The interaction of emotional information from the voice and touch
Rika Oya¹,² and Akihiro Tanaka¹,*

¹Tokyo Woman’s Christian University,
2–6–1 Zempukuji, Suginami-ku, Tokyo, 167–8585 Japan
²Japan Society for the Promotion of Science,
Kojimachi Business Center Building, 5–3–1 Kojimachi, Chiyoda-ku, Tokyo, 102–0083 Japan

(Received 20 April 2022, Accepted for publication 17 May 2022)

Keywords: Emotion, Audio-tactile interaction, Multisensory

1. Introduction
We communicate emotions through nonverbal channels, for example vocal expression. Previous studies have examined the perception of emotions from voice and reported that six basic emotions and some positive emotions can be recognized above chance level [e.g., 1, 2]. Several studies have focused on the perception from multisensory information including voice, most focusing on voice and face [3–7]. Takagi et al. (2015) examined the multisensory interaction in emotional communication, specifically the facilitation effect. The facilitation effect is defined as the difference in the accuracies between multisensory congruent condition (e.g., happy face and happy voice) and unimodal condition (e.g., happy voice). Therefore, the facilitation effect represents whether the unimodal perception of emotions is facilitated by adding emotional information expressed with the other channel. Takagi et al. revealed that the perception of vocal emotions was facilitated by emotionally congruent face.

However, it is not clear whether touch facilitates the perception of vocal emotions. Interpersonal touch plays an important role in our daily communication, for instance that between infants and caregiver [8]. Moreover, recent research revealed that positive emotions can be communicated more correctly through touch than through voice [9]. Given that, it is possible that touch facilitates the perception of positive emotions through voice. Therefore, we examine whether touch facilitates the perception of vocal emotion. Specifically, we hypothesize that emotionally congruent touch facilitates the perception of vocal emotions (i.e., facilitation effect), especially for positive emotions.

2. Method
To examine the effect of touch, we compared the emotion perception from emotionally congruent multimodal stimuli (voice and touch) with that from unimodal vocal stimuli.

2.1. Participants
Twenty-six Japanese women, who were students at Tokyo Woman’s Christian University (M = 20.65 years, SD = 1.06), participated in this study. We recruited only female participants to control for potential gender effects. They provided their verbal consent, and the study was approved by the Ethics Committee of Tokyo Woman’s Christian University.

The participants were randomly assigned to dyads. One member of each dyad was assigned the role of touching (encoder, n = 13), whereas the other member was assigned the role of being touched (decoder, n = 13). The participants did not know any information about their partners.

2.2. Emotion
We examined the perception of four emotions: anger, sadness, happiness, and gratitude. Our previous work showed that each of the four emotions are an emotion in negative and high-aroused (anger), negative and low-aroused (sadness), positive and high aroused (happiness), and positive and low-aroused (gratitude) in two-dimensional space comprised of the emotional valence and arousal [10].

2.3. Stimuli
Vocal stimuli were taken from our previous study [9] including the perception of emotions from vocalization. In the previous study, the encoders freely expressed emotions uttering /e/, which is often used as an interjection in Japanese with emotional prosody [11]. The vocalizations were recorded. We picked up a part of them (the expressions of anger, sadness, happiness, and gratitude) and used as vocal stimuli in this study. The mean accuracy in Oya and Tanaka (2021) was 33.3% (anger), 51.9% (sadness), 44.4% (happiness), and 0% (gratitude).

2.4. Apparatus
A desk (RAC-EC2SN, Sanwa Supply) was divided into two sections using a black curtain. The encoder and decoder were seated on the two opposite sides of the desk with a curtain between them.

An emotion was instructed to the encoder using a monitor. Vocal stimuli were presented to the decoder using a 15-inch laptop (Inspiration 15-7000, Dell), through headphone amplifiers (DAC-HA200, ONKYO) and headphones (HDA300, SENNHEISER) in both sessions. Vocal stimuli were controlled through PowerPoint (Microsoft) in multimodal session and through Psychopy 3.1 [12] in voice session.

2.5. Procedure
In the experiment, there were two sessions included, namely a multimodal session and a voice session. We conducted 16 trials in each session. The encoder and decoder participated multimodal session firstly and then the decoder did a voice session.
2.5.1. Multimodal session
A total of 16 trials (four emotions (anger, sadness, happiness, and gratitude) × four repeats) were conducted. First, the experimenter instructed the encoder to touch the decoder’s arm as far as she could see from behind the curtain in any way she thought appropriate, as long as she did not hurt the decoder. Moreover, the experimenter told the encoder not to speak until the end of the experiment. Next, the experimenter informed the other participant, who was going to play the role of the decoder. We requested her not to speak until the end of the experiment. When the experiment started, the decoder extended her arm to the encoder’s area. The encoder then touched the decoder’s arm after five counts, and tried to convey the instructed emotion through touch. Vocal stimuli were presented at the same time as the encoder’s touching (Fig. 1). Expressed emotion of the vocal stimuli was congruent with that of encoder’s touch for all trial. The order of the expressed emotions was randomized across participants. Once the encoder finished touching, the experimenter handed out a response sheet to the decoder. We instructed the decoders to judge the intended emotion from both voice and touch and the decoder indicated the emotion by choosing one of five response options, which included the names of the 4 emotions and the statement, “None of these terms are correct.” The instruction in the response sheet stated, “Please choose the term that best describes what this person is communicating to you.” The placement of the 4 emotions on the response sheet was randomized across the participants. Interpersonal touch expressing the same emotion was different among each of four trials since it cannot be recorded and presented repeatedly, whereas each of recorded vocal stimuli which represent the same emotion was presented four times.

2.5.2. Voice session
A total of 16 trials (four emotions (anger, sadness, happiness, and gratitude) × four repeats) were conducted. In each trial, vocal stimuli which were the same as used in multimodal session were presented at once. The order of the stimuli was randomized across participants. Then, the decoder indicated the perceived emotion by choosing one of five response options, which is identical with the multimodal session.

3. Results
3.1. Emotion categorization
To examine whether the decoder could perceive the encoder’s expressed emotion, we calculated categorical accuracy showing the degree to which the decoder’s response (perceived emotion) was congruent with the emotion expressed by the encoder (expressed emotion).

Table 1 showed the confusion matrices on the multimodal session and voice session. We set the chance level at 25% because we assumed that the decoder chooses perceived emotions from four quadrants in two-dimensional space. One sample t-tests revealed that the accuracy of anger (p < 0.001), sadness (p < 0.001), and happiness (p < 0.05) were above chance in multimodal session. In voice session, anger and sadness were above chance (p < 0.001).

3.2. The interaction of emotional information from the voice and touch
To examine whether touch facilitates the perception of vocal emotion, an analysis of variance (ANOVA) was performed.

Table 1  Confusion matrices on multimodal and voice session.

<table>
<thead>
<tr>
<th></th>
<th>Perceived emotion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Anger</td>
</tr>
<tr>
<td><strong>Multimodal</strong></td>
<td></td>
</tr>
<tr>
<td>Exp. Emotion</td>
<td></td>
</tr>
<tr>
<td>Anger</td>
<td>78.85***</td>
</tr>
<tr>
<td>Sadness</td>
<td>5.77</td>
</tr>
<tr>
<td>Happiness</td>
<td>11.54</td>
</tr>
<tr>
<td>Gratitude</td>
<td>1.92</td>
</tr>
</tbody>
</table>

| **Voice** |       |         |           |           |      |
| Exp. Emotion |       |         |           |           |      |
| Anger | 69.23*** | 9.62 | 0.00 | 1.92 | 19.23 |
| Sadness | 26.92 | 61.54*** | 3.85 | 0.00 | 7.69 |
| Happiness | 15.38 | 17.31 | 30.77 | 3.85 | 32.69 |
| Gratitude | 19.23 | 30.77 | 17.31 | 5.77 | **26.92** |

Note. Underbar indicates an accuracy of each emotion. Asterisks indicate significantly higher than the chance level (*p < 0.05, **p < 0.01, ***p < 0.001).
Results showed that the accuracies on multimodal emotional information from the voice and touch was stronger than voice session (\(M = 41.83\)), suggesting that the interaction of emotional information from the voice and touch emerged. Two-way interaction between emotion and modality was not significant (\(F(3,36) = 1.30, n.s.\)). However, we conducted post-hoc t-tests between multimodal vs. voice only for each emotion to examine whether the impact of the interaction of emotional information from the voice and touch was stronger in particular positive emotions than in negative emotions. Results showed that the accuracies on multimodal session were higher than that of voice session in sadness (\(t(12) = 2.92, p < 0.05\)) and gratitude (\(t(12) = 4.63, p < 0.01\)), but not in anger (\(t(12) = 1.33, n.s.\)) and happiness (\(t(12) = 1.53, n.s.\)).

4. Discussion

In this study, we investigated whether touch facilitates the perception of vocal emotion. The results revealed that emotions were communicated more accurately through voice and touch than through voice only, but the impact of the interaction of emotional information from the voice and touch was different between emotions.

We hypothesized that emotionally congruent touch facilitates the perception of vocal emotions. Our results showed that the interaction of emotional information from the voice and touch emerged. Although the results are consistent with our hypothesis, we cannot rule out other possibilities. One of the alternative interpretations is that voice suppressed the perception of tactile emotions. To elucidate these possibilities, future studies can employ the cross-modal bias paradigm [13], in which participants are instructed to judge the emotion from one modality (e.g., voice) and ignore another modality (e.g., touch) on multisensory stimulus.

Moreover, we hypothesized that the facilitation effect of touch is stronger for the perception of positive emotions than negative emotions. However, the results did not fully support the hypothesis. Since we examined limited number of emotions, further studies need to focus on the facilitation effect for the perception of various emotions.

Takagi et al. (2015) revealed that the perception of vocal emotions was facilitated by emotionally congruent face [6]. Takagi et al. examined the perception of six basic emotions (anger, disgust, fear, happiness, sadness, and surprise) and found that the facilitation effect of face for the perception of vocal emotion was significant in anger, disgust, and happiness but not in other emotions. In contrast, our results showed that the interaction of emotional information from the voice and touch was significant in sadness and gratitude but not in anger and happiness. Taken together, face and touch may affect differently toward voice.

Acknowledgements

We thank all the participants. We gratefully acknowledge funding from JSPS KAKENHI No. 21J12728 and Grant-in-Aid for Scientific Research on Innovative Areas No. 17H06345 “Construction of the Face–Body Studies in Transcultural Conditions.”

References