotion words. In the following sections, we will present the study design and results in more detail.

2. Data and methods

Our data consist of 123 disyllabic Mandarin Chinese words extracted from the Chinese emotion word list in Lee (2010), which also contained labels of emotion category and emotional intensity for each word. The 123 words represented four basic emotion categories (*joy, anger, sadness, fear*) and three intensity levels (high, mid, low). Table 1 below shows the count of words by category and intensity.

	Joy	**Anger**	**Sadness**	**Fear**
High	N = 4 (e.g. 亢奋 kàngfèn [kʰɑŋ51 fən51]	N = 22 (e.g. 痛恨 tònghèn [tʰoŋ51 hən51])	N = 14 (e.g. 悲恸 bēitòng [pei55 tʰoŋ51])	N = 3 (e.g. 恐慌 kǒnghuāng [kʰoŋ21 hwɑŋ55])
Mid	N = 27 (e.g. 愉悦 yúyùe [y35 ɥɛ51])	N = 8 (e.g. 窝火 wōhuǒ [wɔ55 hwɔ214])	N = 18 (e.g. 沮丧 jǔsàng [tɕy21 sɑŋ51])	N = 7 (e.g. 发怵 fāchù [fa55 ʈʂʰu51])
Low	N = 4 (e.g. 怡和 yíhé [i35 hɤ35])	N = 7 (e.g. 烦心 fánxīn [fæn35 ɕin55])	N= 2 (e.g. 灰心 huīxīn [hwei55 ɕin55])	N = 7 (e.g. 害羞 hàixiū [hai51 ɕiou55])

Table 1 Word counts and examples of Mandarin Chinese emotion words by emotion category and intensity level (High, Mid, Low).

Two tone-related measures were calculated for each disyllabic emotion word: overall tone height (*H*) and slope of tonal contour (*S*). Tonal height is measured by the mean lexical tone level of the two syllables on the 5-point scale (see Table 2). For example, *H* (痛恨 tònghèn [tʰoŋ51 hən51]) = ((5+1)/2+(5+1)/2)/2 = 3. Following the rules of Tone 3 sandhi in Mandarin, if the first syllable of a word bears a Tone 3, it will be converted to either a Tone 2 (when followed by another Tone 3) or a half Tone 3 (when followed by any tone other than Tone 3) before entering the calculation of *H*.

	Tone 1	**Tone 2**	**Tone 3**	**Tone 4**	**Half Tone 3**
Tonal value	55	35	214	51	21
H	(5+5)/2 = 5	(3+5)/3 = 4	((2+1)/2 + (1+4)/2)/2 = 2	(5+1)/2 = 3	(2+1)/2 = 1.5
S	|5-5| = 0	|3-5| = 2	(|2-1| + |1-4|)*0.5 = 2	|5-1| = 4	|2-1| = 1

Table 2. Pitch value, H and S of Mandarin tones on the 5-point system.

Slope of tonal contour is an indicator of the speed of change in tonal height, which in theory depends on both the amount of change in tone height and duration. It has been shown that Mandarin tones correspond to different inherent syllable lengths (Moore

& Jongman, 1997; Xu, 1997), with the largest difference existing between Tone 3 and Tone 4. As both Moore and Jongman (1997) and Xu (1997) showed, syllables of Tone 3 could be almost twice as long as syllables of Tone 4. In order to reflect the durational variation while also keeping the measure simple, the *S* measure of a syllable is approximated with the total amount of pitch change (on the 5-point scale) in the syllable, with a coefficient of 0.5 applied to syllables of Tone 3 only, in recognition of the longer duration (therefore slower pitch change) associated with Tone 3 (see Table 2). The *S* measure of a disyllabic word is the mean *S* of the two syllables. Thus, *S* (痛恨 tònghèn [tʰoŋ51 hən51])= (4+4)/2 = 4, and *S* (窝火 wōhuǒ [wɔ55 hwɔ214]) = (0+2)/2 = 1. Similar to the *H* measure, the *S* measure is calculated based on surface tone values after Tone 3 sandhi.

Each tonal measure was subjected to a two-way analysis of variance (ANOVA) with four levels of emotion category (*fear, anger, fear, sadness*) and two levels of emotion intensity (*high, non-high*). Due to the scarcity of low-intensity items on the word list, mid and low intensity levels were merged into a *non-high* category.

3. Results

3.1 Model on average tone height (H)

The two-way ANOVA on *H* yielded a significant main effect of emotion category ($F(3, 115) = 7.11$, $p < .001$) and a significant interaction between emotion category and emotion intensity ($F(3, 115) = 5.08$, $p = .002$), but no main effect of emotion intensity ($p > .1$). The specific interaction pattern is as follows: when intensity level was high, words expressing *sadness* had significantly higher *H* than words expressing *joy, anger, and fear*; when intensity level was non-high, *H* seemed to be similar across emotion categories. Figure 1 below plots the grouped average *H* by emotion category and intensity. As shown in the plot, cross-emotion-category differences in *H* only exist in the high-intensity condition, but not in the non-high condition. Furthermore, the pattern of *H* variation in the high-intensity condition runs against Prediction 1, as words of *sadness* had higher – rather than lower – pitch height than words of *joy* and *anger*. We will return to this discrepancy in the discussion in Section 4.

3.1 Model on slope of tonal contour (S)

Analysis of variance in *S* also revealed a significant main effect of emotion category ($F(3, 115) = 5.49$, $p = .001$) and a marginally significant interaction between emotion category and emotion intensity ($F(3, 115) = 2.91$, $p = .037$), but no main effect of emotion intensity ($p > .1$). Specifically, when the intensity level was high, words of *joy* and *anger* tended to have greater *S* (e.g. faster pitch change) than words of *sadness* and fear, but the cross-emotion-category differences vanished in the non-high intensity condition. Figure 2 below shows the average *S* of the word grouped by emotion category and intensity. Importantly, the pattern of *S* variation in the high-intensity condition is consistent with Prediction 3.

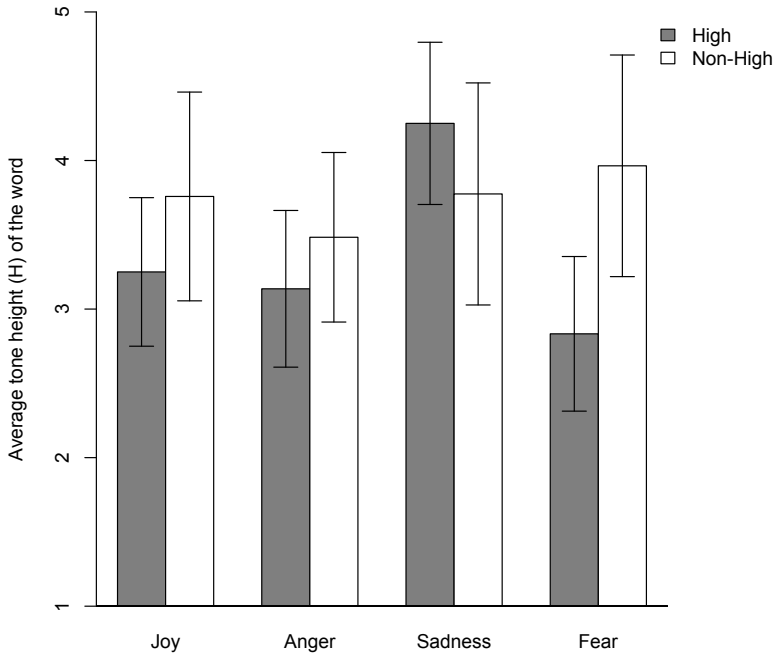

Figure 1. Average tone height (H) of the word by emotion category and emotion intensity (High and Non-High).

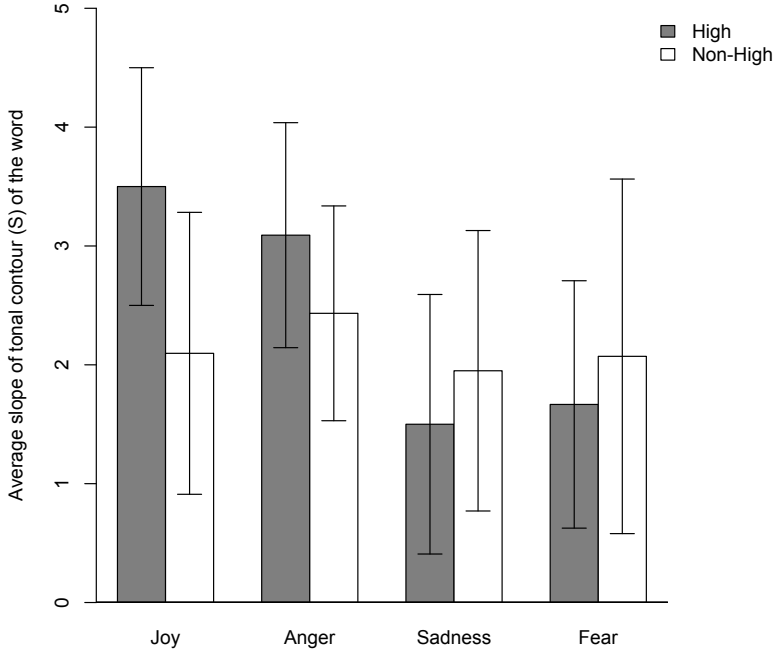

Figure 2. Average slope of tonal contour (S) of the word by emotion category and emotion intensity (High and Non-High).

4. Discussion

4.1 Major findings

In this study, we conducted a statistical analysis of the tonal patterns of Mandarin Chinese emotion words. Our results revealed cross-emotion-category differences in both overall tonal height and slope of tonal contour, but the differences were only reliably observed among high-intensity emotion words. Specifically, when the intensity level is high, we observed that (1) words expressing *sadness* tended to have higher tonal level than those expressing *joy, anger* and *fear*, and (2) words expressing *joy* and *anger* had steeper tonal contour than those expressing *sadness* and *fear*.

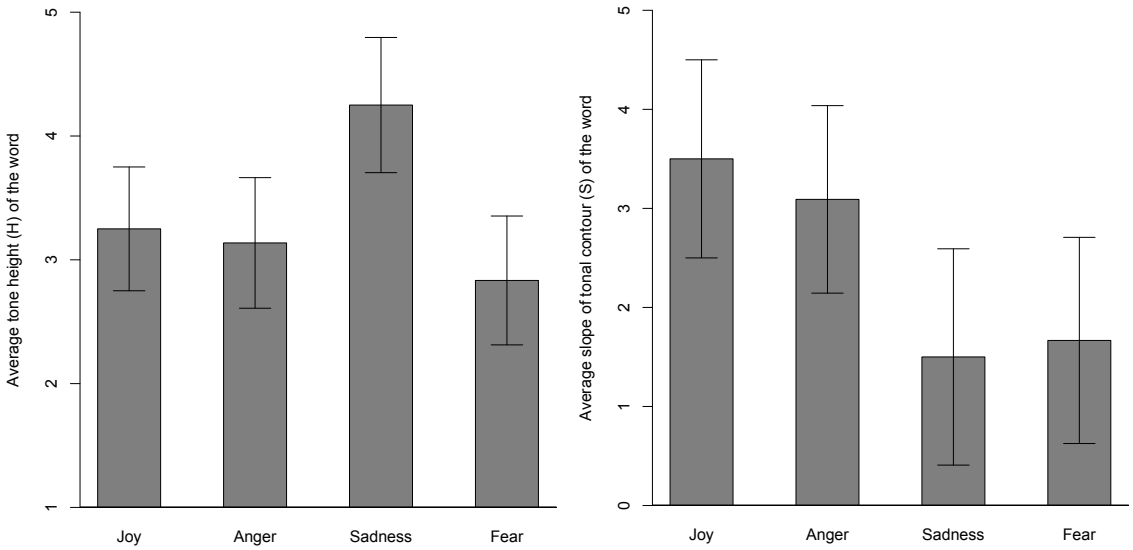

Figure 3. Average tone height (H) (left pane) and slope of tonal contour (S) (right pane) of high-intensity words grouped by emotion category.

If we focus on high-intensity emotion words, the observed tonal patterns (see Figure 3 above) suggest that words of *joy* and *anger* tend to have medium tonal level but fast pitch change, both of which are features of high falling Tone 4 (51); words of *sadness* tend to have high tonal level but slow pitch change, which indicates the wide use of high level Tone 1 (55); words of *fear* have medium tonal level and also slow pitch change, which suggests the use of Tone 2 (35) and Tone 3 (214). To verify these tendencies, we did a post-hoc analysis of the percentage of words containing Tone 1 and Tone 4 syllables in high-intensity emotion words of each category. The results of the post-hoc analysis showed that 100% (4 out of 4) of high-intensity *joy* words and 91% (20 out of 22) *anger* words contain at least one Tone 4 syllable, while as many as half of these words have Tone 4 in both syllables. By contrast, only 64% (9 out of 14) words of *sadness* and 33% (1 out of 3) words of *fear* contain at least one Tone 4 syllable and none of them have two Tone 4 syllables. On the other hand, words of *sadness* have a much

higher percentage of words with Tone 1 syllables (85%) than words of other emotion categories (18% – 33%).

	Joy	**Anger**	**Sadness**	**Fear**
N = total # of words	4	22	14	3
# of words containing at least one Tone 4 syllable (% out of N)	4 (91%)	20 (91%)	9 (64%)	1 (33%)
# of words containing two Tone 4 syllable	3	10	0	0
# of words containing at least one Tone 1 syllable (% out of N)	1 (25%)	4 (18%)	12 (85%)	1 (33%)
# of words containing two Tone 1 syllable	0	0	4	0

Table 3. Number of high-intensity emotion words containing Tone 1 and Tone 4 syllables in each emotion category

The post-hoc analysis above suggested that the tonal patterns observed in the statistical models were majorly implemented by the uneven use of Tone 1 and Tone 4 in emotion words of different categories.

4.2 Revisiting the predictions

Our original hypothesis is that the tonal patterns of Mandarin emotion words would bear some similarity with the prosodic patterns of emotional speech. Three predictions were formulated based on the general hypothesis (repeated below):

Prediction 1: The overall tone height of words expressing *joy*, *anger* and *fear* should be higher than the tone height of words expressing *sadness* and *depression.*

Prediction 2: The overall tone height of words expressing high-intensity emotions should be higher than the tone height of words expressing low-intensity emotions

Prediction 3: The tonal contour of words expressing *joy* and *anger* should have greater slopes (i.e. faster pitch change) than the tonal contour of words expressing *sadness* and *fear*.

Comparing with the original predictions, the current results provided (partial) support for Prediction 3, and the post-hoc analysis showed that the pattern was mostly implemented by the prevailing use of high falling Tone 4 in words of *joy* and *anger*. However, contra to Prediction 1, we found that words of *sadness* had higher tonal height than words of *joy*, *anger* and *fear*, due to the frequent use of Tone 1 in *sadness* words. Is this pattern attributable to the cultural-specific expression of *sadness* in the Chinese culture, or is it a result of some historical tonal change? These speculations are yet to be

verified in subsequent research. Furthermore, the current results provided no support for Prediction 2, as we found no general effects of emotion intensity on tonal height. The lack of such results, however, is not surprising at all. As stated above, the predictions are generated based on mixed findings of emotional prosody. If tone height variation in Mandarin emotion words already encodes emotion category information (i.e. *sadness* vs. *joy*, *anger*, *fear*), it seems less likely for tonal height to also encode emotion intensity.

To sum up, the current results did present evidence for the original hypothesis. Importantly, although only one of the predictions was supported, it should by no means be taken as coincidental. The fact that the observed pattern regarding slope of tonal contour and emotion category conforms to the predicted pattern is in itself significant, given the number of possible patterns that could show in a completely random space.

4.3 Alternative analyses

In order to test the robustness of the major findings presented above, we also built a series of testing models based on alternative ways of modeling the variation phenomena. The first set of alternative models represented a character-based analysis – as distinguished from the word-based analysis in the main models, recognizing the fact that some characters such as 愤 fèn [fən51] appeared in more than one emotion word (e.g. 愤恨, 愤怒, 愤慨, etc.) and were therefore counted multiple times in the word-based analysis. Character-based models were built with the set of unique characters contained in emotion words and were used to adjudicate whether the results of the main models were driven by a small set of multiply-represented characters. The second set of alternative models were built with only emotion words with high usage frequency, in order to verify that the patterns observed in the main models were not driven by low-frequency words that might not even be familiar for average language users. The results of both sets of alternative models showed similar trends regarding the variation of tonal height and slope of tonal contour as the main models, sometimes even to greater degrees (as in the models with high-frequency words only).

5. Conclusion

To conclude, the current study showed that the tonal patterns of Mandarin Chinese emotion words did resemble the prosodic patterns of emotional speech in some aspect. These findings suggest that emotion-related cues may indeed be lexicalized in the lexical forms of emotion words. In our future work, we will extend the research to other Chinese dialects, in order to verify the spread of such lexicalization process and to further investigate the interaction between the lexical tonal system and the preservation of emotional prosody.

REFERENCES

AHMAD, KHURSHID. (Ed). 2008. *Proceedings of the Workshop on Sentiment Analysis: Emotion, Metaphor, Ontology and Terminology* (EMOT-08), In Association with LREC-08, Marrakech, Morocco, May 27.

BÄNZIGER, TANYA and SCHERER, KLAUS. R. 2005. The role of intonation in emotional expressions. *Speech Communication*, 46(3-4): 252-267.

BLASI, ANNA; MERCURE, EVELYN; LLOYD-FOX, SARAH; THOMSON, ALEX; BRAMMER, MICHAEL; SAUTER, DISA; *et al.* 2011. Early specialization for voice and emotion processing in the infant brain. *Current Biology*, 21(14): 1220-1224.

CHUANG, ZE-JING and WU, CHUANG-HEIEN. 2002. Emotion Recognition from Textual Input Using an Emotional Semantic Network. In *Proceedings of the International Symposium on Chinese Spoken Language Processing* (Denver, CO).

DESCARTE, RENÉ. 1649. The Passions of the Soul. *The Philosophical Writings of Descartes Vol.1*, edited by John Cottingham, Robert Stoothoff, Dugald Murdoch, 325-404. Cambridge: Cambridge University Press.

DESMET, PIETER. 2002. *Designing Emotion*. Doctoral Thesis, Delft University of Technology.

HARKINS, JEAN and WIERZBICKA, ANNA. (Eds). 2001. *Emotions in Crosslinguistic Perspective*. Berlin/New York: Mouton de Gruyter.

JAMES, WILLIAM. 1884. What is an Emotion? *Mind*, 9(34): 188–205.

JUSLIN, PATRIK N. and LAUKKA, PETRI. 2003. Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129(5): 770-814.

KÖVECSES, ZOLÁTN. 2000. *Metaphor and Emotion: Language, Culture and Body in Human Feeling*. Cambridge: Cambridge University Press.

LEE, SOPHIA Y-M. 2010. *A Linguistic Approach to Emotion Detection and Classification*. Doctoral Thesis, The Hong Kong Polytechnic University.

LI, AIJUN; FANG, QIANG; and DANG, JIANWU. 2011. Emotional intonation in a tone language: Experimental evidence from Chinese. *Proceedings of ICPhS XVII*, 17-21. Hong Kong.

MOORE, CORINNE. B. and JONGMAN, ALLARD. 1997. Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America*, 102: 1864-77.

ORTONY, ANDREW; CLORE, GERALD; and COLLINS, ALLAN. 1988. *The Cognitive Structure of Emotions*. New York: Cambridge University Press.

PAKOSZ, MACIEJ, 1983. Attitudinal judgments in intonation: Some evidence for a theory. *Journal of Psycholinguistic Research*, 12: 311-326.

PLUTCHIK, ROBERT. 1962. *The Emotions: Fact, Theories and a New Model*. New York: Random House.

PLUTCHIK, ROBERT. 1980. *Emotions: A Psychoevolutionary Synthesis*. New York: Harper & Row.

PLUTCHIK, ROBERT. 1994. *The Psychology and Biology of Emotion*. New York: Harper-Collins.

RUSSELL, JAMES A. 1980. A Circumplex Model of Affect. *Journal of Personality and Social Psychology*, 39: 1161-78.

SCHERER, KLAUS R. 2003. Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40: 227-256.

SOLOMON, ROBERT C. 1976. *The Passions: The Myth and Nature of Human Emotion*. Garden City: Doubleday.

STRAPPARAVA, CARLO and MIHALCEA, RADA. 2008. Learning to Identify Emotions in Text. In *Proceedings of the ACM Conference on Applied Computing ACM-SAC 2008*: 1556-60.

SUBASIC, PERO and HUETTNER, ALISON. 2001. Affect Analysis of Text Using Fussy Semantic Typing. *IEEE Trans*. Fuzzy System, 9: 483-96.

TURNER, JONATHAN H. The Evolution of Emotions in Humans: A Darwinian–Durkheimian Analysis. *Journal for the Theory of Social Behaviour*, 26:1-34.

TURNER, JONATHAN H. 2000. *On the Origins of Human Emotions: A Sociological Inquiry into the Evolution of Human Affect*. Stanford: Stanford University Press.

WIERZBICKA, ANNA. 1992. Defining Emotion Concepts. *Cognitive Science*, 16: 539-581.

XU, YI. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics*, 25: 61-83.