The Effects of Anxiety on the Perception of Emotion in the Face and Voice

Ai Koizumi 1, Akihiro Tanaka 2,4, Hisato Imai 3, Saori Hiramatsu 3, Eriko Hiramoto 3, Takao Sato 1, Beatrice de Gelder 4.

1 Department of Psychology, The University of Tokyo, Japan
2 Waseda Institute for Advanced Study, Waseda University, Japan
3 Department of Psychology, Tokyo Woman’s Christian University, Japan
4 Department of Cognitive Neuroscience, Tilburg University, The Netherlands

Abstract

Anxious individuals have been shown to be hyper-sensitive to cues for others’ negative emotional states. As most studies used facial expressions as emotional cues, we examined whether the trait anxiety affects the cross-modal perception of emotion in the face and voice that were simultaneously presented. The face and voice cues conveyed either matched (e.g., both positive) or mismatched emotions (e.g., positive face and negative voice). Participants indicated the perceived emotion in one of the cues, while ignoring the other. The results showed that highly anxious individuals, compared with low anxious ones, were less able to disregard the to-be-ignored negative cues and misinterpret the attended happy cues as negative. These results were found regardless of the cue modality. The trait anxiety may affect the integration of emotion in the face and voice.

1. Introduction

Accurate perception of others’ emotional states is central to our social communications. Yet, how we perceive others’ emotion appears to vary across individuals depending on our personality traits. Previous studies have shown that individuals with heightened anxiety have a bias to interpret others’ emotional state in a negative manner [1]. Also, anxious individuals have been shown to prioritize processing of negative facial expressions over other available information [e.g., 2]. As such bias towards negative emotional cues could affect our social life (e.g., increase social stress when others appear hostile to you), understanding the nature of such bias is crucial to alleviate it and its negative consequences.

Although most studies have examined the effects of anxiety on the facial expression processing, we often perceive others’ emotional states in other modalities besides the visual modality, especially in the voice. Whether anxious individuals perceive emotion in the voice differently from non-anxious ones remains largely unknown. A few recent studies have shown the effects of anxiety-related personality traits on the neuronal processing of emotion (i.e., anger) expressed in the voice [3, 4]. A remaining important question, however, is whether anxiety affects our perception of others’ emotion in their voices, as alternation of perception could directly affect our social life. One pilot study [5] showed that clinically anxious individuals recognize negative tones of voice more easily and positive ones less easily. Yet, it appears that anxiety level does not always affect the perception of emotion in the voice [4]. Thus, the effect of anxiety on emotional perception in the voice might not be as prominent as that in the face.

However, in the natural environment, we often perceive emotion in the face and voice not in isolation but in combination. Indeed, it has been shown that perception of emotion in the faces is affected by the accompanied voices, and vice versa [6, 7]. For example, it has been shown that we are less likely to perceive negative facial expressions as negative in the presence of positive voice [6]. That is, when we infer others’ emotional states, emotion in the face and voice cues are not processed independently from each other, but are integrated. Therefore, the effects of anxiety on the perception of emotion in the multi-modal cues (face and voice) may differ from those on the perception of unimodal cues (face or voice).

The aim of the study was to examine the effect of anxiety on the multi-modal perception of emotion in the face and voice. In the present study, affective prosodies were presented in synchrony with dynamic facial expressions [8], resembling a naturalistic social situation. We included conditions in which the face and voice conveyed congruous or incongruous emotions (i.e., happiness and anger). This enabled us to examine the effect of trait-anxiety on how we integrate emotional cues across the modalities.

2. Method

2.1. Stimuli

Two Japanese females were audio and video recorded while uttering semantically neutral short sentences such as “What is this?” in either angry or happy emotion. Emotion was expressed through the face and voice. The duration of spoken sentences was matched when expressed in angry ($M = 767$ ms, $SD = 96$ ms) and happy emotion ($M = 771$ ms, $SD = 89$ ms) ($t(15) = 0.34, p = .74$).

The stimuli were prepared from the recorded materials so that emotion (angry or happy) in the face and voice cues was congruous in half the stimuli and incongruous in the other half. Incongruous face and voice cues were integrated so that their asynchrony was within the range of 0 to 80 ms ($M = 27.3$ ms), which was shorter than a detectable audiovisual asynchrony in speech [9, 10]. Thus, it was ensured that participants could not base their response on such subtle, undetectable asynchrony in the stimuli. A total of 32 stimuli were prepared (2 models x 2 emotions x 2 congruency x 4 sentences).

As a preliminary study showed that the perception accuracy of emotion in the face was too high ($M = 98.0\%$) to be modulated by the voice cues, visibility of the face was...
reduced by superimposing random noise [7]. We ensured that the perception accuracy of both face and voice cues would be moderately high (i.e., above 80%), but not too high so that it could be affected by the simultaneously presented cue from the other modality.

2.2. Participants

Twenty-five Japanese college students (15 males and 10 females) were recruited by a word-of-mouth and participated in this study. They were divided into Low (N = 13) and High anxious groups (N = 12) based on their trait-anxiety levels. The trait-anxiety levels were measured with a Japanese-translation [11] of the widely used Spielberger State-Trait Anxiety Inventory [12] following the experimental tasks. High anxious group (M = 37.92, SD = 1.28) had higher trait-anxiety level than Low anxious group (M = 51.42, SD = 2.42) (t(23) = 5.03, p < .01).

2.3. Procedure

There were four tasks in the study: two multi-modal tasks and two single-modal tasks. The single-modal tasks were performed following the multi-modal tasks.

2.3.1. The multi-modal tasks

In the multi-modal tasks, participants were presented with the synchronized face and voice cues. Face was attended in the multi-modal face task, and voice in the multi-modal voice task. The order of these tasks was counterbalanced. The tasks contained 64 trials each (32 congruent) which were presented in a randomized order.

On each trial, participants were presented with a fixation point (1 sec) followed by either the congruent or incongruent stimuli. They were instructed to indicate whether they perceived angry or happy emotion in the attended cue (face or voice) while ignoring the other cue. They were also instructed to perceive the to-be-ignored cues (e.g., do not close eyes or look away from the faces in the voice task), but just not to base their responses on them. Responses were made by key-pressing, and were not speeded.

2.3.2. The single-modal tasks

In the single-modal tasks, participants were presented with either the face or voice cues in separate blocks. The order of the blocks was counterbalanced. The stimuli were identical to the ones in the multi-modal tasks, except that a cue in only one of the two modalities was presented at a time. There were 32 trials in each block which were presented in a random order. They indicated the emotion they perceived in the presented cues as in the multi-modal tasks.

3. Results

3.1. The multi-modal tasks

The accuracy rates are summarized in Table 1. The overall accuracy in the voice task (M = 93.9 %) was higher than that in the face task (M = 71.13 %) (t(24) = 9.06, p < .001).

Importantly, further analysis revealed the effects of anxiety on the perception of multi-modal emotional cues. In order to examine the effects of to-be-ignored emotional cues on the perception of emotion in the attended cues (face or voice), accuracy on the congruous trials was subtracted from that on the incongruous trials (Fig. 1 and 2). Larger calculated differences in accuracy (i.e., congruency effect) would indicate that accuracy in perception was hindered by the to-be-ignored incongruous emotional cues.

The congruency effects were analyzed with the analysis of variance (ANOVA) with Anxiety (low/high) as a between-participant factor and Task (face/voice) and Emotion (angry/happy) expressed in the attended cues as within-participant factors. The analysis revealed a significant main effect of task (F(1, 23) = 43.40, p < .001), showing a larger overall congruency effect in the face task than that in the voice task. More important to our research interest, an interaction between Anxiety and Emotion was significant (F(1, 23) = 11.21, p < .01).

The multiple comparisons (Bonferroni) showed that while Low anxious group was more likely to be affected by the to-be-ignored happy cues than the angry cues (p = .08), High anxious group was affected by the angry cues significantly more than the happy cues (p < .01). This interaction was not modulated by Task (F(1, 23) = 2.07, p = .16), indicating that perception of emotion in High anxious group was affected by the to-be-ignored angry cues regardless of their modality (face or voice).

To further entangle the effects of anxiety in the multi-modal tasks, the congruency effects (i.e., reduced accuracy in the incongruent trials than the congruent trials) in the face and voice tasks were separately analyzed with a two-way ANOVA with Anxiety as a between-participant factor and Valence as a within-participant factor. In agreement with the above mentioned analyses, an interaction between Anxiety and Valence was significant in the face task (F(1,23) = 7.63, p = .01) and marginally significant in the voice task (F(1, 23) = 3.70, p = .07).

In the face task, the multiple comparisons (Bonferroni) showed that High anxious group was more likely to be affected by the to-be-ignored angry voice than Low anxious group (p < .05). Also, while Low anxious group was more likely to be affected by the happy voice cues than the angry voice cues (p < .05), High anxious group showed the opposite trend with larger effect of the to-be-ignored angry voice cues (p < .10). Similarly, in the voice task, it was shown that High anxious group was more likely to be affected by the to-be-ignored angry faces than the happy faces (p < .01), while Low anxious group did not show such trend (p = .88). As the overall frequency of angry responses was equivalent between Low (M = 46.9 %, SD = 1.5%) and High anxious groups (M = 47.6 %, SD = 1.5%) (F(1,23) = .10, p = .76), the above mentioned results cannot be accounted for by a response bias in Low or High anxious group.

3.2. The single-modal tasks

In the face and voice single-modal tasks, the means for the overall accuracy were 80.43 % and 89.25 % respectively. As in the multi-modal tasks, accuracy was higher in the voice task than in the face task (t(24) = 2.77, p < .05).

The accuracy in each task was separately analyzed with a two way ANOVA with Anxiety (low/high) as a between-participant factor and Emotion (angry/happy) expressed in the cues as a within-participant factor. In the face task (Fig. 3), a main effect of Valence was significant (F(1,23) = 13.03, p < .001) with higher accuracy on the perception of the happy faces (M = 87.14, SD = 2.95) than that of the angry faces (M = 73.79, SD = 3.61). A main effect of Anxiety was also significant (F(1,23) = 4.33, p < .05), indicating higher accuracy in Low anxious group (M = 85.48, SD = 3.05) than in High anxious group (M = 74.97, SD = 4.34). The interaction
between Valence and Anxiety was not significant ($F(1, 23) = .15, p = .70$).

Contrary to the face task, none of the factor showed a significant main effect or interaction in the voice task ($p > .05$). That is, neither Valence nor Anxiety modulated the accuracy in the single-modal voice task (Fig. 4).

### Table 1. Accuracy (%) of the emotion perception in Low and High Anxious groups as a function on the target modality, the target emotion, and the congruency in the multi-modal tasks.

<table>
<thead>
<tr>
<th>Target</th>
<th>Congruency</th>
<th>Anxiety Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td><strong>Face</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Angry</td>
<td>Congruent</td>
<td>80.29 (4.74)</td>
</tr>
<tr>
<td></td>
<td>Incongruent</td>
<td>52.40 (7.20)</td>
</tr>
<tr>
<td>Happy</td>
<td>Congruent</td>
<td>87.98 (2.87)</td>
</tr>
<tr>
<td></td>
<td>Incongruent</td>
<td>71.63 (4.45)</td>
</tr>
<tr>
<td><strong>Voice</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Angry</td>
<td>Congruent</td>
<td>96.15 (1.33)</td>
</tr>
<tr>
<td></td>
<td>Incongruent</td>
<td>94.71 (2.90)</td>
</tr>
<tr>
<td>Happy</td>
<td>Congruent</td>
<td>95.19 (1.61)</td>
</tr>
<tr>
<td></td>
<td>Incongruent</td>
<td>93.27 (1.65)</td>
</tr>
</tbody>
</table>

Standard errors of the means are given in parentheses.

### 4. Discussion

The current study examined the effect of trait-anxiety on the perception of emotion in the face and voice, which were presented in synchrony. We showed that, regardless of the cue modality (i.e., face or voice), anxious participants were likely to perceive the happy cues as angry when they were accompanied by the incongruous angry cues. In other words, anxious participants were unable to disregard the to-be-ignored negative emotional cues when judging others’ emotion states. The findings suggest that the bias in anxious individuals to negatively interpret emotional cues [1] is not confined to the visual modality, but prevails across multiple modalities.

Interestingly, despite the above mentioned anxiety effect on voice processing, the perception of voice cues was not affected by the anxiety level when they were presented in isolation (i.e., without face cues). The lack of effect of anxiety on the unimodally presented voice cues is in agreement with a previous report [4]. Therefore, the current results, together with a previous finding [4], suggest that anxious individuals show a biased perception of emotion in the voice when in combination with facial expressions, but not necessarily when presented in isolation.

Our findings that anxiety affects our perception of emotional voice only in the presence of facial expressions can be explained by the integrative nature of emotion processing in the visual and auditory modalities. It has been suggested that simultaneously presented emotional face and voice are processed not independently from each other, but interdependently on each other [6]. Thus, anxiety may affect the integration of emotional information in different modalities more than the single-modal processing of voice. As anxious individuals were likely to be affected by the to-be-
ignored negative cues while attending to the positive cues, trait-anxiety may affect the integration processing by re-weighting two streams of information based on their valence. In other words, highly anxious individuals appear to prioritize negative cues over positive cues even when they are carried through different modalities. This argument is in line with a widely accepted view that anxious individuals prioritize the processing of negative information over that of other information [e.g., 13]. Yet, as this view has been mostly derived from the studies on visual emotional processing, the present study extends the view by showing that prioritization of negative information occurs beyond the boundary of modalities among anxious individuals.

If anxiety were to affect the integration of emotional cues in the face and voice, at which stage of the integrative processing does anxiety exert its effect? A previous study [14] showed that the congruency of emotion in the face and voice modulates the neuronal activity related with conflict processing, which has been shown to be hindered among anxious individuals [see 13]. The hindered conflict processing appears to cause reduced inhibitory control over negative information among anxious individuals [13]. In our study, the voice cues were to be ignored in the multi-modal voice task, while they were to be attended in the single-modal task. As mentioned earlier, anxiety affected the performance only in the former task which required inhibitory control over the to-be-ignored voice. Thus, it is likely that reduced inhibitory control over negative information among anxious individuals is also present in the auditory modality. And this may explain how negative information in the unattended modality is prioritized over positive information processing in the attended modality among anxious individuals.

 Besides the above mentioned effects of anxiety on the emotional voice processing, results from the single-modal face task also give interesting implications. It was shown that highly anxious individuals were less accurate at perceiving facial expressions than low anxious individuals. This hindered perception of the faces among anxious ones was observed regardless of emotion being expressed (i.e., happy and angry). Although this appears to be in disagreement with a previous study showing that anxious individuals interpret facial expressions in a negative manner [1], a closer look yields a different picture. While the previous study [1] used morphed facial expressions conveying different emotions (e.g., fear and surprise), the current study used faces conveying single emotions (i.e., happiness or anger). It may be that anxious individuals show negative interpretation of others’ emotional states especially when a combination of more than one emotion are conveyed in a single modality, as in a previous study [1], or in multi-modalities as in the current study. In line with this possibility, when the effect of anxiety was examined with the faces that convey either one of six basic emotions, no effect of anxiety was found [15]. Of course, the differences in findings across the studies could be also due to other factors such as the target emotions (i.e., fear vs. anger) and the visibility of faces (i.e., with or without noise). Whether these factors interact with the effect of anxiety on the facial expression perceptions is an interesting topic to examine in the future study.

Of note, it is unlikely that inaccurate perception of the facial expressions among anxious individuals would account for their performance in the multi-modal tasks. While anxious individuals were less accurate at perceiving both happy and angry faces in the single-modal task, they were more likely to be influenced by the to-be-ignored negative voice cue, but not positive one, in the multi-modal tasks. If inaccuracy of facial expression perception were to increase the influence of the to-be-ignored voice on the face perception, one would expect that anxious individuals would show increased influence from both happy and angry voices. This was not the case in our study. Thus, anxiety may exert its influence on the integration of emotion in the face and voice in a different manner from that on the single modal emotional processing.

In summary, our findings suggest that a biased processing of others’ emotional states among anxious individuals [1] prevails across modalities. Previously, the effects of anxiety on the facial expression processing have been repeatedly studied. Yet, the effect of anxiety on the emotional voice perception had remained largely unknown. The present study showed that the trait anxiety modulates the integration of emotional processing in the face and voice. In the naturalistic environment, we often perceive others’ emotional states in their face and voice, not separately but in combination. Therefore, the findings of current study could reflect how we vary in the way we perceive others’ emotion in real life, as a function of our anxiety level.

5. Acknowledgements

A.K. was supported by the Japanese Society for the Promotion of Science Research Fellowship for Young Scientists (20-11352). A.T. was supported by a Grant-in-Aid for Specially Promoted Research No. 19001004 from the Ministry of Education, Culture, Sports, Science and Technology, Japan and the Postdoctoral Fellowships for Research Abroad from the Japan Society for the Promotion of Science.

6. References


