

ACOUSTIC CUES AND RECOGNITION ACCURACY IN CROSS-CULTURAL VOCAL EXPRESSION OF EMOTION

Christopher G. Trimmer, Robin Meyer-MacLeod, Lola L. Cuddy and Laura-Lee Balkwill
Dept. of Psychology, Queen's University, Kingston, Ontario, Canada, 4cgt1@qmlink.queensu.ca

1. INTRODUCTION

Telecommunications play a vital role in today's global culture, and as a result, accurate communication of vocal emotion between distinct cultures has become increasingly important. Based on discrepant findings, two alternative theories of cross-cultural vocal emotion transmission have been put forth (Thompson & Balkwill, 2006). The universal cue theory proposes that basic emotions (anger, joy, fear and sadness) are recognized at above chance accuracy regardless of the encoder's culture and language background, whereas an opposing view states that emotion cues are not universal, and emotion recognition is largely dependent on the culture of the decoder (for a review, see Elfenbein & Ambady, 2002). We investigate this theoretical disagreement by testing participants' recognition accuracy for emotional speech in familiar and unfamiliar languages and paralleling the results with acoustic analyses of the same test stimuli.

A feature of the current study is that English-speaking listeners were asked to rate emotional speech presented in two non-European languages—Mandarin and Dari—as well as in English. Since Mandarin and Dari are geographically and linguistically distinct from English, cultural differences in emotional expression (if they exist) should be most evident for these languages. Also, the present study reversed the design of previous research. Most cross-cultural research designs have presented emotions spoken in one language to many groups and compared recognition across groups (Elfenbein & Ambady, 2002). In the present study, each listener decoded emotion from a variety of cultural groups, and as such, acted as her own control.

It was predicted that English-speaking listeners would recognize basic emotions in each language with above chance accuracy due to the existence of universal cues of emotion expression. A variety of acoustic cues were analyzed to directly compare the cues involved in vocal emotion expressions across cultures. It was anticipated that the pattern of acoustic cues for each emotion would generally adhere to the patterns described in Juslin and Laukka's (2003) review of vocal emotion communication, irrespective of culture. However, it was also predicted that the degree of adherence would vary by culture, which would be a mediating influence in recognition accuracy.

2. METHODS

2.1 Participants

Participants were 30 female Queen's University students (*mean age* = 22, *range* = 18 – 53). All reported normal hearing. All were native English speakers and did not speak or read either Mandarin or Dari.

2.2 Materials

Stimuli consisted of the semantically neutral sentence "The bottle is on the table", uttered in English and, in translation, in Mandarin and Dari. There were 120 utterances formed by the cross-classification of four native speakers of each language x two speaker genders x three languages x five emotions (anger, fear, joy, sadness, and neutral). Each speaker recorded the sentence with intent to express each of the five emotions. Each utterance was confirmed to be a typical display of emotion as judged by an independent female rater native to the relevant culture.

A Power Mac G4 computer running PsyScope X b46 software (Cohen, MacWhinney, Flatt, & Provost, 1993) was used for all stimuli presentation and data collection.

2.3 Procedure

Following demographic data collection, the listening portion of the study was conducted in a sound-attenuated booth with the rating scales presented in PsyScope. Participants were instructed to rate each utterance on four 11-point scales (0 = not present and 10 = extremely present) based on the degree of anger, fear, joy and sadness in the utterances. Participants were not informed about the presence of neutral utterances.

The values on the emotional scales were converted into difference scores using a procedure outlined by Resnicow, Salovey and Repp (2004). The difference scores represent the degree to which an intended emotion was successfully transmitted by an emotional utterance (anger, fear, joy or sadness) relative to a participant's ratings of the neutral utterance. Chance accuracy is zero.

Each utterance was analyzed as a wav file using Praat v 4.5.16 software (Boersma and Weenink, 2005). The acoustic cues analyzed in this study were: speech rate (mean syllable per second), fundamental frequency (mean, median and standard deviation), proportion of pauses, proportion of jitter, first formant (mean and bandwidth) and intensity (mean, median, and standard deviation). These are

commonly analyzed acoustic cues in vocal emotion research (for a review, see Juslin and Laukka, 2003).

3. RESULTS

Recognition accuracy

The descriptive statistics for recognition accuracy for each language and emotion are reported in Table 1. Participants recognized all four emotions in each of three languages with above chance accuracy.

Table 1
Descriptive Statistics of Recognition Accuracy (Difference Score) Across Language and Emotion (n= 30)

Factor	Group	M	SD	Confidence Interval (95 %)*
Language	English	.51	.13	.45 - .55
	Mandarin	.22	.08	.19 - .25
	Dari	.10	.05	.07 - .12
Emotion	Anger	.42	.12	.37 - .47
	Fear	.24	.09	.21 - .27
	Joy	.22	.09	.18 - .25
	Sadness	.22	.09	.19 - .25

Note.* Lower bound of all the confidence intervals is greater than zero indicating that emotional recognition was above chance across emotions and languages.

A repeated-measures analysis of variance was conducted on the difference scores with language (English, Mandarin, and Dari), emotion (anger, fear, sadness, and joy), gender (male, female) and speaker (four speakers) as the experimental factors. There was a significant main effect for language, $F(2, 58) = 248.68, p < .01$, and emotion, $F(3, 87) = 45.69, p < .01$. A paired sample t-test revealed that participants recognized emotion in their own language with significantly higher accuracy than Dari, $t(29) = 17.43, p < .01$, or Mandarin, $t(29) = 16.78, p < .01$, though recognition of Mandarin vocal emotion was significantly greater than Dari, $t(29) = 7.60, p < .01$.

Acoustic cues

Eleven separate mixed-model analyses of variance (emotion as the within-speaker factor, language as the between-speaker) were conducted for each measured acoustic cue to determine whether acoustic cues for an emotion differed across languages. Table 2 shows that significant differences across emotions were found for seven of the eleven acoustic cues. In addition, the analyses of variance yielded several significant interactions between emotion and language.

Further analysis of the variation among languages revealed that acoustic cues underlying perception of pitch and rhythm showed similar patterns across emotions for English and Mandarin, but not for Dari utterances. For example, Figure 1 shows that Dari utterances have little variation in fundamental frequency and percent of pauses across emotions relative to English and Mandarin. Given the low accuracy for which participants recognized emotion in Dari speech, the large deviation of acoustic cue patterns in Dari from English appears to be a likely cause for the decreased recognition of emotion.

Table 2
Analyses of Variance for the Main Effects of Emotion on Acoustic Cues

Cue	Value	F(3, 63)	Partial η^2	p
Speech Rate (syllables/s)	Mean	13.20*	0.68	.01
	Median	14.20*	0.69	.01
Pitch	Mean	13.72*	0.68	.01
	Median	5.33*	0.46	.01
	Std. Dev.	16.9*	0.73	.01
Pauses	Mean	2.33	0.27	.11
Jitter	Mean	1.72	2.13	.20
First formant Bandwidth	Mean	3.00	0.32	.06
	Median	12.36*	0.66	.01
Intensity	Mean	6.93*	0.52	.01
	Std. Dev.	0.48	0.07	.70

Note. * $p < .01$

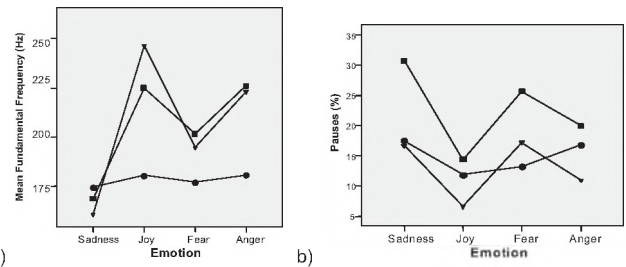


Figure 1 - Measures of (a) mean fundamental frequency (Hz) and (b) percentage of pauses for English, Mandarin, and Dari utterances across the basic emotions [■ English ▼ Mandarin ● Dari]

4. DISCUSSION

All emotions were recognized with above chance accuracy across the three languages. As expected, the English-speaking listeners displayed an in-group advantage in their emotional recognition for English speech compared to Mandarin and Dari. However, the degree to which an individual recognizes emotion conveyed by speech in an unfamiliar culture may be directly related to the acoustic similarities of the expression of emotion in both languages.

REFERENCES

Boersma, P., & Weenink, D. (2005). Praat: Doing phonetics by computer (version 4.3.21) [Computer software]. Amsterdam, Netherlands: University of Amsterdam.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, 25, 257-271.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological Bulletin*, 128, 203-235.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129, 770-814.

Resnicow, J. E., Salovey, P., & Repp, B. H. (2004). Is recognition of emotion in music performance an aspect of emotional intelligence? *Music Perception*, 22, 145-158.

Thompson, W. F., & Balkwill, L.-L. (2006). Decoding speech prosody in five languages. *Semiotica*, 158, 407-424.

ACKNOWLEDGEMENTS

Research supported by NSERC Discovery Grant to LLC.