Effect of species-specificity in auditory-visual intermodal matching in a chimpanzee (*Pan troglodytes*) and humans

L. Martinez*, T. Matsuzawa

Department of Brain and Behavioral Sciences, Primate Research Institute, Kyoto University, Kanrin 41-2, Inuyama, Aichi 484-8506, Japan

ABSTRACT

The goal of this study was to compare the performance of a chimpanzee and humans on auditory-visual intermodal matching of conspecifics and non-conspecifics. The task consisted of matching vocal samples to facial images of the corresponding vocalizers. We tested the chimpanzee and human subjects with both chimpanzee and human stimuli to assess the involvement of species-specificity in the recognition process. All subjects were highly familiar with the stimuli. The chimpanzee subject, named Pan, had had extensive previous experience in auditory-visual intermodal matching tasks. We found clear evidence of a species-specific effect: the chimpanzee and human subjects both performed better at recognizing conspecifics than non-conspecifics. Our results suggest that Pan's early exposure to human caretakers did not seem to favor a perceptual advantage in better discriminating familiar humans compared to familiar conspecifics. The results also showed that Pan's recognition of non-conspecifics did not significantly improve over the course of the experiment. In contrast, human subjects learned to better discriminate non-conspecific stimuli, suggesting that the processing of recognition might differ across species. Nevertheless, this comparative study demonstrates that species-specificity significantly affects intermodal individual recognition of highly familiar individuals in both chimpanzee and human subjects.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Within both the visual and auditory modalities, an advantage in recognizing conspecific individuals compared to non-conspecifics, the so called "species-specific effect", has been established in humans (Bruce and Young, 1986; Vouloumanos and Werker, 2007) and in non-human primates (Pascalis and Bachevalier, 1998; Pascalis et al., 1999). Furthermore, this conspecific superiority effect appears to vary with previous experience with conspecific and non-conspecific cues of individuality (Parr et al., 2006).

Chimpanzee studies focusing on visual cues have led to contradictory results. For example, a chimpanzee subject with prior experience in human face recognition was better at recognizing conspecific than human stimuli (Matsuzawa, 1990). However, other experiments showed that chimpanzees could visually discriminate humans as accurately as conspecifics (Boysen and Berntson, 1989; Parr et al., 1998). Even more, an advantage in discriminating human faces compared to conspecific faces was found in chimpanzee subjects who had had more experience with humans than with conspecifics (Martin-Malivel and Okada, 2007; Tanaka, 2003).

The way previous experience interacts with species-specificity in intermodal associations between auditory and visual cues is par-

ticularly interesting, as faces and voices support the earliest forms of social interactions and contain the most commonly used sensory cues for recognizing known individuals (Kuhl et al., 2003; Matsuzawa, 2006; Myowa-Yamakoshi et al., 2005). Human studies have suggested that prolonged exposure may importantly influence our ability to discriminate the faces or voices of non-conspecific individuals (Diamond and Carey, 1986).

Using preferential-looking measurement techniques, it has been shown that many primates possess the ability to recognize the correspondence between the auditory and visual components of conspecifics' calls (Ghazanfar and Logothetis, 2003). Moreover, non-human primates can also form a cross-modal representation of species other than conspecifics if exposed to that species from birth (Adachi and Fujita, 2007; Adachi et al., 2009; Evans et al., 2005). The way species-specificity affects auditory-visual intermodal discrimination has been poorly explored in chimpanzees, with only a few exceptions (Boysen, 1994; Hashiya, 1999; Hashiya and Kojima, 2001; Izumi and Kojima, 2004; Izumi, 2006; Parr, 2004).

The goal of this study was to assess the effect of species-specificity on the ability of a chimpanzee and humans to perform intermodal individual recognition of conspecifics and non-conspecifics. To address this issue, we tested a chimpanzee and human subjects in an auditory-visual intermodal matching task based on the recognition of familiar chimpanzee and human individuals.

* Corresponding author. Tel.: +81 568 630550; fax: +81 568 622428.
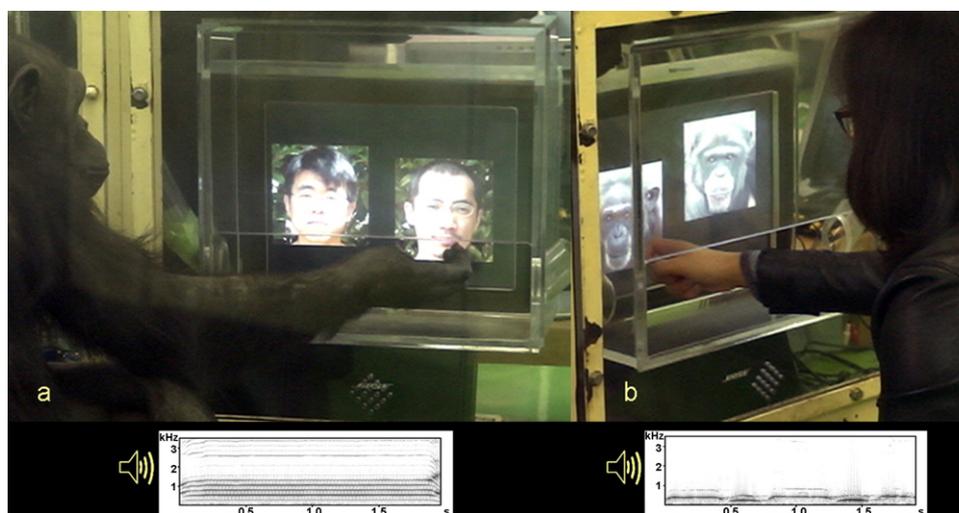E-mail address: laura@pri.kyoto-u.ac.jp (L. Martinez).

Fig. 1. Photographs of the apparatus and two of the subjects performing the AVIM task. (a) Chimpanzee subject tested in the Human stimuli condition. (b) Human subject tested in the Chimpanzee stimuli condition. Sonograms of the auditory samples to be matched are shown at the bottom of each panel.

2. Materials and methods

2.1. Participants

The chimpanzee subject (*Pan troglodytes*) was a 22-year-old female named Pan. Pan had been raised by human caretakers but lived since her infancy within a group of captive chimpanzees at the Primate Research Institute, Kyoto University (KUPRI). Pan had had extensive expertise in auditory tasks: she had been receiving training since her infancy, and eventually succeeded in performing auditory-visual intermodal matching of familiar individuals (Hashiya and Kojima, 2001; Kojima, 2003). In parallel to this experiment, Pan participated in a variety of other visuo-cognitive tasks (Matsuzawa et al., 2006).

The human participants comprised ten students and researchers belonging to the chimpanzee laboratory of KUPRI. They were all highly familiar with both the human and chimpanzee individuals used as stimuli through several hours of daily interactions with them over a period of at least 1 year.

2.2. Apparatus

The experiments were conducted in an experimental booth composed of acrylic panels fixed to a metallic frame. As shown in Fig. 1, a computer monitor with a touch panel system was installed at a suitable height for the subjects. The computer controlling the experiment was connected to a food dispenser and to a pre-amplified speaker located outside the booth, immediately below the monitor.

2.3. Auditory and visual stimuli

The chimpanzee auditory stimuli consisted of segments of pant hoot calls from three members of the KUPRI community. Human stimuli were segments of speech from three researchers of the chimpanzee laboratory, consisting of recordings of a word commonly used in everyday interactions with the chimpanzees ("ooi" in Japanese, meaning "hey"). The chimpanzee and human individuals used as targets were all highly familiar to both Pan and the human subjects. All audio samples were recorded from the edge of the chimpanzees' outdoor compound using a directional microphone. The average length of the vocal samples was comparable between the two species used as stimuli (1989 ms for chimpanzee stimuli and 1848 ms for human stimuli).

Visual stimuli consisted of still images in color, corresponding to the same target individuals as the auditory stimuli. They were taken from video records captured using a digital video camera at the same location and on the same day as the auditory stimuli.

2.4. Procedure

Throughout the study we used an Auditory-Visual Intermodal Matching (AVIM) task with an identical experimental design for all the subjects. The task consisted of matching vocal samples to facial images of the corresponding vocalizers. The human subjects received no indications about the goal of the experiment. A trial was initiated when the subject touched a start key which appeared in the center along the lower edge of the monitor. Each trial consisted of the successive presentation of one auditory sample and two alternative visual targets. Immediately after the end of the voice playback, the two pictures appeared on the monitor at random positions. Of the two visual targets, one always matched the auditory sample (the individual pictured corresponded to the vocalizer), while the other did not. The subject's choice of a visual stimulus (see Fig. 1) was followed by the disappearance of the targets and by an inter-trial-interval (ITI) of 15 s. If the subject correctly touched the picture corresponding to the identity of the vocalizer, the food dispenser was activated, providing positive reinforcement. For correct trials, the chimpanzee subject received a food reward through the food dispenser during the ITI, while human subjects received a proportionate number of sweets at the end of the session. After incorrect trials, subjects were given no reward, and there was no penalty timeout.

In both the Chimpanzee and Human conditions, we used three target individuals. For each target, we prepared four auditory samples and one visual target. A session consisted of 24 trials, divided in two blocks of 12 trials. Within a block, trials were randomized with the restriction that the same target individual was never presented more than three times consecutively. Each of the auditory stimuli appeared twice within a session (once within a block), allowing the correct target to be pitted against both of the incorrect alternatives.

Each subject was tested in both the Chimpanzee and Human stimuli conditions and received two consecutive sessions for each condition. The order of conditions was counterbalanced among the subjects. The subjects performed one session per day, two to four times per week.
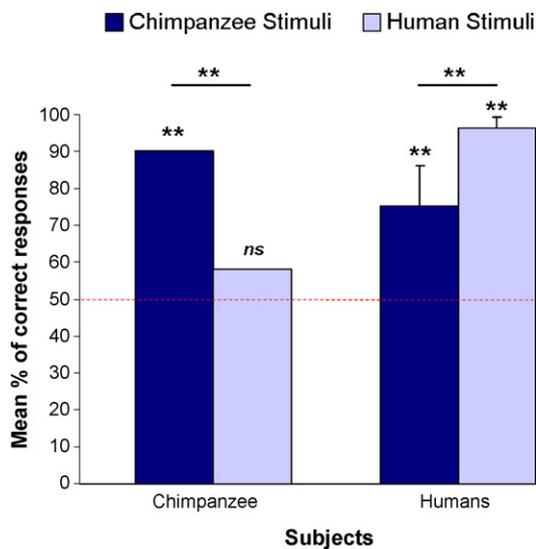
Fig. 2. Mean percentage of correct responses in the Chimpanzee stimuli and in the Human stimuli conditions by the chimpanzee subject Pan ($N = 1$) and the human subjects ($N = 10$). The dotted line shows chance level. *Asterisks* indicate that the subjects performed better than chance level in a particular condition and that their performance was significantly different between the two conditions ($^{**}P < 0.001$).

2.5. Data analysis

For the purposes of analysis, we divided the two sessions carried out in each condition into two blocks of 12 trials. As normality and homogeneity of variance assumptions were met, we used parametric statistics to examine the effects of condition (Chimpanzee stimuli vs. Human stimuli), and the number of blocks carried out (Block 1 to Block 4) on human subjects' performance.

3. Results

As shown in Fig. 2, the chimpanzee and human subjects both performed better at recognizing conspecifics than non-conspecifics.

Pan, the chimpanzee subject, was able to perform recognition of Chimpanzee stimuli better than predicted by chance (binomial test (0.5): $P < 0.001$, $N = 48$). However, with Human stimuli, her performance dropped to chance level (binomial test (0.5): $P = 0.312$, $N = 48$). Pan thus performed significantly better in the Chimpanzee than in the Human condition (Fisher's exact test: $P = 0.0004$, one-tailed, $N = 96$).

We found no statistically significant improvement in Pan's performance between Block 1 and Block 4, either in the Chimpanzee or in the Human stimuli condition (Fisher's exact test: $P = 1$ and $P = 0.793$, respectively, $N = 24$).

Human subjects were able to perform recognition of both Chimpanzee stimuli and Human stimuli better than predicted by chance (one sample t-test (0.5): $t = 7.305$, df = 9, $P < 0.001$; $t = 52.752$, df = 9, $P < 0.001$, respectively). Nine human subjects were able to perform recognition better than predicted by chance in the Chimpanzee condition (binomial test (0.5): $P < 0.001$ for five subjects, $P < 0.05$ for four subjects, $N = 48$), while one subject did not perform better than predicted by chance (binomial test (0.5): $P = 0.665$, $N = 48$). In the Human condition, all ten human subjects were able to perform recognition of Human stimuli better than predicted by chance (binomial test (0.5): $P < 0.001$, $N = 48$).

We conducted a two-way repeated-measures analysis of variance (ANOVA) with condition (Chimpanzee stimuli vs. Human stimuli) × block (Block 1 vs. Block 2 vs. Block 3 vs. Block 4). The main effects of condition and block were significant ($F(1,9) = 32.688$, $P < 0.001$, and $F(3,27) = 8.140$, $P = 0.001$, respectively). These results

indicate that the human subjects performed significantly better in the Human than in the Chimpanzee condition. We found no statistically significant interaction between condition and block ($F(3,27) = 2.635$, $P = 0.070$). In the Chimpanzee stimuli, post hoc pairwise multiple comparisons showed a significant improvement of human subjects' performance between Block 1 and Block 4 and between Block 1 and Block 3 (Bonferroni's test, $P = 0.045$, $P = 0.050$, respectively). In contrast, in the Human stimuli conditions, we found a significant increase in accuracy only between Block 3 and Block 4 ($P = 0.031$). Thus, human subjects improved their performance in both conditions, but the learning effect was evident between Sessions 1 and 2 only for the chimpanzee stimuli.

4. Discussion

Our results revealed that in both chimpanzee and human subjects, the ability to perform intermodal matching of conspecifics and non-conspecifics varied according to the species to be discriminated.

This species-specific effect was clearly evident: the chimpanzee and the human subjects, sharing comparable degrees of acquaintance with the stimuli, both performed better at recognizing conspecifics than non-conspecifics.

A study focusing on the vocal development of Pan presumed that early exposure and training to human-like vocal interactions could have strongly influenced her vocal behavior, as well as her vocal maternal interactions with her own daughter (Kojima, 2003). Our results suggest that Pan's early experience did not seem to favor a perceptual advantage in better discriminating familiar humans compared to familiar conspecifics, quite the reverse: Pan was significantly better at recognizing conspecifics. The results from human subjects also indicated that prolonged acquaintance with chimpanzees could not compensate for an earlier acquired advantage in recognizing conspecifics.

Further analysis of the human subjects' performance revealed a clear learning effect in the chimpanzee condition, and a weaker learning trend in the human condition. In the chimpanzee condition, more accurate performance was observed at the end of the experiment (Block 4) compared to the beginning (Block 1). This improvement may be attributed to a rapid learning of the perceptual intermodal association between a specific vocal sample and the corresponding facial picture. If this were true, the subjects would have learned the association within an extremely short period of time and on the basis of a very limited number of stimulus presentations (each vocal sample was heard four times in total). As an alternative, we suggest that the improvement of accuracy in the chimpanzee condition might be attributed to the choice of an efficient cognitive strategy, rather than to an improvement of perceptual discrimination. At the end of the experimental sessions, human subjects were requested to comment on their results. While the ten subjects confirmed that they could always identify the target individuals depicted on the pictures, six subjects reported a difficulty in identifying the three familiar chimpanzees from their voice. These subjects were very confident at recognizing the voice of at least one chimpanzee. They thus based some of their responses on a choice by exclusion ("I know this was not the voice of A, so I will choose the picture of B").

In contrast, we found no evidence of a learning effect in the case of Pan. Interestingly, a previous study had revealed that Pan matched facial pictures of herself in response to her own vocalizations by exclusion, and did not show evidence of vocal self-recognition (Kojima et al., 2003).

Designing new procedures suitable for training and testing a larger number of chimpanzee subjects will be beneficial for further comparisons of auditory-visual cognitive skills in chimpanzees and

humans. Our present findings nonetheless highlight the influence of species-specificity in the context of interspecies auditory-visual recognition.

Acknowledgments

References

Adachi, I., Fujita, K., 2007. Cross-modal representations of human caretakers in squirrel monkeys. Behav. Process. 74, 27–32.

Adachi, I., Kuwahata, H., Fujita, K., Tomonaga, M., Matsuzawa, T., 2009. Plasticity of ability to form cross-modal representation in infant Japanese macaques. Dev. Sci. 12, 446–452.

Boysen, S.T., 1994. Individual differences in the cognitive abilities of chimpanzees. In: Wrangham, R.W., McGrew, W.C., de Waal, F.B.M., Heltne, P.G. (Eds.), Chimpanzee Cultures. Harvard University Press, Cambridge, MA, pp. 335–350.

Boysen, S.T., Berntson, G.G., 1989. Conspecific recognition in the chimpanzee (*Pan troglodytes*): cardiac responses to significant others. J. Comp. Psychol. 103, 215–220.

Bruce, V., Young, A., 1986. Understanding face recognition. Brit. J. Psychol. 77, 305–327.

Diamond, R., Carey, S., 1986. Why faces are and are not special: an effect of expertise. J. Exp. Psychol. Gen. 115, 107–117.

Evans, T.A., Howell, S., Westergaard, G.C., 2005. Auditory-visual cross-modal perception of communicative stimuli in tufted capuchin monkeys (*Cebus apella*). J. Exp. Psychol. Anim. Behav. Process. 31, 399–406.

Ghazanfar, A.A., Logothetis, N.K., 2003. Facial expressions linked to monkey calls. Nature 423, 937–938.

Hashiya, K., 1999. Auditory-visual intermodal recognition of conspecifics by a chimpanzee (*Pan troglodytes*). Primate Res. 15, 333–342.

Hashiya, K., Kojima, S., 2001. Hearing and auditory-visual intermodal recognition in the chimpanzee. In: Matsuzawa, T. (Ed.), Primate Origins of Human Cognition and Behaviour. Springer, Tokyo, pp. 155–189.

Izumi, A., 2006. Auditory-visual crossmodal representations of species-specific vocalizations. In: Matsuzawa, T., Tomonaga, M., Tanaka, M. (Eds.), Cognitive Development in Chimpanzees. Springer-Verlag, Tokyo, pp. 330–339.

Izumi, A., Kojima, S., 2004. Matching vocalizations to vocalizing faces in a chimpanzee (*Pan troglodytes*). Anim. Cogn. 7, 179–184.

Kojima, S., 2003. A Search for the Origins of Human Speech: Auditory and Vocal Functions of the Chimpanzee. Kyoto University Press, Kyoto, 195 pp.

Kojima, S., Izumi, A., Ceugniet, M., 2003. Identification of vocalizers by pant hoots, pant grants and screams in a chimpanzee. Primates 44, 225–230.

Kuhl, P.K., Tsao, F.M., Liu, H.M., 2003. Foreign-language experience in infancy: Effects of short-term exposure and social interaction in phonetic learning. Proc. Natl. Acad. Sci. U.S.A. 100, 9096–9101.

Martin-Malivel, J., Okada, K., 2007. Human and chimpanzee face recognition in chimpanzees (*Pan troglodytes*): role of exposure and impact on categorical perception. Behav. Neurosci. 121, 1145–1155.

Matsuzawa, T., 1990. Form perception and visual acuity in a chimpanzee. Folia Primatol. 55, 24–32.

Matsuzawa, T., 2006. Evolutionary origins of mother-infant relationship. In: Matsuzawa, T., Tomonaga, M., Tanaka, M. (Eds.), Cognitive Development in Chimpanzees. Springer-Verlag, Tokyo, pp. 127–141.

Matsuzawa, T., Tomonaga, M., Tanaka, M., 2006. Cognitive Development in Chimpanzees. Springer-Verlag, Tokyo, 522 pp.

Myowa-Yamakoshi, M., Yamaguchi, M.K., Tomonaga, M., Tanaka, M., Matsuzawa, T., 2005. Development of face recognition in infant chimpanzees (*Pan troglodytes*). Cognit. Dev. 20, 49–63.

Parr, L.A., Dove, T., Hopkins, W.D., 1998. Why faces may be special: evidence of the inversion effect in chimpanzees. J. Cognit. Neurosci. 10, 615–622.

Parr, L.A., 2004. Perceptual biases for multimodal cues in chimpanzee (*Pan troglodytes*) affect recognition. Anim. Cogn. 7, 171–178.

Parr, L.A., Heintz, M., Akamagwuna, U., 2006. Three studies on configural face processing by chimpanzees. Brain Cogn. 62, 30–42.

Pascalis, O., Bachevalier, J., 1998. Face recognition in primates: a cross species study. Behav. Process. 43, 87–96.

Pascalis, O., Petit, O., Kim, J.H., Campbell, R., 1999. Picture perception in primates: the case of face perception. Curr. Psychol. Cognit. 18, 889–921.

Tanaka, M., 2003. Visual preference by chimpanzees (*Pan troglodytes*) for photos of primates measured by a free choice-order task: implication for influence of social experience. Primates 44, 157–165.

Vouloumanos, A., Werker, J.F., 2007. Listening to language at birth: evidence for a bias for speech in neonates. Dev. Sci. 10, 159–171.