



Short communication

Visual and auditory conditional position discrimination in chimpanzees (*Pan troglodytes*)

L. Martinez*, T. Matsuzawa

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

ARTICLE INFO

Article history:

Received 15 December 2008

Received in revised form 12 March 2009

Accepted 22 March 2009

Keywords:

Auditory
Chimpanzee
Discrimination
Visual

ABSTRACT

Chimpanzee cognition has been studied predominantly through the visual modality, and much less through the auditory modality. The aim of this study was to explore possible differences in chimpanzees' processing of visual and auditory stimuli. We developed a new conditional position discrimination (CPD) task requiring the association between a stimulus (from either the auditory or the visual modality), and a spatial position (left or right). The stimuli consisted of the face and voice of two individuals well known to the subjects (one chimpanzee and one human). Six chimpanzees participated in both the visual and the auditory conditions. We found contrasting results between the two conditions: the subjects acquired the CPD more easily in the visual than in the auditory condition. This supports previous findings on the difficulties encountered by chimpanzees in learning tasks involving auditory stimuli. Our experiments also revealed individual differences: the chimpanzee with the most extensive experience in symbolic visual matching tasks showed good performance in both conditions. In contrast, the chimpanzee expert in an auditory–visual intermodal matching task showed no sign of learning in either condition. Future work should focus on finding the most appropriate procedure for exploring chimpanzees' auditory–visual cognitive skills.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

In the laboratory, advanced cognitive abilities of birds and primates have been extensively investigated through the visual modality, whereas considerably fewer studies have focused on the auditory modality (Harrison, 1984; Kojima, 2003; Wright et al., 1990; Wright, 1999). Despite general difficulties in training animals to learn tasks involving complex auditory discrimination, researchers ultimately succeeded – after several modifications of the procedure – in training macaques and pigeons to perform auditory serial probe recognition and sequential discrimination of the abstract “same/different” concept (Murphy and Cook, 2008; Wright et al., 1990; Wright, 1998). There is also a growing interest in investigating cross-modal competence using auditory–visual stimuli. Experiments have shown, using a violation of expectancy procedure, that several primate species share a common cross-modal representation of familiar individuals and are able to recognize the correspondence between the auditory and visual components of conspecific calls or human speech (Adachi and Fujita, 2007; Adachi et al., 2009; Evans et al., 2005; Ghazanfar and Logothetis, 2003). Other studies have also demonstrated

monkeys' cross-modal numerical capacities using naturalistic or non-ecologically relevant features, which require the ability to form a conceptual equivalence between physical stimuli originating in different sensorial modalities (Jordan et al., 2005, 2008).

In contrast, literature on chimpanzees' auditory discrimination is lacking, although chimpanzees predominantly rely on the visual and auditory modalities to perceive and interact with their physical and social environment (Kojima, 2003). Only a limited number of experimental studies have investigated auditory–visual intermodal discrimination in chimpanzees (Bauer and Philip, 1983; Boysen, 1994; Parr, 2004; Savage-Rumbaugh et al., 1988; Izumi, 2006). Hashiya and Kojima (1997, 2001a,b) succeeded after intensive efforts to train a chimpanzee to perform auditory–visual intermodal matching of familiar objects and individuals. However, intermodal performance remained less accurate than intramodal visual matching (Hashiya and Kojima, 2001a). These findings support the view that intermodal auditory–visual matching and complex auditory discrimination are particularly difficult tasks to learn under experimental conditions – not only for chimpanzees, but also for monkeys and rats – and call for newly designed procedures to better explore auditory cognition in non-human animals (D'Amato and Colombo, 1985; Hashiya and Kojima, 2001b; Mercado et al., 2005).

The aim of this study was to explore possible differences in the cognitive processing of auditory and visual stimuli. We developed a new unimodal task requiring the simple association between a

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

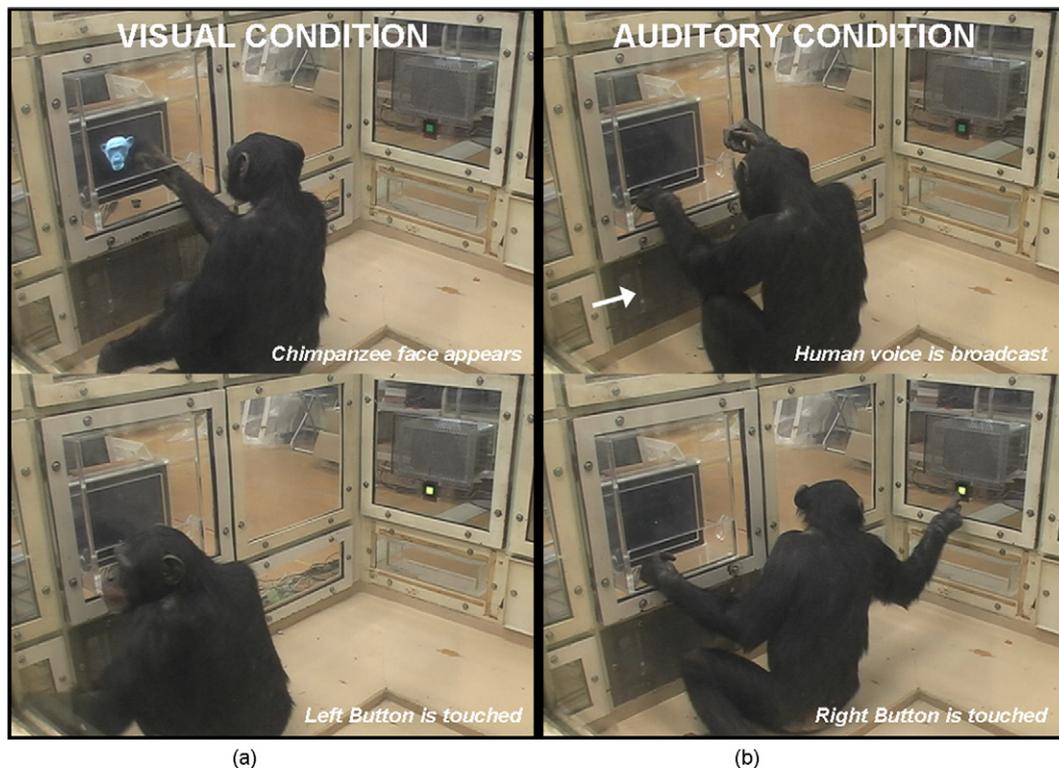


Fig. 1. Still images taken from video footage recorded during experimental sessions. (a) In the visual condition, the sample appeared on the screen and subjects were required to touch this stimulus. In the trial shown here, the correct response was to press the left button, associated with the chimpanzee face. (b) In the auditory condition, the sample was broadcast on the central speaker (see arrow). In the trial shown here, the correct response was to press the right button, associated with the human voice.

stimulus from either the auditory or the visual modality and a specific spatial position (left or right). This conditional position discrimination (CPD) task aimed to elucidate whether the difficulty in performing intermodal matching could be attributed to a deficiency in processing auditory stimuli. If so, subjects would be expected to perform better in the visual condition than in the auditory condition.

2. Materials and methods

2.1. Subjects

The six chimpanzee subjects (*Pan troglodytes*) were three mother–juvenile pairs. The mothers, named Ai, Chloe, and Pan (29, 25, and 22 years old, respectively) and the juveniles, named Ayumu (male), Cleo, and Pal (females), all aged 7 years, lived together at the Primate Research Institute of Kyoto University, participating daily in a variety of socio-cognitive tasks (Matsuzawa et al., 2006). Only Pan had had extensive experience with auditory tasks, having received training since the age of 8 years (Hashiya and Kojima, 2001b; Kojima, 2003). The other five chimpanzees had never been successfully trained on any auditory tasks.

2.2. Apparatus

The experiments were conducted in an experimental booth (1.5 m wide × 2 m deep × 1.8 m high) composed of acrylic panels fixed to a metallic frame. In the central acrylic panel, a 21-in. computer monitor with a touch panel system was installed at a suitable height for the subjects (see Fig. 1). Next to the monitor, a food dispenser was set up to deliver rewards on the same acrylic panel. We mounted a pre-amplifier speaker (BOSE MMS-1SP) immediately below the monitor, covered with a frameless metal grille allowing sound to be easily transmitted inside the booth. We symmetri-

cally mounted illuminated pushbuttons (28 mm × 28 mm) on the left and right walls of the booth perpendicular to the wall containing the monitor. The two buttons were identical, both located 70 cm from the central panel. Thus, the buttons were within easy reach of a subject sitting in front of the monitor, but could not be touched simultaneously.

2.3. Auditory and visual stimuli

We used one female chimpanzee and one male human as the target individual stimuli. Both targets were individuals highly familiar to the subjects. The voices were recorded opportunistically using a directional microphone (Sennheiser ME66) and a DAT walkman (Sony TCD-D100). The chimpanzee auditory stimulus, of 1566 ms duration, was a segment of a scream call captured in the chimpanzees' outdoor compound. The human auditory stimulus, of 1370 ms duration, was a segment of spoken Japanese uttered by the human target during a friendly interaction with a chimpanzee, meaning “[come] down, down, good” (*shita, shita, sodaa*). The sound pressure level was 80 dB on average. The visual stimuli consisted of still images from video records, depicting frontal views of the target individuals' faces, which were captured using a digital video camera (Sony DCR-HC90) at the same location and on the same day as the auditory stimuli. We used one picture for each individual. The two images were processed to standardize size (404 × 404 pixels), brightness, and to remove all background.

2.4. Procedure

The chimpanzees voluntarily entered the testing room where they received incentive food rewards. One experimental session was conducted per day, typically six times a week. The mother–offspring pairs were tested within the same experimen-

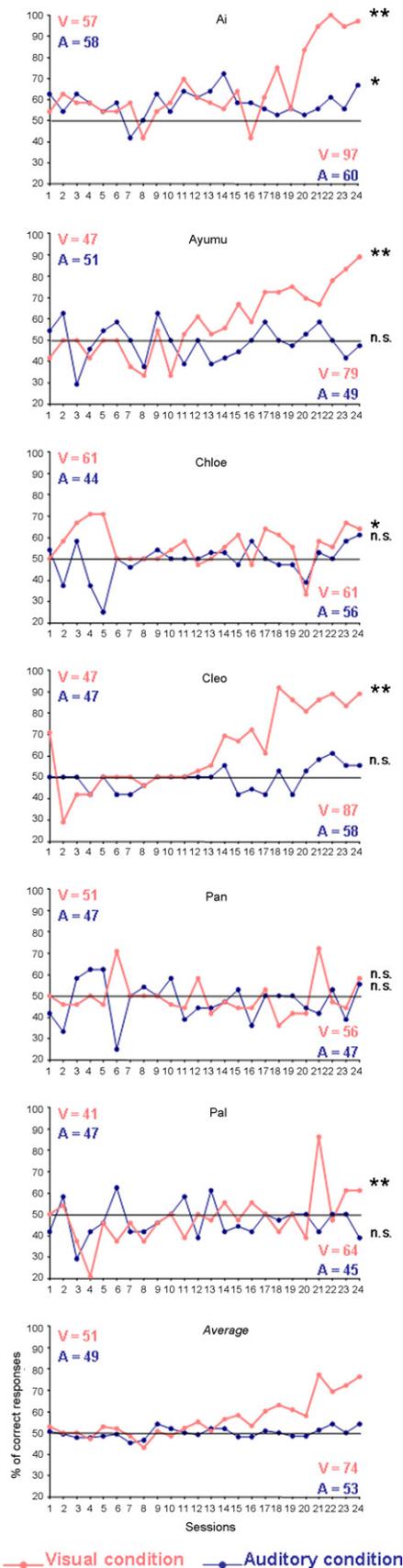


Fig. 2. Percentage of correct responses obtained per session by each subject and on average; in the visual and auditory conditions across the 24 sessions of the testing phase (chance level was 50%). Numbers in the upper left of each graph indicate the mean percentage of correct responses in the first sessions, and numbers in the bottom right indicate the mean percentage in the last sessions (V for visual condition

tal slot. While one subject was being tested on the CPD task, the other subject was tested in an adjacent booth on a visual identity matching-to-sample task.

The CPD task consisted of associating each target individual to a specific spatial position. In both the visual and auditory conditions, the chimpanzee target was associated with the left position, while the human target was associated with the right position. The left/right assignment for stimulus type was identical across subjects. The six subjects participated in both conditions in parallel and we daily alternated the order in which the two conditions were tested. A session consisted of trials of only one modality, either the auditory modality (auditory condition) or the visual modality (visual condition). In both modalities, a trial was initiated when the subject repeatedly touched a start key which appeared three times on the monitor at random positions. The last start key stimulus always appeared in the center near the bottom edge of the screen, ensuring that the subject's hand always began a trial from a central position. A response to the last start key resulted in the immediate presentation of the sample. In the visual condition, the sample appeared on the screen and the subjects were required to touch this stimulus, thereby confirming that their attention was focused on the monitor. In the auditory condition, the sample was broadcast on the speaker (see Fig. 1), the source of the sound similarly drawing the subjects' attention to a central location. Sample presentation was in turn immediately followed by the illumination of the two buttons. The subject's touch of a button only had an effect when the buttons were illuminated. If the sample was a chimpanzee stimulus, the subject was required to touch the button on the left, whereas a human stimulus required the right button to be touched. The subject's choice was followed by the extinction of the button illumination and by an inter-trial-interval (ITI) of 2 s. For correct trials, the subjects received a food reward during the ITI. After incorrect trials, the subjects were given no reward, and a time-out of 3 s was used as penalty.

We first conducted five training sessions (24 trials each) for each sensory modality, in which we alternated blocks of trials presenting either the human or the chimpanzee stimulus. In the testing phase, trials were randomized with the restriction that the same stimulus was never presented more than three times consecutively, and that each of the stimuli appeared with equal frequencies within a session. A total of 24 sessions was carried out in each condition. The first 10 sessions comprised 24 trials each, while the last 14 sessions were 36 trials long. Thus, each subject performed 744 trials in each modality.

2.5. Data analysis

For the purposes of analysis, we combined the number of correct responses obtained by each subject in the first six sessions (first sessions: 24 trials per session, $N=144$ trials) and in the last four sessions (last sessions: 36 trials per session, $N=144$ trials) of the testing phase. Two-tailed binomial tests were conducted to determine whether the accuracy of each subject was significantly greater than that expected by chance. As normality and homogeneity of variance assumptions were met, we used parametric statistics to test the effects of condition (visual vs. auditory), the number of sessions carried out (first vs. last) and stimulus type (chimpanzee vs. human) on subjects' performance.

and A for auditory condition). Asterisks indicate that the subject performed better than predicted by chance in the last sessions (binomial test, ** $P < 0.001$ and * $P < 0.05$; $N = 144$ trials).

3. Results

Fig. 2 shows the percentage of correct responses per session for each subject, as well as the subjects' average accuracy across the 24 sessions of the testing phase. Fig. 2 also indicates the mean percentage of correct responses in both the visual and auditory conditions in the first and last sessions.

In the last sessions of the visual condition, five subjects (Ai, Ayumu, Chloe, Cleo and Pal) were able to perform CPD better than that predicted by chance (binomial test (0.5): $P < 0.001$, $P < 0.001$, $P = 0.010$, and $P < 0.001$ and $P = 0.001$, respectively) while one subject (Pan) did not perform above chance level ($P = 0.211$). These results indicate that five subjects eventually succeeded in correctly associating each visual stimulus with the corresponding spatial position.

In the last sessions of the auditory condition, one subject (Ai) was able to perform CPD better than that predicted by chance (binomial test: $P = 0.024$). In contrast, the other five subjects (Ayumu, Chloe, Cleo, Pan and Pal) remained at chance level ($P = 0.934$, $P = 0.211$, $P = 0.080$, $P = 0.560$ and $P = 0.279$, respectively). These results suggest that only one subject was able to learn the task in the auditory condition.

We conducted a three-way repeated-measures analysis of variance (ANOVA) with condition (visual vs. auditory) \times session (first vs. last) \times stimulus type (chimpanzee vs. human). The main effects of condition and session were significant ($F(1,5) = 39.632$, $P = 0.001$, and $F(1,5) = 13.887$, $P = 0.014$, respectively). These results indicate that the subjects performed better in the visual than in the auditory condition. It also shows that subjects improved their performance throughout the testing phase, as their accuracy increased significantly between the first and the last sessions. In contrast, the main effect of stimulus type was not significant ($F(1,5) = 0.873$, $P = 0.392$), despite marked preferences for one stimulus type (or side) in all subjects at some stages of the training. We found no statistically significant interaction between condition and session ($F(1,5) = 6.130$, $P = 0.056$), between condition and position ($F(1,5) = 0.092$, $P = 0.774$), and between session and position ($F(1,5) = 1.809$, $P = 0.236$). Although the interaction condition \times session was not significant, the results of the binomial tests suggest that the effect of session was stronger in the visual condition than in the auditory condition.

4. Discussion

In this study we compared chimpanzees' ability to perform CPD in two conditions, visual and auditory, using the face and voice of a chimpanzee and a human as stimuli. In the visual condition, five out of six subjects successfully learned the association between the pictorial representation of an individual and a specific spatial position. Conversely, in the auditory condition, the accuracies obtained by the six subjects remained low. As predicted, the subjects more easily acquired the correct CPD when the stimuli presented were visual rather than auditory.

4.1. Methodological considerations

In their natural habitat, chimpanzees possess the ability to acoustically recognize individuals and to accurately determine their location (Goodall, 1986). Thus, the lack of accurate performance in our auditory condition is not likely to be accounted for by a lack of chimpanzee discriminatory ability in this modality. Instead, we argue that the difficulties encountered under experimental conditions can be attributed, to some extent, to a failure to spontaneously perceive and process recordings broadcast through a speaker as equivalents of natural vocalizations rather than artificial sounds. Just as subjects have to learn to process pictures as two-dimensional

representations of real objects, they also have to learn the equivalence between a natural auditory stimulus and its recorded version.

In addition, the auditory and visual conditions in our study differed in one important aspect: subjects had to touch the visual stimulus on the monitor, but they were not required to touch the speaker broadcasting the auditory stimulus. It has been proposed that physical contact with the sound source may strongly enhance the acquisition of auditory tasks in monkeys and avian species (Harrison, 1984; Wright et al., 1990). Thus, differences in levels of attention may have biased subjects' discrimination of visual versus auditory stimuli (Harrison et al., 1977). However, in a previous attempt to train chimpanzees to perform an intermodal auditory–visual matching task, we designed a procedure where subjects had to localize the auditory stimulus and touch a button adjacent to the sound source (two lateral speakers) before choosing the corresponding visual target on a central monitor (a study partly inspired by Wright et al., 1990). After an 11-month training period and more than 6500 trials, none of the six subjects of the present study had learned to perform the task above chance level (Martinez and Matsuzawa, unpublished data). Therefore, while we do not rule out possible influences of contact with the auditory stimulus or the need for a longer training period to acquire the auditory CDP task, these factors cannot by themselves explain our results. Nevertheless, future investigations would likely benefit from incorporating systematic contact with the auditory source, a larger number of exemplars in each stimulus type, and a progressive replacement of natural sounds by their recorded versions (Harrison, 1984; Hashiya and Kojima, 2001a; Wright et al., 1990).

4.2. Individual differences

Individual differences in the testing histories of the subjects appeared to be of major importance in our experiment. Ai was the only chimpanzee to perform well in both the visual and the auditory conditions. Compared to our other subjects, Ai had had similar auditory exposure, since all the subjects live together in the same physical and social environment and had had close daily interactions with human caretakers since their infancy. Thus, differential exposures to natural auditory stimuli may not account for the discrepancy in performance.

Interestingly, on the other hand, of all our subjects it is Ai who has achieved success in the widest variety of cognitively highly demanding tasks, demonstrating her ability to learn concepts and to perform symbolic visual matching (Matsuzawa, 2003). Thus, her unique testing history, particularly her extensive and continuous experience in dealing with a large variety of highly demanding cognitive tasks, may have underpinned the successful learning process observed in this study. We might also have expected a positive learning curve for Pan, at least in the auditory condition, since she had had extensive experience in performing intermodal auditory–visual matching, demonstrating a functional understanding of the equivalence between a real sound and the corresponding recording (Hashiya and Kojima, 2001a,b). Yet, Pan was among the least successful subjects in the CPD task. This suggests that her remarkable success in intermodal matching was not spontaneously transferable to a new task, even when the latter required a simpler unimodal conditional association. Her capacities may then reflect less flexibility than that seen in Ai in successfully learning novel cognitively demanding tasks.

All our six subjects, including Pan, have been far more exposed to tasks in the visual domain than in the auditory domain. Thus, past testing experience in dealing with stimuli from one sensory modality may also greatly influence any attempt to train subjects in tasks related to another sensory modality. This study highlights the importance of meticulously examining differences among individuals sharing a common sensory environment but diverg-

ing in their testing histories—an issue highly relevant to the recent debate concerning the utility of cognitive studies with captive chimpanzees in understanding the mind of the chimpanzee (Boesch, 2007; Tomasello and Call, 2008).

Taken together, these results provide new evidence for the difficulties encountered by chimpanzees during laboratory tasks involving auditory stimuli compared with tasks entirely in the visual domain, and reveal individual differences related to knowledge acquired in previous experiments.

Acknowledgments

This study was supported by Grant-in-aid for Scientific Research #16002001 and #20002001 from MEXT (Japan), JSPS-ITP-HOPE, JSPS-GCOE (D07 for Psychology and A06 for Biodiversity) to T. Matsuzawa and by a MEXT scholarship to L. Martinez. We are grateful to our colleagues at the Language and Intelligence Section and at the Center for Human Evolution Modeling Research. We thank D. Biro, N. Granier and JB. Leca for help in preparing and editing the manuscript. We also thank two anonymous reviewers for their valuable comments. The use of the chimpanzees adhered to the *Guide for the Care and Use of Laboratory Primates* (2002) of the Primate Research Institute, Kyoto University.

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.*, doi:10.1111/j.1467-7687.2008.00780.x.
- Bauer, H.R., Philip, M., 1983. Facial and vocal individual recognition in the common chimpanzee. *Psychol. Res.* 33, 161–170.
- Boesch, C., 2007. What makes us human (*Homo sapiens*)? The challenge of cognitive cross-species comparison. *J. Comp. Psychol.* 121, 227–240.
- 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 Massachusetts, pp. 335–350.
- D'Amato, M.R., Colombo, M., 1985. Auditory matching-to-sample in monkeys (*Cebus apella*). *Anim. Learn. Behav.* 13, 375–382.
- 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.
- Goodall, J., 1986. *The Chimpanzees of Gombe: Patterns of Behavior*. Harvard University Press, Cambridge Massachusetts, p. 673.
- Harrison, J.M., Iversen, S.D., Pratt, S.R., 1977. Control of responding by location of auditory stimuli: adjacency of sound and response. *J. Exp. Anal. Behav.* 28, 243–251.
- Harrison, J.M., 1984. The functional analysis of auditory discrimination. *J. Acoust. Soc. Am.* 75, 1848–1854.
- Hashiya, K., Kojima, S., 1997. Auditory–visual intermodal matching by chimpanzee (*Pan troglodytes*). *Psychol. Res.* 39, 182–190.
- Hashiya, K., Kojima, S., 2001a. Acquisition of auditory–visual intermodal matching-to-sample by a chimpanzee (*Pan troglodytes*): comparison with visual–visual intramodal matching. *Anim. Cogn.* 4, 231–239.
- Hashiya, K., Kojima, S., 2001b. Hearing and auditory–visual intermodal recognition in the chimpanzee. In: Matsuzawa, T. (Ed.), *Primate Origins of Human Cognition and Behavior*. Springer, Tokyo, pp. 155–189.
- Kojima, S., 2003. *A Search for the Origins of Human Speech: Auditory and Vocal Functions of the Chimpanzee*. Kyoto University Press, Kyoto, p. 195.
- 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.
- Jordan, K.E., Brannon, E.M., Logothetis, N.K., Ghazanfar, A.A., 2005. Monkeys match the number of voices they hear to the number of faces they see. *Curr. Biol.* 15, 1034–1038.
- Jordan, K.E., MacLean, E.L., Brannon, E.M., 2008. Monkeys match and tally quantities across senses. *Cognition* 108, 617–625.
- Matsuzawa, T., 2003. The Ai project: historical and ecological contexts. *Anim. Cogn.* 6, 199–211.
- Matsuzawa, T., Tomonaga, M., Tanaka, M., 2006. *Cognitive Development in Chimpanzees*. Springer-Verlag, Tokyo, p. 522.
- Mercado III, E., Orduña, I., Nowak, J.M., 2005. Auditory categorization of complex sounds by rats (*Rattus norvegicus*). *J. Comp. Psychol.* 119, 90–98.
- Murphy, M.S., Cook, R.G., 2008. Absolute and relational control of a sequential auditory discrimination by pigeons (*Columba livia*). *Behav. Process.* 77, 210–222.
- Parr, L.A., 2004. Perceptual biases for multimodal cues in chimpanzee (*Pan troglodytes*) affect recognition. *Anim. Cogn.* 7, 171–178.
- Savage-Rumbaugh, S., Sevcik, R.A., Hopkins, W.D., 1988. Symbolic cross-modal transfer in two species of chimpanzees. *Child Dev.* 59, 617–625.
- Tomasello, M., Call, J., 2008. Assessing the validity of ape–human comparison: a reply to Boesch (2007). *J. Comp. Psychol.* 122, 449–452.
- Wright, A.A., Shyan, M.A., Jitsumori, M., 1990. Auditory same/different concept learning by monkeys. *Anim. Learn. Behav.* 18, 287–294.
- Wright, A.A., 1998. Auditory list memory in rhesus monkeys. *Psychol. Sci.* 9, 91–98.
- Wright, A.A., 1999. Auditory list memory and interference processes in monkeys. *J. Exp. Psychol. Anim. Behav. Process.* 25, 284–296.