PRODUCTION AND PERCEPTION OF VOCAL EMOTIONS: A COMPARISON OF MANDARIN CHINESE AND GERMAN EMOTIONAL PROSODY

Huan Wei[1], Mathias Scharinger[1,2], Ulrike Domahs[1,2]

[1]Department of German Linguistics, Philipps-University Marburg, Marburg, Germany
[2]Center for Mind, Brain, and Behavior, Universities of Marburg and Giessen, Marburg, Germany
weih@students.uni-marburg.de

ABSTRACT

Previous findings have demonstrated that prosodic cues accompanying expressed emotions are similar in different languages. Still, it is unclear how those suprasegmental properties vary in different languages and lead to different mappings between expressions and emotional meanings. The present study compared the production and perception of vocal emotions (*NEUTRAL*, *HAPPINESS*, *PLEASANT SURPRISE*, *SADNESS*, and *DISGUST*) in Mandarin Chinese and German. We found that positive emotions in both languages were produced with significantly higher pitch, while negative emotions were expressed with longer duration. However, the pitch contours and durations vary between the two languages. In a perception task measuring response times for the categorization of emotions, 21 German and 21 Mandarin native speakers performed best for *NEUTRAL* with faster response times. Mandarin speakers performed worse for positive than for negative emotions. The group differences in both the production and perception tasks suggest language-specific effects in the processing of emotional prosody.

Keywords:
emotional prosody production, recognition of emotional prosody, cross-linguistic processing

1. INTRODUCTION

In discourse, the emotional states of a given statement are often expressed by means of prosodic cues (e.g., pitch, intensity, and duration). Cross-cultural studies [1], [2] point out that speakers without knowledge about a language are able to recognize emotional prosodies successfully in this respective language, suggesting that emotional prosody exhibits a core set of acoustic-perceptual features which promote accurate cross-language recognition of emotions. However, [3]–[6] report an in-group advantage in recognizing emotional prosody for native speakers in contrast to L2- or foreign speakers and pointed to culture- and language-specific paralinguistic patterns affecting the encoding and decoding of vocal expressions of emotions.

Xu and colleagues [7] proposed emotional meanings to be encoded along benefit-oriented bio-informational dimensions, which involve both segmental and prosodic aspects of the vocal signal. Although previous findings [8]–[10] have shown that the prosodic cues used for the production of emotions in different languages are identical, it is unclear how these suprasegmental properties vary in different languages and lead to different mappings of expressions to emotional meanings. To follow up on this issue, the present study compared the production of vocal emotions in the tonal language Mandarin Chinese and the non-tonal language German. It was assessed how language (Mandarin Chinese and German) and culture (Asian and Western) affect the production and perception of vocal emotions in Mandarin Chinese and German. If emotional prosody is produced universally, it can be expected that the acoustic features of the emotional prosodies in both languages are similar. If the perception of emotions in the speech is unaffected by the culture, it can be expected that the native speakers of both languages could successfully recognize all the emotional prosodies in their native languages.

2. PRODUCTION OF VOCAL EMOTIONS

2.1. Stimulus material

We selected ten bisyllabic German nouns from the semantic category »food«, each with initial syllable stress and no reduced syllables (e.g., *Kaffee* "coffee"). According to the EmoInt-2017-Database [11], the mean emotional valence of the stimuli was 5.695 (SD = 0.52). Additionally, we selected ten

Chinese nouns related to »food«, built from two characters and covering all five tones in Mandarin Chinese (e.g., *苹果*, *píng-guǒ*, "apple").

Three female native speakers of German and Mandarin Chinese were instructed to record each stimulus conveying four emotions (*HAPPINESS*, *PLEASANT SURPRISE*, *SADNESS*, *DISGUST*) and a *NEUTRAL* mode. All recordings were conducted in a sound-proof studio and used as stimuli for the acoustic analysis and perception experiments.

2.2. Acoustic features

The word duration and intensity of the stimuli (Figure 1) were analyzed using *Praat* [12], and time-normalized pitch contours of each utterance (10 points per syllable) were generated using a Praat script of *ProsodyPro* [13] (Figure 2). Statistical analyses were conducted using ANOVAs and post-hoc pairwise comparisons of emotions with Bonferroni adjustment in the R [14].

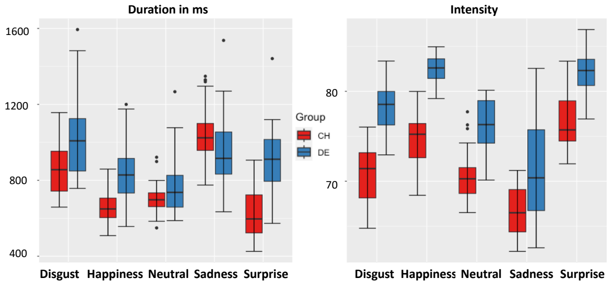


Figure 1: Average word duration and intensity

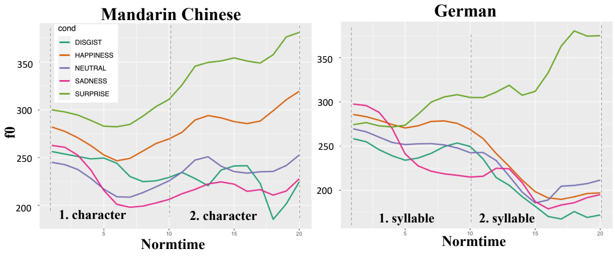


Figure 2: Average pitch contours of stimuli

conditions	emotion	pitch	duration	intensity
	results of post-hoc test within Chinese group			
NEUTRAL	HAPPINESS	< 0.001	-	< 0.001
	SURPRISE	< 0.001	-	< 0.001
	SADNESS	-	< 0.001	< 0.001
	DISGUST	-	< 0.001	-
HAPPINESS	SURPRISE	< 0.001	-	-
	SADNESS	< 0.001	< 0.001	< 0.001
	DISGUST	< 0.001	< 0.001	< 0.001
SURPRISE	SADNESS	< 0.001	< 0.001	< 0.001
	DISGUST	< 0.001	< 0.001	< 0.001
DISGUST	SADNESS	-	< 0.001	< 0.001
	results of post-hoc test within German group			
NEUTRAL	HAPPINESS	-	-	< 0.001
	SURPRISE	< 0.001	-	< 0.001
	SADNESS	-	< 0.01	< 0.001
	DISGUST	< 0.05	< 0.001	-
HAPPINESS	SURPRISE	< 0.001	-	-
	SADNESS	< 0.001	< 0.05	< 0.001
	DISGUST	< 0.001	< 0.001	< 0.001
SURPRISE	SADNESS	< 0.001	-	< 0.001
	DISGUST	< 0.001	-	< 0.001
DISGUST	SADNESS	-	-	< 0.001
Chinese	German	between two languages		
NEUTRAL	*NEUTRAL*	-	< 0.05	< 0.001
HAPPINESS	*HAPPINESS*	< 0.001	< 0.001	< 0.001
SURPRISE	*SURPRISE*	-	< 0.001	< 0.001
DISGUST	*DISGUST*	< 0.05	< 0.01	< 0.001
SADNESS	*SADNESS*	-	-	< 0.05

Table 1: Significant results of the post-hoc tests

The ANOVA results (Table 1) showed significant differences between emotions in pitch, word duration, and intensity for the Chinese and German stimuli: pitch (F(1, 5998) = 26.11, $p < 0.001$), word duration (F(1,298) = 25.42, $p < 0.001$), and intensity (F(1, 598) = 244.4, $p < 0.001$). The emotional prosodies in German had longer word duration and higher intensity, while Mandarin Chinese had higher pitch. These findings suggest that the suprasegmental properties of emotional prosody differ between Mandarin Chinese and German, which may affect the decoding of vocal expressions of emotions.

3. PERCEPTION OF VOCAL EMOTIONS

3.1. Methods

Twenty-one native Mandarin speakers (12 females, mean age of 31.33 years, age range 23-41 years) judged on the stimuli with varying emotional prosodies in Mandarin Chinese, and twenty-one native German speakers (12 females, mean age of 28.86 years, age range 21-54 years) on those in German. Participants of both groups were healthy and did not report any hearing impairments.

The online study consisted of two parts, an online survey on *SoSci Survey* (version 3.4.03) [15] for the demographic data and a *PsychoPy* experiment (version v2021.2.3) on *Pavlovia* [16]. The online experiment consisted of six blocks with 25 trials each. Each trial started with a white fixation point, followed by a 300ms inter-stimulus interval (ISI). Then, an auditory stimulus was presented with a white fixation cross for up to 4000ms until a response was made. Participants had to judge the emotional prosody of each word in their native language and press one of six response options (five emotional conditions plus one for 'others'). An inter-trial interval lasted for 1000 ms after each response.

3.2. Results

The response times were analyzed utilizing linear mixed regression models and the accuracies with logistic mixed regression models. The fixed factors included the variables CONDITION and GROUP, and the base model included both PARTICIPANTS and SPEAKERS as random intercepts.

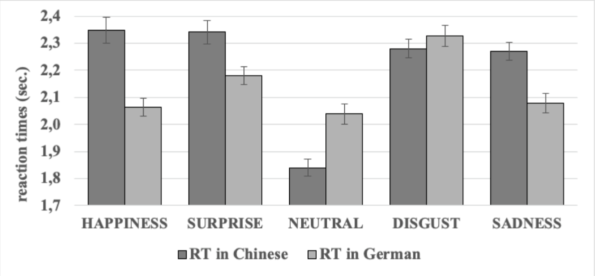


Figure 3: Mean reaction times of the correct answers in the categorization of vocal emotions in Mandarin Chinese and German by the native speakers

The results showed that German native speakers had shorter reaction times (*mean* = 2.13s, *SD* = 0.73) compared to Chinese native speakers (*mean* = 2.25s, *SD* = 0.72) (Figure 3). Although the statistical analysis did not show significant effects of the conditions or interactions on response times between the two groups, within-group analyses revealed significant differences between conditions in response times for both Chinese ($\chi^2 = 132$, $p < 0.001$) and German group ($\chi^2 = 100.46$, $p < 0.001$). Table 2 displays the post-hoc comparisons of conditions within each group. The Chinese native speakers had significantly shorter reaction times in recognizing the *NEUTRAL* mode than emotional prosodies. However, their reaction times for recognizing positive emotional prosodies *HAPPINESS* and *SURPRISE* were longer than for the negative emotional prosodies *DISGUST* and *SADNESS*. In contrast, German native speakers recognized the *NEUTRAL* mode with the shortest reaction times, but they needed significantly longer reaction times to recognize emotional prosody *DISGUST* compared to the other emotional prosodies.

conditions	emotion	RT	accuracy
	results of Chinese group		
NEUTRAL	HAPPINESS	t = 1.98 *	z = -14.05 ***
	SURPRISE	6.48 ***	-12.66 ***
	SADNESS	7.06 ***	-6.47 ***
	DISGUST	7.71 ***	-4.1 ***
HAPPINESS	SURPRISE	4.52 ***	-
	SADNESS	5.07 ***	8.62 ***
	DISGUST	5.74 ***	10.85 ***
SURPRISE	SADNESS	-	6.98 ***
	DISGUST	-	9.28 ***
DISGUST	SADNESS	-	-2.5 *
	results of German group		
NEUTRAL	HAPPINESS	-	-4.05 ***
	SURPRISE	2.46 *	-2.82 **
	SADNESS	-	-
	DISGUST	7.72 ***	-5.37 ***
HAPPINESS	SURPRISE	3.9 ***	-
	SADNESS	2.03 *	4.26 ***
	DISGUST	9.16 ***	-
SURPRISE	SADNESS	-	3.02 **
	DISGUST	5.31 ***	-2.61 **
DISGUST	SADNESS	-7.12 ***	5.57 ***
Chinese	German	between two languages	
NEUTRAL	*NEUTRAL*	-	-
HAPPINESS	*HAPPINESS*	-	-2.8 **
SURPRISE	*SURPRISE*	-	-2.58 **
DISGUST	*DISGUST*	-	-
SADNESS	*SADNESS*	-	-

Table 2: Significant results of the post-hoc tests

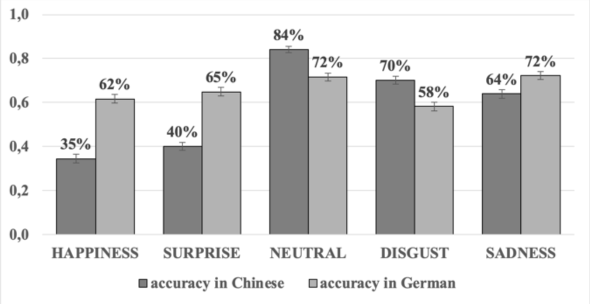
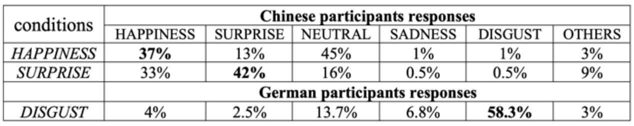


Figure 4: Accuracies of the categorization of vocal emotions in Mandarin Chinese and German by the native speakers

The analysis of overall accuracies (Figure 4) revealed a significant interaction between CONDITION and GROUP ($\chi^2 = 186.68$, $p < 0.001$). Specifically, German native speakers had significantly higher accuracies in categorizing positive emotional prosodies *HAPPINESS* and *SURPRISE* compared to the Mandarin native speakers (Table 2). Furthermore, within-group comparisons showed significant effects of CONDITION for both Chinese ($\chi^2 = 414.43$, $p < 0.001$) and German group ($\chi^2 = 48.4$, $p < 0.001$). The Mandarin Chinese native speakers performed significantly worse when categorizing positive emotions *HAPPINESS* and *SURPRISE*, while German native speakers had lower accuracy in categorizing the emotion *DISGUST*.

4. GENERAL DISCUSSION

The production study found that the acoustic features of the emotional prosody (pitch, word duration, and intensity) were similar in both languages, with both expressing negative emotions *DISGUST* and *SADNESS* with a longer word duration, and the positive emotions *HAPPINESS* and *SURPRISE* with higher pitch and intensity. However, the analysis also revealed significant differences in pitch, word duration, and intensity between the two languages, indicating that the suprasegmental properties of emotional prosody vary across different languages. The results of the perception study showed that both the Mandarin and German native speakers were equally sensitive to emotional prosody in their native language. However, the response patterns differed slightly depending on the specific emotions and language. To be precise, the Mandarin native speakers needed significantly longer reaction times to recognize the positive emotions *HAPPINESS* and *SURPRISE*, and had lower accuracy scores for these emotions. These findings are consistent with the results of [10] that Mandarin listeners had lower accuracy in recognizing positive emotions. On the other hand, the German native speakers had the most difficulties recognizing the emotion *DISGUST*, as evidenced by significantly longer reaction times and lower accuracy scores.



conditions	Chinese participants responses					
	HAPPINESS	SURPRISE	NEUTRAL	SADNESS	DISGUST	OTHERS
HAPPINESS	**37%**	13%	45%	1%	1%	3%
SURPRISE	33%	**42%**	16%	0.5%	0.5%	9%
	German participants responses					
DISGUST	4%	2.5%	13.7%	6.8%	**58.3%**	3%

Table 3: Response matrices for the categorization

We analysed participants' error patterns to investigate the categorization difficulties in the Chinese and German participants (Table 3). Our findings revealed that the Chinese listeners were generally able to distinguish between positive and negative emotions, but struggled with distinguishing between *HAPPINESS*, *SURPRISE,* and *NEUTRAL*. This suggests that the acoustic features of these emotions in Mandarin may have contributed to the confusion. Specifically, we found that the word durations of positive emotions *HAPPINESS* and *SURPRISE* in Mandarin were significantly shorter than in German or negative emotions in Mandarin Chinese, but there were no significant differences between the positive and neutral emotions in Mandarin Chinese. Furthermore, while there were significant differences in pitch between *SURPRISE*, *HAPPINESS*, and *NEUTRAL*, the pitch contours of these conditions were similar (Figure 2). These similarities in pitch contours and the shorter word duration may have made it challenging for the Chinses listeners to differentiate between these emotions. In contrast, the German group did not show such a performance difference between positive and negative emotions. However, the analyses of the error patterns in their responses to the emotion *DISGUST* revealed that 13.7% of the stimuli were classified as *NEUTRAL*, 6.8% as *SADNESS*. These findings suggest that the acoustic features of *DISGUST*, *SADNESS*, and *NEUTRAL* in German may have contributed to the confusion. Specifically, the lack of intensity differences between *DISGUST* and *NEUTRAL*, and the similarity in word duration and pitch contours of *DISGUST* and *SADNESS* may have made it challenging for German speakers to differentiate between these emotions. In addition to the acoustic explanation, Asian culture has a commonly observed norm to hide negative emotions in the communication of messages [6]. This culture-specific rule may strengthen Chinese listeners' sensitivity to acoustic features of negative emotions even more than to positive emotions. The group differences in both the production and the perception tasks suggest language-specific effects on the processing of emotional prosody, at least when the words varying in emotional prosody are presented without any contextual information.

5. CONCLUSIONS

The present study investigated whether prosodic cues of vocal emotions vary in Mandarin and German and how they may lead to different mappings of emotional prosodies to emotional meanings. We found that both languages produced positive emotions with higher pitch than neutral and negative emotions. In contrast to positive emotions, negative emotions were expressed by changes in duration. The behavioral study on the same stimuli with a categorization task showed that the Mandarin group performed better with negative than positive emotions. Further studies need to show whether Chinese listeners may have more difficulties using pitch information for emotional categorization than German listeners because pitch plays a crucial role in Chinese lexical processing.

6. REFERENCES

[1] W. F. Thompson and L.-L. Balkwill. 2006. “Decoding speech prosody in five languages,” *De Gruyter Mouton*, 407–424. doi: https://doi.org/10.1515/SEM.2006.017.

[2] M. D. Pell, L. Monetta, S. Paulmann, and S. A. Kotz,. 2009. “Recognizing Emotions in a Foreign Language,” *J Nonverbal Behav*, vol. 33, no. 2, 107–120. doi: 10.1007/s10919-008-0065-7.

[3] K. R. Scherer, R. Banse, and H. G. Wallbott,. 2001. “Emotion Inferences from Vocal Expression Correlate Across Languages and Cultures,” *J Cross Cult Psychol*, vol. 32, no. 1, 76–92. doi: 10.1177/0022022101032001009.

[4] H. Wei, Y. He, C. Kauschke, M. Scharinger, and U. Domahs,. 2022. “An EEG-study on L2 categorization of emotional prosody in German,” *Speech Prosody 2022*, 629–633. doi: 10.21437/speechprosody.2022-128.

[5] P. Liu, S. Rigoulot, and M. D. Pell,. 2015. “Cultural differences in on-line sensitivity to emotional voices: comparing East and West,” *Front Hum Neurosci*, vol. 9, 311. doi: 10.3389/fnhum.2015.00311.

[6] S. Paulmann and A. K. Uskul,. 2013. “Cross-cultural emotional prosody recognition: Evidence from Chinese and British listeners,” *Cognition Emot*, vol. 28, no. 2, 230–244. doi: 10.1080/02699931.2013.812033.

[7] Y. Xu, A. Kelly, and C. Smillie,. 2013. “Emotional expressions as communicative signals,” *Prosody and iconicity*, 33–60.

[8] M. D. Pell, S. Paulmann, C. Dara, A. Alasseri, and S. A. Kotz,. 2009. “Factors in the recognition of vocally expressed emotions: A comparison of four languages,” *J Phonetics*, vol. 37, no. 4, 417–435. doi: 10.1016/j.wocn.2009.07.005.

[9] S. Paulmann, M. D. Pell, and S. A. Kotz,. 2009. “Comparative processing of emotional prosody and semantics following basal ganglia infarcts: ERP evidence of selective impairments for disgust and fear,” *Brain Res*, vol. 1295, 159–169. doi: 10.1016/j.brainres.2009.07.102.

[10] P. Liu and M. D. Pell,. 2012. “Recognizing vocal emotions in Mandarin Chinese: A validated database of Chinese vocal emotional stimuli,” *Behav Res Methods*, vol. 44, no. 4, 1042–105. doi: 10.3758/s13428-012-0203-3.

[11] M. Köper, E. K. Klinger, and R. Klinger,. 2017. “IMS at EmoInt-2017: Emotion Intensity Prediction with Affective Norms, Automatically Extended Resources and Deep Learning,” *Association for Computational Linguistics*.

[12] Boersma and Weenink,. 2018. “Praat: Doing phonetics by computer [Computer program]. Version 6.0.37.”. https://www.fon.hum.uva.nl/praat/ (accessed Oct. 21, 2021).

[13] Y. Xu,. 2013. “ProsodyPro - A Tool for Large-scale Systematic Prosody Analysis,” *TRASP*.

[14] R. C. Team,. 2018. “R: The R Project for Statistical Computing, version 1.1.419.”. https://www.r-project.org/ (accessed Oct. 21, 2021).

[15] D. Leiner,. 2014. “SoSci survey, version 3.4.03”. https://www.soscisurvey.de

[16] J. W. Peirce,. 2007. “PsychoPy—Psychophysics software in Python,” *J Neurosci Meth*, vol. 162, no. 1–2, 8–13. doi: 10.1016/j.jneumeth.2006.11.017.