Title	Integration of auditory and visual information in human face discrimination in humans and pigeons
Sub Title	
Author	渡辺, 茂(Watanabe, Shigeru) 辻井, 岳雄(Tsujii, Takeo) 山本, 絵里子(Yamamoto, Eriko)
Publisher	Centre for Advanced Research on Logic and Sensibility The Global Centers of Excellence Program, Keio University
Publication year	2008
Jtitle	CARLS series of advanced study of logic and sensibility Vol.1, (2007.) ,p.7- 34
JaLC DOI	
Abstract	Social stimuli are often multi-sensory. Animals recognize a con specific often through visual, auditory and olfactory sensory cues. Human face recognition has also audio-visual interaction, for example, McGurk effects. Using morphing stimuli consisting of human and monkey faces, we found that presentation of human or monkey voice affected judgment of visually ambiguous face stimuli (Experiment 1). Human ERP study using the morphing stimuli and two voices showed facilitative effect of human voice on visual ERP (Experiment 2). Pigeons were trained on discrimination between human and monkey faces then tested with morphing images of the human and the monkey faces (Experiment 3). Then, the morphing images were presented with auditory stimuli, namely, vocalization of human or monkey. Facilitative effect of human voice on human face discrimination was demonstrated but that of the monkey voice was not. The subjects had been exposed to the human voice in the living cages but not to the monkey face or monkey vocalization. This difference in experience should result in difference in effects of auditory stimuli on visual discrimination. Anatomical study (Experiment 4) provided evidence that visual and auditory pathways were independent at the level of primary sensory area in the telencephalon (entopallium and field L). Thus, the visual and auditory systems should integrated in higher associative area, such as neopallium caudo-lateralis (NCL). These results agreed to human ERP study, which suggested integration of two sensory systems in later stage of sensory processing.
Notes	Part 1: Brain and Evolution
Genre	Research Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO12002001-20080331-0007

1

Integration of Auditory and Visual Information in Human Face Discrimination in Humans and Pigeons

Shigeru Watanabe[1,2] *Takeo Tsujii*[2], *and Eriko Yamamoto*[2]

[1] Department of Psychology, Keio University
[2] Centre for Advanced Research on Logic and Sensibility (CARLS), Keio University

Summary

Social stimuli are often multi-sensory. Animals recognize a conspecific often through visual, auditory and olfactory sensory cues. Human face recognition has also audio-visual interaction, for example, McGurk effects. Using morphing stimuli consisting of human and monkey faces, we found that presentation of human or monkey voice affected judgment of visually ambiguous face stimuli (Experiment 1). Human ERP study using the morphing stimuli and two voices showed facilitative effect of human voice on visual ERP (Experiment 2). Pigeons were trained on discrimination between human and monkey faces then tested with morphing images of the human and the monkey faces (Experiment 3). Then, the morphing images were presented with auditory stimuli, namely, vocalization of human or monkey. Facilitative effect of human voice on human face discrimination was demonstrated but that of the monkey voice was not. The subjects had been exposed to the human voice in the living cages but not to the monkey face or monkey vocalization. This difference in experience should result in difference in effects of auditory stimuli on visual discrimination. Anatomical study (Experiment 4) provided evidence that visual and auditory pathways were independent at the level of primary sensory area in the telencephalon (entopallium and field L). Thus, the visual and auditory systems should integrated in higher associative area,

such as neopallium caudo-lateralis (NCL). These results agreed to human ERP study, which suggested integration of two sensory systems in later stage of sensory processing.

Humans have a well-developed ability to recognize many different human faces. Because the elements of the face such as, the eyes, nose and mouth are always positioned in a similar manner, discrimination between faces requires analysis at a more refined level (see Barton, Keenan, & Bass, 2001). Neuropsychological studies of prosopagnisia suggest a particular region of the brain (fusiform gyrus) is involved in human face perception (Moskovich, Winocur, & Behrman, 1997). In addition, studies have shown that human face recognition is different from recognition of non-face objects. Specifically, face recognition depends on configural processing of the individual features within the face, while the non face recognize depends more on individual featural analysis (Farah, Wilson, Drain & Tanaka, 1998; Masuda, 2004). Human face recognition by non-human primates has been examined both behaviorally and physiologically (Vermeire & Hamilton, 1998; Phelps & Roberts, 1994; Desimone, 1991).

Because the human face is composed of several individual elements, it has been used as example of complex discriminative stimulus in operant discrimination studies with birds (for example, Troje, Huber, Loidolt, Aust & Fieder, 1999; Makino & Jitsumori, 2000; Loidolt, Aust, Meran, & Huber, 2003). Crows have been shown to discriminate between human faces and human face expressions (Kusayama, 2004). While humans' recognition of human faces is impaired by upside-down presentation (the inversion effect), chimpanzees (Pan troglodytes) do not show such effect (Tomonaga, Itakura, & Matsuzawa, 1993). Is this face inversion effect unique to humans? The human face is a stimulus that is most frequently viewed by humans in its upright position. It can be argued that we become experts at viewing human faces in their normal orientation. If this is true, recognition deficits would be seen for other frequently viewed stimuli when they are presented upside down. In fact, dog experts showed a stimulus inversion effect in a dog recognition task (Diamond & Carey, 1986).

Visual conspecific recognition has been examined with not only humans and other primates but also with birds (Chickens: Ryan & Lea, 1994, Budgerigars: Brown & Dooling, 1992, pigeons: Watanabe & Ito, 1991, Java sparrow; Watanabe & Jian, 1993). In each of these experiments, birds were able to discriminate individual conspecifics by visual information. Using operant conditioning, Stoddard, Beecher and Leosche (1992) successfully trained white crown sparrows to discriminate many conspecific song repertoires. Clearly, many bird species have the ability to discriminate individual conspecifics with visual or auditory cues.

One interesting topic in human face recognition is audio-visual interaction. Typical example of audio-visual interaction in human face-voice cognition is McGurk effects (McGurk & MacDonald, 1976). The heard syllable/ba/ when synchronized with a speaker seen to be saying /ga/ was often heard as /da/. Additional research supports the notion that speech perception can be facilitated by visual display of the face (Vatikiotis-Bateson, 1998; Munhall & Tohkura, 1998). Kamachi, Hill and Lander (2000) demonstrated cross-modal matching of human face and voice also. These results suggest integration of multi-sensory processing of the human faces by humans. However, modification of visual perception by auditory stimulus has been not reported. Probably the visual cues dominate the auditory cues in humans.

In natural settings, social stimuli are often multi-sensory. For instance, animals recognize a conspecific often through visual, auditory and olfactory sensory cues (Rowe, 1999). In multimodal information processing, a lack of one modality may be compensated by other sensory channels. When conveying socially relevant information, many birds exhibit both visual auditory signals. Recently, Partan, Yelda, Price and Shimizu (2005) showed to female pigeons, videotaped playback of male courtship display. Some of the playbacks contained auditory information without the video image, while other playbacks contained only visual information. The results showed an enhancement of elicited male courtship display beyond the individual effects of visual or auditory stimuli when both auditory and visual signals were presented together. Watanabe and Jian (1993) trained Bengalese finches on an individual conspecific discrimination task in which still images and contact calls were simultaneously presented. The subject birds appeared to rely on visual cues more so than the auditory cues when both signals were from the same individual. However, when chimeras of visual images (e.g., a head of one individual connected with a body of other individual) were presented the subjects used the contact call for the discrimination. Thus, a dominance order of the sensory modality was visual then auditory but the second one compensates the first one when the first sensory modality did not provide enough information. In this case, multi-sensory system is a kind of insurance system of conspecific recognition.

Computer generated stimuli are a versatile alternative to videotaped playbacks and can be used for the investigation of visual cognition in animals (see Watanabe and Troje, in press). One specific form of computer-generated stimuli involves morphing, which has been used to create composite face (Steyvers, 1999). By changing morphing rate, we can produce stimuli along likeness continuum, namely from complete stimulus A to complete stimulus B. Inter-medium stimuli have features of both A and B depending on morphing rate. Previously, we made morphing images of pigeon and starling and reported gradient of responding along the morphing rate scale after discriminative training with pigeon and starling faces in pigeons (Watanabe & Furuya, 1997). There needs to be a sentence or two here discussing the conclusions of this study and what is still unclear. For instance, auditory signals were not included in that study. Therefore, it is still unclear if the auditory signal can facilitate conspecific recognition in a morphing procedure.

Experiment 1: Human behavioral study

In the first experiment, we examined effects of voices of human or monkey on visual recognition of these two species by human observers. By using morphing stimuli, we presented several ambiguous stimuli in addition to the complete human or monkey faces.

Methods

Participants

10 adults humans (5 males, 5 females, averaged age 26.7 ranging from 22 to 58). They had normal visual acuity with naked eyes or corrected by glass. They had neither any deficit in auditory system.

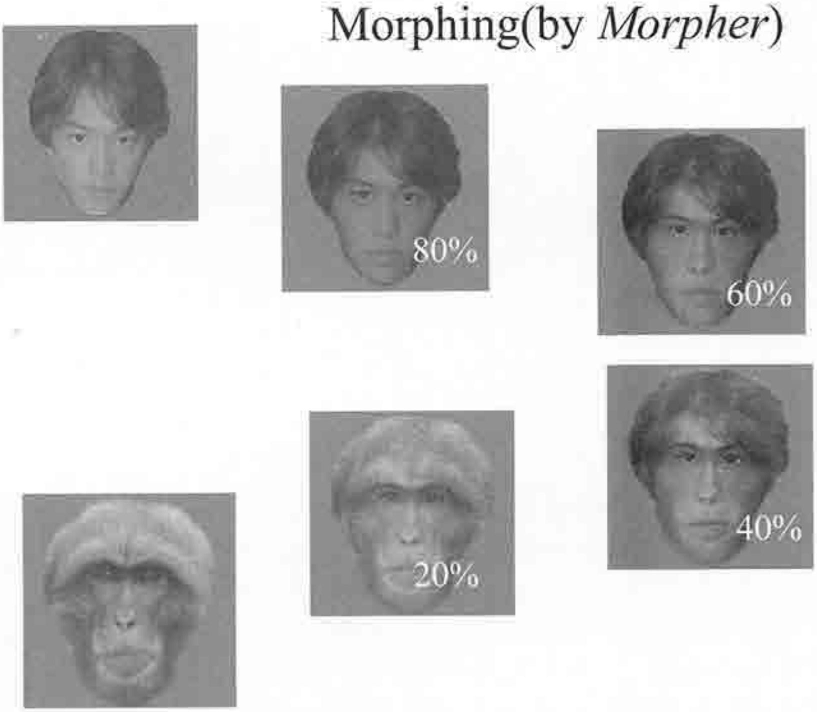

Figure 1. Morphing stimuli. The left top and bottom are original human face and monkey face. Others are morphing stimuli from 80% human to 20 % monkey.

Apparatus

The participants watched the stimuli on a monitor of an iMac (Power PC G4 with liquid crystal display). The distance between the monitor and the participant's eyes was approximately 45 cm but no chin rest was used. Stimuli were presented on the monitor of the iMac using power point software. There were two response keys, that were pecking key for pigeon (MED associates) connected to a computer. A computer with a MED-SKED system controlled the experiment.

Stimuli

We used monochromatic images of a human face and a Japanese macaque (Macaca fuscata) face as discriminative stimuli. The human face was an image

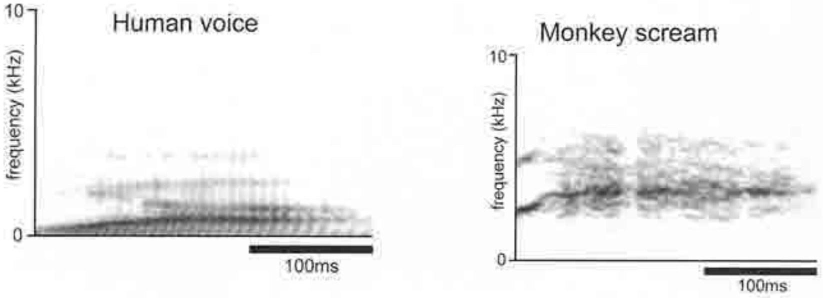

Figure 2. Auditory stimuli. Sonographs of human voice and monkey voice. Intensity and duration of both stimuli are approximately matched. These stimuli are continuously presented during the presentation of the visual stimuli in the tests.

of young male of unknown age. He was completely unfamiliar for the subjects. The size of the stimuli was approximately 12 x 12cm when displayed on the monitor screen. Three different modifications of the stimuli were carried out to produce stimuli for testing. The first modification employed the morphing technique using software Morpher (http://www.asahi-net.or.jp/~FX6M-FJMY/mop00j.html). 50 points in each of the human and monkey faces were selected to produce the morphing images. Mixing ratio was 0, 20, 40, 60 and 80 % as shown in Figure 1.

Auditory stimuli were also employed. Human voice recorded on an auditory tape was edited to audio-clip on Power point. In each instance, the auditory (voice) signal did not match the individual visual image. Japanese monkey voice was obtained from the primate research Institute of Kyoto University. The voice was not voice of the monkey of the visual stimuli. Sonogrphs of the auditory stimuli can be seen in Figure 2. Loudness of the auditory stimuli was 60 dB. The auditory stimuli were repeated during presentations of visual stimulus.

Procedure

The participants were instructed to discriminate images on the monitor into "human" or "monkey" by pressing one of two response keys. No information about the auditory stimuli was given. 6 morphing stimuli with two auditory stimuli (total 12 stimuli) were randomly presented according to a Latin Square.

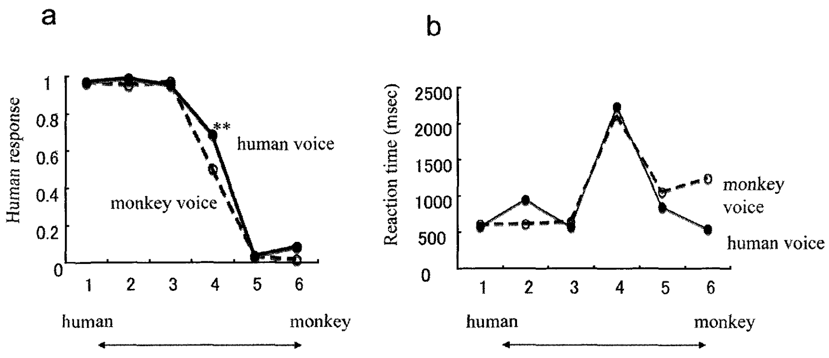

Figure 3. Results of Experiment 1. a: Ratio of hunan response to the morphing stimuli. Closed and open circles indicate response to human voice and that to monkey voice respectively. b: Mean reaction times for the morphing stimuli. ** $p<0.05$

Intertrial interval was 3 sec. Reaction time was also measured. After the experiment, each subject received interview and asked possibility of effects of auditory stimuli on her/his judgment.

Results and Discussion

Figure 3a shows mean number of the human key choice. The participants clearly choose the human key until 60% mixture with monkey face regardless of the auditory stimuli. From 20 to 0 % mixtures they choose the monkey key again regardless of the auditory stimuli. They choose, however, either the human or monkey key at the mixing rate of 40% and the choice was affected by the auditory stimuli. The subjects tended to choose the human key when the face was accompanied with human voice and choose the monkey key when accompanied with the monkey voice. There was a significant difference between the two auditory conditions (paired two tails t-test, $t(10)=2.4$, $p<0.05$).

Figure 3b shows mean reaction time to each key. The subjects showed prolonged reaction time at 40% mixing rate regardless of the auditory stimuli. There was a significant difference between the 40% and other stimuli ($t(19)=2.9$ to 4.7, $p<0.01$).

In the interview after the experiment, some subjects said possible effects of the auditory stimuli on their judgment but some did not. There were no relationships between the subjective reports and their performance.

These results clearly demonstrated effects of auditory stimuli on visual species recognition in humans. However, if the visual stimulus gives clear information, the auditory stimuli did not affect the recognition. Thus, the effects of the auditory stimuli were observed only when the visual stimuli gave ambiguous information.

Experiment 2: Human event related potential(ERP) study

Primate research demonstrated that visual species-specific brain region was activated by species-specific vocalization (Gil-da-Costa et al., 2004) suggesting multi-sensory integration in social cognition in monkey (Ghazanfar et al., 2005). In human brain imaging study, neighboring regions of posterior superior temporal cortex was activated by either animal image or vocalization (Beauchamp et al., 2004). Human face elicited a particular negative ERP after 170-200ms after onset of stimuli (Allison, et al., 1994). Puce et al (2007) recorded ERP to congruent and incongruent stimuli consisted of human face, monkey face, human voice and monkey voice. They found that N140 in auditory ERP was enhanced by human face, but did not observed modification of N170 in visual ERP by auditory stimuli. Effect of the incongruent stimuli was observed only in the case of mismatch of human face with monkey voice.

Here, we examined effect of the voice on visual N170 component like Puce et al. (2007). We modified the procedure in two points. First, we presented the voice stimuli immediately before face stimuli, whereas Puce et al. (2007) presented face and voice simultaneously. The simultaneous voice presentation may not affect early visual processing because it will take some time to process voice information. Second, we employed morphing stimuli used in Experiment 1. Multi-sensory interaction may be more clearly observed when one sensory modality does not give enough information to identify the stimuli.

Methods

Participants

Eleven healthy volunteers (six males, five females) aged 21-31 years (mean age 24.18±10.2 years) participated in the study. There were ten right-handers and one left-hander. Subjects had no previous history of neurological abnormalities, had normal or corrected to normal vision, and reported having normal hearing. Written informed consent was obtained before the experiment.

Procedures

Participants were seated in a dimly lit sound attenuated cabin, with response buttons under their left and right hands. A computer screen was placed 70 cm in front of the participant's eyes. Stimuli were gray-scale photographs of an unfamiliar human male face, monkey face, and 50 % morphed image of the human and monkey face (middle face). They were presented on a computer monitor in front of a gray background. They occupied a visual angle of approximately 8.17° X 8.17° (viewing distance = 70 cm).

Each face was presented 120 times throughout the experiment, half of which was presented immediately after a human voice and the other half a monkey voice. The duration of voice presentation was 300 ms, which was preceded by fixation cross for 700 ms. The subjects were required to discriminate race of the face (human or monkey) and press buttons as fast as possible. They were also instructed to ignore the race of voice. The experiment took about 30 minutes in total.

EEG Recordings

Scalp EEG recordings were made from 19 electrodes (Fz, Cz, Pz, Fp1, Fp2, F3, F4, F7, F8, C3, C4, T3, T4, T5, T6, P3, P4, O1, O2) according to the international 10-20 system. A ground electrode was positioned on the forehead. They were referenced by averaged potentials of all electrodes. The vertical electrooculogram was recorded from electrodes positioned above the left eye to monitor eye movements during experiment. Electrode impedance was kept blow 5k ohm. The EEGs were sampled with a digitization rate of 1,000 Hz through a band pass filter of 1.0 to 25.0 Hz. The EEGs were averaged for each material condition during epochs from 100 ms pre-stimulus to 400 ms post-stimulus, after discarding trials with eye blinks or other large noise. We only reported results of the left (T5) and right (T6) lateral temporal electrodes, because of the page limit. Peak amplitudes and latencies of N170 were estimated for each condition.

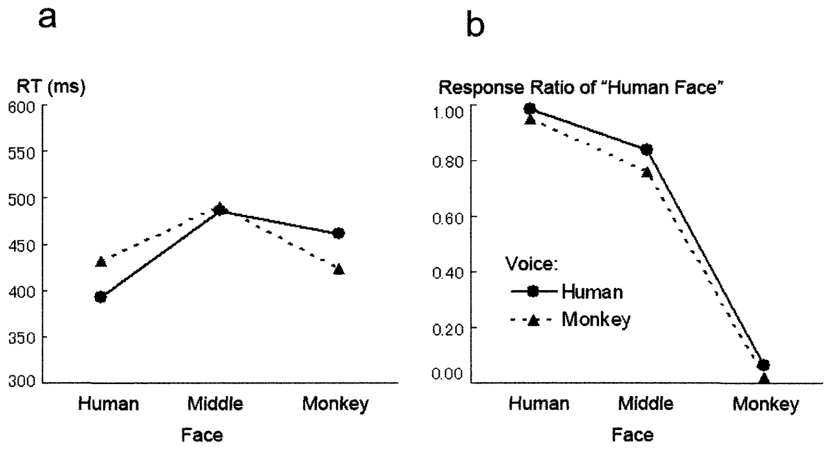

Figure 4. Mean reaction times (RTs) and ratios of "human face" response in Experiment 2 (face-race discrimination task).

Results and Discussion

Behavioral Analysis

Figure 4a summarized the mean reaction times (RTs) for each condition in the face-race discrimination task. A two-way Analysis of Variance (ANOVA) with within-subject factors of Face (human, middle, and monkey) and Voice (human and monkey) showed that there was significant Face effect, F (2, 20) = 25.93, $p < 0.01$. Paired comparison analysis (Ryan's method) revealed that RTs for human face was significantly faster than those for monkey face, t (20) = 2.58, $p < 0.05$, which was significantly faster than middle face, t (20) = 4.55, $p < 0.01$. Face X Voice interaction was also significant, F (2, 20) = 13.88, $p < 0.01$, suggesting that human voice significantly reduced RTs for human face, F (1, 30) = 18.66, $p < 0.01$, whereas monkey voice significantly reduced RTs for monkey voice, F (1, 30) = 16.48, $p < 0.01$. There was no significant voice effect in the middle face, F (1, 30) = 0.32, $p < 0.57$.

Mean ratios of the "human face" response were summarized in Figure 4b. A two-way ANOVA showed that there was significant main effect of Voice, F (1, 10) = 9.81, $p < 0.05$, and Face, F (2, 20) = 207.51, $p < 0.01$. Human voice produced higher response ratios than monkey voice. Paired comparison analysis showed that human face produced higher response ratios of middle face, t (20) = 3.23, $p < 0.01$, which produced higher response ratios of monkey voice, t

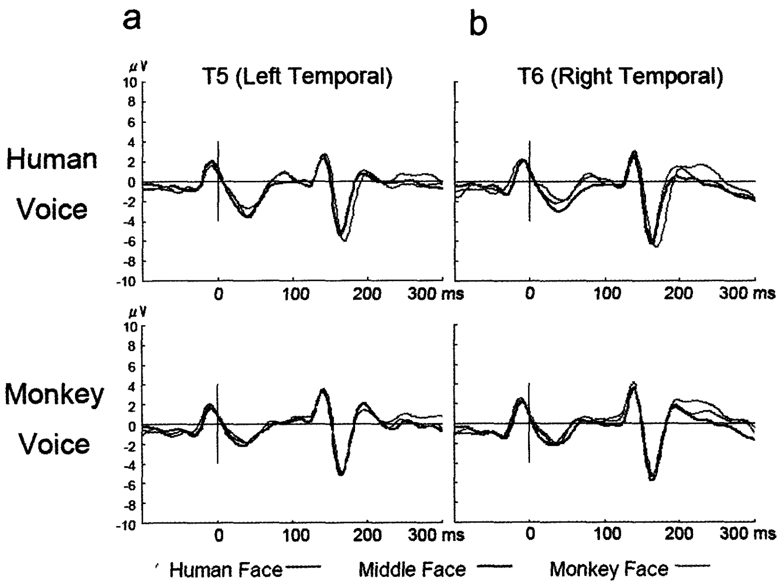

Figure 5. ERP waveforms at the left (T5) ad right (T6) lateral temporal electrodes for each material condition in Experiment 2.

(20) = 15.80, $p < 0.01$. There was no significant Face X Voice interaction, F (2, 20) = 2.14, $p = 0.14$.

EEG Analysis

Figure 5 presented ERP waveforms were summarized for each material condition in lateral temporal electrodes (T5 and T6). We observed clear N170 components for all conditions. Figure 6 summarized peak amplitudes and latencies for each condition. A 3-way ANOVA with within-subject factors of Face (human, middle, and monkey), Voice (Human, Monkey), and Hemisphere (LH, RH) on the latency showed the significant Face effect, F (2, 20) = 21.03, $p < 0.01$. Paired comparison analysis revealed that N170 to monkey faces was significantly slower than human face, t (20) = 5.98, $p < 0.01$, and middle face, t (20) = 5.17, $p < 0.01$. There was no significant difference between human face and middle face, t (20) = 0.80, $p = 0.43$.

The latency analysis also revealed significant Face X Voice interaction, F (2, 44) = 5.31, $p < 0.05$, suggesting that significant Face effect was observed for

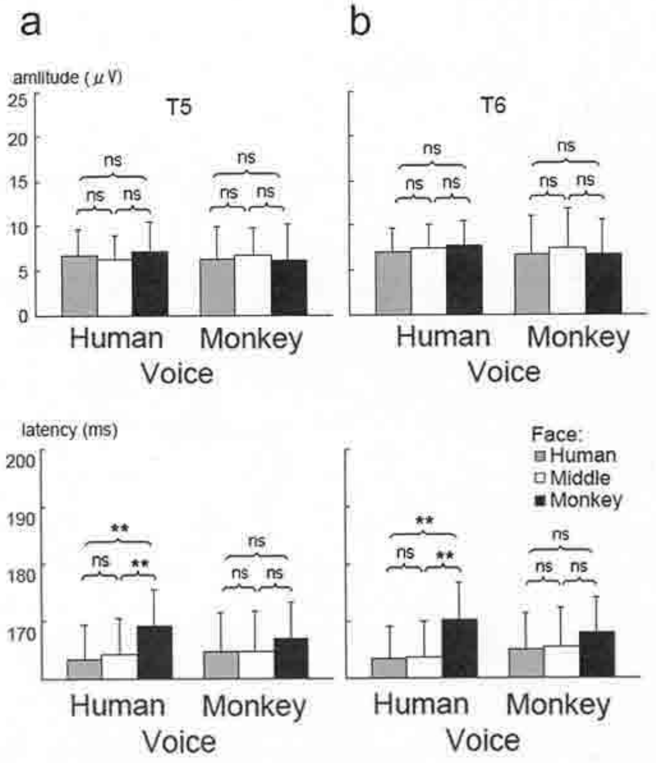

Figure 6. Mean amplitude and latency of N170 in the left (T5) and right (T6) lateral temporal Electrodes for each material condition in Experiment 2.

human voice, F (2, 40) = 23.14, $p < 0.01$, but not for monkey voice, F (2, 40) = 2.55, $p = 0.09$. The interaction also means that human voice significantly delayed N170 latency for monkey face, F (1, 30) = 7.29, $p < 0.05$, but not for human face, F (1, 30) = 0.1.82, $p = 0.19$, and for middle face, F (1, 30) = 3.45, $p = 0.07$. The amplitude analysis did not show any significant main effect and interaction ($p > 0.05$).

Experiment 2 demonstrated that the visual face processing was modulated by auditory voice presentation in the human brain. When human voice was presented, N170 to human face was elicited faster relative to monkey face. This is consistent with previous ERP findings without voice presentation (de Haan et al., 2002; Carmel & Bentin, 2002). In contrast, when monkey voice was presented, there was no N170 latency difference between human and monkey face. Itier & Taylor (2004) demonstrated that repeated presentation reduced N170 latency in face perception task. Presumably, the priming effect of voice presentation may contribute the present results. In fact, monkey voice reduced N170 latency for monkey face but not for human and middle face.

The present ERP findings are contradicted with previous ERP study, which examined the voice effect on early face processing (Puce et al., 2007). Puce et al. (2007) suggested that voice presentation did not affect N170 component for face material. There are several procedural differences between the present experiment and Puce et al's experiment (2007). In addition to detailed difference of visual and auditory stimuli, we presented the static image whereas Puce et al used apparent motion picture caused by presenting open and closed mouth. We presented auditory stimulus 300 ms before presenting the visual stimuli, whereas Puce et al presented visual and auditory stimuli simultaneously. Those procedural differences produced the different findings between the present study and Puce et al's study (2007).

Experiment 3: Pigeon behavioral study

Humans have a well developed visual system and most of birds are another animals which have also well developed visual system. Here, we trained pigeons on discrimination between human and monkey faces then presented them morphing images with auditory stimuli. Because the pigeons were familiar with human but not with monkey, the human voice should enhance human-likeness of morphing images that have visual features of both the human and monkey faces, if the birds have a visio-auditory integrating system for human face recognition. On the other hand, monkey vocalization should not have strong effects on visual discrimination of monkey face, because the pigeons do not have cross-modal experience of the monkey.

Methods

Subjects

Eigh experimentally naïve pigeons (Columba livia) obtained from the Japanese Association of Racing Pigeons were used in this study. Subject birds were maintained at 80% of their free-feeding weights. Water and grit were freely available in the cages. The temperature of the animal room was maintained at 23° C and the light cycle was maintained at 12L: 12D.

Apparatus

Standard operant chambers were used (30 x 25 x 30 cm, MED). The front panel contained a rectangular transparent pecking key (10 x 7 cm) through which the subject could see an iMac (Power PC G4 with liquid crystal display) computer monitor. An electronic liquid shutter (UM glass, Tokyo) was placed between the key and the monitor. The distance between the key and the monitor was 15 cm. Stimuli were presented on the monitor of the iMac using power point software. A computer with a MED-SKED system controlled the experiment.

Stimuli

We used monochromatic images of a human face and a Japanese macaque (Macaca fuscata) face as discriminative stimuli. These were the same stimuli used in Experiment 1 and 2. The size of the stimuli was approximately 12 x 12cm when displayed on the monitor screen. Three different modifications of the stimuli were carried out to produce stimuli for testing. The first modification was the morphing stimuli as Experiments 1 and 2. The second modification involved the exchange of specific elements of the face. Eyes, mouth or nose was exchanged between the human and monkey faces. These part exchanged stimuli were produced by Photoshop software. The third modification involved chimera stimuli. The upper or lower half of the original faces was combined to make the chimera stimuli. The left or right half was also used to make chimera stimuli. The width and length of the faces were adjusted to make these stimuli, Auditory stimuli used in Experiments 1 and 2 were also employed. Loudness of the auditory stimuli was 60 dB. The auditory stimuli were repeated during presentations of visual stimulus.

Procedures

Pretraining: The subjects were first trained to peck a transparent key. They could see the computer monitor through the key, on which the images were displayed. Subjects were then trained on a reinforcement schedule with a variable interval (VI) with a mean of 5 s, 10s, and finally 20 s (VI 5 s, VI 10 s, VI 20 s).

Discriminative Training: The subjects were divided into two groups of four. The human group was trained to peck at the human face but not to monkey face, while the monkey group was trained to peck at the monkey face but not to the human face. Pecking to the S+ was rewarded by a 4 s period of access to a feeder after a variable interval with a mean of 25 s. Responses to the S- was extinguished. S+ and S– were presented in a pseudo-random sequence, constrained so that images of S+ or S- never appeared more than three times in succession. There was a 5 s blackout period made by an electric shutter between the presentations of the stimuli. During this blackout period differential reinforcement of zero rate (DRO) was effective, that is, a peck prolonged the dark period for 5 s. Daily sessions consisted of 40 trials, 20 with S+, and 20 with S-. Training continued until 80% of discrimination ratio calculated by dividing the number of responses to S+ by total number of responses to S+ and S- in two successive sessions.

Test 1: Morphing

After the discriminative training, subjects were tested with the morphing stimuli. 2 original stimuli (100% human face and 100% monkey face) and morphing stimuli with 4 different mixture rates (20%, 40%, 60%, or 80%) were presented during the test in random order. Each stimulus was presented 6 times each lasting 25 s separated by a 5 s blackout period. DRO schedule was effective during the blackout period. Response to any stimuli was reinforced with VI 25 s. First the subjects received the test without auditory stimulus. Then, after 2 sessions of the ordinal discriminative training, they received a test with auditory stimuli. Either a monkey or human voice was presented during the presentation of the visual stimuli. Each visual stimulus was presented with each auditory stimulus 3 times. The order of presentation was quasi-random (i.e., no 3 successions of each auditory stimulus).

Test 2: Exchange of internal face features

Then, the subjects received the discriminative training again to maintain their discriminative behavior. After subjects reached criterion again, the feature-exchange test was carried out. In this test, internal features (i.e., eyes, nose and mouth) of one discriminative stimulus were exchanged with the external features of the other discriminative stimulus. For example, the eyes, nose or mouth of the human face were exchanged with the monkey face. The 6 feature-exchanged stimuli and the original 2 faces were presented 5 times each. Each stimulus presentation period was 25 s separated by 5 s blackout period? DRO schedule was effective during the blackout period. Reinforcement was available on VI25 for responding to any stimuli.

Test 3: Chimera test

Then, the subjects received the discriminative training again to maintain their discriminative behavior. After subjects reached the criterion again, the chimera test was carried out. Chimera stimuli were created by pasting the upper, lower, left or right half of a monkey face to the lower, upper, right or left half of a human face. Four chimera stimuli and 2 original stimuli were presented 6 times each in random order. Each stimulus presentation period was 25 s separated by 5 s blackout period. DRO schedule was effective during the blackout period. Reinforcement was available on VI25 for responding to any stimuli. Then, after 2 sessions of the ordinal discriminative training, they received a test with auditory stimuli. Either a monkey or a human voice was presented during the presentation of the visual stimuli. Each visual stimulus was presented with each auditory stimulus 3 times. The order of presentation was quasi-random (i.e., no 3 successions of each auditory stimulus).

Results and Discussion

The human group reached the criterion within 45.3 sessions (range from 39 to 54) and the monkey group by 25.3 sessions (range from 14 to 34). There was a statistically significant difference in number of sessions required to reach criterion between the two groups (t-tailed t-test, $t(6) = 3.25$, $p<0.05$). Thus, it was easier to learn to respond to the monkey face rather than to the human face.

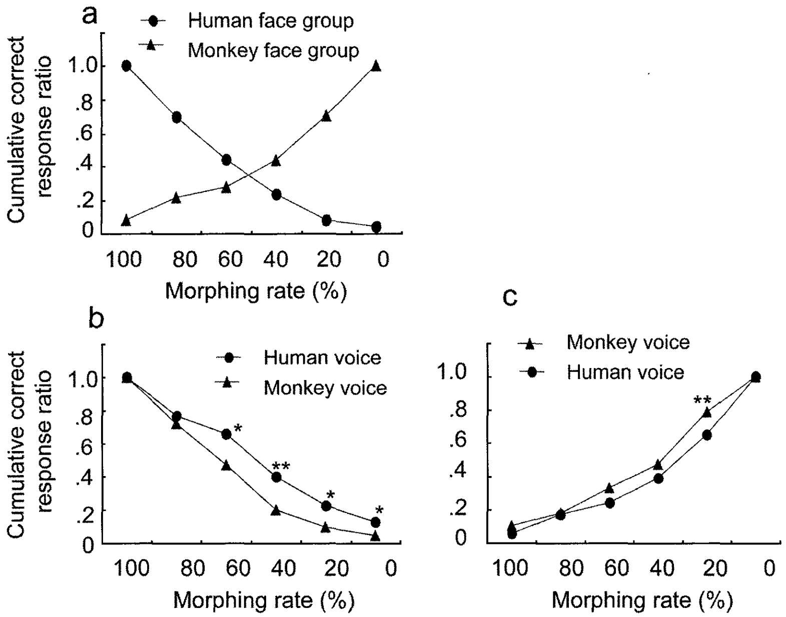

Figure 7. Results of the morphing test. Top panel shows results of the morphing test without auditory stimuli. Lower left (human group) and right (monkey group) show results of the morphing tests with auditory stimuli. ** $p<0.05$ * $p<0.1$.

Test 1: Morphing

Figure 7a shows cumulative relative responses to morphing stimuli without auditory stimuli. Both groups showed gradual increment in responding. Figure 7b shows the results of the human group. The human voice increased responding to the morphing stimuli. A two factor ANOVA (morphing x sound) reveals a significant effect of morphing (F (5/36) = 86.04, MSE=121.93, $p<0.00001$) and sound (F (1/36) = 11.9, MSE=3.37, $p<0.05$) but not of interaction (F (5/36) =1.05). Paired t-test gives a significant difference in responding between the human and the monkey voice at morphing rate 40% (t (3) = 6.40, $p<0.01$). There is a tendency of difference in 60% (t (3) = 2.61, $p<0.1$), 20% (t (3)=2.42, $p<0.1$) and 0%(t (3) = 2.64, $p<0.01$). Thus, the human voice facilitates responding. Figure 7c shows results of the monkey group, two factors ANOVA (morphing x sound) reveals a significant effects of morphing (F (5/36)=352.83, MSE=118.07, $p<0.00001$) and sound (F (1/36)=15.70, MSE=1.05, $p<0.001$) but not of interaction (F (5/36)=2.00).

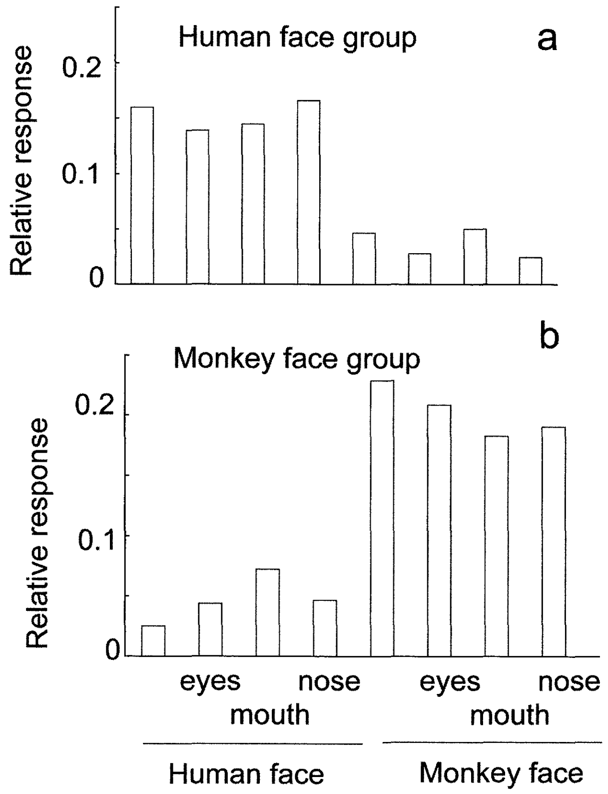

Figure 8. Exchange of the internal features. In each panel, the left three columns present human faces with monkey internal features and the right monkey faces with human internal features. The upper and lower panels show the human group and monkey group.

Paired t-test gives a significant difference in responding between the human and the monkey voice at morphing rate 20% ($t(3) = 5.28$, $p<0.05$) but not at other morphing rate. Hence, facilitative effects of auditory stimulus were stronger in responding to human face than to the monkey face.

Test2: Exchange of internal features

Figure 8 shows results of the exchange of internal features test. Both groups clearly maintained facial discrimination even when the eyes, nose or mouth was exchanged. Thus, none of such internal features was crucial for the facial discrimination.

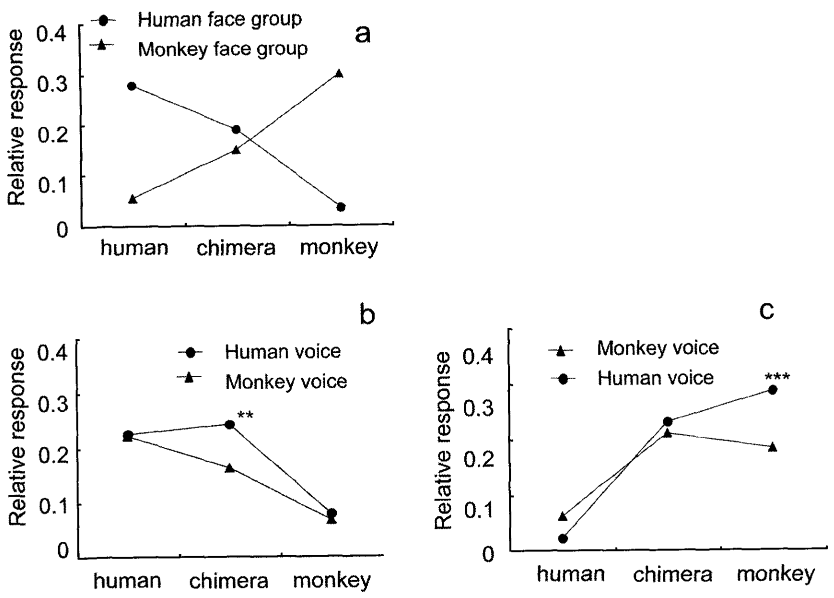

Figure 9. Chimera test. Top panel shows results without auditory stimuli. Lower left (human group) and right (monkey group) show results of the chimera tests with auditory stimuli. ** $p<0.05$ *** $p<0.01$.

Test 3. Chimera

Figure 9a presents the chimera tests without sound stimuli. Because there was no statistically significant difference in responses between the vertical chimera stimuli, or between the horizontal chimera stimuli (paired t-test), responses for 4 different chimera stimuli were combined into the category of chimera. Responding to the chimera stimuli was intermediate between the original faces. There was no significant difference to the chimera stimuli between the human and monkey groups ($t(3)=0.91$).

Figure 9b presents results of the chimera test with sound in the human group. Neither the human voice or monkey voice affected the responses to the original faces but increased response to the chimera stimuli (paired t-test, $t(3)=3.70$, $p<0.05$). Figure 9c presents results of the chimera test with sound in the monkey group. Neither the monkey voice or human voice affected responding to the original human faces nor the chimera faces but the human voice increased responding to the original monkey face and the monkey voice decreased the response. There was a significant difference between the monkey voice and human voice (t (3)=19.56, p<0.01).

The present experiments show an interaction of audio-visual information in face discrimination by pigeons. While enhancing effect of human voice is clear, that effect of monkey voice is not clear.

Dimensional stimulus control by morphing: The morphing tests demonstrated gradient of responding that corresponds to the morphing rate. The gradient of the human and monkey groups are almost symmetrical. If the gradients without sound were converted to gradient along the original S+ to the original S-, two gradients were almost identical. There was no statistical difference in each morphing rate between the two groups (two-tailed t-test). Thus, the slope of the gradient did not dependent upon the kinds of the original S+.

The sound stimuli affected the gradient in the human group but did not much on that in the monkey group. The human voice enhanced responding in the human group. This effect was clearly observed in more ambiguous stimuli (morphing rate of 60%). In the monkey group, the enhancing effect was observed only at 20% morphing rate. Because the subjects lived in the laboratory, they must have heard human voices quite often, whereas monkey face and voice should be novel stimuli for the subjects. Because the subjects were not exposed to the auditory stimuli during the discriminative training, the enhancing effects the difference in the previous exposure to human stimuli may have contributed to the enhancing effect observed in the human group.

Even in the human groups, the enhancing effect did not occur for the original stimulus. They ignored auditory stimuli when they saw complete S+ or complete S-. Therefore, the enhancing effect of auditory stimuli was observed only when the visual stimulus did not provide enough information. Previously, I reported that Bengalese finches used contact calls as discriminative stimulus only when the vision did not give enough information (Watanabe, 1993). The present experiments agree with this result.

External and internal features: Some of the face-specific neurons in the monkey inferotemporal cortex show preference for a particular face feature, such as eyes (Perrett, Rolls, & Caan, 1982). However, most the neurons in this brain region prefer holistic configurations of the face (for review, Desimone, 1991). Although both global and local cues were used for face recognition in humans (see, Bruce, 1988), humans showed preference for the internal features when the face is familiar one. In the current study, pigeons clearly showed selective responses to the external features in Test 2. When pigeons were trained on pigeon vs. starling discrimination, they showed selective stimulus control by feathers but not by the beak or eyes in the partial feature exchange test (Watanabe & Furuya, 1997). The starling had white feathers on its cheeks that the pigeon did not have such coloration. Differences between the human and monkey might be not so salient in each internal cue. On the other hand, difference in the external cues might be large enough to facilitate discrimination between these two species. The exchange of the internal features did not give ambiguous stimuli for pigeons.

Effects of auditory stimuli on chimera stimuli. There were no significant response differences among the four different chimeras that were tested. Spatial configuration of the eyes, nose and mouth was different between the monkey face and human face. This fact resulted in asymmetrical configuration of these elements in the chimera image. Even the free moving subjects were able to see the stimuli from different positions; the images might be seen as completely unknown stimuli. In other words, any half of the stimuli did not have dominant effect over other half. Human observers showed better performance when two halves of different human faces were misaligned than aligned (Young, Hellawall & Hay, 1987).

Effects of auditory stimuli in the human face group were plausible. The human voice increased responding when the stimuli were ambiguous. Results of the monkey face group were unpredictable. The fact that no clear effects of monkey voice on responding to the chimera stimuli were found seems to match the results of the morphing test. Again the monkey voice did not have effect on the discrimination perhaps because the subjects did not have experience of the monkey voice. One unpredicted result was facilitative effect of the human voice for response to the original monkey face. Because the human voice was a kind of familiar stimulus through experience, it may have provided a general facilitative effect on responding.

Experiment 4: anatomical study

Experiment 3 showed interaction of visual and auditory information in pigeons. One visual information flow in pigeons starts from retina to optic tectum, then from there to nucleus rotundus in thalamus, and finally to entopallium in telencephalon. Another flow directly goes to dorsal thalamus, then from there to hyperpallium. Several lesion studies demonstrated that the former pathway is crucial to fine visual cognition including social cognition (Watanabe, 1996). Auditory flow starts from cochlea nucleus to nucleus ovoidalis in thalamus, the finally to field L of Rosa in neopallium. In telencephalon, both visual and auditory information go to neopallium caudo-lateralis (NCL), however, the visual fibers go to middle part of NCL (Husband & Shimizu, 1999) and the auditory fibers to the ventral parts of NCL (Wild et al., 1993). Interaction between the two sensory information may occur in NCL. Another possibility in convergence of two sensory information at lower level, namely auditory projection to the entopallium or visual projection to the field L. Here, we examined possibility of such convergence in entopallium and filed L by injecting retrograde tracer to the entopalium of one hemisphere and field L in the other hemisphere.

Methods

Subjects

33 pigeons were used. They lived in individual cages and had free access to food and water. They were treated in accordance with Guidelines of Animal Experiment of Japanese Society of Animal Psychology and a vet unary doctor checked their health conditions.

Tracer injection

The subjects were anesthetized with an intramuscular injection of pentobarbital (Nembutal, Dainippon Yumitomo Pharmaco), and then placed in a stereotaxic apparatus (David Kopf). Local anesthesia (Xylocaine, AstrZanica) was injected in incision site. The coordinates of the injection site were determined in accordance with an atlas of pigeon brain (Karten & Hodos, 1967). Small holes were made with a drill fixed in a micromanipulator. Then, 1.0 μl syringe (Hamilton,) fixed on a micro drive (Micro 4th, World Precision Instruments) was inserted into the brain. 100 nl of 10% solution of biotinylated dextran amine (BDA, MW3000) was injected (5nl/min). The needle was placed for 5-10min after the injection. The BDA was injected into the entopallium in one hemisphere and into the field L in the other hemisphere.

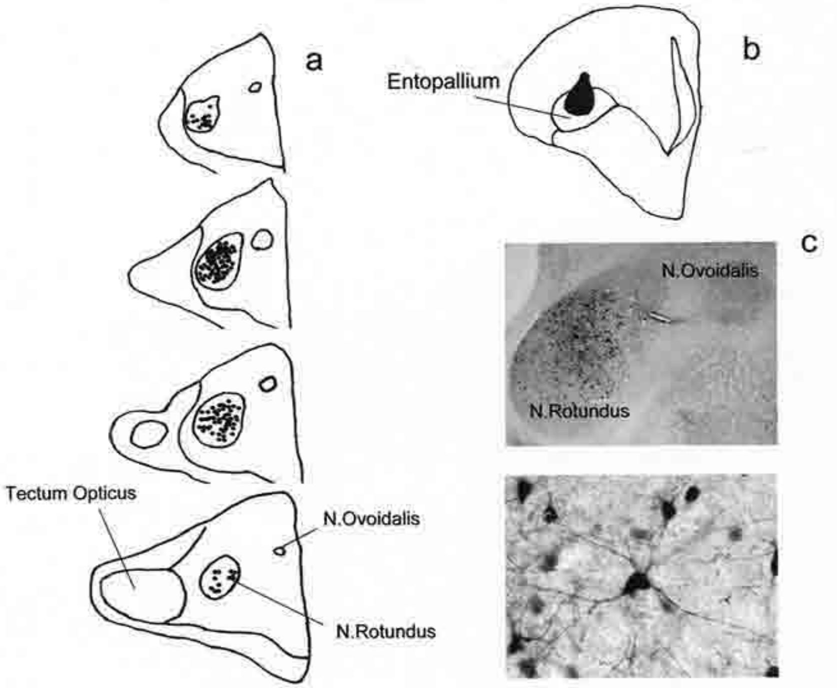

Figure 10. a: Distribution of BDA labeled cells. b: Injection site of BDA. c: Photos of the nucleus rotundus and BDA labeled cells,

Histology

After 5-7 days of survival period, the birds were deeply anesthetized with overdose injection of pentobarbital, the perfused with a 0.9% saline followed by 50 ml of 6% dextran in 0.1M PB (pH7.4), 1000ml of 4% paraformaldehyde with 0.1M lysine and 0.01M sodium periodate in 0.1M PBS (pH7.4). The brains were removed from the skull and post fixed in the fixative above for one night at 4C, then placed in 30% sucrose for one night at 4C. The brains were frozen in a Cryostat (HM505E, Microm) and cut at 50μm frontal section. Every second sections were washed 3 times in 0.1M PB, then placed in 25% methanol for 30min, washed 3 times in PB, then incubated in avidin-biotinylated HRP (Vectastain ABC kit, Vector) for 60 min. The sections were reacted in a solution of 0.05%DAB (diaminobenzedine) with 0.01% H2O2 or DAB substrate kit with metal enhancement (Vector). Sections were mounted

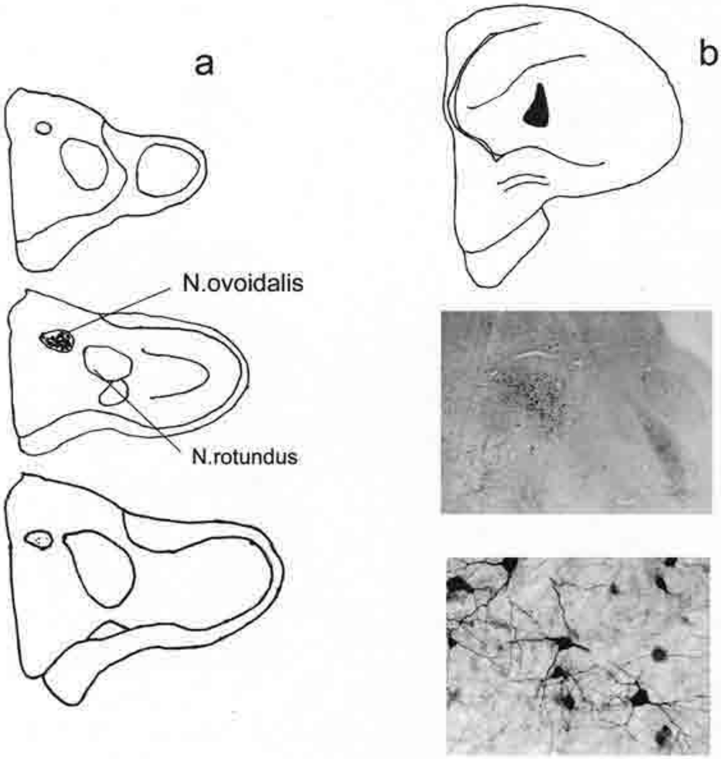

Figure 11/ Figure 10. a: Distribution of BDA labeled cells. b: Injection site of BDA. c: Photos of the nucleus ovoidalis and BDA labeled cells,

on glass slides, dehydrated in an ethanol and xylen series and coverslipped with (Eukitt, O.Kindler). The tissues were examined under a microscope.

Results and Discussion

Entopallium injection hemisphere: Figure 10 shows the injection site (b) and distribution of BDA labeled cells (a and C). Injection site spread to both core and belt of the entopallium. In the telencephalon no cells in the field L were labeled. There were also no labeled cells in NCL suggesting no projections from NCL to the entopallium. Ventromedial part of NCL received auditory input (Wild et al., 1993) and the neostriatum caudomedilalis (NCM) also received projection from the field L (Vates et al., 1996). Retrograde injection into the entopallium did not results in labeled cells in NCL, NCM nor the field L. Therefore, the entopallium did not received auditory information from auditory area in the telencephalon. In the diencephalons the labeled cells were observed in the nucleus rotundus. The labeled cells were distributed in the nucleus triangularis but the number of the labeled cells was small in comparison to those in the nucleus rotundus. No labeled cells were found in the nucleus ovoidalis, therefore, no connection between the ovoidalis and the entopallium. The entopallium sends efferent projection to the neopallium caudolateralis (NCL) (Husband & Shimizu, 1999).

Field L injection hemisphere: Figure 11 shows injection site (b) and distribution of BDA labeled cells (a and c). Due to no clear landmark of the field L in pigeon brain, it is hard to clarify range of spreading of the injection even though the injection site clearly contained the field L. There was no BDA labeled cells in the entopallium. Thus no direct connection between the entopallium and the field L. (telencephalon). The labeled cells were observed in the nucleus ovoidalis but not in the nucleus rotundus.

Possible area of visio-auditory integration: The present results suggest that primary visual system and primary auditory system are independent pathways. Thus, multi-sensory integration must occur in the lower level such as thalamus or higher level after the entopallium or the field L. Korzeniewska (1987) reported multi sensory nucleus in thalamus of pigeons. Cell population in the nucleus dorsolateralis posterior thalami (DLP) was activated by somatosensory, auditory and visual stimulation. DLP sends fibers to the neopallium including NCM (Gamlin and Cohen, 1986). The present study did not show retrograde tracing from the entopallium to DLP. Taken together, the visual and the auditory systems were rather independent from the thalamus to the entopallium and the field L.

Possible area of visio-auditory integration: The present results suggest that primary visual system and primary auditory system are independent pathways. Thus, multi-sensory integration may occur in the lower level such as thalamus or higher level after the entopallium or the field L. Korzeniewska (1987) reported multi sensory nucleus in thalamus of pigeons. Cell population in the nucleus dotsolateralis posterior thalami (DLP) was activated by somatosensory, auditory and visual stimulation. DLP sends fibers to the neopallium including NCM (Gamlin and Cohen, 1986). The present study did not show retrograde tracing from the entopallium to DLP. Other possible multi-sensory area is

NCL in the telencephalon. Because no anterograde tracer was used in the present experiment, we cannot provide evidence of connection from the entopallium or the field L to NCL but observation that no labeled cells in NCL in either hemisphere suggests no direct projection from the NCL to the entopallium or the field L. The visual and auditory projection reaches to the different area of NCL(Husband & Shimizu, 1999; Wild et al., 1993). Thus, there may be short connecting fibers between the visual and auditory areas within NCL.

Conclusion

Both humans and pigeons demonstrated multi-sensory integration in species cognition (Experiments 1 and 3). The visual stimuli contained conspecific for human subjects but not for pigeons. In pigeon study, modification of visual cognition by auditory stimulus was stronger in human voice. This observation suggests the multi-sensory integration depending on experience. Previously we reported morphing test after the pigeon vs. starling discrimination. The results show a steep gradient of responding along morphing rate (Watanabe & Furuya, 1997). Although the software making the morphing stimuli differed from that used in the present experiment, the difference in the results suggests peculiarity in discrimination of own species. Human ERP study(Experiment 2) suggested that the multi-sensory integration occurred in the higher processing stage not in early stage of visual perception. Anatomical study with pigeons (Experiment 4) also showed no direct connection between the visual and auditory information at the primary sensory system. Thus, in both humans and pigeons multi-sensory processing of species recognition occurs after the primary sensory processing.

References

Barton J. J., Keenan, J. P., & Bass, T. (2001) Discrimination of spatial relations and features in faces: Effects of inversion and viewing duration. British Journal of Psychology, 92, 527-549.

Bruce,V. (1988) Recognizing face. Lawrence Erlbaum Associates (London).

Brown, S., & Dooling, R. J. (1992) Perception of conspecific faces by budgerigars (Melopsittacus undulatus): I. Natural faces. Journal of Comparative Psychology, 106, 203-216.

Carmel, D., & Bentin, S. (2002). Domain specificity versus expertise: factors influencing distinct processing of faces. Cognition, 83, 1-29.

Desimone, R. (1991) Face-selective cells in the temporal cortex of monkeys. Journal of Cognitive Neuroscience, 3, 1-8.

de Haan, M., Pascalis, O., & Johnson, M. H. (2002). Specialization of neural mechanisms underlying face recognition in human infants. Journal of Cognitive Neuroscience, 14, 199-209.

Diamond, R., & Carey, S. (1986) Why faces are and are not special: An effect of expertise. Journal of experimental psychology: general. 115, 107-117.

Ellis, H. D., Shepard,J.W., & Davies, G.M. (1979) Identification of familiar and unfamiliar faces from internal and external features; Some implication for theories of face recognition. Perception, 8, 431-439.

Farah, M., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998) What is 'special' about human face perception? Psychological Review, 105, 482-498.

Gamlin, P. D. R., & Cohen, D. H. (1986) A second ascending visual pathway from the optic tectum to the telencephalon in the pigeon (Columba livia). Journal of Comparative Neurol0gy, 250, 296-310.

Huber, L. (2003) Categorization of human faces in pigeons. In Watanabe, S.(Ed) Comparative analysis of mind. Keio Univ. Press (Tokyo) pp. 101-118.

Husband, S. A., & Shimizu, T. 1999 Efferent projections of the ectostriatum in the pigeon (Columba livia). Journal of Comparative Neurology, 406, 329-345.

Itier, R. J., & Taylor, M. J. (2004) Effects of repetition learning on upright, inverted and contrast-reversed face processing using ERPs. Neuroimage, 21, 1518-1532.

Kamachi, M. Hill, H., & Lander K. (2003) Putting the face to the voice: matching identity across modality. Current Biology, 13, 1709-1714.

Korzeniewska, E. (1987) Multisensory convergence in the thalamus of the pigeon (Columba livia). Neuroscience Letters., 80, 55-60.

Kusayama, T. (2004) Cognition of human facial expression by crows. In Watanabe, S.(Ed) Comparative analysis of mind. Keio Univ. Press (Tokyo), pp. 137-152.

Loidolt, M., Aust, U., Meran, I., & Huber, L. (2003) Pigeons use item-specific and category-level information inn the identification and categorization of human faces, Journal of Experimental Psychology: Animal Behavior Processes, 29, 261-276.

Mcgurk, H. & Macdonald, J. (1976) Hearing lips and seeing voices. Nature, 264, 746-748.

Makino H., & Jitsumori, M. (2000) Category learning and prototype effect in pigeons: A study by using morphed images of human faces. Japanese Journal of Psychology, 71, 477-485 (in Japanese).

Masuda, S. (2004) Configural and featural processing in face recognition in humans. In Watanabe, S.(Ed) Comparative analysis of mind. Keio Univ. Press (Tokyo), p119-136.

Moskovich, M., Winocur, G., & Behrman, M. (1997) What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. Journal of Cognitive Neuroscience, 9, 555-604.

Munhall, K. G. & Tohkura, Y. (1998) audiovisual gating and the time course of speech perception. Journal of Acoustical Society of America, 104, 530-539.

Partan, S., Yelda, S., Price, V., & Shimizu, T . (2005) Female pigeons, Columba livia, respond to multisensory audio/video playbacks of male courtship behavior. Animal Behavior, 79, 957-966.

Perrett, D. L., Rolls, E. T., & Caan, W. (1982) Visual neurons responsive to face in the monkey temporal cortex. Experimental Brain Research, 47, 329-342.

Phelps, M. T. & Roberts, W. A. (1994). Memory for pictures of upright and inverted primate faces in humans (Homo sapiens), squirrel monkeys (Saimiri sciureus), and pigeons (Columba livia). Journal of Comparative Psychology, 108, 114-125.

Rowe, C. (1999) receiver psychology and the evolution of multicomponent signals. Animal Behaviour, 58, 921-931.

Ryan, C. M. E., Lea, S. E. G. (1994) Images of conspecific as categories to be discriminated by pigeons and chickens: Slides, video tapes, stuffed birds and live birds. Behavioural Peocesses, 33, 155-175.

Steyvers, M. (1999) Morphing techniques for manipulating face images. Behavior research Methods, Instruments & Computers, 31, 359-369.

Stoddard, P. K., Beecher, M. D., Leosche, P. (1992) Memory does not constrain individual recognition in a bird with song repertoires. Behaviour, 122, 274-287.

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

Troje, N., Huber, L., Loidolt, M., Aust, U., & Fieder, M. (1999) Categorical learning in pigeons: the role of texture and shape in complex static stimuli. Vision Research, 39, 353-366.

Vatikiotis-Bsteson, E., Eigsti, I, & Yano, S. (1998) Eye movement of perceivers during audiovisual speech perception. Perception & Psychophysics, 60, 926-940.

Vermeire, B. A., & Hamilton, C. R. (1998) Inversion effect for faces in split brain monkeys. Neuropsychologia, 36, 1003-1014.

Watanabe, S. (1996) Effects of ectostriatal lesions on concepcific, species and familiar object discrimination in pigeons. Behavioural Brain Research,81,183-188.

Watanabe, S. and Ito,Y. (1991) Individual recognition in pigeon. Bird Behaviour, 36, 20-29.

Watanabe, S., & Furuya, I. (1997) Usage of video display for study of visual cognition in pigeons. International Journal of Comparative Psychology, 10, 111-127

Watanabe, S. & Jian, T. (1993) Visual and auditory cues in conspecific discrimination learning in Bengalese finches. Journal of Ethology, 11, 111-116.

Watanabe, S., & TrojeN. F. (2006) Towards a "virtual pigeon": A new technique to investigate avian social perception. Animal Cognition, 9, 271-279.

Wild, J. M., Karten, H, J, & Frost, B. J. (1993) Connections of the auditory forebrain in the pigeon (Columba livia). Journal of Comparative Neurology, 337, 32-62.

Young, A., Hellawaii, D. J. & Hay, D. C. (1987) Configural information in face perception. Perception, 16, 747-759.

(This paper was submited to *Social Neuroscience*)