Anim Cogn (2009) 12 (Suppl 1):S71–S85
DOI 10.1007/s10071-009-0269-1

ORIGINAL PAPER

Auditory–visual intermodal matching based on individual recognition in a chimpanzee (*Pan troglodytes*)

Laura Martinez · Tetsuro Matsuzawa

Received: 20 January 2008 / Revised: 20 July 2009 / Accepted: 22 July 2009 / Published online: 22 August 2009

Abstract The ability to recognize familiar individuals with different sensory modalities plays an important role in animals living in complex physical and social environments. Individual recognition of familiar individuals was studied in a female chimpanzee named Pan. In previous studies, Pan learned an auditory–visual intermodal matching task (AVIM) consisting of matching vocal samples with the facial pictures of corresponding vocalizers (humans and chimpanzees). The goal of this study was to test whether Pan was able to generalize her AVIM ability to new sets of voice and face stimuli, including those of three infant chimpanzees. Experiment 1 showed that Pan performed intermodal individual recognition of familiar adult chimpanzees and humans very well. However, individual recognition of infant chimpanzees was poorer relative to recognition of adults. A transfer test with new auditory samples (Experiment 2) confirmed the difficulty in recognizing infants. A remaining question was what kind of cues were crucial for the intermodal matching. We tested the effect of visual cues (Experiment 3) by introducing new photographs representing the same chimpanzees in different visual perspectives. Results showed that only the back view was difficult to recognize, suggesting that facial cues can be critical. We also tested the effect of auditory cues (Experiment 4) by shortening the length of auditory stimuli, and results showed that 200 ms vocal segments were the limit for correct recognition. Together, these data demonstrate that auditory–visual intermodal recognition in chimpanzees might be constrained by the degree of exposure to different modalities and limited to specific visual cues and thresholds of auditory cues.

Keywords Chimpanzee · Individual recognition · Auditory–visual intermodal matching · Face recognition · Pant hoot vocalization

Introduction

Primates, like other animals, have the ability to recognize individuals based on cues from species-specific vocalizations (Snowdon and Cleveland 1980; Cheney and Seyfarth 1990; Rendall et al. 1996; Fitch and Fritz 2006). Individual recognition can also occur based on visual features, especially the face of conspecifics (Pascalis and Bachevalier 1998; Dufour et al. 2006; Neiworth et al. 2007). The ability to process auditory–visual information should play an important role in social interactions and group cohesion because it helps in the recognition of individuals and their emotional states.

One essential factor in individual recognition should be the degree of exposure/interaction with target individuals usually living in the same social group. Familiarity in this paper is defined as subjective knowledge of a target individual derived from interactive experiences or observations on a daily basis. Thus, the process of familiarization is shaped through exposure to sensory features such as voice, face, smell and touch, among others. Human studies have explored the neuro-cognitive mechanisms of the familiarity effect, showing that distinctive features in each sensory modality are expected to strongly determine individual

This contribution is part of the Supplement Issue "The Chimpanzee Mind" (Matsuzawa 2009).

L. Martinez (✉) · T. Matsuzawa
Department of Brain and Behavioral Sciences,
Primate Research Institute, Kyoto University,
Kanrin 41-2, Inuyama, Aichi 484-8506, Japan
e-mail: laura@pri.kyoto-u.ac.jp

recognition (e.g., in face processing: Dubois et al. 1999; Ellis and Lewis 2001; in voice processing: Van Lancker and Kreiman 1987; Denes and Pinson 1993; Nakamura et al. 2001). In the same manner, Porter et al. (1983) have revealed that human mothers are able to recognize 2–6-day-old neonates by odor cues alone, even with an average pre-test contact of only 2.4 h with their baby. Although the similarity of the infant's odor to that of the mother may mediate this ability, it proves that a familiarization process leading to identity recognition can occur in the very first few days of life.

Auditory recognition

In chimpanzees, lifelong social relationships are strengthened by affiliative behaviors including subtle changes in vocalizations, facial expressions or gestures (Goodall 1986). It is well known that chimpanzees can recognize each individual member of their community from unique facial and/or body characteristics, or distinctive vocal features. The most marked vocalization in chimpanzees is the pant hoot, which is a loud and complex sound. This call is used in a variety of occasions and contexts, but one of its main functions is long distance communication. For example, the pant hoot can convey information on who is where (Mitani and Nishida 1993; Clark Arcadi 1996; Notman and Rendall 2005). There is a high degree of individual difference in the pant hoot of individuals (Marler and Hobbet 1975; Mitani 1994; Clark Arcadi et al. 1998), but there are consistent differences in the structure of male and female pant hoots (Mitani and Brandt 1994; Mitani et al. 1996). In addition, there is a developmental change in pant hoot, with it being used more frequently with increasing age (Marler and Tenaza 1977). Thus, pant hoot makes it possible to identify an invisible vocalizer and facilitates species-specific fission–fusion of sub-grouping (Clark and Wrangham 1994; Mitani et al. 1996).

Visual recognition

Chimpanzees are also able to use visual cues such as face specificity to recognize individuals. Chimpanzee faces, just like those of humans, possess distinctive morphological characteristics and salient individual differences. In addition, specific facial expressions, such as a relaxed, open mouth or grimace showing teeth, elicit species-typical responses from early stages of development (Tomonaga et al. 1993; Parr et al. 2000; Vokey et al. 2004; Myowa-Yamakoshi 2006). Faces are so salient and individualized that chimpanzees can learn to pair specific faces with corresponding arbitrary symbols, such as letters of the alphabet, in a discrimination learning task (Matsuzawa 1990).

Multimodal individual recognition

The most distinctive and common cues of individual recognition in humans also consist of the face and voice of the particular persons (Bruce and Young 1986; Belin et al. 2000). Recent neuro-imaging studies have emphasized the existence of a brain region responsible for multimodal face–voice integration (for a review: Campanella and Belin 2007). Auditory–visual speech perception studies confirmed that combined information extracted from face and voice features gives access to a faster and more robust recognition than unimodal cues (Ellis et al. 1997; Calvert 2001; Kamachi et al. 2003). Thus, auditory and visual information are complementary and effective in enhancing individual recognition when the two different cues are combined (Von Kriegstein and Giraud 2006). A relatively unexplored question is whether other species, especially our closest evolutionary neighbors, the chimpanzees, also possess the ability of auditory–visual intermodal matching (AVIM).

Bauer and Philip (1983) conducted the first successful attempt at voice-to-face matching of familiar individuals in chimpanzees. In that study, three infant chimpanzees were able to match different vocal recordings and facial portraits of familiar conspecifics. Savage-Rumbaugh et al. (1986, 1988) carried out similar experiments in bonobos (*Pan paniscus*) and chimpanzees (*P. troglodytes*). Subjects with previous experience with visual lexigrams and human speech successfully associated spoken English words with corresponding pictures or symbols. More recently, Boysen (1994) investigated the ability of chimpanzees to match visual and vocal representations of familiar conspecifics and humans, using facial pictures and corresponding chimpanzee recordings of bark calls or human greetings. The four chimpanzees trained in that study were capable of performing intermodal recognition and maintained accurate performance when novel stimuli were presented. Despite these few studies proving that chimpanzees were able to perform intermodal individual recognition, these types of studies were rarely replicated because training chimpanzees to perform AVIM was not an easy task.

Kojima et al. carried out the most systematic investigation concerning AVIM in the last two decades. The first series of studies aimed to investigate the auditory perception of chimpanzees (for a review: Kojima 2003). These studies explored the perception of human speech, such as vowels and consonants, in three chimpanzees (Kojima and Kiritani 1989; Kojima et al. 1989). Based on speech sound discrimination data, the researchers succeeded in training one juvenile female, named Pan, to perform AVIM of familiar objects (Hashiya and Kojima 1997, 2001a). Later, she learned to match familiar human voices with the corresponding face (Hashiya 1999; Hashiya and Kojima 2001b).

Moreover, she succeeded in performing intermodal matching of conspecifics with three different types of vocalizations (pant hoots, pant grunts and screams), and also matched the facial expression depicted in a still picture regardless of the vocalizer's identity (Kojima et al. 2003). Pitch shifts had a detrimental effect on the performances, suggesting that at least pitch is an important cue for voice-related individual recognition. In addition, Pan did not show vocal self-recognition, and correctly match her own picture in response to her vocalizations only by exclusion (Kojima et al. 2003). The performance was consistent across static and dynamic (video) visual stimuli depicting neutral or emotional facial expressions (Izumi and Kojima 2004; Izumi 2006). Taken together, these results reflect that auditory and visual characteristics of familiar objects or individuals became equivalent for Pan. However, it must be noted that her performance on the AVIM task was relatively low in comparison to that on an intramodal visual matching task (Hashiya and Kojima 2001a). The limited number of reports on auditory–visual intermodal abilities in chimpanzees indicates that these types of studies were rarely replicated in experimental conditions, most probably because AVIM was not an easy task for chimpanzees. Thus, there is as yet little understanding of what kind of constraints or limits exist to chimpanzees' ability to perform AVIM.

The present study takes advantage of Pan's unique training in an AVIM task and attempts to promote our understanding of chimpanzee AVIM for individual recognition. We hypothesized that familiarity plays an essential role in voice–face intermodal processing. We predicted that auditory and visual cues of individuality produce distinctive effects and designed a series of experiments to test whether each of these cues distinctively influence intermodal processing of individual recognition. A potential species-typical effect was taken into account by testing familiar conspecifics and humans.

In Experiments 1 and 2, a new series of auditory and vocal samples were used to test the generalization of Pan's abilities to identify chimpanzees and humans of both sexes, varying in age and in the degree of exposure to the subject. In 2000, three babies were born in Pan's group, and this provided a unique opportunity to expand the panel of chimpanzees used as target stimuli (Matsuzawa et al. 2006). Furthermore, Pan's ability at infant individual recognition remained unexplored in the previous auditory–visual studies. In Experiment 3, visual stimuli were modified to investigate the effect of visual cues, especially facial cues, on intermodal individual recognition. Pictures depicting the same individuals in different visual perspectives were tested. Experiment 4 explored the effect of auditory cues. The length of the auditory stimuli was varied to determine the minimum length of vocal segments needed to successfully perform intermodal individual recognition.

General materials and methods

Subject

The subject was a 22-year-old female chimpanzee (*Pan troglodytes*) named Pan. She was born at the Primate Research Institute, Kyoto University in Inuyama, Japan, and raised by human caretakers from birth. However, she lived since her infancy with a group of 14 chimpanzees. Pan is presently the youngest adult in the group (see Table 1). The group also comprises now three infants, including Pan's daughter, Pal, and two other infants who were born in 2000. This group of chimpanzees was housed in a semi-indoor residence connected to an outdoor compound (Matsuzawa 2006). Pan is the only chimpanzee in this group who has been extensively trained to perform AVIM tasks since the age of 8 years old (Kojima 2003). When the present experiment began in July 2005, 4 months had passed since the last time she had performed this kind of task. She has also been a subject in different auditory tasks and in a variety of socio-cognitive tasks (Tanaka 2001; Matsuzawa 2001, 2003, 2009; Matsuzawa et al. 2006). The subject was not food-deprived and voluntarily entered the experimental room with her daughter where she received incentive food rewards. 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.

Apparatus

The experiments were conducted in an experimental booth (2.4 m wide × 2.0 m deep × 1.8 m high) mainly composed of acrylic panels fixed to a metallic structure and connected to a twin experimental booth. The experimental room was located in a reinforced concrete building in which noise levels were low enough to not disturb the experiments.

As shown in Fig. 1, a 21-inch computer monitor with a touch panel system (Pro-Tect, model PD-105TP15; resolution in pixels: 1,024 width × 768 height) was placed behind one of the acrylic panels at a suitable height for the subject. The experiment was controlled by a personal computer (equipped with sound card E-MU 0404) connected to a food dispenser (Biomedica) and a pre-amplifier speaker (BOSE MMS-1SP). A Borland C# Builder 1.0 program was customized to operate the task and collect the data. The booth was not soundproofed and sounds could be easily transmitted to the subject. To improve the quality of the auditory stimuli transmission, small holes were pierced in the acrylic panel in front of the speaker located outside the booth, right below the monitor.

Table 1 Auditory stimuli presented to the subject in Experiment 1

Chimpanzees		Humans	
Name (sex, age[a])	Average length of pant hoot in ms (SD)	Name (sex, years[b])	Average length of speech segments in ms (SD)
Reiko (♀, 39)	2046 (69)	KK (♂, 22)	1980 (54)
Gon (♂, 39)	1845 (158)	TM (♂, 22)	1972 (82)
Puchi (♀, 39)	1994 (78)	NM (♂, 14)	2165 (84)
Akira (♂, 29)	1950 (173)	TO (♀, 9)	2045 (127)
Mari (♀, 29)	2092 (277)	TI (♀, 6)	1977 (150)
Ai (♀, 29)	1997 (7)	AK (♀, 4)	2098 (142)
Pen (♀, 28)	2135 (221)	MH (♀, 4)	2103 (93)
Popo (♀, 23)	2074 (277)	YM (♂, 4)	2076 (95)
Reo (♂, 23)	1960 (263)	TT (♀, 3)	2139 (82)
Ayumu (♂, 5)	2088 (279)	SI (♀, 2)	2002 (146)
Cleo (♀, 5)	1856 (24)	SY (♂, 2)	2034 (112)
Pal (♀, 5)	1976 (179)	SW (♂, 1)	2107 (79)
Average	2001 (91)	Average	2058 (66)

The list shows, for chimpanzees and for humans, the name of the vocalizer and the average length of the auditory stimuli. The dotted line in the Chimpanzees' column corresponds to the boundary between adult and infant chimpanzees. The dotted line in the Humans' column corresponds to the boundary between longer length of interaction (4–22 years interacting daily) and shorter length (1–3 years interacting daily) with target individuals

[a] The age of the chimpanzees at the point of voice-recording in 2005. The subject Pan was 22 years old, and she knew all the adult chimpanzees since her infancy. The chimpanzee named Chloe (female, 24-year old) was living in the same social group. She was omitted from the target individuals used as stimuli because she seldom uttered pant hoot calls. Pan herself was also omitted from the target individuals

[b] Years refer to the number of years the subject spent in daily contact with humans until 2005

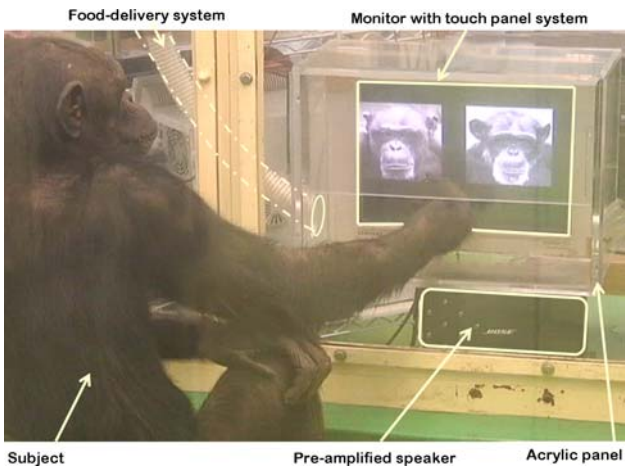

Fig. 1 The apparatus and the subject performing the AVIM task

Auditory and visual stimuli

The auditory stimuli consisted of segments of chimpanzee and adult human voices. The chimpanzee voices used were pant hoot calls from members of Pan's group. The targeted chimpanzees included nine adult (three males and six females) and three infant chimpanzees (one male and two females aged 5 year-old, including the daughter of the subject). Human voices were a segment of speech from caretakers and researchers, including six men and six women, who were in daily close contact with Pan at the moment of the experiment. Human speech was recorded opportunistically while the humans were interacting with the chimpanzees and their voices emitted in a friendly context. The words differed across samples and individuals. However, all the speech segments consisted of familiar Japanese words or sentences commonly used in everyday interactions with the chimpanzees, e.g., "well done", "good boy", "show me your hand please", "come on, come on". Chimpanzee vocalizations were recorded in the outdoor compound and the human voices in the caretakers' indoor area. All auditory samples were opportunistically recorded in the usual acoustic conditions in 2005, less than 6 months before the series of experiments were conducted. Thus, the three infants were nearly of the same age when the stimuli were recorded and when they were tested.

In previous experiments, the subject was already requested to recognize the voice of the adult chimpanzees and four of the adult humans used as target individuals (Hashiya and Kojima 2001b; Kojima 2003). However, the entire stimuli set used in this experiment stemmed from new recordings. A directional microphone (Sennheiser ME66) and a DAT walkman (Sony TCD-D100) at 48-kHz

sampling rate and AGC volume were used for the recordings. Selected segments of chimpanzee pant hoots and human speech were digitalized, with Audio/MIDI Interface Edirol UA-25 and sound card E-MU 0404, and then edited using Adobe Audition 1.5 software (48 kHz sampling rate and 16 bit precision). Background noises were removed from the stimuli as long as they did not overlap with the voice frequencies, but some sounds (such as flowing fountain, insect or bird songs) were sometimes audible after editing. Nonetheless, only the voice of one chimpanzee or human was audible in each stimulus.

In the process of editing stimuli, we carefully selected segments that preserved exhalation or inhalation phases of pant hoots. We usually omitted the climax part and instead preferred segments of introduction and buildup to homogenize the samples collected from adult and infant chimpanzees. Indeed, most of the pant hoots uttered by the three infants did not contain a climax phase, or other chimpanzees simultaneously pant hooting masked the climax phase. In the same way, human speech sequences were segmented between morphemes, usually short Japanese words (see examples in Figs. 2, 6). The auditory level was calibrated with a sound level meter (Rion NL-22, Japan) from the position commonly used by the subject during experimental sessions. The sound pressure level was 80 dB on average.

Visual stimuli consisted of still images from video records captured using a digital video camera (Sony DCR-HC90) in the same places and on the same days as auditory stimuli. Each footage was edited with Ulead Video Studio 8 and Adobe Photoshop Elements 3.0. The natural background (e.g., wall, corridor door, metallic structure) was only partially removed, just as the background noises from the auditory stimuli were. The size of each picture used on the monitor was 404 × 404 pixels (Fig. 2).

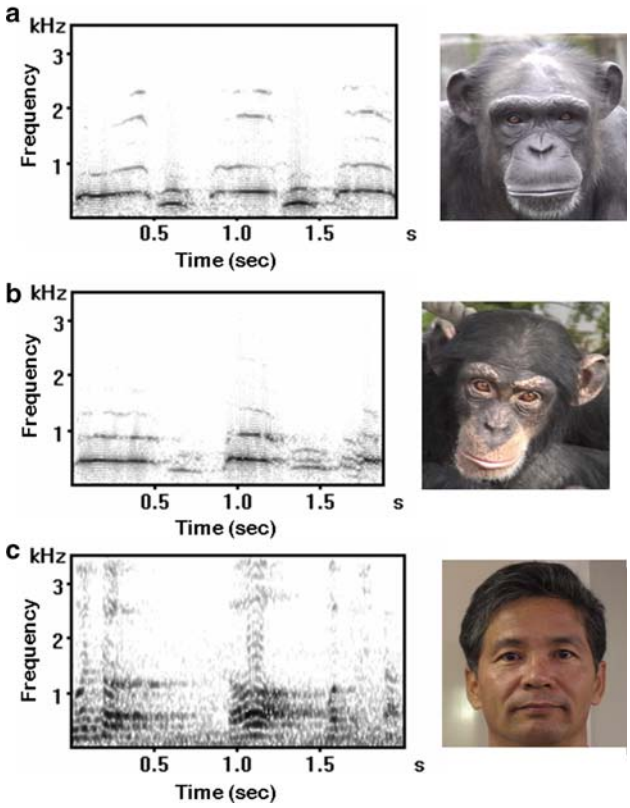

Fig. 2 Sonograms of auditory samples (pant hoot and human speech segments) and visual targets (frontal picture of face) used as stimuli in Experiment 1. **a** Adult chimpanzee, **b** infant chimpanzee, **c** adult human

General procedure

Throughout the study, we used an AVIM task. To focus Pan's attention on the monitor, before each trial, she was first required to consecutively touch a start key three to six times (red rectangle 5-cm wide × 3-cm high), which appeared at random positions on the monitor. The trial was initiated when the subject touched the last start key, which always appeared at the bottom-center of the monitor to keep the subject's hand in that position for measuring response time.

Each trial consisted of the successive presentation of one auditory sample and two alternative visual stimuli (matching individual was pitted against non-matching individuals). Immediately after the end of the voice playback, the two pictures appeared on the monitor at random positions. If Pan correctly chose the corresponding individual picture, she received a variety of food rewards through the food dispenser (Fig. 1). The intertrial interval (ITI) was 15 s. If Pan chose the non-matching individual picture, she got no reward and the trial was followed by ITI. We did not use any penalty, such as time out. To maintain the subject's motivation, this differential reinforcement was sustained throughout the trials.

One experimental session was conducted per day, and usually five times per week. The AVIM task was given right after a series of tests where Pan had to visually discriminate numbers and other non-pictorial shapes (Inoue and Matsuzawa 2007). Her daughter Pal was always in the adjacent twin booth also performing visuo-cognitive tasks. All statistical analyses were carried out on SPSS 16.0 software.

Experiment 1

Experiment 1 was designed to test whether Pan was able to generalize her abilities to match familiar individuals using new sets of voice and face stimuli. The experiment also aimed to test whether this transfer could extend to the voice and face of infant chimpanzees. For that purpose, in addition to familiar adult chimpanzees and familiar adult humans, we tested the recognition of newly introduced

infant chimpanzees. Adult humans differed in the number of years of daily proximity with the subject, and we therefore also tested the influence of that factor on intermodal individual recognition.

Vocal samples: pant hoot and human speech

The pant hoot vocalizations and human speech used were segmented to a constant length of about 2 s (average for pant hoot segments: 2,001 ms, SD = 91; and for human speech segments: 2,058 ms, SD = 66). Table 1 shows the average length of each pant hoot vocalization and human speech segment used as auditory stimuli.

Visual targets: pictures of faces

The pictures depicted a clear frontal view of the face, from the forehead to the chin, with direct gaze and a fairly neutral facial expression. Only one picture was used for each of the 24 target individuals (Fig. 2). By using one single visual target per individual, it was possible to finely test the individual recognition from each specific vocal sample. Moreover, a single visual target encouraged Pan to pay attention to the vocal samples, preventing possible confusions with other pure visual matching tasks.

Procedure

The 24 target individuals, consisting of 12 chimpanzees (3 adult males, 6 adult females, 1 infant male and 2 infant females) and 12 adult humans (6 men and 6 women) were used. The length of daily interaction between the subject Pan and target human individuals was used to evaluate the degree of exposure (Table 1).

The target individuals were divided into three stimulus conditions ("adult chimpanzees", "infant chimpanzees" and "adult humans") that were tested separately (within condition). In a first series of tests, we divided the target adult chimpanzees and adult humans into fixed trios of individuals (three trios of adult chimpanzees and four trios of adult humans) to maintain consistency with the sole trio of infant chimpanzees. Each trio was tested in two sessions. In each session, there were 24 trials, consisting of eight trials to test recognition of each of the three individuals in the trio. For each individual, the eight trials consisted of four different vocal samples that were pseudo-randomly presented twice (meaning each vocal sample was heard four times in total). As there were eight trios to be tested, there were a total of 16 experimental sessions and 48 trials in total per trio.

In the second test series, for "adult chimpanzee" and "adult human" stimulus conditions, all possible pairs of individuals were exhaustively tested. Thus, the subject was still requested to discriminate one out of three individuals, but the same pairs were never repeated within a session, except for the "infant chimpanzee" condition, which remained unchanged due to the small sample size. There was a total of 21 experimental sessions: 15 sessions for "adult chimpanzees + infant chimpanzees" and 6 sessions for "adult humans". In the "adult chimpanzee + infant chimpanzee" sessions, trials corresponding to the "adult chimpanzee" condition were pseudo-randomly mixed with trials corresponding to the "infant chimpanzee" condition. Each vocal sample was heard four or five times in total and a session consisted of 24 trials (63 trials per trio on average). Trials where the subject was troubled by distractions independent of the experiment were removed from analyses.

Results

Pan was able to perform individual recognition of familiar adult chimpanzees, infant chimpanzees and adult humans better than that predicted to occur by chance alone (two-tailed binomial test (0.5): $P < 0.001$ ($N = 333$), $P = 0.037$ ($N = 222$) and $P < 0.001$ ($N = 336$), respectively). Figure 3a shows the percentage of correct responses for each stimulus condition.

We found a statistically significant effect of the stimulus conditions on the frequency of correct versus incorrect responses (2×3 contingency table: $N = 888$ trials, $\chi^2 = 60.59$, $df = 2$, $P < 0.001$). We found no statistical difference between "adult chimpanzees" and "adult humans" (χ^2 with continuity correction; 2×2 contingency table: $N = 666$ trials, $\chi^2 = 1.88$, $df = 1$, $P = 0.170$). However, the frequency of correct versus incorrect responses was significantly different between "infant chimpanzees" condition and "adult chimpanzees" condition ($N = 552$ trials, $\chi^2 = 50.47$, $df = 1$, $P < 0.001$), and between "infant chimpanzees" condition and "adult humans" condition ($N = 558$ trials, $\chi^2 = 33.78$, $df = 1$, $P < 0.001$). Figure 3 shows the percentage of correct responses for each stimulus condition.

We found no statistically significant effect of sex of the target individuals in the frequency of correct versus incorrect responses in either the "adult chimpanzees" condition (χ^2 with continuity correction; 2×2 contingency table: $N = 330$ trials, $\chi^2 = 0.23$, $df = 1$, $P = 0.628$) or in the "adult humans" condition ($N = 336$ trials, $\chi^2 = 0.47$, $df = 1$, $P = 0.492$). The percentage of correct responses in each sex category was: chimpanzee males 75%, chimpanzee females 71%, human males 78% and human females 82%.

Pan showed discrimination performances above chance in the early phase of data collection for adult chimpanzee and adult human conditions. The mean percentage of correct responses on the first four trials in this experiment was 85% in the "adult chimpanzees" condition, 79% in the "adult humans" condition and 48% in the "infant chimpanzees"

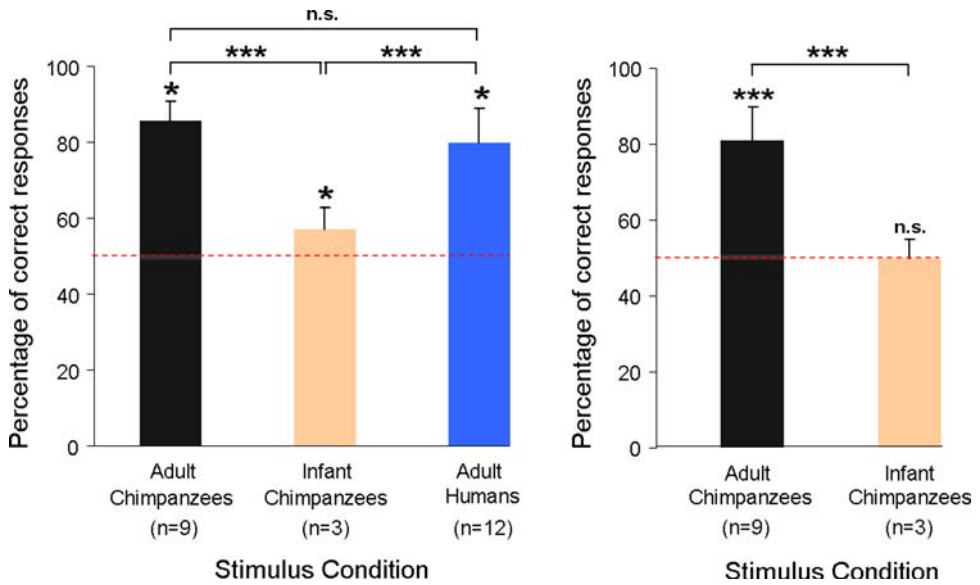

Fig. 3 Percentage of correct responses in Experiments 1 and 2. **a** Experiment 1: for the three categories of individuals used as stimuli: adult chimpanzees, infant chimpanzees (5 year-old) and adult humans. *Asterisks* indicate that the subject performed better than that predicted by chance in the three conditions (*$P < 0.05$) and indicate significance effect of the stimulus condition (***$P < 0.001$). **b** Experiment 2: for adult and infant chimpanzees. *Asterisks* indicate that the subject performed better than that predicted by chance for "adult chimpanzees" condition (***$P < 0.001$) and indicate significant effect of the stimulus condition (***$P < 0.001$)

condition. Pan's performance remained consistent through the sessions. The mean percentage of correct responses on the last four trials in this experiment was 83% in the "adult chimpanzees" condition, 82% in the "adult humans" condition and 56% in the "infant chimpanzees" condition. We found no statistically significant difference between the performance obtained during the first four trials and those obtained during the last four trials in either the "adult chimpanzees" condition (χ^2 with continuity correction; 2×2 contingency table: $N = 288$ trials, $\chi^2 = 0.105$, $df = 1$, $P = 0.746$), the "adult humans" condition ($N = 384$ trials, $\chi^2 = 0.410$, $df = 1$, $P = 0.522$) or the "infant chimpanzees" condition ($N = 96$ trials, $\chi^2 = 0.376$, $df = 1$, $P = 0.540$). Thus, Pan's performance remained consistent through the sessions indicating that the discrimination was not due to associative learning but to spontaneous recognition. However, the performance in the "infant chimpanzees" condition also showed that repeated exposure may have a positive effect on Pan's discrimination.

In sum, Pan could perform intermodal individual recognition of familiar adult chimpanzees and adult humans equally well. Thus, no species-specific effect was observed. On the other hand, Pan showed difficulty in performing individual recognition of infant chimpanzees. In terms of social proximity and degree of daily exposure, infant chimpanzees were closer to the subject than adult humans. Therefore, the daily length of interaction cannot explain the better performance of individual recognition for adult humans. Instead, the age of the target chimpanzees appears to be a key factor.

To investigate whether the duration of the relationship (in years) could cause this "age factor", we tested the effect of the duration of the relationship between the subject and the target humans. We divided the group of humans into two sub-groups (comparable in size to chimpanzees groups, adults versus infants) of eight individuals who had spent a longer period of time in interacting daily with the subject (4–22 years) and four individuals who had spent a shorter length of time in interacting daily with Pan (1–3 years) (Table 1). We found no statistical difference in the frequency of correct versus incorrect responses between the sub-groups (χ^2 with continuity correction; 2×2 contingency table: $N = 336$ trials, $\chi^2 = 0.19$, $df = 1$, $P = 0.662$). The percentage of correct responses was 81% for longer duration of interaction and 79% for shorter duration of interaction. These results suggest that the duration of relationship between Pan and the target humans did not affect Pan's ability to perform intermodal individual recognition and, consequently, the deficiency in infant chimpanzee recognition cannot be explained by a shorter duration of social interaction.

Experiment 2

This experiment was designed to assess whether the individual recognition of familiar chimpanzees observed in Experiment 1 could be generalized to new auditory samples. It predicts that if generalization occurs, the introduction of new auditory samples should not affect performance.

Set of stimuli and procedure

We used the same 12 familiar chimpanzees (9 adults and 3 infants) as in Experiment 1 as target individuals. For each of them, we doubled the number of auditory stimuli to eight vocal samples, namely the four vocal samples already used in Experiment 1 and four new vocal samples never used before. The visual stimuli were the same facial pictures that were used in the previous experiment to specifically test the effect of the introduction of new vocal samples. All the possible pairs of individuals within each stimulus condition ("adult chimpanzees", "infant chimpanzees") were exhaustively tested. One session consisted of 24 trials pseudo-randomly presented, six trials for "infant chimpanzees" condition, and 18 trials for "adult chimpanzees" condition. For each target individual, two different vocal samples were tested in a session (one vocal sample already used in Experiment l and one new vocal sample), so that a vocal sample was used only once within a session.

There were a total of 17 experimental sessions. In the "adult chimpanzees" condition, each vocal sample was heard four times (288 trials) and five times (120 trials) in the "infant chimpanzees" condition.

Results

As in Experiment 1, we found a statistically significant effect of stimulus condition on the frequency of correct versus incorrect responses (Chi-square with continuity correction; 2×2 contingency table: $N = 408$ trials, $\chi^2 = 38.45$, $df = 1$, $P < 0.001$). The introduction of new auditory samples did not affect the individual recognition of familiar adult chimpanzees ("adult chimpanzees" condition: binomial test (0.5), $P < 0.001$). In contrast, the individual recognition of familiar infants dropped to that predicted by chance alone (binomial test (0.5), $P > 0.99$) (Fig. 3b). Based on a larger set of auditory stimuli, which includes samples previously used and new vocal samples, we confirmed that Pan had difficulties in discriminating 5-year-old chimpanzees including her daughter.

We found no significant difference in the frequency of correct versus incorrect responses between the samples previously used and the new samples (χ^2 with continuity correction; 2×2 contingency table: $N = 528$ trials, $\chi^2 = 0.03$, $df = 1$, $P = 0.861$). The novelty of the sample did not to affect individual recognition, and the repetition of a previously heard vocal sample did not facilitate the recognition either. Thus, a potential learning association seemed to play a limited role in Pan's performance.

Experiment 3

This experiment was designed to test the effect of visual cues on the generalization of intermodal individual recognition. For that purpose, we introduced new visual targets depicting different pictorial perspectives of the same chimpanzees previously tested. In addition to pictures depicting a frontal view of the face, Pan was asked to match a chimpanzee's pant hoot to pictures depicting a frontal view of the whole body, a lateral view of the face or the back view of the body.

Visual targets: different pictorial perspectives

Four pictures were used for each of the 12 target chimpanzees. The pictures clearly depicted either a frontal view of the face with direct gaze and neutral facial expression, i.e., the same picture as in Experiment 1 ("frontal" condition), a frontal body view in a sitting position with a frontal view of the face ("body" condition), a lateral view of the face with neutral facial expression ("lateral" condition) or a back view of the body without facial cues ("back" condition) (Fig. 4). The auditory stimuli were the same as in Experiment 1.

Procedure

We used the same 12 familiar chimpanzees (9 adults and 3 infants) as in Experiments 1 and 2. For each individual, we used six vocal samples already tested in Experiments 1 and 2.

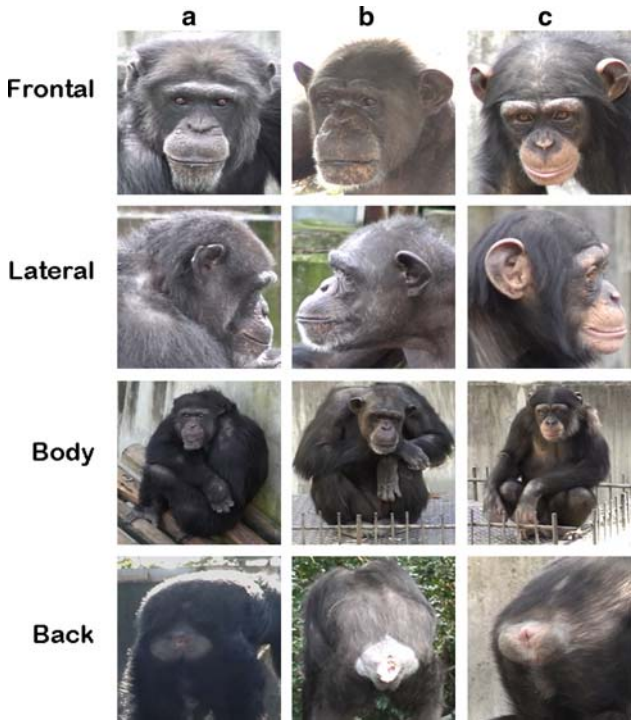

Fig. 4 Shows examples of the pictorial targets used in Experiment 3. Frontal, body, lateral and back correspond to the four perspectives tested for each targeted individual. **a** Adult male chimpanzee, **b** adult female chimpanzee, **c** infant male chimpanzee

We also used the four kinds of visual targets described above, i.e., "frontal", "body", "lateral" and "back" conditions. The target individuals were grouped into the same fixed trios of individuals as in Experiment 1 (three trios of adult chimpanzees, one trio of infant chimpanzees and four trios of adult humans).

The new visual targets were progressively introduced as probe trials with differential reinforcement. For each session, 2/3 of the trials presented the baseline perspective as visual targets ("frontal" condition), and 1/3 presented the new perspectives ("body", "lateral" or "back" conditions). A total of 13 sessions, consisting of 27 trials each were carried out (234 trials for frontal condition and 39 probe trials for each of "body", "lateral" and "back" conditions). Since the subject's performance appeared to be above that predicted by chance in the probe trials, additional sessions were conducted with the three new perspectives only. In this phase, the sessions were conducted for each of the three new conditions ("lateral", "body" and "back"). Each session included two trials for each of 12 individual that were pseudo-randomly mixed. A total of 14 sessions were carried out. In total, 830 trials were analyzed in this series of tests.

Results

Pan succeeded in intermodal individual recognition better than that predicted by chance alone in the four conditions using adult chimpanzees as target individuals (binomial test (0.5): $P < 0.001$ in the "frontal", "body" and "lateral" conditions and $P = 0.045$ for "back" condition). This suggests that Pan could generalize intermodal recognition of adult chimpanzees to new visuals targets, including back pictures.

When infant chimpanzees were target individuals, Pan failed to generalize to new visual stimuli. She did not perform better than that predicted by chance alone using any of the four conditions (binomial test (0.5): $P = 0.215$ in the "frontal" condition, $P = 0.405$ in the "body" and "lateral" conditions $P = 0.581$, and $P > 0.99$ in the "back" condition). The data from this experiment confirm that Pan was unable to correctly perform intermodal recognition of infant chimpanzees, regardless of the visual targets used.

Figure 5a shows the percentage of correct responses for each condition. Statistically significant differences were found between "back" and "frontal" conditions ($N = 404$, $\chi^2 = 14.14$, $df = 1$, $P < 0.001$), and between "back" and "body" conditions ($N = 313$, $\chi^2 = 12.15$, $df = 1$, $P < 0.001$) when adult chimpanzees were the target individual. Differences were not significant between "back" and "lateral" conditions, but showed a similar pattern ($N = 161$, $\chi^2 = 3.40$, $df = 1$, $P = 0.065$). Likewise, there was no statistically significant difference among "frontal", "lateral" and "body" conditions (2×2 contingency table with continuity correction: frontal–lateral: $N = 385$, $\chi^2 = 0.57$, $df = 1$, $P = 0.450$; frontal–body: $N = 537$, $\chi^2 = 0.00$, $df = 1$, $P > 0.99$; body–lateral: $N = 294$, $\chi^2 = 0.46$, $df = 1$, $P = 0.497$).

When infant chimpanzees were the target individuals, there were no statistically significant differences across all the four condition combinations (2×2 contingency table with continuity correction; frontal–lateral: $N = 78$, $\chi^2 = 0.00$, $df = 1$, $P > 0.99$; frontal–body: $N = 101$, $\chi^2 = 0.00$,

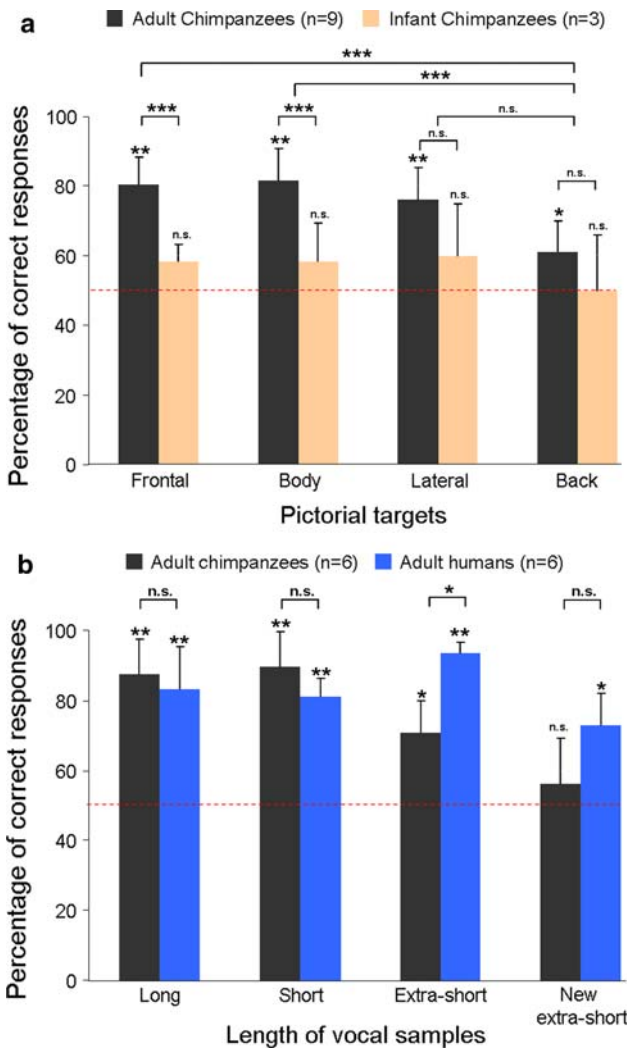

Fig. 5 Percentage of correct responses in Experiments 3 and 4. **a** Experiment 3: for adult chimpanzees and infant chimpanzees as target individuals. The *dotted line* shows chance level. *Asterisks* indicate that the subject performed better than that predicted by chance (**$P < 0.01$ and *$P < 0.05$). There was a significant difference in adult chimpanzees' recognition between "frontal" and "back" and between "body" and "back" conditions (***$P < 0.001$). Finally, there was a significant difference between adult and infant chimpanzees recognition in the "frontal" and "body" conditions (***$P < 0.001$). **b** Experiment 4: for adult chimpanzees and adult humans as target individuals. *Asterisks* indicate that the subject performed better than that predicted by chance (***$P < 0.001$ and **$P < 0.01$) and indicate a significant effect of the target individuals in the "extra-short" condition (**$P < 0.01$)

$df = 1$, $P > 0.99$; body–lateral: $N = 49$, $\chi^2 = 0.00$, $df = 1$, $P > 0.99$; back–frontal: $N = 83$, $\chi^2 = 0.14$, $df = 1$, $P = 0.710$); back–body: $N = 54$, $\chi^2 = 0.84$, $df = 1$, $P = 0.771$; back–lateral: $N = 31$, $\chi^2 = 0.07$, $df = 1$, $P = 0.786$).

Overall, visual targets depicting back seem to negatively affect intermodal individual recognition, although there was no significant difference between "lateral" and "back" conditions. In the back perspective, no facial cues were visible, whereas in the three other perspectives, the face was clearly visible. These data suggest that intermodal individual recognition is facilitated by the presence of facial cues in the visual targets, but this hypothesis still needs to be confirmed in other experiments.

Experiment 4

The goal of this experiment was to test the effect of auditory cues in the generalization of intermodal individual recognition. For that purpose, we tested whether the amount of auditory information provided in the vocal samples could affect individual recognition. Experiment 1 showed that Pan was able to discriminate equally well between familiar adult chimpanzees and adult humans. In this experiment, the length of the pant hoot and human speech segments previously used were shortened from an approximate average of 2,000 ms to an average length of 600 ms and then again to an average length of 200 ms.

Vocal samples: shortened pant hoot and human speeches

We used a set of six adult chimpanzees and six adult humans whom Pan had already been asked to recognize. For each individual, the frontal pictures used in Experiment 1 were cropped and a black background replaced the natural background. A preliminary test showed that these new pictures did not influence the subject's performance.

For each individual, three sets of auditory stimuli were prepared. The "long" set consisted of vocal samples of an approximate length of 2,000 ms (average for pant hoot segments: 1,983 ms, SD = 66; and for human speech segments: 2,027 ms, SD = 29). This set constituted a baseline set, since these vocal samples were all previously tested in experiment 1, 2 and 3.

In addition to the "long" set, two other new sets of auditory samples were prepared by shortening each vocal sample. The "short" set consisted of segments with an approximate length of 600 ms (average for pant hoot segments: 682 ms, SD = 27; for human speech segments: 652 ms, SD = 21). The "short" samples usually included sequence of two to three exhalation and inhalation phases for pan hoot segments and four to five morphemes for human speeches segments. The "extra-short" set consisted of segments with an approximate length of 200 ms (average for pant hoot segments: 225 ms, SD = 12; for human speech segments: 204 ms, SD = 4). The "extra-short" samples usually included sequence of one to two exhalation and inhalation phases for pant hoot segments and one to three morphemes for human speeches segments. The "new extra-short" samples were similar in length and structure to "extra-short" samples, except that those samples were prepared from other vocal samples than those used in the three other conditions. Figure 6 shows examples of sonograms for each set of stimuli.

Procedure

We used six familiar adult chimpanzees (three males and three females) and six familiar adult humans (three males and three females) as target individuals. We used the same set of vocal samples tested in Experiment 2. From this set, four vocal samples for each individual were used in the "long" condition, and later progressively shortened in the "short" and "extra-short" conditions. Thus, these three conditions were related since some physical cues of the "long" vocal samples were also present in the "short" and "extra-short" vocal samples. In contrast, in the "new extra-short" condition, we also used vocal samples from Experiment 2, but in this case they consisted of four vocal samples for each individual that were not used in the three previous conditions. Thus, any physical feature from the vocal samples used in "long", "short" and "extra-short" was replicated in the "new extra-short" vocal samples. All four conditions were tested chronologically.

A limited number of sessions were conducted to minimize a possible effect of facilitation caused by an extensive repetition of the stimulus presentation. Each vocal sample was used only once within a session and each vocal sample was heard twice in this experiment. The pair of individuals used as alternative choices was combined according to a Latin square. Thus, within a condition (tested separately), all the possible pairs were compared at least once. A total of 16 sessions of 12 trials were carried out (four sessions for each "long", "short", "extra-short" and "new extra-short" conditions) for adult chimpanzees as well as for adult humans. Thus, a total of 384 trials were conducted (192 trials for adult chimpanzees and 192 trials for adult humans). The visual stimuli were the same frontal face pictures as that used in Experiment 1.

Results

Pan succeeded in intermodal individual recognition of both adult chimpanzees and adult humans. In the case of adult humans as target individuals, she performed better than that predicted by chance alone with all four types of auditory

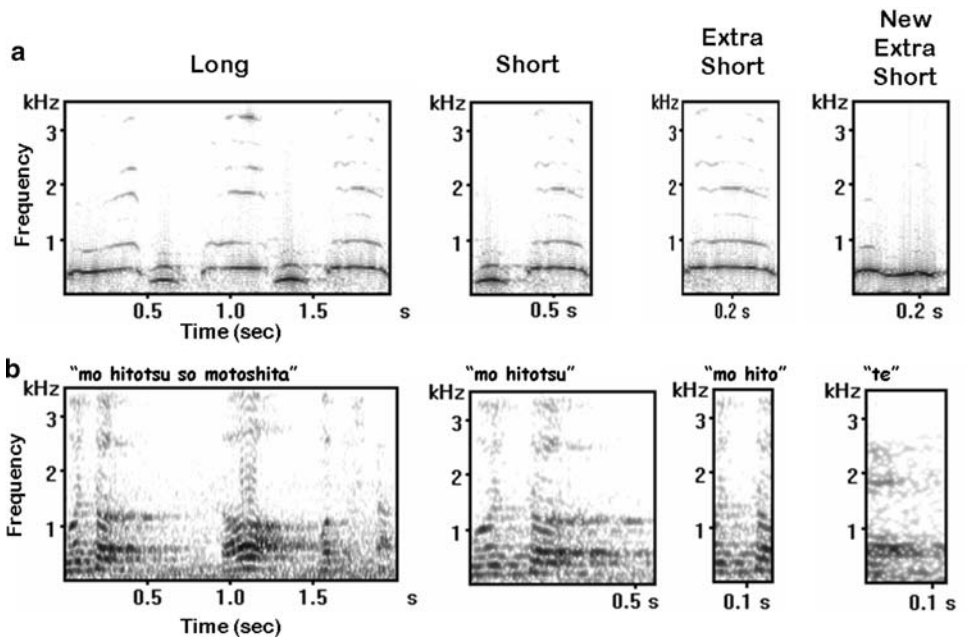

Fig. 6 Sonograms of auditory samples (pant hoot call and human speech segments) used as stimuli in Experiment 4. Length of auditory samples: "long" condition, about 2,000 ms; "short" condition, about 600 ms; "extra-short" and "new-extra short" conditions, about 200 ms. **a** Adult chimpanzee, **b** adult human, with corresponding Japanese words *above*

samples (binomial test (0.5): $P < 0.001$ in the "long", "short" and "extra-short" conditions and $P = 0.002$ in the "new extra-short" condition). This suggests that Pan can generalize intermodal recognition of adult humans with increasingly shortened vocal samples as stimuli.

When adult chimpanzees were the target individuals, Pan performed better than that predicted by chance alone with three types of auditory samples (binomial test (0.5): $P < 0.001$ in the "long", "short" and $P = 0.006$ in the "extra-short" conditions). However, her performance was not significantly better than that predicted by chance alone in the "new extra-short" condition ($P = 0.471$). Pan was clearly successful in generalizing intermodal recognition of adult chimpanzees with shortened vocal samples, but her performance deteriorated when vocal samples with an approximate length of 200 ms were not repeatedly heard before in a longer version (Fig. 5b).

We found a statistically significant influence of the length of the vocal samples on the frequency of correct *versus* incorrect responses for individual recognition of chimpanzees (2×4 contingency table: Chi-square: $N = 192$, $\chi^2 = 19.326$, $df = 3$, $P < 0.001$). However, we did not find a significant influence of length on the frequency of correct versus incorrect responses for individual recognition of humans (2×4 contingency table: $N = 192$, $\chi^2 = 7.428$, $df = 3$, $P = 0.059$). These results suggest that, for Pan, human voices were more consistently recognized in very short vocal samples than chimpanzee voices.

We found a statistically significant influence of species (i.e., chimpanzee target individuals versus human target individuals) in the "extra-short" condition (2×2 contingency table with continuity correction: $N = 96$, $\chi^2 = 7.148$, $df = 1$, $P = 0.008$), but not in the other three conditions (2×2 contingency table with continuity correction; "long": $N = 96$, $\chi^2 = 0.084$, $df = 1$, $P = 0.772$; "short": $N = 96$, $\chi^2 = 0.753$, $df = 1$, $P = 0.386$; "new extra-short": $N = 96$, $\chi^2 = 2.231$, $df = 1$, $P = 0.135$).

Finally, Pan's overall frequency of correct versus incorrect responses was not affected by the species (chimpanzees: 75%, humans: 88%), or their sex (chimpanzee males: 75%, human males: 95% and chimpanzee females: 75%, human females: 82%) (2×2 contingency table: $N = 192$, $\chi^2 = 0.243$, $df = 1$, $P = 0.622$).

General discussion

Auditory–visual intermodal matching

This series of experiments was designed to explore the abilities of a female chimpanzee named Pan, who became an expert in AVIM task during her infancy. The first experiments confirmed that she was able to perform AVIM with a larger number of familiar individuals and with new sets of vocal samples. Thus, Pan's abilities to identify familiar conspecifics are not constrained to a limited collection of vocal samples (Kojima 2003). Pan demonstrated the capacity to spontaneously match a vocal sample with the picture of the corresponding vocalizer from the very first few trials. Likewise, she could also easily recognize the voice of familiar humans who did not take part in previous experiments. If Pan's performance were mainly based on associative learning, we would expect a clear increase of performance for samples repeatedly presented, but it did

not occur. In Experiment 1, the results showed that repeated exposure to the stimuli did not have a significant effect on Pan's performance. However, since we ran a large number of trials, our results should be interpreted with caution. In addition to Pan's ability to perform spontaneous discrimination, Pan may have performed associative learning when she could not easily find the correct response, as it may have occurred in Experiment 4.

The pant hoot is typically a male call (Mitani and Brandt 1994). Therefore, one might expect that Pan should recognize males from pant hoot vocalizations more easily. However, our results show that Pan recognized familiar chimpanzees from both sexes equally well. Clearly, the pant hoot can be used as an auditory cue useful in the recognition of all familiar conspecifics by chimpanzees.

This study also tested a chimpanzee's ability to recognize adult human speech recorded in "naturalistic" condition, i.e., from opportunistic recordings in everyday interactions with the chimpanzees. This method made the human speech recordings more comparable to the pant hoot recordings. Although there was a lot of morphological variation in human speech samples, Pan was good at performing individual recognition. Many studies have demonstrated the existence of a strong species-specificity effect in face recognition (Matsuzawa 1990; Fujita 1987). However, data from this study did not show such species specificity probably because the subject was raised by humans right after birth.

Overall, Pan's performance showed that chimpanzees are able to have intermodal representation of familiar individuals. In their natural habitat, chimpanzees spend most of their time in a partially occluded environment and often split in sub-groups. Thus, it makes sense that Pan, as all the members of her species, possesses the ability to acoustically recognize an invisible individual (Goodall 1986). However, up to now, Pan is the only chimpanzee that has been successfully trained to perform AVIM based on individual recognition. The AVIM task is an extremely demanding cognitive task, and it is probably among the most difficult tasks for a chimpanzee to learn in experimental conditions (Matsuzawa 2009). The reasons explaining this difficulty still remain obscure. Five other chimpanzees, experts in complex visual tasks and living in the same environment as Pan, have been trained to perform several auditory–visual intermodal tasks. Yet, after 1 year of continuous training, their performances remain at the chance level (L. Martinez and T. Matsuzawa, unpublished data).

Individual recognition of infant chimpanzees

In contrast to Pan's nearly perfect individual recognition of familiar adults of both species, we found a clear decrease in her success performing individual recognition of 5-year-old infant chimpanzees. This difficulty continued even after repeatedly using the same samples of pant hoot segments across sessions and conditions. The three target infant chimpanzees were living in Pan's social group for at least for 4 years prior to the experiment. Why, then, was it difficult for Pan to perform intermodal individual recognition of infant chimpanzees? Several explanations may account for this result.

During our collection of auditory stimuli, we observed that infants usually emit pant hoot simultaneously or just after one or several adults have started to pant hoot. This is consistent with observations done in free-living chimpanzees (Marler and Tenaza 1977; Plooij 1984). An infant's pant hoot is uttered with a lower acoustic intensity and often omits some parts of a typical adult pant hoot sequence, especially the climax part, which is known to contain highly distinctive individual cues (Clark Arcadi et al. 1998; Riede et al. 2004). Nonetheless, it must be noted that Pan was able to utilize the non-climax building part as a cue for individual recognition of adult chimpanzees. Thus, repeated opportunities for hearing the pant hoot uttered by one voice without interferences from other voices may also be critical for individual recognition.

Another possible explanation for Pan's difficulty in performing intermodal individual recognition of infant chimpanzees is that mother chimpanzees might be able to discriminate their offspring's voice using other call types in the early stages of development. For example, more commonly uttered and socially important vocalizations such as staccatos, screams or whimpers may be more important for individual recognition than the pant hoot, which is developed much later in life. Thus, individual recognition of infants based on the pant hoot call may be limited.

The data presented in this study showed that Pan had difficulty in individual recognition of infant chimpanzees using AVIM tasks, but was able to recognize familiar adult humans. Yet, the three infants and especially Pan's daughter lived in the same social group as Pan and spent far more time with her than humans. Clearly, familiarity based on social and physical proximity cannot be the only determinant of auditory–visual intermodal recognition.

Visual cues

Data from Experiment 3 demonstrated that intermodal individual recognition was consistent when pictures were changed. However, a discrepancy of performance was observed between "frontal" and "back" conditions, and this is congruent with the highly specific functional mechanism for intermodal face and voice processing found in humans. The rich repertoire of chimpanzee vocalizations, their fission–fusion social organization and their lifelong social

relationships may have enhanced a similar bias to preferentially perform intermodal individual recognition by specifically associating the cues of the face and the voice.

Auditory cues

Intermodal individual recognition by Pan could occur even when previously heard vocal samples are gradually shortened to 200 ms ("extra-short" condition). However, her performance decreased on the novel 200-ms human stimuli and did not show spontaneous generalization to newly introduced 200-ms chimpanzee vocal samples ("new extra-short" condition). Thus, her ability to successfully perform recognition from 200-ms samples appears to be unsteady. However, this value might be one indication of a minimal length, or a detection threshold in pant hoot inhalation and exhalation phases, necessary to efficiently recognize a familiar chimpanzee based on its pant hoot. Alternatively, repeated exposure to "long" and "short" samples may have helped Pan to discriminate the "extra-short" samples, since these three types of samples shared common physical cues. Therefore, Pan could have been simply responding to an association between specific physical cues presented across the vocal samples and the target picture. Yet, both alternative explanations suggest that only one or two inhalation and/or exhalation phases in the introduction or buildup parts of the pant hoot may contain enough acoustic information to identify a familiar voice or familiar physical cues.

Similarly, one or two morphemes of 200 ms of human speech seemed to be sufficient for Pan to successfully perform recognition. The slight superiority of human voice discrimination in comparison to chimpanzees' voices might be due to species differences in the acoustic structure of the target voices. For example, human voices are produced by voluntary control over the vocal cords and results in the production of formant transitions (Denes and Pinson 1993). In one single phoneme, such as "te" or "so", the human voice contains several individuality-related cues such as pitch or formant transitions. The kind of variation in acoustic cues evident in human communication might have helped Pan in human individual recognition. Previous investigations have shown a significant influence of pitch in the individual recognition in chimpanzees, but the absence of articulated speech may limit the acoustic features of their voice (Kojima 2001; Kojima et al. 2003). However, some chimpanzee vocalizations such as "huu" or "hoo" calls can be much shorter than 200 ms (Marler and Tenaza 1977). Therefore, the individual recognition of a familiar individual signaling puzzlement or distress to invisible social partners might be a highly adaptive ability that chimpanzees most probably possess. Further investigations will be necessary to better determine whether Pan can successfully recognize a familiar individual using vocalization segments of about 200 ms in length from calls other than pant hoot.

In summary, the present study has shown that a chimpanzee can successfully perform intermodal auditory–visual matching and can generalize this ability to new stimuli, but with some limitations. Previous studies have shown that familiarity is an important factor in unmistakably recognizing individuals based on either their voice or face (Van Lancker and Kreiman 1987; Assmann and Summerfield 2004; Ellis and Lewis 2001; Martin-Malivel and Okada 2007). This may also be true in the ability of chimpanzees to perform individual recognition. The most innovative findings in neuroimagery have emphasized the existence of specific multimodal face–voice integration regions that seem to play an important role from earlier stages of identity processing (Campanella and Belin 2007). Our data suggest that simultaneous exposure to different modalities, such as voices and faces, may be essential for chimpanzees to develop a multimodal, multi-layered knowledge of familiar individuals. As has been suggested in humans, this almost permanent face–voice simultaneous exposure could have possibly led to the development of similarly high specialized brain regions in chimpanzees as well (Campanella and Belin 2007; Calvert 2001).

Our study also demonstrates that there are limits to visual cues and thresholds to auditory cues necessary for successful intermodal individual recognition. Pan's difficulties in discriminating some stimuli associations may also reflect a chimpanzee's natural ability to learn some relationships more easily than others. There may be numerous influential factors shaping the auditory–visual intermodal recognition in a chimpanzee. Factors such as the consistency of the physical cues, the social relevance of the stimuli, the intermittence of the stimulus occurrence and differential perceptual capabilities among individuals merit further investigations.

Acknowledgments We would like to acknowledge: for their help in running experiments, D. Biro, S. Inoue, E. Nogami and T. Takashima; for their voice/face contributions, K. Hashiya, M. Hayashi, S. Hirata, T. Imura, S. Inoue, A. Kato, K. Kumazaki, N. Maeda, T. Matsuno, E. Nogami, T. Ochiai, T. Takashima, M. Tanaka, M. Tomonaga, S. Watanabe, S. Yamamoto and S. Yamauchi; for the daily care of chimpanzees, all the staff of the Center for Human Evolution Modeling Research, especially to R. Hirokawa, K. Kumazaki, N. Maeda and S. Watanabe; for technical help in recording, programming and interfacing, A. Izumi, A. Lemasson, S. Nagumo and M. Tanaka; for scientific advice, K. Hashiya, A. Izumi, S. Kojima and M. Tomonaga. This study was supported by Grant-in-Aid for Scientific Research, Ministry of Education, Science, Sports and Culture (MEXT, Japan) No. 16002001 and No. 20002001 and JSPS-ITP-HOPE, JSPS-GCOE (D07 for Psychology and A06 for Biodiversity) to T. Matsuzawa and by a MEXT scholarship to L. Martinez. We would like to express our sincere gratitude to J.B. Leca for his precious help in analyzing the

data, reading and comments on earlier drafts of this manuscript, to A.D. Hernandez for his valuable reading and English corrections and to C. Garcia, N. Granier and C.A.D. Nahallage for their help. We also thank the three anonymous reviewers for their constructive comments and suggestions.

References

Assmann P, Summerfield Q (2004) The perception of speech under adverse conditions. In: Greenberg S, Ainsworth WA, Popper AN, Fay RR (eds) Speech processing in the auditory system. Springer handbook of auditory research. Springer, New York, pp 231–308

Bauer HR, Philip M (1983) Facial and vocal individual recognition in the common chimpanzee. Psychol Res 33:161–170

Belin P, Zatorre R, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human cortex. Nature 403:309–311

Boysen ST (1994) Individual differences in the cognitive abilities of chimpanzees. In: Wrangham RW, McGrew WC, de Waal FBM, Heltne PG (eds) Chimpanzee cultures. Harvard University Press, Cambridge, pp 335–350

Bruce V, Young A (1986) Understanding face recognition. Br J Psychol 77:305–327

Calvert GA (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb Cortex 11:1110–1123

Campanella S, Belin P (2007) Integrating face and voice in person perception. Trends Cogn Sci 11:535–543

Cheney DL, Seyfarth RM (1990) How monkeys see the world: inside the mind of another species. The University of Chicago Press, Chicago

Clark AP, Wrangham RW (1994) Chimpanzee arrival pant hoots: do they signify food or status? Int J Primatol 15:185–205

Clark Arcadi A (1996) Phrase structure in wild chimpanzee pant hoots: patterns of production and interpopulation variability. Am J Primatol 39:159–178

Clark Arcadi A, Robert D, Boesch C (1998) Buttress drumming by wild chimpanzees: temporal patterning, phrase integration into loud calls, and preliminary evidence for individual distinctiveness. Primates 39:505–518

Denes PB, Pinson EN (1993) The speech chain: the physics and biology of spoken language. W.H. Freeman and Company, New York

Dubois S, Rossion B, Schiltz C, Bodart JM, Michel C, Bruyer R, Crommelinck M (1999) Effect of familiarity on the processing of human faces. Neuroimage 9:278–289

Dufour V, Pascalis O, Petit O (2006) Face processing limitation to own species in primates: a comparative study in brown capuchins, tonkean macaques and humans. Behav Processes 73:107–113

Ellis HD, Lewis MB (2001) Capgras delusion: a window on face recognition. Trends Cogn Sci 5:149–156

Ellis HD, Jones DM, Mosdell N (1997) Intra- and intermodal repetition priming of familiar faces and voices. Br J Psychol 88:143–156

Fitch WT, Fritz JB (2006) Rhesus macaques spontaneously perceive formants in conspecific vocalizations. J Acoust Soc Am 120:2132–2142

Fujita K (1987) Species recognition by five macaque monkeys. Primates 28:353–366

Goodall J (1986) The chimpanzees of Gombe: patterns of behavior. Harvard University Press, Cambridge

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

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

Inoue S, Matsuzawa T (2007) Working memory of numerals in chimpanzees. Curr Biol 17:R1004–R1005

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

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

Kamachi M, Hill H, Lander K, Vitikiotis-Bateson E (2003) Putting the face to the voice: matching identity across modality. Curr Biol 13:1709–1714

Kojima S (2001) Early vocal development in a chimpanzee infant. In: Matsuzawa T (ed) Primate origins of human cognition and behavior. Springer, Tokyo, pp 190–196

Kojima S (2003) Mapping the origins of human speech. A search for the origins of human speech. Auditory and vocal functions of the chimpanzee. Kyoto University Press, Kyoto

Kojima S, Kiritani S (1989) Vocal–auditory functions in the chimpanzee: vowel perception. Int J Primatol 10:199–213

Kojima S, Tatsumi IF, Kiritani S, Hirose H (1989) Vocal–auditory functions of the chimpanzee: consonant perception. Hum Evol 4:403–416

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

Marler P, Hobbet L (1975) Individuality in a long-range vocalization of wild chimpanzees. Z Tierpsychol 38:97–109

Marler P, Tenaza RC (1977) Signaling behavior of apes with special reference to vocalization. In: Sebeok TE (ed) How animals communicate. Indiana University Press, Bloomington, pp 965–1033

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 (2001) Primate origins of human cognition and behavior. Springer, Tokyo

Matsuzawa T (2003) The Ai project: historical and ecological contexts. Anim Cogn 6:199–211

Matsuzawa T (2006) Sociocognitive development in chimpanzees: a synthesis of laboratory work and fieldwork. In: Matsuzawa T, Tomonaga M, Tanaka M (eds) Cognitive development in chimpanzees. Springer, Tokyo, pp 3–33

Matsuzawa T (2009) Chimpanzee mind: looking for the evolutionary roots of the human mind. Anim Cogn. doi:10.1007/s10071-009-0277-1

Matsuzawa T, Tomonaga M, Tanaka M (2006) Cognitive development in chimpanzees. Springer, Tokyo

Mitani JC (1994) Ethological studies of chimpanzee vocal behaviour. In: Wrangham RW, McGrew WC, de Waal FBM, Heltne PG (eds) Chimpanzee cultures. Harvard University Press, Cambridge, pp 241–254

Mitani JC, Nishida T (1993) Contexts and social correlates of long-distance calling by male chimpanzees. Anim Behav 45:735–746

Mitani JC, Brandt KL (1994) Social factors influence the acoustic variability in the long-distance calls of male chimpanzees. Ethology 96:233–252

Mitani JC, Gros-Louis J, Macedonia JM (1996) Selection for acoustic individuality within the vocal repertoire of wild chimpanzees. Int J Primatol 17:569–581

Myowa-Yamakoshi M (2006) Development of facial information processing in nonhuman primates. In: Matsuzawa T, Tomonaga M, Tanaka M (eds) Cognitive development in chimpanzees. Springer, Tokyo, pp 142–154

Nakamura K, Kawashima R, Sugiura M, Kato T, Nakamura K, Hatano K, Nagumo S, Kubota K, Fukuda H, Ito K, Kojima S (2001) Neural substrates for recognition of familiar voices: a PET study. Neuropsychologia 39:1047–1054

Neiworth JJ, Hassett JM, Sylvester CJ (2007) Face processing in humans and new world monkeys: the influence of experiential and ecological factors. Anim Cogn 10:125–134

Notman H, Rendall D (2005) Contextual variation in chimpanzee pant hoots and its implications for referential communication. Anim Behav 70:177–190

Parr LA, Winslow JT, Hopkins WD, de Waal FBM (2000) Recognizing facial cues: individual recognition in chimpanzees (*Pan troglodytes*) and rhesus macaques (*Macaca mullata*). J Comp Psychol 114:47–60

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

Plooij FX (1984) The behavioral development of free-living chimpanzee babies and infants. Ablex, Norwood

Porter RH, Cernoch JM, McLaughlin FJ (1983) Maternal recognition of neonates through olfactory cues. Physiol Behav 30:151–154

Rendall D, Rodman PS, Emond RE (1996) Vocal recognition of individuals and kin in free-ranging rhesus monkeys. Anim Behav 51:1007–1015

Riede T, Owren MJ, Clark Arcadi A (2004) Nonlinear acoustics in pant hoots of common chimpanzees *(Pan troglodytes)*: frequency jumps, subharmonics, biphonation, and deterministic chaos. Am J Primatol 64:277–291

Savage-Rumbaugh S, McDonald K, Sevcik RA, Hopkins WD (1986) Spontaneous symbol acquisition and communicative use by pygmy chimpanzee (*Pan paniscus*). J Exp Psychol Gen 115:211–235

Savage-Rumbaugh S, Sevcik RA, Hopkins WD (1988) Symbolic cross-modal transfer in two species of chimpanzees. Child Dev 59:617–625

Snowdon CT, Cleveland J (1980) Individual recognition of contact calls by pygmy marmosets. Anim Behav 28:717–727

Tanaka M (2001) Discrimination and categorization of images of natural objects by chimpanzees (*Pan troglodytes*). Anim Cogn 4:201–211

Tomonaga M, Itakura S, Matsuzawa T (1993) Superiority of conspecific faces and reduced inversion effect in face perception by a chimpanzee. Folia Primatol 61:110–114

Vokey JR, Rendall D, Tangen JM, Parr LA, de Waal FBM (2004) Visual kin recognition and family resemblance in chimpanzees (*Pan troglodytes*). Comp Psychol 115:194–199

Van Lancker D, Kreiman J (1987) Voice discrimination and identification are separate abilities. Neuropsychologia 25:829–834

Von Kriegstein K, Giraud AL (2006) Implicit multisensory associations influence voice recognition. PLoS Biol 4:326