Multisensory Perception of Emotion for Human and Chimpanzee Expressions by Humans

Marina Kawase[1,3], *Ikuma Adachi*[2], *Akihiro Tanaka*[3]

[1]Department of Behavioral Science, Graduate School of Letters, Hokkaido University, Japan
[2]Primate Research Institute, Kyoto University, Japan
[3]Department of Psychology, Tokyo Woman's Christian University, Japan
mkawase@eis.hokudai.ac.jp

Abstract

We examined how we human perceive multimodal affective expressions in chimpanzees and whether the underlying cognitive systems are similar to those for human expressions. In the experiment, we presented audiovisual stimuli in which face and voice express congruent or incongruent emotions. Participants were instructed to ignore vocal emotion but to judge facial emotion. The results showed that in the chimpanzee stimuli, accuracy was marginally lower in the incongruent stimuli than in the congruent stimuli. Moreover, the congruency effect on positive human faces was marginally stronger than on positive chimpanzee faces. The gaze behavior results showed that human participants focused around the eye area when presented with positive and negative facial expressions of humans and negative facial expression of chimpanzees, while they focused both around the eye and the mouth areas when presented with positive facial expressions of chimpanzees. Our findings suggest that humans perceive affective expressions of other species multisensorily, and that underlying cognitive systems are not similar to those for humans.

Index Terms: multisensory perception, emotion perception, human-animal interaction

1. Introduction

We humans live in a relatively large and complex social group in animal kingdom. It is, therefore, important to recognize others' emotion and to change own behavior flexibly in order to coexist with others. Previous studies (e.g., [1-3]) have shown that emotions are perceived not from a single sensory modality but from multiple modalities (e.g., audition and vision). de Gelder and Vroomen [1] examined whether facial and vocal emotional information interact with each other. In their experiments, the stimuli were presented by combining facial and vocal expressions, which express either congruent or incongruent emotion. Participants were instructed to judge the emotion of either sensory modality and to ignore the other. The results showed that the accuracy of the incongruent condition was lower than that of the congruent and unisensory conditions, indicating that the emotion perception was influenced by to-be-ignored sensory modality. These results suggest that auditory and visual affective information interact with each other.

Multisensory emotion perception has been reported not only in humans but also in animals (e.g., [4,5]). Izumi and Kojima [4] examined whether chimpanzees possess cross-modal representations of affective expressions. In the experiment, a chimpanzee was presented with species-specific auditory expressions of conspecifics (e.g. pant-hoot, which is often observed when chimpanzees call other individuals in distance), which accompany certain meanings and emotional states. After that, two facial movies, which express congruent (e.g. pant-hoot) or incongruent (e.g. scream, which is often observed when chimpanzees are in fear, frustration, or aggression) emotions, were presented successively. Then the chimpanzee chose one of the movies. The results showed that a chimpanzee could match the vocal and facial expressions correctly except one of the three combinations.

Many studies have investigated multisensory perception of emotion in conspecies. Several studies examined multisensory perception of affective expressions of dogs and rhesus monkeys by human infants [6,7]. However, little is known about multisensory perception of heterospecies by human adults. It is important to investigate whether multisensory perception of emotions for conspecies and heterospecies use the same or different processes. In this study, we compared human multisensory perception of emotion between conspecies and heterospecies by using positive and negative expressions of humans and chimpanzees. We chose chimpanzees because they are phylogenetically closest to humans and the arrangement of facial muscles is similar between humans and chimpanzees. In addition to the accuracy and the congruency effects, we recorded the gaze behavior to facial expression by using an eye tracker. Analyses of the gaze behavior would enable to investigate whether diagnostic parts are shared between conspecies and heterospecies faces when judging facial emotion.

2. Material and Methods

2.1. Participants

The participants were 81 female university students in Japan (M = 20.37, SD = 1.36).

2.2. Stimuli

Chimpanzees' expressions were recorded at Primate Research Institute, Kyoto University. We recorded spontaneous and induced chimpanzees' negative expressions ("display" which is often observed when chimpanzees show own power, and "scream"), and induced chimpanzees' positive expression ("food call" which is often observed when chimpanzees find food). As chimpanzees have nonverval vocal expression, humans meaningless facial and vocal expressions were used as human stimuli. We selected facial and vocal expressions which we judged as "natural" from TV programs we recorded. Facial expressions were static pictures. We used angry expressions as negative stimuli, and laughter or positive surprise as positive stimuli. All expressions were emitted by Japanese. Based on the

10.21437/AVSP.2017-22

results of preliminary experiment, we finalized which facial and vocal expressions for the test. In the preliminary experiment, human participants rated emotional valence of facial and vocal affective expressions by 5-point scaling. We used 32 multisensory stimuli in total (4 samples × 2 emotions × 2 modalities × 2 species). The mean ratings were matched between facial and vocal expressions of the same emotions. In chimpanzee expressions, we tried to match the mean ratings of positive and negative emotions. However, the mean ratings of negative expressions were much different from the neutral while the difference was smaller in positive expressions. In human expressions, the mean ratings of male and female expressions were matched. The facial and vocal expressions were combined within conspecies. In addition, human expressions were combined within the same sex.

In each combination, ratings were closely matched between facial and vocal expressions in order to avoid the situation that more salient information dominates the other [3,8]. Consequently, each facial expression was not combined with a vocal expression of the same individuals.

In total, 16 multisensory stimuli (8 congruent and 8 incongruent emotions) were made from humans and chimpanzees. The duration of the movies depended on that of the vocal expressions. The facial expressions were presented for 2000 ms when the duration of the vocal expressions was less than 2000 ms.

2.3. Apparatus

Presentation of stimuli and recording of the gaze of participants were controlled by gaze measuring device (T60 Tobii Technology). The sounds were presented through an amplifier (ONKYO, DAC-HA200) and headphones (SENNHEISER, HDA 200).

2.4. Procedure

On average, it took 15 minutes for participants to complete all tasks. The experiment consisted of two blocks of multisensory session (humans and chimpanzees) and four blocks of unisensory session (presenting only facial or vocal stimuli of humans or chimpanzees). Multisensory sessions always preceded unisensory sessions in order not to use the memory cues in multisensory sessions, that can be learned in unisensory sessions. The orders of the human and chimpanzee blocks were counterbalanced across pariticpants. The stimuli in each block were presented in random order. In multisensory sessions, participants were instructed to judge facial expressions as either negative or positive and to ignore vocal emotion expressions. In each trial, stimuli were presented after 1000 ms fixation point. Participants reported their answers verbally. Participants were instructed to report stronger emotion if they received both negative and positive emotions simultaneously. We instructed the participants to judge what they perceived from the expressions, but not what they felt from the expressions.

3. Results

3.1. Accuracy

We calculated the mean accuracy in congruent and incongruent conditions in multisensory sessions for each participant in order to examine whether humans perceive humans' and chimpanzees' emotions multisensorily. We also calculated the mean accuracy in unisensory sessions in order to examine whether the accuracy was different between multisensory and unisensory perception. We conducted 2 (facial emotion: negative or positive) × 2 (congruency: congruent or incongruent) within subject factor analyses of variance (ANOVA) for human and chimpanzee blocks (Figure 1).

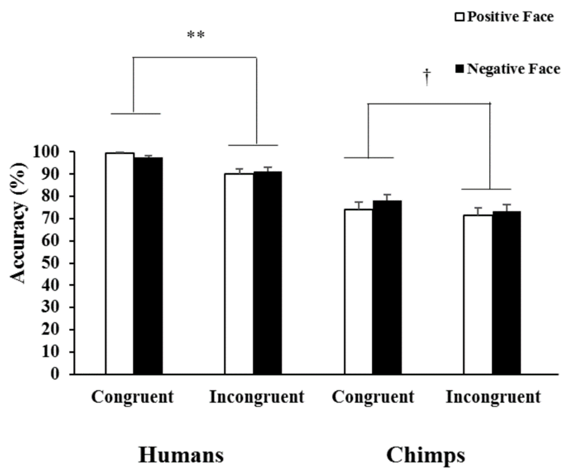

Figure 1 : *Mean accuracies in multisensory sessions as a function of species and congruency. Error bars represent standard errors. Asterisks and dadder indicate significant differences between groups (**$p < .01$, †$p < .10$).*

In human block, the main effect of congruency was significant ($F(1,80) = 19.04$, $p < .01$). The main effect of emotion ($F(1,80) = 0.12$, n.s.) and the interaction ($F(1,80) = 1.48$, n.s.) were not significant. In chimpanzee block, the main effect of congruency was marginally significant ($F(1,80) = 3.07$, $p < .10$). The main effect of emotion ($F(1,80) = 0.70$, n.s.) and the interaction ($F(1,80) = 0.32$, n.s.) were not significant. These results indicate that when facial expressions are incongruent with vocal expression, their judgement is less accurate than when they are congruent in humans' expressions, although the effect is only marginal in chimpanzees' expressions.

We conducted separate ANOVAs in order to examine the effects of congruency for each emotion. The results showed that the congruency effects were significant in both negative ($F(1,80) = 11.01$, $p < .01$) and positive ($F(1,80) = 16.19$, $p < .01$) facial expressions of humans. On the other hand, in chimpanzee expressions, the congruency effect was marginally significant in negative ($F(1,80) = 2.98$, $p < .10$), but not in positive ($F(1,80) = 0.83$, n.s.) facial expressions.

3.2. Congruency Effects

We calculated the congruency effect by subtracting incongruent accuracy from congruent accuracy, and conducted one sample *t*-tests in each emotion in order to examine whether the congruency effects are different between species (Figure 2). When positive facial expressions were combined with negative vocal expressions, the congruency effects is marginally larger in human than chimpanzee expressions ($t(80) = 1.75$, $p < .10$), indicating that human positive facial expressions were more influenced by negative vocal expressions.

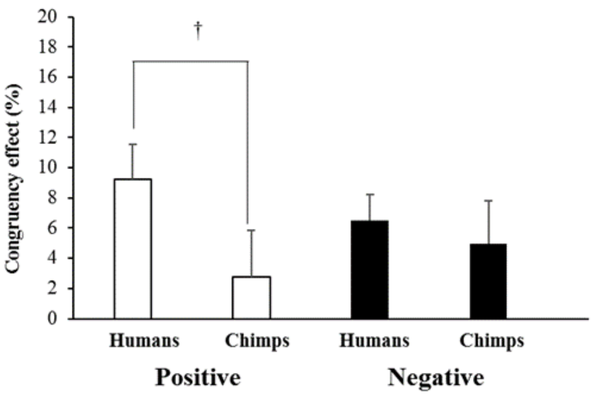

Figure 2 : *Face-voice congruency effects on positive and negative expressions of humans and chimps. Error bars represent standard errors. Dadder indicates significant differences between groups (†p < .10).*

3.3. Eye-tracking analyses

We calculated average gaze frequency in congruent and incongruent conditions in each facial emotion and each species (Figure 3).

We conducted 2 (species : humans or chimpanzees) × 2 (facial emotions : negative or positive) × 2 (congruency : congruent or incongruent) within subject factor analysis of variance (ANOVA). The results showed that species × emotions interaction was significant ($F(1,80) = 159.07$, $p < .01$). Simple main effect analyses showed that in chimpanzee facial expressions, humans gaze mouth area more frequently in positive than negative facial expressions ($F(1,80) = 195.02$, $p < .01$). On the other hand, in human facial expressions, the frequency was not different between facial emotions ($F(1,80) = 0.96$, n.s.).

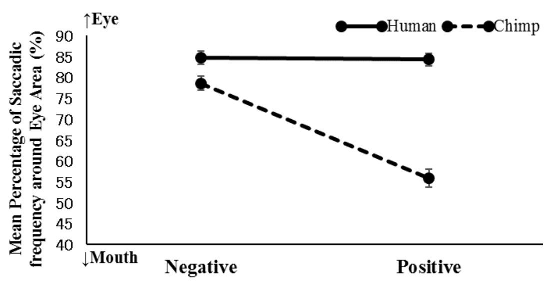

Figure 3 : *Saccadic frequency around the eye area when observing negative and positive faces of humans and chimps*

4. Discussion

In this study, we compared multisensory perception of emotion between humans' and chimpanzees' expressions by human participants. In chimpanzees' expression, the accuracy in incongruent condition was marginally lower than that of congruent condition. The results suggest that humans perceive emotions multisensorily for heterospecies although the effects were smaller than humans' expressions, and the tendency was observed only in negative facial expressions.

Positive facial expressions were not affected by negative vocal expressions. One of the possible factors which influenced the results in positive facial expressions is vocal duration. The vocal duration of negative expressions in chimpanzees was much longer than those of positive expressions. Therefore, participants observed positive facial expressions for a long time when they were combined with negative vocal expressions. If audiovisual integration of emotion occurs at an early stage just after stimulus presentation, there would be stronger influence by negative vocal expressions than observed in this study. However, because the vocal duration was longer, the effects may have been cancelled through time even if there had been a stronger influence of negative vocal expressions.

Our results showed that the effect of combining positive facial expressions with negative vocal expressions was larger in humans' than chimpanzees' expressions. This pattern is in line with findings in cross-cultural study. Tanaka and colleagues [9] reported that the congruency effects are different between in-group and out-group expressions. In their study, the congruency effects were larger for in-group than out-group stimuli when the positive facial expressions were combined with negative vocal expressions. The authors pointed out that Japanese people tend to hide their own negative expressions and express positive facial expression. However, it is difficult to hide negative vocal expressions. Consequently, Japanese people tend to perceive positive facial expressions as negative if vocal expressions are negative because they often experience such situation in daily life. Therefore, Japanese people learn to perceive negative emotion through experience when Japanese positive facial expressions are combined with negative vocal expressions even though they understand that they have to judge the facial expressions.

Smaller congruency effects in foreigners' expressions found in their study [9] parallel with our findings in chimpanzees' stimuli. Foreigners' are similar to chimpanzees in that they are unfamiliar. It could be that congruency effects were smaller because participants have insufficient contact experience for foreigners and chimpanzees. If this is true, we could predict that results for chimpanzees' expressions would be similar to those for humans' expressions with more contact experience with chimpanzees.

One of alternative interpretations is that the results were influenced by their knowledge or impression for chimpanzees. Our supplementary investigation showed that humans have impression that it is difficult for chimpanzees to hide their own emotions. It might be that humans feel strange when positive facial expressions are combined with negative vocal expressions of chimpanzees, and that humans do not perceive incongruent expressions of chimpanzees in the same way as those of humans. Facial expressions of chimpanzees were less influenced by vocal expressions because audiovisual integration was inhibited by the top-down impression that chimpanzees do not hide their own emotions.

Results from humans' expressions showed multisensory perception of emotions in facial and vocal expressions taken from realistic situations. Most previous studies [1-3] investigated multisensory perception of emotion by acting expressions. Our study investigated multisensory perception of emotion by using not acting but natural expressions, and demonstrated that humans perceive emotions multisensorily in our daily life.

Eye-tracking results showed no difference between negative and positive human facial expressions. On the other hand, human participants focused around the eye area when presented with negative facial expressions of chimpanzees, while they focused both around the eye and the mouth areas equally when presented with positive facial expressions of chimpanzees. These results suggest differential diagnostic areas for positive and negative facial expressions of chimpanzees.

In this study, we investigated multisensory perception of emotion in humans' and chimpanzees' expression. This is the first study on multisensory perception of emotion in humans and chimpanzees, and has many suggestions for the future study.

First, human facial expressions used in the unisensory session were very easy for human participants to perceive and caused the ceiling effect, since we used ecologically valid stimuli which appear in daily situation. We could not investigate whether the accuracy of the congruent multisensory expressions is higher than that of unisensory expressions. In future studies, it is important to manipulate the difficulty (e.g. by adding noises to the stimuli) in order to lower the accuracy of unisensory expressions and to examine the congruency effects in an equal condition.

Second, we employed only a limited number of species-specific expressions of chimpanzees (i.e., "display" and "scream" as negative and "food call" as positive). It would be interesting to examine whether our results apply to wider range of negative and positive emotions by including other species-specific expressions (e.g., laughter, whimper, etc.)

Third, we did not control the duration of negative and positive vocal expressions because we assumed that the duration of vocalizations is one of the important factors, which is essential for the species-specific expressions of chimpanzees. Future studies can examine the possibility that the congruency effects may change as a function of durations of vocal expressions.

Fourth, future studies can use other species' expressions in order to investigate whether our results also apply to any species. In this study, we used the expressions of chimpanzees which are phylogenetically closest to humans but yet not familiar to our participants. If they use expressions of heterospecies which are more familiar to humans and not phylogenetically close (e.g., dogs), the results might be different from our results. Thus, we could examine whether multisensory perception of emotion for heterospecies require familiarity or phylogenetically closeness to humans by using various species.

5. Acknowledgements

This work was supported by JSPS KAKENHI (No. 15H02714).

6. References

[1] de Gelder B. & Vroomen J. (2000) The perception of emotions by ear and by eye. cognition and emotion.14 (**3**), 289–311

[2] 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. *Psychological Science*, 21(**9**), 1259–1262.

[3] Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., & Lassonde, M. (2008) Audio-visual integration of emotion expression. *Brain Research*, **1242**, 126–135.

[4] Izumi A. & Kojima S. (2004) Matching vocalizations to vocalizing faces in a chimpanzee (Pan troglodytes) *Animal Cognition* **7** : 179–184

[5] Parr, L.A. (2004) Perceptual biases for multimodal cues in chimpanzee (Pan troglodytes) affect recognition. *Animal Cognition,* **7**, 171–178.

[6] Flom R, Whipple H, & Hyde D. (2009) Infants' intermodal perception of canine (Canis familairis) facial expressions and vocalizations. *Developmental Psychology*, **45**, 1143–51.

[7] Lewkowicz, D. J., & Ghazanfar, A. A. (2006) The decline of cross-species intersensory perception in human infants. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 6771– 6774.

[8] Ernst, M.O., & Bank, M.S. (2002) Human integrate visual and haptic information in a statistically optimal fashion. *Nature*, **415**, 429-433.

[9] Tanaka, A., Koizumi, A., Imai, H., Hiramatsu, S., Hiramoto, E., & de Gelder, B. (2010) Cross-Cultural Differences in the Multisensory Perception of Emotion. *Proceedings of the International Conference on Auditory-Visual Speech Processing pp*.49-53