How quickly do listeners recognize emotional prosody in their native versus a foreign language?

Marc D. Pell [1], *Jessica Robin* [1,2], *Silke Paulmann* [3]

[1] School of Communication Sciences and Disorders, McGill University, Canada
[2] Department of Psychology, University of Toronto, Toronto, Canada
[3] Department of Psychology, University of Essex, United Kingdom

marc.pell@mcgill.ca, jessica.robin@utoronto.ca, silke.paulmann@essex.ac.uk

Abstract

This study investigated whether the recognition of emotions from speech prosody occurs in a similar manner and has a similar time course when adults listen to their native language versus a foreign language. Native English listeners were presented emotionally-inflected pseudo-utterances produced in English or Hindi which had been gated to different time durations (200, 400, 500, 600, 700 ms). Results looked at how accurate the participants were to recognize emotions in each language condition and explored whether particular emotions could be identified from shorter time segments, and whether this was influenced by language experience. Results demonstrated that listeners recognized emotions reliably in both their native and in a foreign language; however, they demonstrated an advantage in accuracy and speed to detect some, but not all emotions, in the native language condition.
Index Terms: speech prosody, emotions, culture

1. Introduction

Accumulating research supports the idea that 'basic' emotions, such as anger or happiness, can be recognized from the face and from vocal attributes of speech across diverse cultures through the application of universal principles [1]. For example, Scherer and colleagues [2] played emotionally-inflected pseudo-sentences produced in German to speakers of different languages from nine countries. The participants identified four emotions (happy, sad, anger and fear) as well as neutral prosody at accuracy levels well above chance, and similar confusion patterns were observed for listeners from the nine countries. At the same time, it was found that the German listeners demonstrated the highest recognition rates, reflecting an 'in-group' advantage when listeners were presented vocal emotions in their native language.

Thompson and Balkwill [3] conducted a similar study but varied the language of the utterance instead of the listener. They played utterances in five languages (English, German, Chinese, Japanese and Tagalog) to native English speakers and asked them to identify the emotion expressed. The English speakers identified the four emotions (joy, anger, fear and sadness) at above-chance levels in every language, with the highest recognition rates for angry and sad utterances. They also exhibited an in-group advantage, achieving the highest accuracy for the English utterances. The authors concluded that both culture-specific and universal factors are involved in emotion recognition from the voice. This conclusion is supported by a more recent study by Pell, Monetta, Paulmann & Kotz [4] who presented pseudo-sentences in English, Spanish, German and Arabic to a group of monolingual Argentine Spanish speakers; whereas Spanish participants could identify five emotions (happy, sad, fear, anger and disgust) at levels three to four times above chance in each of the four languages, they were significantly better when listening to Spanish (i.e., their native language). Collectively, this research implies that there are universal principles which allow for the recognition of emotions from speech, even when listeners have no experience with the language, but that cultural factors are also important since listeners routinely show an in-group advantage to detect emotions in their native language.

Since it has been established that emotion recognition occurs reliably from speech cues, even when listeners are exposed to a foreign language, a next step is to determine how and when the recognition process occurs. In order to study the time course of vocal emotion recognition, Pell and Kotz [5] have used the auditory gating paradigm to determine *how much* prosodic information is needed to recognize vocal expressions of emotion in a listener's native language. Gating involves presenting an auditory stimulus in segments of increasing duration in order to isolate the exact point where a target meaning is recognizable to listeners. In [5], English-speaking participants were presented emotional pseudo-utterances that increased by one syllable in duration across seven gating conditions; their goal was to identify the emotion conveyed by each stimulus in a forced-choice paradigm. Results demonstrated that fear, sadness, anger, and neutral prosody were recognized more accurately at short gate intervals (i.e., sentences containing 1-3 syllables) than happiness, and particular disgust. When the gate associated with the "emotion identification point" of each stimulus was calculated, data indicated that fear (M = 517 ms), sadness (M = 576 ms), and neutral (M = 510 ms) expressions were identified from shorter acoustic events than the other emotions. Anger and happiness were recognized somewhat later (M = 710 and 977 ms, respectively), whereas disgust required much more exposure to prosodic cues to be recognized (M = 1486 ms). These findings highlighted emotion-specific differences in the time required to process meaning from prosodic cues in English when listeners were exposed to utterances spoken in their native language.

Despite previous evidence that listeners can recognize emotions in a foreign language at high accuracy levels when full utterances are presented, it is unclear whether emotion-specific differences in how quickly listeners recognize basic emotions in the voice would show similarities when listeners process vocal cues in a *foreign* language. As well, it is unknown whether the recognition of vocal emotions in a foreign language would demonstrate a general lag or delay when compared to the native language condition; this possibility is suggested by data reported by Pell and Skorup [6]. In that study which investigated whether emotional priming is produced by emotional utterances gated to specific durations (primes lasting 300, 600, or 1000 ms), the authors

found that listeners required increased exposure to utterances in a foreign language in order to activate vocal emotional meanings when compared to utterances in their native language, suggesting that the time course of emotional processing may differ when processing foreign versus native speech. However, there is little data to corroborate this view.

In this study, we explored the relative time course for recognizing emotions in the listeners' native versus a foreign language by again adopting the auditory gating paradigm. Here, a group of English-speaking listeners heard gated pseudo-utterances produced in their native language, English, and in a foreign language, Hindi. This will allow us to explore how differences in emotion recognition in foreign and native language contexts evolve over time. We focused on recognition of four emotions (happy, sad, fear, anger) and neutral expressions to compare the relative timing of emotion recognition and to determine if the same patterns are observed between the two language conditions (English, Hindi).

Based on previous research, we predicted that the overall recognition rates for the foreign language utterances would be lower than for English, consistent with an in-group advantage. In addition, we expected that emotion recognition would occur faster overall in one's native language than in a foreign language [6]. Finally, we hypothesized that similar patterns of emotion recognition would occur between languages, with fear and sadness having the fastest and most accurate recognition rates and happiness having the least accurate and slowest recognition [1,3,5].

2. Methods

2.1. Participants

Seventeen participants (Mean age = 21.1 years) completed the study (8 female, 9 male). All were native speakers of Canadian English with no knowledge of Hindi.

2.2. Stimuli

The stimuli were 160 audio recordings of pseudo-sentences selected from a published inventory [7] and then edited to different gate durations to create a total of 960 sound files. Eighty Hindi pseudo-sentences and 80 English pseudo-sentences (16 sentences x 5 emotion types in each language) were chosen. Each sentence was produced in four emotional tones (happiness, sadness, fear, anger) and in a neutral manner; all sentences had obtained high consensus rates about the emotion being conveyed according to a group of native listeners of each language who judged the corresponding stimulus set (see [7]). We sought to match the mean recognition rates of each emotion across the two languages, while simultaneously controlling for the sentence type, the length of the sentence, and the sex of the speaker (two male speakers produced the sentences in each language).

Gate Construction – The selected pseudo-sentences were edited using Praat speech analysis software into five audio files of increasing duration. The intervals chosen for each gate were: 200ms, 400ms, 500ms, 600ms, 700ms and the full utterance. We chose to edit the stimuli based on time rather than number of syllables in order to control the length of the stimuli across language conditions. On average, a syllable in English was slightly longer (0.25 s vs. 0.19 s) than a syllable in Hindi, and listeners would therefore receive more acoustic information from the same number of syllables in English versus Hindi stimuli. The durations of the gates were chosen based on the results of [5] who found that, on average, most of the basic emotions were identified between 300 and 800 ms (with the exception of happiness, which was identified at closer to 1000 ms). We chose gates that provided several distinctions within this crucial range in order to further elucidate the time course of emotion recognition. The full utterance was played as the sixth gate, with the aim of measuring the overall recognition rates. In sum, the use of 6 gates, 2 languages, 5 emotional tones, 2 speakers per language and 8 sentences per speaker resulted in a total of 960 audio stimuli used in the experiment.

2.3. Experimental design and procedures

Participants were tested in a quiet laboratory, in a single testing session of approximately 2 hours in duration. Before starting the experiment, participants were told that they would hear audio stimuli that may sound like nonsense or a foreign language, but told to focus on the emotions rather than the meanings of the stimuli. The experiment was designed and run using Superlab 4.0 software. Audio files were played to the listener via headphones, and following the stimulus, the participants performed two forced-choice tasks. The first task, an emotion identification task, presented the participant with five emotions on the screen (anger, happy, sad, fear and neutral) and the participant was asked to select the one expressed in the audio file by clicking on it. The position of the emotions on the screen was randomized and varied across participants. Next, a confidence rating scale appeared and the participant was asked to rate how sure they were of their previous choice on a scale from 1 (least sure) to 7 (most sure). Stimuli were presented in a blocked design, starting with the shortest gate (200 ms) and increasing incrementally to the full utterance. Within the blocks, stimuli were presented in a unique random order intermingling the five emotions and the two languages.

2.4. Statistical analyses

Accuracy scores were determined based on the total number of correct emotional identifications at each gate for each of the 17 participants. This yielded a percent correct score for each item that was then averaged by emotion to compare accuracy across language conditions. Next, the "identification point" was determined for each item for all participants following the procedure described in [5]. These data were used to compare the time course of recognition for each emotion across the native and foreign language conditions.

3. Results

3.1. Accuracy

Mean accuracy scores for each emotion at each gate, expressed as the percentage of correct target responses in each condition, are displayed in Table 1. Qualitative inspection reveals that, in general, accuracy scores increased as the duration of the stimulus increased and the highest accuracy rates were achieved upon hearing the full utterance. All emotions in both language conditions, except for anger in Hindi, reached accuracy levels of at least three times chance (chance level was 20% since there were five choices). Angry Hindi pseudo-utterances reached 54.8% accuracy at gate 6, which was still well above two times chance. Accuracy rates were generally higher for English stimuli throughout, with the

exception of happy utterances, which had consistently higher recognition rates in Hindi until gate 6.

A 6 x 2 ANOVA with repeated measures of gate duration (1-6) and language (English and Hindi) was performed for each emotion condition (anger, fear, happiness, neutral, sadness). All five ANOVAs yielded significant main effects for language and for gate duration. The interaction of language and gate duration was significant for all emotion conditions except anger, Anger: F (5, 80) = 1.47, p = 0.208, Fear: F (5, 80) = 2.59, p = 0.0316, Happiness: F (5, 80) = 13.66, p < 0.0001, Neutral: F (5, 80) = 5.169, p = 0.0004, Sadness: F (5, 80) = 3.63, p = 0.0052.

Every emotion displayed a distinct pattern of accuracy scores across the language conditions and the six gates. Post hoc (Tukey's) analyses of these interactions (p < 0.1) examined how accuracy differed between language conditions at each gate. At every gate, accuracy rates for the recognition of anger were significantly higher in English than in Hindi. Fearful stimuli, however, generated accuracy rates that were not different across English and Hindi until the full utterance was played (gate 6), at which accuracy in English was significantly higher than in Hindi. For happy stimuli, accuracy rates did not differ significantly at the first or last gate but Hindi recognition rates were significantly higher from gate 2 to gate 5. Accuracy for neutral stimuli did not differ across language conditions until gate 3 and then remained different for the three final gates, with English rates higher than those for Hindi. Finally, accuracy for recognition of sad stimuli differed significantly across language conditions only at gates 3 and 5, where the recognition rate for English stimuli was significantly higher. Overall, the accuracy scores for angry and happy stimuli were the most different in English versus Hindi whereas recognition of fearful stimuli was the most similar between native and foreign language conditions.

3.2. Emotion Identification Points

The identification point of each item was calculated for each participant by identifying the gate at which the participant correctly identified the target emotion and maintained the correct response over all subsequent gates. One exception to this criterion was if the participant correctly identified the target emotion twice in succession and only made one error following this. This exception was included in order to allow for some participant error. Once these identification points were determined, they were translated into time values in milliseconds based on the durations of each gate. These data were then averaged to produce a new dependent measure: the mean amount of time required to identify each of the five emotions in each language condition.

Since English and Hindi utterances were slightly different in duration, instances when participants identified the target emotion only at the final gate were omitted from this analysis. The longer English stimuli could potentially bias results by giving the false impression that more time was required to identify emotions in English compared with Hindi. By only using data from the first five gates, we could be assured that participants were exposed to exactly the same duration of the stimulus in both language conditions.

Using this new data representing the time required to identify each emotion, a 5 x 2 ANOVA was performed with repeated measures of emotion (anger, fear, happiness, neutral and sadness) and language (English and Hindi). This analysis revealed no significant main effect of language, F (1, 10) = 0.815, p = 0.388, but a significant main effect of emotion, F (4, 40) = 7.23, p = 0.00018. In addition, the interaction of language and emotion was significant, F (4, 40) = 14.34, p < 0.0001. Post hoc (Tukey's) analysis (p < 0.1) further revealed that the identification points for anger and happy significantly differed across language conditions but those for fear, neutral and sadness did not. Anger was identified significantly faster in English than the foreign language condition (English: M ¬= 356 ms, SD = 77 ms; Hindi: M = 471 ms, SD = 89 ms) whereas happiness was identified significantly faster in Hindi than English (English: M = 592 ms, SD = 89 ms; Hindi: M = 483, SD = 94 ms). In general, all emotions were identified between 300 and 500 milliseconds (between gates 1 and 3), except for happiness in English which took closer to 600 milliseconds to correctly identify.

Table 1: *Mean accuracy scores (percent correct target responses) for each of the five emotions at each gate duration for English and Hindi.*

	Gate Duration (ms)					
Emotion	G1 (200)	G2 (400)	G3 (500)	G4 (600)	G5 (700)	G6 (full)
English (Native Language Condition)						
Anger	54.8	69.9	77.9	82.7	81.6	89.0
Fear	39.3	44.9	55.9	54.4	62.1	79.4
Happy	8.8	12.1	9.6	11.0	9.2	67.3
Neutral	59.2	70.6	76.1	82.0	77.9	90.1
Sad	44.9	60.3	71.0	61.8	69.5	83.1
Hindi (Foreign Language Condition)						
Anger	34.6	40.8	52.2	54.4	59.2	54.8
Fear	33.1	44.1	48.9	48.5	52.2	60.3
Happy	18.8	33.5	39.0	41.2	43.8	65.8
Neutral	62.5	57.4	60.7	62.1	53.7	65.4
Sad	47.4	56.3	53.7	56.3	56.3	76.8

4. Discussion

The results demonstrate that, as expected, listeners can accurately recognize emotions at levels significantly above chance in both native and foreign language conditions [2,3,5]. Also as expected, accuracy of emotion recognition for the full utterance was consistently higher in the native language condition, demonstrating an in-group advantage. Consistent with findings reported by Pell and Kotz [5], accuracy of emotion recognition increased steadily with increased exposure to a vocal stimulus across both language conditions.

Each emotion studied revealed a unique pattern of recognition accuracy over the six gates. We hypothesized that the emotions would show similar recognition patterns in the

native and foreign language conditions, with overall slower recognition in a foreign language. Qualitative examination of the accuracy rates over time reveals that the trends in recognition across the six gates were fairly similar across the language conditions. Recognition rates for neutral, sad and especially fearful stimuli increased along similar trajectories in the foreign and native language conditions. Angry stimuli were consistently more accurately recognized in English, but the rate of accuracy increase is comparable in the two language conditions. Happy stimuli differed the most between the language conditions, although in both English and Hindi recognition rates increased dramatically at gate 6.

The emotional identification point data show similar results in that neutral, fear and sadness were the most alike across language conditions whereas anger and happiness differed significantly. The average identification points for fearful, sad and neutral stimuli were not significantly different in the foreign and native language conditions. Angry stimuli were recognized significantly faster in the native language condition and happy stimuli were recognized significantly faster in the foreign language condition. Thus, we did not find that more time was consistently required to recognize emotions in a foreign language as compared to one's native language. This finding was surprising and contrary to what was expected based on the results of Pell and Skorup [6] who found that longer vocal stimuli were necessary to activate emotional information in a foreign language versus a native language. The present study found that this was only true in the case of angry stimuli, and in fact, less time was needed to identify happy stimuli in the foreign language condition. The reason for these differences in results could be due to the fact that the two studies used very different experimental paradigms. In [6], the focus was on implicit emotional processing studied via priming effects whereas this study used an explicit forced-choice task. It is possible that implicit and explicit emotion processing involve separate mechanisms that respond differently in the foreign language condition as a function of task demands. Alternatively, the different results could be due to the specific languages chosen for the experiment since Arabic was employed for the foreign language condition in the Pell and Skorup study, while the present study used Hindi.

Based on previous studies of emotion recognition in foreign and native languages, we predicted that fearful and sad stimuli would have the highest accuracy rates and be recognized the fastest, while happy stimuli would have the lowest accuracy and slowest recognition. We found that in the native language condition angry and neutral emotions were recognized the most accurately as well as the fastest, while in the foreign language condition sadness was recognized the most accurately and fear, sadness and neutral were recognized the fastest. Happiness was recognized the slowest and least accurately in the native language, as predicted, but while it was also the slowest emotion to be recognized in the foreign language condition, anger was the least accurate. In order to determine whether these patterns can be generalized to all foreign languages or whether these effects are unique to Hindi alone, more research using different foreign and native language conditions will need to be conducted.

5. Conclusions

The results of this study support theories that suggest the existence of universal factors involved in vocal emotions, allowing them to be identified cross-culturally, as well as culture-specific cues which account for in-group advantages in vocal emotion recognition [2,3,5]. In addition, this study demonstrates unique patterns of emotion recognition over time for the five emotions studied in the two language conditions. While the time course of recognition varied significantly across native and foreign language conditions for certain emotions, for others it was very similar. The emotion that displayed the most consistent pattern of recognition across language conditions was fear, which was also one of the emotions recognized the fastest in the foreign language condition. Some suggest that certain emotions are better recognized than others due to evolutionary significance (Thompson and Balkwill, 2006). If this is the case, then it is possible that the recognition of fear has served a more important evolutionary role and thus has certain features that make its recognition consistent across languages. Other emotions, such as happiness, may be less crucial for survival and therefore have been shaped to a greater extent by cultural factors, accounting for less consistent and slower recognition across cultures. More research will be necessary to determine whether the results from this study are due to factors specific to Hindi alone or whether they can be applied to the foreign language condition in general.

6. Acknowledgements

We thank Catherine Knowles for help with testing. This research was financed by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada.

7. References

[1] Ekman, P., "An argument for basic emotions", Cognition and Emotion, 6: 169-200, 1992.

[2] Scherer, K.R., Banse, R., and Wallbott, H., "Emotion inferences from vocal expression correlate across languages and cultures", Journal of Cross-Cultural Psychology, 32: 76-92, 2001.

[3] Thompson, W., and Balkwill, L.-L., "Decoding speech prosody in five languages", Semiotica, 158(1/4): 407-424, 2006.

[4] Pell, M.D., Monetta, L., Paulmann, S., Kotz, S.A., "Recognizing emotions in a foreign language", Journal of Nonverbal Behavior, 33(2): 107-120, 2009.

[5] Pell, M.D. and Kotz, S.A., "On the time course of vocal emotion recognition", PLoS ONE, 6 (11): e27256. doi: 10.1371/journal.pone.0027256, 2011.

[6] Pell, M.D., and Skorup, V., "Implicit processing of emotional prosody in a foreign versus native language", Speech Communication, 50: 519-530, 1998.

[7] Pell, M.D., Paulmann, S., Dara, C., Alasseri, A., and Kotz, S.A., "Factors in the recognition of vocally expressed emotions: a comparison of four languages", Journal of Phonetics, 37: 417-435, 2009.