Exp Brain Res
DOI 10.1007/s00221-011-2668-1

RESEARCH ARTICLE

The effects of anxiety on the interpretation of emotion in the face–voice pairs

Ai Koizumi · Akihiro Tanaka · Hisato Imai · Saori Hiramatsu · Eriko Hiramoto · Takao Sato · Beatrice de Gelder

Received: 17 October 2010 / Accepted: 28 March 2011

Abstract Anxious individuals have been shown to interpret others' emotional states negatively. Since most studies have used facial expressions as emotional cues, we examined whether trait anxiety affects the recognition of emotion in a dynamic face and voice that were presented in synchrony. The face and voice cues conveyed either matched (e.g., happy face and voice) or mismatched emotions (e.g., happy face and angry voice). Participants with high or low trait anxiety were to indicate the perceived emotion using one of the cues while ignoring the other. The results showed that individuals with high trait anxiety were more likely to interpret others' emotions in a negative manner, putting more weight on the to-be-ignored angry cues. This interpretation bias was found regardless of the cue modality (i.e., face or voice). Since trait anxiety did not affect recognition of the face or voice cues presented in isolation, this interpretation bias appears to reflect an altered integration of the face and voice cues among anxious individuals.

A. Koizumi (✉) · T. Sato
Department of Psychology, The University of Tokyo,
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
e-mail: bellkoizumi@gmail.com;
ll087023@mail.ecc.u-tokyo.ac.jp

A. Tanaka
Waseda Institute for Advanced Study,
Waseda University, Tokyo, Japan

H. Imai · S. Hiramatsu · E. Hiramoto
Department of Psychology,
Tokyo Woman's Christian University, Tokyo, Japan

A. Tanaka · B. de Gelder
Department of Medical Psychology and Neuropsychology,
Tilburg University, Tilburg, The Netherlands

Keywords Anxiety · Emotion · Face · Voice · Multisensory

Introduction

Recognition of others' emotional states is central to social communication. Yet, how people perceive others' emotions appears to vary across individuals and may depend on their personality traits. Previous studies have shown that individuals with heightened anxiety trait have a bias to interpret others' emotional states in a negative manner (Richards et al. 2002). When a face conveys mixed emotions (e.g., fear and surprise), anxious individuals tend to recognize this as negative (Richards et al. 2002). Similarly, clinically and sub-clinically anxious individuals tend to make negative interpretations of neutral faces that are ambiguous in their emotional messages (Mohlman et al. 2007; Yoon and Zinbarg 2008). Thus, when others express ambiguous emotions, allowing for multiple interpretations of their emotional states (e.g., angry or happy), anxious individuals tend to choose a negative interpretation over other possible ones (c.f., Bishop 2007).

While many studies examined the bias of anxious individuals in the recognition of facial expressions, few studies have examined the effect of anxiety on the recognition of emotional prosodies. One pilot study (Quadflieg et al. 2007) showed that clinically anxious individuals recognize negative tones of voice more easily and positive tones less easily. A few other studies have found no effect of heightened anxiety on the recognition of emotional prosody, although anxiety was shown to enhance neuronal responses in the area involved in the evaluation of negative prosody stimuli (i.e., orbito-frontal cortex; Quadflieg et al. 2008). Thus, anxiety trait may enhance the interpretation

Published online: 13 April 2011

and/or processing of the cues conveying negative emotions, not only in the visual modality (i.e., face), but also in the less investigated auditory modality (i.e., voice). While the effect of anxiety traits on the interpretation of emotional voices still needs further examination, studies on the interpretation of multisensory emotional cues are even scarcer. This is surprising considering that we often express emotions or perceive others' emotions not in one modality alone but in combinations of different modalities. A bias in the interpretation of multisensory cues, if found, could genuinely reflect how anxiety may have an impact on daily social life.

Moreover, examination of a biased interpretation of multisensory emotional cues could elucidate a novel aspect of the effect of anxiety that cannot be inferred from studies on single sensory modality processing. Accumulating evidence suggests that emotions expressed through multisensory cues are processed interdependently of each other. Investigations have showed that the perception of emotion in faces is affected by the accompanying voices, and vice versa (Collignon et al. 2008; de Gelder and Vroomen 2000). For example, people are less likely to perceive sad facial expressions as sad in the presence of a happy voice (de Gelder and Vroomen 2000). The interdependent nature of multisensory emotional cue processing has been demonstrated even with voluntary control in ignoring the cue from either modality (de Gelder and Vroomen 2000; Experiment 2 and 3) or with no conscious awareness in the visual modality (de Gelder et al. 2002). Thus, emotion expressed in the face and voice appears to be integrated in an automatic fashion. Recent studies have advanced our understanding of the neuronal mechanisms underlying the multisensory processing of emotional cues. Integration of the face and voice not only involves a circuit for multisensory integration, such as the superior temporal gyrus, but also involves altered processing in the modality-specific regions (Kreifelts et al. 2007). Since multisensory emotional processing has a unique aspect, the effect of trait anxiety on multisensory emotional cue processing needs to be examined directly, rather than indirectly inferred from studies targeting either facial expressions or emotional prosodies alone.

The current study examines the effect of trait anxiety, measured with the Spielberger Trait-State Anxiety Inventory (Spielberger et al. 1970), on the interpretation of others' emotional states involving simultaneously presented multisensory cues (i.e., face and voice). As mentioned earlier, previous studies of the recognition of facial expressions suggest that anxious individuals have a preference for negative interpretations when conflicting interpretations are available (e.g., Richards et al. 2002). The current study particularly focused on whether anxious individuals would show a negative interpretation bias even when ambiguity arises from the conflict in the emotions conveyed from the cues in two different modalities (e.g., happy face and angry voice). To our knowledge, this is the first study to examine the effect of trait anxiety on the interpretation of others' emotions using multisensory cues. One recent study reported that individuals with heightened levels of anxiety and depression showed enhanced neuronal activities for emotional face–voice pairs (e.g., sad face and sad voice) compared with neutral pairs (Campanella et al. 2010). Although their results suggest the presence of an effect of anxiety on multisensory emotional processing, the effect on recognition of multisensory cues remains to be examined. Whether anxiety affects recognition of multisensory emotional cues is particularly important in understanding how heightened anxiety affects, or even hinders, people's social lives.

In Experiment 1, affective prosodies were presented in synchrony with dynamic facial expressions (Tanaka et al. 2010), resembling a naturalistic social situation. We used the immediate cross-modal bias paradigm (Bertelson and de Gelder 2004) that included conditions in which the face and voice conveyed congruent or incongruent emotions (i.e., happiness and anger). As in earlier studies (de Gelder and Vroomen 2000, Experiment 2 and 3; Tanaka et al. 2010), one of the modalities was to-be-attended and the other was to-be-ignored. We were particularly interested in how anxious individuals would interpret emotional cues in one modality (e.g., happy face) in the presence of incongruent emotional cues in another modality (e.g., angry voice). This methodology enabled us to examine the effect of trait anxiety on the weighting of the cues based on their valence and modality when judging others' emotions. In Experiment 2, dynamic facial expressions and affective prosodies were presented separately, in order to examine the effect of trait anxiety on single sensory modality processing. This allows us to evaluate whether the effect of trait anxiety on the integration of the face–voice cue is merely due to the altered perception of the face or voice cues presented in isolation.

Experiment 1

Method

Participants

Japanese undergraduate or graduate students ($N = 39$; 23 males) were recruited by word-of-mouth to participate in this study. Their trait anxiety levels were measured with a Japanese translation (Shimizu and Imae 1981) of the Spielberger State-Trait Anxiety Inventory (Spielberger et al. 1970) following the experimental tasks. They were divided into low- ($n = 22$) and high-anxious groups ($n = 17$) based on their median level of trait anxiety for the purposes of the data analyses. The high anxious group ($M = 52.41$, $SD = 7.91$) had a higher trait anxiety level than the low-anxious group ($M = 37.77$, $SD = 4.97$) ($t(37) = 7.07$, $P < .001$).

Stimuli

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

The stimuli were prepared from the recorded materials so that the emotions (angry or happy) in the face and voice cues were congruent in half the stimuli and incongruent in the other half. In congruent face and voice cues were combined with their asynchrony ranging from 0 to 80 ms ($M = 27.3$ ms), which was smaller than a detectable audiovisual asynchrony in speech (Munhall et al. 1996; van Wassenhove et al. 2007). Thus, it was ensured that participants could not base their responses on subtle, undetectable asynchrony in the stimuli. A total of 32 stimuli were prepared (2 models × 2 emotions × 2 congruency × 4 sentences). Since a preliminary study showed that the accuracy of perceiving emotion in the face was too high ($M = 98.0\%$) to be modulated by voice cues, the visibility of the face was reduced by superimposing random noise (Tanaka et al. 2010).

Procedure

In the multisensory tasks, participants were presented with the synchronized face and voice cues. The face was to be attended in the multisensory face task and the voice in the multisensory 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 s) 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 not to base their responses on them. Responses were made by a key-press and were not speeded.

Results

The overall accuracy for correctly perceiving the emotion in the voice task ($M = 92.91\%$) was higher than in the face task ($M = 72.40\%$; $t(38) = 19.33$, $P < .001$). The mean accuracies for the congruent and incongruent trials in the voice and face tasks among the low- and high-anxious groups are summarized in Table 1.

Importantly, further analysis revealed the effects of anxiety on the interpretation of multisensory emotional cues. In order to examine the effects of to-be-ignored emotional cues on the perception of emotion for the attended cues (face or voice), the accuracy on the congruent trials was subtracted from the incongruent trials (Fig. 1a, b). Larger calculated differences in accuracy (i.e., congruency effect) would indicate that the perception of emotion for the attended cues was more likely to be affected by the to-be-ignored incongruent emotional cues.

The congruency effects were analyzed using an analysis of variance (ANOVA) with anxiety (low/high) as a between-participants 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 ($F(1, 37) = 60.13$, $P < .001$), showing a larger overall effect of voice cues in the face task than of face cues in the voice task. More important to our research interest, the interaction between anxiety and emotion was significant ($F(1, 37) = 7.69$, $P < .01$).

Multiple comparisons (Bonferroni method) showed that the high anxious group was affected by the angry cues significantly more than by the happy cues ($P < .05$). In contrast, the low-anxious group showed a non-significant trend toward a larger effect of the to-be-ignored happy cues than of the angry cues ($P < .10$). The interaction between

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 multisensory tasks

Target	Congruency	Anxiety level	
		Low	High
Face			
Angry	Congruent	80.39(8.94)	78.62(14.91)
	Incongruent	52.56(9.63)	56.58(9.66)
Happy	Congruent	86.93(5.99)	89.14(4.99)
	Incongruent	69.32(6.76)	60.86(9.65)
Voice			
Angry	Congruent	94.03(15.82)	88.97(13.66)
	Incongruent	93.18(25.64)	90.44(16.25)
Happy	Congruent	94.89(10.19)	96.69(6.90)
	Incongruent	92.90(17.13)	91.18(19.50)

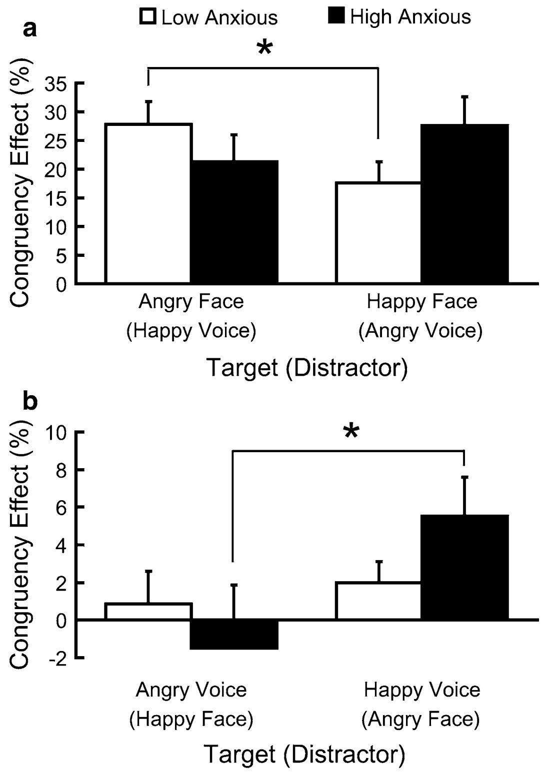

Fig. 1 The difference in the congruency effect (%) between the low- and high-anxious group for the multisensory face task (**a**) and the voice task (**b**). *Error bars* represent the standard error of the means. *$P < .05$

anxiety and emotion was not modulated by task ($F(1, 37) = 2.11$, $P = .15$), indicating that the high anxious group interpreted others' emotional states by prioritizing the to-be-ignored angry cues regardless of their modality (face or voice).

Although there was no three-way interaction among anxiety, task, and emotion, we further conducted multiple comparisons to elucidate a more detailed picture of the effect of anxiety in the face and voice tasks. In the face task, while the low-anxious group was less likely to be affected by the to-be-ignored angry voice cues than by the happy voice cues ($P = .02$), the high anxious group did not show such a reduced effect of angry voice cues ($P = .20$). In the voice task, while the low-anxious group was affected by angry and happy face cues to a similar extent ($P = .69$), the high anxious group was more likely to be affected by the to-be-ignored angry face cues than the happy face cues ($P = .04$). Thus, the high anxious group was characterized by lesser attenuation (i.e., relative enhancement) of the effect of the to-be-ignored angry voices, as well as by a relative enhancement of the effect of the to-be-ignored angry faces. However, the lack of a significant interaction among anxiety, task, and emotion limits any conclusions based on these multiple comparisons. Since the overall frequency of angry responses between the low ($M = 47.02\%$, $SD = 6.08\%$) and high ($M = 47.43\%$, $SD = 4.09\%$) anxious groups was equivalent ($t(37) = .24$, $P = .81$), the above findings cannot be accounted for by a response bias in the low- or high-anxious group.

Discussion

The aim of Experiment 1 was to examine the effect of trait anxiety on the interpretation of emotional cues (e.g., happy voice) that were simultaneously presented with the cues in a different modality (e.g., angry face). The results showed that, when recognizing emotions in the cues from the attended modality, anxious individuals were more likely to be affected by the angry cues than by the happy cues that were presented in the to-be-ignored modality. Therefore, higher trait anxiety appears to affect the multisensory emotional cue processing by putting more weight on the to-be-ignored negative cues, irrespective of their modalities.

As mentioned earlier, previous studies have shown that anxious individuals tend to interpret ambiguous facial expressions as negative (e.g., Richards et al. 2002; Yoon and Zinbarg 2008). The current results add to these previous findings by suggesting that higher anxiety prioritizes the processing of negative cues even when they are presented in the ignored modality. However, the previously demonstrated effects of anxiety on the recognition of only face cues (Yoon and Zinbarg 2008) or only voice cues (Quadflieg et al. 2007) suggest another possible interpretation. High anxious individuals might have shown an enhanced effect of the to-be-ignored angry cues not because they integrated them differently than do low-anxious individuals, but simply because they perceived them more negatively. To examine this possibility, we conducted Experiment 2 where high and low-anxious individuals were compared in their recognition of the face or voice cues presented in isolation.

Experiment 2

Method

The same participants from Experiment 1 took part in Experiment 2. The stimuli were identical to Experiment 1, except that a cue from only one of the two modalities was presented at a time. The participants were presented with the face or voice cues in separate blocks. The order of the blocks was counterbalanced. There were 32 trials in each block, which were presented in a randomized order. The participants indicated the emotion they perceived in the presented cues in the same manner as in Experiment 1.

Results

In the single sensory tasks, the means for the overall accuracy of identifying the emotion were 81.41% for the face cues and 92.26% for the voice cues. The overall accuracy was higher in the voice task than in the face task ($t(38) = 5.36$, $P < .01$).

The accuracy in each task was separately analyzed with a two-way ANOVA with anxiety (low/high) as a between-participants factor and emotion (angry/happy) expressed in the cues as a within-participant factor. In the face task (Fig. 2a), the main effect of emotion was significant ($F(1,37) = 13.68$, $P = .001$), with higher accuracy for the perception of happy faces ($M = 87.18$, $SD = 12.98$) than angry faces ($M = 75.64$, $SD = 18.79$). Importantly, the effect of emotion was not modulated by anxiety ($F(1,37) = 1.03$, $P = .32$). The main effect of anxiety was also non-significant ($F(1,37) = .46$, $P = .50$). In the voice task (Fig. 2b), there were no significant main effects or interactions ($ps > .05$). Therefore, anxiety did not modulate the accuracy of identifying emotions in the single sensory face and voice tasks.

Discussion

Experiment 2 examined the effect of trait anxiety on the recognition of emotions in the separately presented face cues or voice cues, which had been presented together in Experiment 1. Overall, the happy face was more easily recognized than the angry face, replicating the findings of earlier studies (e.g., de Gelder et al. 1998). The enhanced recognition of happiness in the face may also be explained by the use of female faces in the current study, since happiness is more easily recognized in female faces than in male faces (Becker et al. 2007). The happy voice was recognized no more accurately than the angry voice, which is also consistent with a previous report (Vroomen et al. 1993). Important to our purpose, the results showed that high anxious individuals did not differ from low-anxious individuals in recognition of the face and voice cues presented in isolation.

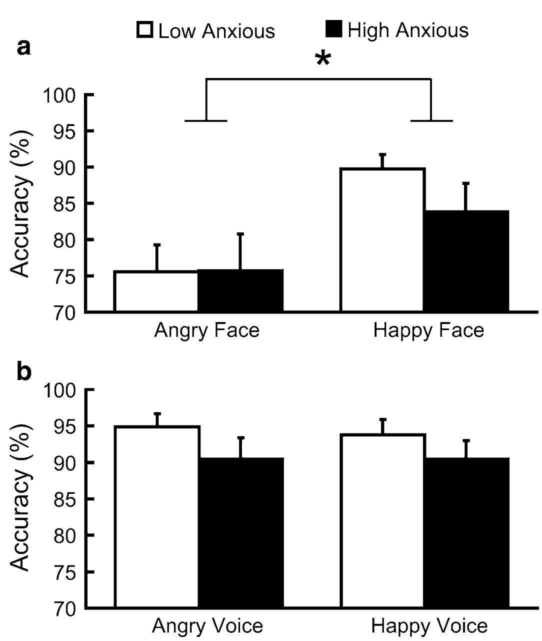

Fig. 2 Accuracy (%) in the single sensory face task (**a**) and the voice task (**b**). *Error bars* represent the standard errors of the means. $*P < .05$

Although some previous studies have found an effect of anxiety on the recognition of single sensory emotional cues, the difference between the earlier studies and the present study appears to arise from the choice of clinical versus non-clinical anxiety samples. Although clinically anxious individuals were shown to perceive angry faces more accurately than the controls (Mohlman et al. 2007), individuals grouped as “high anxious” according to their trait anxiety level, as in the current study, showed no such enhancement in negative face recognition (Cooper et al. 2008). For voice recognition, Quadflieg et al. (2007) found hindered recognition of a happy voice among clinically anxious individuals. Since high trait anxiety individuals showed no such effect for the happy voice recognition in our present study, the recognition of voice cues, as well as face cues, presented in isolation may be modulated only by clinical levels of anxiety.

Although the effect of anxiety on the recognition of single sensory emotional cues is an interesting topic by itself, our main interest was to test whether the enhanced effect of the angry cues in the integration of the face and voice cues, observed among high trait anxiety individuals in Experiment 1, was merely due to an altered perception of the single sensory cues. Regarding this point, the lack of an effect of anxiety on the perception of the single sensory cues ruled out this possibility. Thus, trait anxiety appears to exert its influence particularly in the course of the integration of multisensory emotional cues. We further discuss the possible mechanisms underlying the effect of anxiety on the multisensory emotional processing in the “General discussion”.

General discussion

The current study (Experiment 1) examined the effect of trait anxiety on the recognition of emotion in the face and voice which were presented in synchrony. The face and voice cues conveyed either congruent or incongruent emotions (e.g., happy face and angry voice), and participants were instructed to recognize the emotion in one of the cues while ignoring the other. We showed that, regardless of the cue modality (i.e., face or voice), participants with heightened trait anxiety levels were more likely to be affected by the to-be-ignored angry cues than the happy cues when judging emotions in the attended cues. In contrast, low trait anxiety individuals were more likely to be affected by the to-be-ignored happy cues than the angry ones, although this trend missed statistical significance. Thus, when judging others' emotions, more anxious participants, unlike less anxious participants, were unable to disregard the negative emotional cues even when they were presented in a to-be-ignored modality. It has been shown in other studies that emotional cues in one modality (e.g., face) affect the recognition of cues presented in a different modality (e.g., voice) even with instructions to ignore one modality of the cues (de Gelder and Vroomen 2000; Tanaka et al. 2010). Therefore, although parallel processing of the face and voice cues appears to affect each other in general, trait anxiety appears to have an additional influence on multisensory integration processing by prioritizing negative information. Previous studies have suggested that anxious individuals prefer negative interpretations over other possible ones when facial expressions convey conflicting information (e.g., Richards et al. 2002; Yoon and Zinbarg 2008). The current study showed for the first time that individuals with higher levels of trait anxiety show a negative interpretation bias even when the cues from the two different modalities (face and voice) convey conflicting and thus ambiguous emotions.

Contrasts between the performance of high trait anxiety individuals in the multisensory tasks (Experiment 1) and in the single sensory tasks (Experiment 2) suggest that trait anxiety particularly affects the course of multisensory integration of the emotional cues. First, despite the finding that recognition of the face and voice cues was not affected by the trait anxiety level when they were processed in isolation (Experiment 2), high trait anxiety individuals, compared with low-anxiety individuals, were more likely to be affected by the to-be-ignored angry cues when processing the face and voice cues in parallel (Experiment 1). Thus, trait anxiety appears to affect the extent to which the angry cues that are presented outside of the attended modality are processed or attenuated. Second, high trait anxiety individuals showed a larger effect of the to-be-ignored angry face than of the happy face in Experiment 1, despite the finding that they recognized the angry faces less accurately than the happy faces in Experiment 2. Thus, even though the angry faces were not as recognizable as the happy faces, the high trait anxiety individuals put extra weight on these less reliable cues when interpreting others' emotions in the voices. Taken together, when interpreting others' emotions from the multisensory cues, high trait anxiety appears to prioritize the processing of angry cues even when these cues are delivered outside of the focused modality and they are less recognizable cues for others' emotional states when presented alone.

One plausible mechanism underlying the prioritization of angry cues among higher trait anxiety individuals is their reduced inhibitory control over negative information processing. Previous reports about the negative interpretation bias among anxious individuals have been explained as their reduced inhibitory control over the to-be-ignored negative information (Bishop 2007; Mathews and MacLeod 1998). At the neuronal level, the reduced inhibitory control among anxious individuals has been related to their hindered prefrontal activity, including the anterior cingulate cortex (Bishop et al. 2004). The anterior cingulate cortex is activated in response to incongruency in emotional information processed in parallel (e.g., negative face and written positive word; Etkin et al. 2006; Haas et al. 2007), and it attenuates the processing of to-be-ignored emotional information (Egner et al. 2008; Etkin et al. 2006). Although these studies have targeted the processing of visually presented information, the anterior cingulated cortex has also been shown to respond to conflicts of emotions between the face and voice cues (e.g., fearful face and happy voice; Pourtois et al. 2002). Thus, hindered processing of emotional conflicts in a common neuronal region (i.e., the anterior cingulated cortex) may explain the reduced control among anxious individuals over the to-be-ignored negative cues irrespective of their modality.

Yet, as repeatedly mentioned in this manuscript and many others, the parallel processing of emotional cues in the face and voice engages wider neuronal activities than the processing alone of either modality cues (e.g., Kreifelts et al. 2007). The simultaneous processing of face and voice cues has been shown to modulate many stages of processing including visual processing (Ethofer et al. 2006; Kreifelts et al. 2007), auditory processing (Kreifelts et al. 2007), as well as prefrontal control activity (Pourtois et al. 2002). Therefore, mechanisms underlying the effect of anxiety on the interpretation of face and voice cues should be directly tested, rather than indirectly inferred from studies targeting either one of these two modalities.

The current findings that the negative interpretation bias among higher trait anxiety individuals prevails across modalities have implications for future studies. Besides the bias in the interpretation of others' emotions, as in our study, anxiety has been also shown to affect many stages of emotional information processing such as attention and working memory (see Eysenck et al. 2007, for a review). Thus, future research should examine whether anxiety-related bias in different processing stages can be observed across modalities.

Another important issue that remains to be examined is whether the addition of noise to the face cues but not to the voice cues in the current study may have affected the findings. This asymmetrical manipulation of noise between the cue modalities may explain some of the current findings. First, it may explain the more accurate recognition of emotions in the voice than the face cues. Second, it could also explain the larger effect of to-be-ignored voice cues on face emotion recognition than the effect of face cues on voice emotion recognition, which appears to reflect the law of inverse effectiveness in multimodal integration (Meredith and Stein 1983). Yet, more importantly, the asymmetrical manipulation of noise between the face and voice cues cannot explain the main finding in our study that high trait anxiety individuals are more likely to be affected by the angry cues irrespective of their modality. Although the asymmetrical manipulation of noise may have made the voice cues more reliable than the face cues, more anxious individuals prioritized the angry cues irrespective of their modality. This could support the persistent nature of anxious individuals to prioritize negative information. Yet, it is worth examining whether the current findings would hold up when adding noise to voices as well as faces.

In summary, our findings suggest that a biased processing of others' emotional states among anxious individuals prevails across modalities. The effects of anxiety on the processing of facial expressions have been repeatedly studied. Yet, the effect of anxiety on multisensory emotional processing has remained largely unknown. The present study shows that trait anxiety modulates the interpretation of emotions in face and voice cues in a valence-dependent manner as well as a modality-independent manner. In a naturalistic environment, people often perceive others' emotional states in their face and voice, not separately but in combination. Therefore, the findings of the current study may reflect how people vary in the way they perceive others' emotion in real life, as a function of their trait anxiety level.

Acknowledgments 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, the Postdoctoral Fellowships for Research Abroad from the Japan Society for the Promotion of Science, and a SCOPE Grant for Young ICT Researchers No. 102103011 from the MIC, Japan.

References

Becker DV, Kenrick DT, Neuberg SL, Blackwell KC, Smith DM (2007) The confounded nature of angry men and happy women. J Per Soc Psychol 92:179–190

Bertelson P, de Gelder B (2004) The psychology of multimodal perception. In: Spence C, Driver J (eds) Crossmodal space and crossmodal attention. Oxford University Press, Oxford, pp 151–177

Bishop SJ (2007) Neurocognitive mechanisms of anxiety: an integrative account. Trends Cogn Sci 11:307–316

Bishop S, Duncan J, Brett M, Lawrence AD (2004) Prefrontal cortical function and anxiety: controlling attention to threat-related stimuli. Nat Neurosci 7:184–188

Campanella S, Bruyer R, Froidbise S, Rossignol M, Joassin F, Kornreich C, Noel X, Verbanck P (2010) Is two better than one? A cross-modal oddball paradigm reveals greater sensitivity of the P300 to emotional face-voice associations. Clin Neurophysiol 121:1855–1862

Collignon O, Girard S, Gosselin F, Roy S, Saint-Amour D, Lassonde M, Lepore F (2008) Audio-visual integration of emotion expression. Brain Res 1242:126–135

Cooper RM, Rowe AC, Penton-Voak IS (2008) The role of trait anxiety in the recognition of emotional facial expressions. J Anxiety Disord 22:1120–1127

de Gelder B, Vroomen J (2000) The perception of emotions by ear and by eye. Cogn Emot 14:289–311

de Gelder B, Vroomen J, Bertelson P (1998) Upright but not inverted faces modify the perception of emotion in the voice. Cahiers de Psychologie Cognitive 17:1021–1032

de Gelder B, Pourtois G, Weiskrantz L (2002) Fear recognition in the voice is modulated by unconsciously recognized facial expressions but not by unconsciously recognized affective pictures. Proc Natl Acad Sci USA 99:4121–4126

Egner T, Etkin A, Gale S, Hirsh J (2008) Dissociable neural systems resolve conflict from emotional versus nonemotional distracters. Cereb Cortex 18:1475–1484

Ethofer T, Anders S, Erb M, Droll C, Royen L, Saur R, Reiterer S, Grodd W, Wildgruber D (2006) Impact of voice on emotional judgment of faces: an event-related fMRI study. Hum Brain Mapp 27:707–714

Etkin A, Egner T, Peraza DM, Kandel ER, Hirsch J (2006) Resolving emotional conflict: a role for the rostral anterior cingulate cortex in modulating activity in the amygdala. Neuron 51:871–882

Eysenck MW, Derakshan N, Santos R, Calvo MG (2007) Anxiety and cognitive performance: attentional control theory. Emotion 7:336–353

Haas B, Omura K, Constable RT, Canli T (2007) Emotional conflict and neuroticism: Personality-dependent activation in the amygdala and subgenual anterior cingulated. Behav Neurosci 121:249–256

Kreifelts B, Ethofer T, Grodd W, Erb M, Wildgruber D (2007) Audiovisual integration of emotional signals in voice and face: An event-related fMRI study. NeuroImage 37:1445–1456

Mathews A, MacLeod C (1998) A cognitive model of selective processing in anxiety. Cogn Ther Res 22:539–560

Meredith MA, Stein BE (1983) Interactions among converging sensory inputs in the superior colliculus. Science 221:389–391

Mohlman J, Carmin CN, Price RB (2007) Jumping to interpretations. Social anxiety disorder and the identification of emotional facial expressions. Behav Res Ther 45:591–599

Munhall KG, Gribble P, Sacco L, Ward M (1996) Temporal constraints on the McGurk Effect. Percept Psychophys 58:351–362

Pourtois G, Debatisse D, Despland PA, de Gelder B (2002) Facial expressions modulate the time course of long latency auditory brain potentials. Brain Res Cogn Brain Res 14:99–105

Quadflieg S, Wendt B, Mohr A, Miltner WHR, Straube T (2007) Recognition and evaluation of emotional prosody in individuals with generalized social phobia: A pilot study. Behav Res Ther 45:3096–3103

Quadflieg S, Mohr A, Mentzel HJ, Miltner WHR, Straube T (2008) Modulation of the neural network involved in the processing of angry prosody: The role of task-relevance and social phobia. Biol Psychol 78:129–137

Richards A, French CC, Calder AJ, Webb B, Fox R, Young AW (2002) Anxiety-related bias in the classification of emotionally ambiguous facial expressions. Emotion 2:273–287

Shimizu H, Imae K (1981) State-Trait Anxiety Inventory no nihongoban (daigakuseiyo) no sakusei [Development of the Japanese edition of the Spielberger State-Trait Anxiety Inventory (STAI) for student use]. Japanese Journal of Educational Psychology 29:62–67

Spielberger CD, Gorsuch RL, Lushene RE (1970) Manual for the state-trait anxiety inventory. Consulting Psychologists Press, Palo Alto

Tanaka A, Koizumi A, Imai H, Hiramatsu S, Hiramoto E, de Gelder B (2010) I feel your voice: Cultural differences in the multisensory perception of emotion. Psychol Sci 21:1259–1262

van Wassenhove V, Grant KW, Poeppel D (2007) Temporal window of integration in bimodal speech. Neuropsychologia 45:598–607

Vroomen J, Collier R, Mozziconacci S (1993) Duration and intonation in emotional speech. Proceedings of the Third European Conference on Speech Communication and Technology, Berlin 577–580

Yoon KL, Zinbarg RE (2008) Interpreting neutral faces as threatening is a default mode for socially anxious individuals. J Abnorm Psychl 117:680–685