Gender effects on perception of emotional speech- and visual-prosody in a second language: Emotion recognition in English-speaking films

Suzanne Verheul1, Adriana Hartman2, Roselinde Supheert3, Aoju Chen4

1Scotch & Soda, The Netherlands
2Willem van Oranje College, The Netherlands
3Institute for Cultural Inquiry, Utrecht University, The Netherlands
4Utrecht Institute of Linguistics-OTS, Utrecht University, The Netherlands

s.verheul@hotmail.com, adrianahartmann@me.com
r.g.j.l.supheert@uu.nl, aoju.chen@uu.nl

Abstract

Speakers use both speech prosody and visual prosody (facial expressions, gestures, body postures) to express emotion. Receivers register and recognise emotion via both types of prosodic cues. In this study, we examined gender differences in both recognition of type of emotion (e.g. anger vs. joy) and perceived emotionality (e.g. the degree of anger) expressed via speech prosody and visual prosody in a second language (L2).

In a perception experiment using film scenes, proficient Dutch learners of English rated the emotionality of each protagonist and identified the specific type of emotion expressed by each protagonist in each scene in both the visual-only and audio-only modality. We have found no evidence for gender-related differences in perceived emotionality, possibly due to potential difficulty of participants in identifying with the protagonists portrayed in a different society. However, the female Dutch learners of English were more accurate in recognising type of emotion than the male Dutch learners of English from both speech prosody and visual prosody. These findings suggest that there is transfer of learners’ ability in recognising type of emotion in the native language to L2 and that female L2 learners may be better at learning cues in speech prosody to emotion in L2.

Index Terms: emotion recognition, speech prosody, visual prosody, L2, Dutch, English

1. Introduction

Speakers vary both speech prosody [e.g., 1, 2, 4, 5] and visual prosody (i.e. facial expressions, hand and arm gestures, body postures) to attach emotional information to words [3, 6-8]. Receivers register and recognise emotions “by ears and eyes” [9]. Emotions are more difficult to recognise from speech prosody alone than from facial expressions [10]. Various studies have demonstrated the importance of a coordination between gestures and speech prosody [11, 12] and between facial expressions and speech prosody [13, 14] in emotion recognition. As argued by [13], “when there is inconsistency between verbally and implicitly expressed attitude, the implicit portion will dominate in determining the total message”.

Gender differences have been reported for recognition of emotions expressed via visual prosody. For example, [15] and [16] found that women performed better than men in emotion recognition based on either facial expressions or body language. Other studies [17, 18], however, showed that there were no gender differences in the recognition of facially expressed emotions. [19] suggested that the inconsistency in the findings across studies could be explained by differences in the nature of stimuli; studies using emotional expressions of high intensity showed fewer differences between male and female raters than those using more subtle expressions. Relatedly, [20] analysed gender differences in emotional contagion, and found that women shared the targets’ emotional states more readily than men, although no distinction was made between emotions expressed via speech prosody and visual prosody. [21] investigated the influence of emotional speech prosody on men’s and women’s processing speed when they were visually confronted with positive and negative target words. The female participants showed a priming effect with a smaller interval between the prosodic prime and the visual target word than men. This result indicated that women make a “faster use of emotional prosody during language processing” than men.

Emotion recognition via speech- and visual-prosody in a second language (hereafter L2) can be rather demanding due to cross-linguistic and cross-cultural differences in the expression of emotion [32]. For example, [22] found that Dutch learners of Mandarin Chinese need to be extensively exposed to Mandarin Chinese in order to accurately identify emotions expressed via speech prosody in Mandarin Chinese. [33] found that L2 proficiency also affects the identification of emotions via speech prosody. However, little is known about gender differences in the interpretation of emotional speech prosody and visual prosody in L2. Communicating via L2 has become the norm for more and more language users. Will the ‘female superiority’ in recognising facially expressed emotions in one’s native language [15, 16] be transferred to L2? Will men and women recognise emotions expressed in speech prosody equally well? Past work on emotion recognition is primarily concerned with recognising the type of emotion (e.g. anger vs. joy). There is little research on the interpretation of emotionality (i.e. degree of a certain type of emotion) expressed via speech prosody and visual prosody. Can there be differences in perceived emotionality between men and women in L2?

The present study examines gender differences in emotion recognition in L2 by examining the ability of proficient Dutch
learners of English in not only recognition of type of emotion but also interpretation of emotionality expressed through speech prosody and visual prosody in English. Previous studies on emotion recognition typically use still images or short video clips of one individual displaying a specific emotion, limiting the implications of the findings on emotion recognition in real life. To circumvent this limitation, we use scenes from costume drama films adopted from Charlotte Brontë’s 1847 novel Jane Eyre about the growth of the penniless orphan, who encounters true and false guides on her road to adulthood, acquires relations, a fortune and, finally, finds marital love in her marriage to Mr. Rochester. Film can be a highly suggestive, realistic medium. Costume drama combines naturalistic elements with technical artifice, aiming to manipulate the viewer into emotional identification and response [23, 24].

Using scenes from costume drama films can thus shed light on how L2 users recognise emotion in real-life situations.

Based on the finding of women’s stronger readiness to resonate with others’ emotional states [20], we hypothesise a gender difference in perception of emotionality (Hypothesis 1) and predict that female Dutch learners of English will perceive a higher degree of emotionality in the protagonists than male Dutch learners of English from both visual and speech prosody. In line with the assumption of positive transfer of ‘female superiority’ in recognising types of emotion expressed via visual prosody [15, 16], we hypothesise ‘female superiority’ in similar tasks in L2 (Hypothesis 2). Our prediction is that female Dutch learners of English would be more accurate than male learners in recognising type of emotion from visual prosody in L2. Considering the reported difficulty in recognising type of emotion from speech prosody [10] and the differences in the perception of attributes such as ‘emphatic’ and ‘surprised’ expressed via speech prosody between Dutch and British English (e.g., Native speakers of Dutch perceive a higher degree of surprise emphasis than native speakers of British English given the same pitch span in a speaker) [25], we hypothesise a lack of female superiority in recognising types of emotion from speech prosody (Hypothesis 3). Our prediction is that female Dutch learners of English will encounter similar difficulty to male Dutch learners of English in learning the use of speech prosody to express emotions in English and consequently they perform similarly in recognising type of emotion from speech prosody.

2. Method

We conducted a perception experiment with proficient Dutch learners of English. In this experiment, the participants rated the emotionality of the protagonists and identified the specific type of emotion expressed by the protagonists in the selected film scenes in both the visual-only and audio-only modality.

2.1. Participants

Thirty native speakers of Dutch (age range: 20 ~ 25 years, 15 male, 15 female) took part in the study. They were undergraduate students at Utrecht University at the time of testing and their level of proficiency in English was at least B2 level (intermediate to advanced) in listening, speaking and writing and C1 level (advanced) in reading according to the Common European Framework of Reference [26]. The participants were familiar with neither the novel Jane Eyre nor the film adaptations prior to the experiment, as shown by their responses in the background questionnaire (section 2.4).

2.2. Stimuli

The experimental materials were drawn from two film adaptations of Jane Eyre produced in the UK, i.e. the 2011 [27] and 1997 [28] adaptations. These two adaptations were chosen for two reasons. First, they were filmed largely in the same period, i.e. at around the millennium. Second, they differed in the actors’ expressiveness in terms of speech prosody and visual prosody. In the 2011 version, the two protagonists used subdued

Figure 1. The marriage proposal scene (upper panel: 2011, lower panel: 1997)

body language and a narrow pitch in general; in the 1997 version, the protagonists’ speech had more pitch dynamics and was accompanied by very frequent use of hand gestures and body movements. The 2011 version was thus considered less expressive in its use of prosody and body language than the 1997 version. Figure 1 illustrates the difference in the use of visual prosody between the two film adaptations.

Three scenes that were faithful to the relevant passages in the novel and important to the theme of the novel were chosen per adaptation, leading to three pairs of nearly identical scenes. The first pair of the scenes featured Mr. Rochester and Jane’s final reunion, the second centred on the lines by Jane “because I am poor, plain and little, I am soulless and heartless? You think wrong! - I have as much soul as you, - and full as much heart!”, and the third featured the conversation between Jane and Mr. Rochester starting from his “Do you think me handsome?” and ending with “Come, speak to me”. The first scene portrayed an intense joyful emotion, the second scene an intensely sad emotion, and the third scene a neutral expression according to the novel. Both protagonists were featured to a similar degree in these scenes.

The six scenes were clipped so that they were of similar durations (100s, 94s, 52s in the 2011 adaptation, 129ms, 81s, 60s in the 1997 adaptation), and then made available in two modalities, i.e. visual-only and audio-only.

2.3. Experimental tasks

The participants completed two tasks. In one task, they rated how emotional they perceived each protagonist to be in each
scene on a five-point equal appearing interval scale. The score ‘1’ stood for ‘not emotional’ and the score ‘5’ stood for ‘very emotional’. In the other task, they identified the type of emotion expressed by each protagonist by choosing one out of three options, namely sadness, happiness and neutral.

2.4. Procedure
The participants did the experiment individually in a quiet room. Prior to the experiment, they were asked to fill out a questionnaire on their language background and familiarity with the novel Jane Eyre and the film adaptations. They then proceeded to the experiment and familiarised themselves with the tasks via written instructions. The participants were first presented with the six scenes via PowerPoint presentation in the visual-only modality in which they focused on Mr. Rochester (part 1), then the same six video clips in the visual-only modality in which they focused on Jane (part 2), then these scenes in the audio-only modality in which they focused on Mr. Rochester (part 3), and lastly these scenes in the audio-only modality in which they focused on Jane (part 4). The order in which the scenes were presented to the participants was randomised for each part of the experiment to minimise the chance of any biases in the ratings. We chose to present the participants first with the scenes in the visual-only modality in order to minimise the risk that they might relate the protagonists’ dialogues with images and pay less attention to the images in the visual-only modality. In each part of the experiment, the participants played each scene once and rated first the degree of emotion and then identified the specific type of emotion for the protagonist they were supposed to focus on in each scene, using an answer sheet (Figure 2). The experiment was self-paced, and lasted for about thirty minutes per participant.

<table>
<thead>
<tr>
<th>not emotional</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>very emotional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happiness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Sadness ( )</td>
</tr>
<tr>
<td>Neutral</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Neutral ( )</td>
</tr>
</tbody>
</table>

Figure 2. The scale for emotionality and options for type of emotion for one trial.

3. Analysis and results
We conducted separate analyses on the emotionality scores and judgements on type of emotion to assess the effects of the participants’ gender, modality of prosody, version of film adaptations and the protagonists’ gender on each outcome variable.

3.1. Perception of emotionality
We first analysed the effects of the experimental variables on the emotionality scores, using the lme4 package [29] in the R environment [30]. We included four fixed factors, i.e. GENDER (of participants), with two categories (male vs. female), MODALITY (of prosody), with two categories (speech prosody, visual prosody), FILM, with two categories (1997, 2011), and GENDER_PROTAGONIST (male, female), and one random factor PARTICIPANT. Seven models were built, starting with a model with only the constant and random factors and adding one more term (i.e. a main effect or an interaction) to each subsequent model. The ‘anova’ function in R was used to compare model fit of different models. The model with the best-fit was determined by comparing each model with the winning model in the preceding model comparisons. The best-fit model was the model including the main effects and the interactions between three of the fixed factors (GENDER, MODALITY, FILM) and the main effect of the fixed factor GENDER_PROTAGONIST. But only the main effect of GENDER_PROTAGONIST was statistically significant ($\beta=0.294, SE=0.086, t=3.396, p < 0.001$) in the best-fit model. Both the male and female participants assigned a higher emotionality score to the female protagonist than the male protagonist (3.32 vs. 3.09 in the female participants; 3.27 vs. 2.92 in the male participants).

3.2 Recognition of emotion type
We used the mixed-effect (binary) logistic regression model in SPSS [31] to analyse the participants’ judgements on type of emotion. The outcome variable of the model was correctness of emotion type, including two categories (correct, incorrect). The fixed factors included four main effects, i.e. GENDER (male vs. female), MODALITY (speech prosody, visual prosody), FILM (1997, 2011), and GENDER_PROTAGONIST (male, female). In addition, one random variable was added to the model, i.e. PARTICIPANT. The mixed-effect model revealed only a main effect of GENDER ($\beta=0.445, SE=0.217, t=2.046, p < 0.05$). Specifically, the female participants (61.4%) were significantly more likely to correctly identify the type of emotion of the protagonists than the male participants (54.7%).

4. Discussion and conclusions
Using materials from the film adaptations of Charlotte Brontë’s novel Jane Eyre, we have examined whether proficient male and female Dutch learners of English differed in recognition of type of emotion and perceived emotionality in English (L2). We have found no evidence for gender-related differences in perceived emotionality in L2. Both the male and female Dutch learners of English perceived a higher degree of emotionality in the female protagonist than in the male protagonist, independent of the modality of the prosodic cues and version of the film. This contradicts our prediction of a higher perceived degree of emotionality in the female participants based on women’s stronger readiness to identify with other’s emotional states [20]; Hypothesis 1 is thus not supported. A possible explanation could be that the film adaptations portrayed a society set in the 1840’s with different norms and values, especially regarding women, and the female participants may have felt more distant towards the protagonists than otherwise.

Furthermore, the female Dutch learners of English were more accurate in recognising type of emotion, independent of modality of the prosodic cues, version of the film, gender of the protagonist. This finding is in line with our prediction regarding visual prosody, suggesting transfer of L2 learners’ ability to recognise type of emotion from visual prosody in the native language to L2 and supporting Hypothesis 2.

Unexpectedly, the female Dutch learners of English also outperformed the male Dutch learners of English in recognising type of emotion from speech prosody. This finding would seem to suggest that female L2 learners might be better at picking up cues in speech prosody to type of emotion in English than male L2 learners, contra Hypothesis 3.

To conclude, our study has revealed interesting differences and similarities between proficient male and female L2 learners of English in emotion recognition in English. Future research
using films portraying contemporary society can answer the question as to whether proficient female Dutch learners of English will perceive a higher degree of emotionality in English, as they would do in their native language. More research is needed to assess the generalisability of our finding that female L2 learners are better at recognising type of emotion from not only visual prosody but also speech prosody in L2.

5. Acknowledgements

We wish to thank Nuzha Moritz for useful discussion in the early stage of the study and our participants for their collaboration. This study was supported by a ‘Dynamics of Youth’ grant awarded to Aoju Chen by Utrecht University.

6. References


http://www.erlk.twitter.com/erkenvedo/


