Acoustic communication in Bengalese finches:
Studies Using Visual stimulus

2008 年 1 月

千葉大学大学院自然科学研究科
多様性科学専攻 理学
関 義正
Contents

Abstract................................................................................................................................. ii
Acknowledgements................................................................................................................ iii

General Introduction ............................................................................................................. 1

Chapter 1: Perception of AudioVisual stimulus

   Section 1: Sex difference in the audio-visual discrimination learning by Bengalese finches (*Lonchura striata* var. *domestica*)................................................................. 4
   Section 2: The effect of sound location in a song discrimination task in Bengalese finches (*Lonchura striata* var. *domestica*)................................................................. 33

Chapter 2: Production of Acoustic signals and Visual stimulus interruption

   Section 1: Song production pattern of the Bengalese finch (*Lonchura striata* var. *domestica*) is consistent with the statistical “second-order” song unit... 47

Chapter 3: Looking for Neural visual pathway into Song production system

   Section 1: Effects of visual stimulation on the auditory responses of the HVC song control nucleus in anesthetized Bengalese finches................................. 86
   Section 2: Functional evidence for internal feedback in the songbird brain nucleus HVC......................................................................................................................... 110

General Discussion.............................................................................................................. 128

References............................................................................................................................ 135
Abstract

Songbirds, including Bengalese finches, often use acoustic signals in social communication. One type of signal is called “song” and is a learned vocalization. Most studies of birdsong have attended to vocalization and its perception. However, because most birds have well-developed visual systems, studies of acoustic communication in songbirds can include visual stimuli. This dissertation is based on five behavioral–neuroscientific studies from such a viewpoint. In the first chapter, I examine sex differences in an audiovisual discrimination task and the importance of correspondence between visual attention and sound source in an auditory discrimination task. In the second chapter, I show that songs have a kind of hierarchical structure in the production of sound sequences using visual stimuli. In the third chapter, I discuss visual input to the song control nervous system based on electrophysiological experiments.

These studies provide new insights about acoustic communication in songbirds. Moreover, I note some similarities between the acoustic communication of Bengalese finches and that of humans. This dissertation contributes to furthering understanding of universal acoustic communication in various systems.
Acknowledgements

I am grateful to my research advisor, Dr. Kazuo Okanoya, for supporting my study and guiding me to Dr. dissertation.

I thank my advisory committee members: Profs. Yuichi Ochiai, thank my advisory committee members: Profs. Yuichi Ochiai, Yoh Kohori, Masakazu Iwasaka.

I appreciate to Drs. Tetsu Okumura, Yumiko Yamazaki, and Kohta Kobayasi for providing many advices. I also thank past and the present members of Lab. for Biolinguistics at RIKEN-BSI for many discussions and assistance.

Finally, I express my deep appreciation to my family and Chihiro Ikeda for supporting my life.

This research was partly supported by the Sasakawa Scientific Research Grant from the Japan Science Society.
General Introduction

Many animal species, from flies to elephants, communicate with each other using vocalizations or vibrations (Taub & Eberl 2003; Poole et al. 2005). Although the velocity of acoustic signals is slower than that of visual signals, these kinds of communications are available anywhere if the vehicle exists, even when visual signals cannot be used. Moreover, acoustic communication is available in one-to-many situations. Humans have a well-developed visual system, but we communicate with each other using acoustic signals even in situations in which visual information is available.

Songbirds

Animals that use acoustic communication can be categorized into two groups. The first group uses only innate vocalizations and the second group can use learned vocalizations. Songbird species are remarkable examples of the second group. In Japan, Hill Mynas (Gracula religiosa) are good imitators of human voices. In Europe and America, mockingbirds (Mimus polyglottos) are renowned for their mimicry of various sounds. As these examples show, many songbirds are good vocal learners. In many songbird
species, juvenile birds learn vocalizations from tutors (in many cases, the tutors are their fathers) and use the learned vocalization to communicate with each other (Immelmann 1969). Because the learning ability of songbirds is unique and rearing songbirds is relatively easy, many behavioral (e.g. Marler 1970) and neuroscientific (e.g. Nottebohm et al. 1976) studies have used songbirds. Therefore, I used Bengalese finches (*Lonchura striata var. domestica*), a songbird species, to investigate acoustic communication.

**Acoustic communication and visual information**

In general, acoustic communication studies attend to auditory information processing and vocal patterns. However, in cross-modal or multimodal studies, information processing of auditory or vibration signals is related to the processing of visual signals (e.g., spiders, Vandersal & Hebets 2007; rats, Komura et al. 2005; humans, McGurk & MacDonald 1976; Bushara et al. 2003). Additionally, visual stimuli have been used in studies of not only acoustic signal perception but also vocal production (Cynx 1990; ten Cate & Ballintijn 1996; Riebel & Todt 1997; Franz & Goller 2002; Miller et al. 2003).

This dissertation contains five studies using Bengalese finches. The remarkable feature of this dissertation is that several visual stimuli were used in studies of acoustic communication. These studies show a sex difference in audiovisual discrimination.
learning and that a lack of correspondence between visual attention and sound source location affects acoustic discrimination learning. I also demonstrate the hierarchical structure of the vocal pattern in Bengalese finches and argue two possible visual pathways to the song control nervous system of songbirds.

This dissertation provides natural history knowledge and helpful suggestions for cross-modal or multimodal studies in various animals, including humans.
Chapter 1: Perception of AudioVisual stimulus

Section 1:

Sex difference in the audio-visual discrimination learning by Bengalese finches (Lonchura striata var. domestica)

Abstract

Both visual and auditory information are important for songbirds, especially in developmental and sexual contexts. To investigate bimodal cognition in songbirds, we conducted audiovisual discrimination training in Bengalese finches. We used two types of stimulus: an "artificial stimulus," which is a combination of simple figures and sound, and a "biological stimulus," consisting of video images of singing males along with their songs. We found that while both sexes predominantly used visual cues in the discrimination tasks, males tended to be more dependent on auditory information for the biological stimulus. Female responses were always dependent on the visual stimulus for both stimulus types. Only males changed their discrimination strategy according to stimulus type. Although males used both visual and auditory cues for the biological stimulus, they responded to the artificial stimulus depending only on visual information, as the females did. These findings suggest a sex difference in innate auditory
sensitivity.

**Keywords:** Bengalese finch, birdsong, multi-modal discrimination, operant conditioning, sex differences.

**Introduction**

Many avian species are highly dependent on visual cues. Because of this characteristic, pigeons and other avian species have been used as appropriate subjects for many visual discrimination tasks. Bengalese finches (*Lonchura striata var. domestica*) also have sufficient visual capacity to be used for discrimination tasks in operant conditioning. For example, Watanabe et al. (1993) showed that this species could discriminate the video images of conspecific individuals. Moreover, video images could elicit courtship behavior from male Bengalese finches. Ikebuchi and Okannoya (1999) and Takahasi et al. (2005) reported that male Bengalese finches sang to images of females.

Auditory stimulus is also important for many birds, particularly songbirds. Researchers using operant conditioning techniques have demonstrated that Bengalese finches can clearly discriminate differences between conspecific distance calls (Okanoya & Kimura, 1993) and conspecific songs (Ikebuchi et al., 2000; Okanoya
Songbirds are known to be good vocal learners. (For a review on song learning, see Brainard & Doupe, 2002.) In many songbirds, including Bengalese finches, only the males sing. Males develop their own songs when young by listening to their tutors' songs. This auditory experience is necessary for songbirds to learn to sing normal songs (Marler, 1970). Moreover, in zebra finches (Taeniopygia guttata), which are closely related to Bengalese finches, auditory input and visual information seem to be important for song learning (Bolhuis et al., 1999). Bolhuis et al. (1999) suggested that both auditory and visual information might also be important for male Bengalese finches. Clayton (1988a) reported that visual cues might be more important than vocal cues in the song tutor choice of male zebra finches and Bengalese finches. These auditory and visual signals may also be very important to females in their choice of mating partners, as male courtship behavior involves songs and dances.

Several behavioral studies have been conducted on zebra finches to investigate relationships between auditory and visual cues, especially from the biological perspective of "female choice" in sexual selection. Using two subspecies of zebra finch, Clayton (1990) found that both song and breast-band size could play a role in mate choice. Brazas and Shimizu (2002), using live birds and their songs as visual and
auditory stimuli, reported that visual cues alone were effective in initiating choice behavior in females, while auditory cues facilitated the choice based on visual information. Collins et al. (1994) showed that a high song rate is more important than vivid beak color in females' choices, although these two characteristics are highly correlated. These results indicated that not only visual but also auditory cues are important in courtship contexts.

Zebra finches have prominent sexual dimorphism in their visual characteristics. This is in contrast to Bengalese finches, which have few visible dimorphic characteristics. In a study using combinations of static images and contact calls of conspecific birds as discriminative stimuli, Watanabe et al. (1993) showed that Bengalese finches tended to rely on visual rather than on auditory cues, as long as the visual cue was salient enough. This may mean that visual cues are more important than auditory cues for this species to be able to discriminate between individuals, as is true in zebra finches. Although the researchers used contact calls as an auditory stimulus, songs can also investigate the role of auditory cues in cognition among individual Bengalese finches. Because Bengalese finches have few sexually dimorphic features that are visible, songs may play a more important role in the mate choices of this species than they do for zebra finches. This means that if songs instead of contact calls had been
used for discrimination, females might use the signals, even in experiments with operant conditioning.

This research used an operant conditioning method to examine how Bengalese finches recognize audiovisual information. We prepared two types of audiovisual compound stimuli. One type, the "artificial stimulus," consisted of computer-made simple artificial figures and sounds. The other type, termed the "biological stimulus," was a combination of video movies of singing conspecific birds and their songs. These different kinds of stimuli allowed us to clarify how biological signals affect the task of discrimination. In addition, we examined whether there is a difference between the sexes in the usage of two different sensory stimuli (visual and auditory). Moreover, we examined whether the type of stimulus leads to a difference in sensory dominance. We also sought to discover how males and females recognize singing neighbors, and to understand the interaction of visual and auditory inputs in songbird cognition.

**Methods**

**Subjects**

We began this study using four male and four female Bengalese finches. However, one
female bird was retired from the experiment during the training period because she failed to stay on the observing perch during the provided intervals. Consequently, we added two females to the subject pool. As a result, complete experiments were conducted with four males and five females, referred to below as the nine subjects of the study.

The birds were kept in an aviary at Chiba University, Japan. All subjects were adults (12–50 months old) and weighed between 14 and 21 g. The temperature and the humidity of the aviary were maintained at about 25°C and 60%, respectively, and the light/dark cycle was 13/11 hours. During the experimental period, subjects were kept together in a steel cage (D 25.5 × W 31.5 × H 40.0 cm) and were fed only 3 or 4 hours before the daily training and testing, although vitamin-enhanced water was available ad libitum. Prior to this study, two males and two females had participated in auditory operant experiments, where they pecked at a key in order to obtain food reinforcement. No other subjects had experimental experience.

Apparatus

The experiment chamber was a mesh cage (D 35.5 × W 18.5 × H 26.0 cm) for small birds. One of the walls was removed and a transparent acrylic plate was attached in its
place. A 14.5-inch thin film transistor (TFT) liquid crystal display monitor (resolution: 1024 × 748 pixels), equipped with two loudspeakers (MT145X; Akia, Tokyo, Japan), was placed in front of the acrylic plate of the cage, and visual stimuli were presented on the monitor. The luminance (cd/m²) and chromaticity (CIE 1931; Y, x, y) of the monitor, measured with a colorimeter (CS-100A; Minolta, Tokyo, Japan), was 85.1 (0.274, 0.327) when displaying white color. In behavioral investigations, TFT monitors are better substitutes for real objects than are CRT monitors (Ikebuchi & Okanoya, 1999). There were two perches in the chamber, located at distances of 8 cm and 19.5 cm from the monitor and a height of 4 cm from the mesh floor. The perch nearer to the monitor was used for response and the other perch was used to observe the stimuli. To detect whether the bird was on a perch, photo-sensors (MP-L50A-DC24V; MATSUSHITA, Osaka, Japan) were attached to both sides of the perches. A feeder (Okubo Sokkoki, Tokyo, Japan) was placed at the right side of the acrylic plate, and a food cup was attached between the response perch and the monitor. When the feeder was used, one or two grains at a time would come into the food cup. The chamber was lit with a miniature bulb.

The experiment chamber was placed in a sound attenuated box (SC-2: W 70.0 × D 88.5 × H 95.0 cm; Music cabin, Tokyo, Japan). Inside the chamber, the brightness
was 15.70 (vertical) / 11.66 (horizontal) lx when the light was on, and background noise was measured at 22 dB. A personal computer was used to control the experiment.

**Stimulus**

We used two types of stimulus: a "biological stimulus" and an "artificial stimulus."

**Biological stimulus** We used video movies of two male Bengalese finches as biological visual stimuli. The birds displayed in the stimuli were unfamiliar to the subjects. These video movies had been recorded while birds were singing undirected songs (Sossinka & Bohner, 1980) in a cage. Video movies were taken by a digital video camera (VL-MR1PRO; Sharp, Osaka, Japan), and these records were captured as ".avi" movie files (frame rate: 29.97 frames/sec, frame size: 720 × 480) and edited using Premiere Pro software (Adobe systems, Tokyo, Japan). One bird had white feathers and the other had brown feathers. For the head area of the brown bird, the luminance and the chromaticity of these stimuli on the monitor were 4.91(0.290, 0.334); for the head area of the white bird, the luminance and chromaticity were 15.5(0.253, 0.278). The songs of these stimulus birds were recorded with a DAT recorder (DTZ-ZA5ES; Sony, Tokyo, Japan) through a microphone (MS957; Sony, Tokyo, Japan) as biological auditory
stimuli. These songs were converted to Windows PCM files (".wav" format, sampling rate: 22.05 kHz) using software (Avisoft SASlab Pro; Avisoft bioacoustics, Berlin, Germany). These movie files and sound files were synthesized to the "avi" format using Premier Pro. The sound pressure level of the songs was adjusted to 60–70 dB at the observing perch.

*Artificial stimulus*  Simple figures served as artificial visual stimuli. One was a circle (220 pixels in diameter) and the other was a cross (210 × 270 pixels). Both images were drawn with black lines on a gray background. Two types of auditory stimuli were used as the artificial auditory stimulus; one was a 3 kHz pure tone, and the other was white noise. These were made with SASLab Pro. The sound pressure level was also adjusted to about 65 dB at the observing perch. These were presented in the same way as the biological auditory stimuli.

*Procedure*

After they had adapted to the feeder, the subjects were given food when they moved from the observing perch to the response perch. In the next step, each subject was trained to stay on the observing perch for more than 1 s before jumping to the response
perch. The length of time they stayed on the observing perch was gradually increased, until it reached 2–3 s.

Each subject was given two types of tasks that pertained to the discrimination of both biological and artificial stimuli. Subjects were randomly divided into two groups. One group started with the biological discrimination task, followed by the artificial task; in the other group, the order of these tasks was reversed. Both tasks were carried out with the same training procedures. However, because subjects had already learned to wait on the observing perch in the first task, some of the early training procedures were omitted in the second task.

The subjects were then trained in a Go–NoGo type of discrimination. In these training sessions, subjects were trained to stay on the observing perch for more than 4 s. If the subject stayed on the observing perch for 4 s, a visual stimulus appeared on the TFT monitor and an auditory stimulus was played from the loudspeakers. These auditory and visual stimuli were presented simultaneously and lasted for 3 s. Two different combinations of visual and auditory stimuli were used. The response to one combination (V1A1: S+) was reinforced by a food reward (Hit) consisting of two or three pieces of grain from the feeder. The response to the other combination (V2A2: S-) was followed by a blackout that lasted for 16 s (False Alarm; FA). The stimulus
combinations were balanced among the subjects. The limited hold for a response was 2 s. When \( S^+ \) was presented, the subject was not given any food if it did not move within the period (Miss). If the subject stayed on the observing perch in response to \( S^- \), the next trial was started (Correct Rejection; CR). When the subject moved to the response perch before the stimulus had stopped being played, the visual stimulus disappeared and the auditory stimulus stopped simultaneously; the bird then had to return to the observing perch and again stay there for 4 s. In cases of Hits and FAs, the subjects returned to the observing perch after the reinforcement or the punishment, and the next stimulus was started 4 s after that point of time. In cases of Misses and CRs, the next stimulus was started 4 s after the offset of the previous stimulus (limited hold + 2 s).

One session consisted of 25 trials of \( S^+ \) and 25 trials of \( S^- \). The two types of stimuli were presented randomly. If the subjects made incorrect responses, correction trials were repeated until subjects made correct responses. The criterion of acquisition was set at 90% correct performances over two consecutive sessions. The percentage of correct responses was calculated by the formula \( \frac{\text{Hits} + \text{CRs}}{\text{Hits} + \text{Misses} + \text{FAs} + \text{CRs}} \times 100 \).

When the birds attained the criterion of acquisition, a test session was conducted on the next day. In a test session, six probe test trials were randomly
interspersed within 54 training trials (27 each of S+ and S-). Therefore, each test session consisted of 60 trials. Three types of probe stimulus were used: auditory stimulus alone (A1, A2); visual stimulus alone (V1, V2); and mixed stimuli (V1A2; V2A1). Each probe stimulus was presented once per test session. Neither food nor blackouts were given for responses to probe stimuli. The test sessions were conducted five times for each subject, and training sessions were inserted between the test sessions to examine whether the criteria were maintained. If the subjects showed over 90% correct performance in two training sessions inserted between test sessions, the next test session was conducted. However, if a subject failed to show such performance, retraining was conducted until the bird reached the criterion, and then the bird was tested again. In this way, we obtained results for 30 probe test trials (six types × five test sessions) per subject, and we analyzed the number of Go-responses to the probe stimuli.

Statistical Analysis.

We used a two-way repeated-measures analysis of variance (ANOVA; stimulus type [biological or artificial] × sex) to analyze the number of training sessions required to achieve the discriminative criterion; a three-way repeated-measures ANOVA (probe stimulus [6 types] × stimulus type [biological or artificial] × sex [male or female]) to
analyze the Go-responses to the probe stimuli; and a multiple comparison test (Tukey's honestly significant difference [HSD], confidence interval: CI = 0.95) to compare the means for the levels of each factor.

The effects of task order (i.e., which type of task was performed first) and experimental history prior to the present study were also analyzed.

Results

Training sessions

All but one subject reached the discriminative criterion (>90% correct responses in two consecutive sessions) for all tasks within 4 to 15 training sessions. For one male bird, the criterion was lowered to >85% correct responses in two consecutive sessions. This bird did not reach the criterion until the thirteenth session, although he could reach >85% correct responses in several sessions; we adopted the lowered criterion for this bird to avoid over training. We then continued the test with the bird. The scores improved after the first test session. Therefore we regarded the performance of this bird to be very similar to that of other subjects. In the discrimination learning of the biological task, more training sessions were required than for the artificial task [F(1, 14) = 4.862, p < .05], however, there was no sex difference in the number of training
sessions needed to reach the criterion \( [F(1, 14) = 0.128, p = .726] \) and no interaction between sex and the stimulus type ["biological" or "artificial" stimulus; \( F(1, 14) < 0.001, p = .987 \)].

In most training sessions inserted between the test sessions, the results showed that the birds maintained the learning criteria (>90% correct in two sessions). In a few cases, the scores were <90% (twice in a male in the artificial task, once in a female in the biological task); these birds were trained until they achieved scores of >90% in two consecutive sessions. Three retraining sessions for the male and five retraining sessions for the female were required to reach the criterion; we then continued the tests with these birds.

**Test sessions**

Because V2A2 was associated with punishment in the training sessions, Go-responses to V2 and A2 were rare in both males and females for both the biological (V2, males: 0%, females: 4%; A2, males: 0%, females: 0%) and artificial tasks (V2, males: 5%, females: 4%; A2, males: 0%, females: 0%), as expected. If the probe stimulus included V1, Go-responses were elicited by the artificial stimulus in almost all tests in both males (V1A2: 100%, V1: 95%) and females (V1A2: 100%, V1: 100%), whereas for the
biological stimulus, there were fewer Go-responses by males (V1A2: 85%, V1: 60%), and the females' responses were similar to those for the artificial stimulus (V1A2: 100%, V1: 92%). In contrast, if the probe stimulus included A1, males showed more Go-responses to the biological stimulus (V2A1: 45%, A1: 25%) than to the artificial one (V2A1: 15%, A1: 10%), although there was no apparent difference in females' responses between the two stimulus types (bio: V2A1: 0%, A1: 8%; art: V2A1: 0%, A1: 0%). The following paragraphs present the results of statistical tests of these data.

The statistical tests revealed no effect for the factors "stimulus type" ["biological" and "artificial" stimulus; F (1, 84) = 0.09, p = .76] and "sex" [F (1, 84) = 1.19, p = .28]. This shows that the activation levels or the response strengths of the subjects did not change in these factors. As expected, the difference among the six types of "probe stimulus" was significant [F (5, 84) = 223.65, p < .001]. The multiple comparison test revealed significantly more Go-responses to V1A2 and V1 (these included V1) than to V2A1, A1, V2, and A2 (p < .001). There were also more Go-responses to V2A1 than to A2 (p < .05; See Table 1 in the supplement).

The interaction of stimulus type × sex was not significant [F (1, 84) = 0.116, p = .734]; however, the stimulus type × probe stimulus was significant [F (5, 84) = 4.216, p < .01]. The multiple comparison test showed significantly more Go-responses to
V1A2 and V1 (these included V1) than to V2A1, A1, V2, and A2 for the biological, artificial, and biological–artificial stimuli ($p < .001$). There were more Go-responses to V1A2 in the artificial task than to V1-only in the biological task ($p < .05$). Subjects gave 100% Go-responses to V1A2 in the artificial task, but subjects sometimes did not respond to V1 in the biological task (See Table 2 in the supplement).

The interaction of sex $\times$ probe stimulus was also significant [$F (5, 84) = 8.026, p < .001$]. The multiple comparison test revealed many more differences than the above instances. The Go-responses to V1A2 and V1 (these included V1) were also significantly more than to V2A1, A1, V2, and A2 in males, females, and males–females ($p < .001$); males responded with fewer Go-responses to V1 than did females to V1A2 ($p < .05$). This difference reflects the fact that V1 elicited relatively fewer responses from males than from females, while females responded to all V1A2. Moreover, there were significantly more male Go-responses to V2A1 than female Go-responses to V2A1 ($p < .001$), A1, V2 ($p < .01$), and A2 ($p < .001$) and male Go-responses to V2, A2 ($p < .01$). This suggests that V2A1 elicited more responses from males (See Table 3 in supplement).

The interaction of stimulus type $\times$ sex $\times$ probe stimulus was significant [$F (5, 84) = 2.662, p < .05$]. The multiple comparison test clarified which relations affected the
results of the statistical tests shown above. Both stimulus tasks for the females but only
the artificial task for the males showed a simple tendency for more Go-responses to
V1A2 and V1 than to V2A1, A1, V2, and A2 ($p < .001$); no significant differences were
found among other levels. However, the males’ Go-responses to the biological stimulus
revealed some other differences. Males showed significantly more Go-responses to
V2A1 in the biological task than did females ($p < .001$; Fig. 1a), as well as for A1 ($p <
.01$), V2 ($p < .01$), and A2 ($p < .001$) in the biological task, and V2A1, A1 ($p < .001$),
V2 ($p < .01$), and A2 ($p < .001$) in the artificial task. Male Go-responses were also
greater for V2A1 than for V2, A2 in the biological task ($p < .001$) and A1 ($p < .05$), V2
($p < .01$), and A2 ($p < .001$) in the artificial task. Interestingly, the females did not
respond to V2A1 at all, even in the biological task. This suggests that A1 could elicit
many more Go-responses from males in the biological task, if the stimulus included a
visual stimulus. There were fewer male than female Go-responses to V1 in the
biological task ($p < .05$; Fig. 1a) and in the artificial task for both males ($p < .05$) and
females ($p < .01$). Moreover, there was no significant relation of male Go-responses
between V2A1 and V1 and between V1A2 and V1 in the biological task. This result
suggests that male responses had lower dependence on V1 than did female responses
only in the biological task. These interesting differences were not observed at all in the
artificial task (Fig. 1b). Thus, males’ responses were more auditory-information dependent than were females’ responses only in the biological task (See Table 4 in the supplement).

*Differences originated in experiences*

We calculated a three-way ANOVA to examine the effect of experimental order on the results. The analysis showed no differences between the groups for all probe stimuli in both stimulus types [stimulus type × groups: F (1, 84) = 0.498, p = .482; probe stimulus × groups: F (5, 84) = 0.414, p = .414; and stimulus type × groups × probe stimulus: F (5, 84) = 0.519, p = .761].

The same analysis was conducted to examine the effect of the subjects’ previous operant experience. The results did not show any significant differences [stimulus type × experimental history: F (1, 84) = 0.164, p=.687; probe stimulus × experimental history: F (1, 84) = 0.831, p = .531; and stimulus type × experimental history × probe stimulus: F (5, 84) = 1.739, p = .135].

These results showed that the females consistently used the visual cues for both biological and artificial stimuli. However, the males were more strongly affected by
auditory cues than were the females, especially for the biological stimulus. The females showed a high average response to the probe stimulus V1 (>90% average) for both biological and artificial stimuli, whereas the males' responses to the stimuli (60%) decreased significantly only for the biological stimulus. The males responded to V2A1 with a considerably high probability (45%), although the females did not respond at all to that stimulus. These results indicated that the females' responses were strongly V1 dependent regardless of the stimulus type, but the males' judgments were affected by whether the stimulus included A1 for the biological stimulus.

Discussion

This research clearly shows that different stimulus types elicit different responses. All data analyses revealed that males and females tended to be more dependent on visual information than they were on auditory information in experiments testing both biological and artificial discrimination. These results are consistent with a previous study of Bengalese finches (Watanabe et al., 1993). Some food-reward operant studies have shown that auditory stimuli are comparatively less effective than visual stimuli, or that visual information overshadows auditory information in discrimination learning using light and tone/noise, in rats (Colwill & Rescorla, 1990) and pigeons (Kraemer &
Roberts, 1985). In our study, although the auditory Go-stimulus (A1) could elicit Go-responses, the response was much less than expected if the subjects had truly learned to 'Go' only in the auditory stimulus. This result may be related to differences in visual and auditory learning capabilities. Interestingly, auditory stimuli were more effective than light stimuli in stimulus control for shock avoidance in pigeons (Foree & LoLordo, 1973) and rats (Field & Boren, 1963). These studies demonstrated that visual stimuli are not always more effective than auditory stimuli in discrimination learning and that the priority of such senses can be influenced by context. In our experiment, the males' responses varied, appearing to be dependent on the stimulus (V1 or A1) when the stimulus type (biological or artificial) was changed. In contrast, the females' responses did not vary in this way; they relied on visual cues at all times.

*Sex difference in audio-visual discrimination*

The results suggest that males used audiovisual information differently from females. In this experiment, visual information was relatively important in order for both males and females to discriminate stimuli. The females responded strongly to probe stimuli if V1 was included, regardless of auditory stimulus. However, for the biological stimulus, the males' responses were much more A1 dependent than were those of the females. If the
stimulus included A1, the males responded to the stimulus even though the visual information (V2) had been associated with punishment in the trained stimulus combination (V2A2). This result is not consistent with a previous audiovisual playback experiment in pigeons (Ryan & Lea, 1994), which reported that males reacted to video images, but did not react at all to cooing sounds. However, this may reflect a difference between species. The Bengalese finch is a songbird, while pigeons are not. As we will discuss later, male songbirds have some good reasons to be sensitive to auditory signals.

As shown in Figure 1a, males responded more to V2A1 than they did to A1 alone. This suggests that while the auditory stimulus alone could elicit Go-responses from males, their responses would be elicited more efficiently if A1 was combined with a visual stimulus. In contrast, females’ responses to the probe stimuli in both biological and artificial experiments indicated that they depended only on visual cues.

The results also indicate that the males’ discrimination behaviors changed according to the stimulus type, whereas the females did not show this tendency. The statistical analyses suggest that males might use songs as a discrimination cue more efficiently than do females in the biological task. This result seems to be consistent with the findings of a previous operant study (Cynx & Nottebohm, 1992), in which zebra
finch females required more trials to reach the criterion of acquisition than did males in the song discrimination task; however, we did not achieve these results in similar studies of the Bengalese finch.

Below, we discuss some possible explanations for our results, associated with song learning, the nervous system, and overshadowing based on the ethological background of songbird species.

*The song control system and song learning in male songbirds*

Adret (1993) found that juvenile male zebra finches could learn a taped song with no visual information, although Houx and ten Cate (1999a) reported that social interactions facilitate song learning. While all subjects in our study were adults, these findings may explain why males were more sensitive to auditory input.

The brain structure of songbirds shows conspicuous sexual dimorphism, which could have caused these results. As discussed above, as in many other birds, only male Bengalese finches can learn and sing songs. Male brains possess a remarkably developed structure called the song control system, which is composed of nerve nuclei; in female brains, these structures are much smaller or vestigial (Nottebohm & Arnold, 1976). The brains of female Bengalese finches, in particular, lack some song nuclei
(Tobari et al., 2004). The lesions of some song nuclei affect song discrimination in some birds (Scharff et al., 1998; Burt et al., 2000; Gentner et al., 2000; Okanoya et al., 2001). The volume of a song nucleus is also involved in processes of song perception and discrimination (Leitner & Catchpole, 2002). These studies have shown that the song control system is critically related to song recognition. The sex differences in the neural structure may also relate to the results of the present audiovisual discrimination.

*Mating behavior and imprinting in females*

Our results suggest that females attended less to the songs than did males. However, it seems reasonable to expect that females would listen carefully to conspecific songs, which are an important component in male courtship behavior and female mate choice. An audiovisual video playback experiment using pigeons showed that the audio channel effectively elicited female courtship responses (Partan et al., 2005). Although pigeons are not songbirds, this result supports the hypothesis that auditory information is an important cue for females to be stimulated in a sexual context. In our experiment, however, females did not seem to attend to songs as a discriminative cue.

One possible explanation is that this experimental condition is too far removed from normal situations in which females observe singing males. A previous study of
female cowbirds that investigated whether listening to songs in the sexual context differed from listening in a food-related context (Johnsrude et al., 1994) seems to support this idea. For females, courtship behavior is quite different from the discrimination of audiovisual stimulus in operant tasks. Male birds, however, may use auditory cues for discrimination as their strategy in nature.

The other possibility is that females at first visually recognize conspecific males, and only then will listen to the songs. Because some songbird species can even copy the songs of other species, to choose a mating partner based on only auditory information is inappropriate for females, and visual information could take precedence over auditory signals. From the viewpoint of imprinting, auditory imprinting starts over 25 days after hatching (Clayton, 1988b). Early exposure to songs affects the preferences of zebra finch females (Lauay et al., 2004; Riebel et al., 2002), while early social imprinting begins with visual information only 10 days after hatching; social imprinting then guides the development of sexual preference (Bischof, 2003). Thus, visual imprinting may be stronger than auditory imprinting, and the visual stimulus could overshadow the auditory stimulus. Without being able to see the singer, females attend to sounds more carefully. Some physiological and ethological studies have suggested that females distinguish between songs and that even recorded songs can elicit females'
responses (Clayton & Pröve, 1989; Ikebuchi et al., 2003; Okanoya, 2004). Although males are also imprinted by visual information earlier, males might pay more attention to auditory signals than do females because males have a strong tendency to do so in nature.

Another explanation and suggestions for future studies

There is another possible interpretation of the results. The stimulus sets we used were chosen to simplify the experimental design; however, as a result, the “biological” and the “artificial” stimulus sets were not equivalent. While the biological stimuli were very rich, the artificial stimuli were very simple. This difference may have caused some differences in the attention given to the stimuli by the male and female finches. The males may have paid more attention to the complicated acoustic features than did the females (even if the stimulus was not a biological one). On the other hand, if the females paid more attention to the complicated visual stimulus (regardless of whether the stimuli were biological) than did the males, the rich visual stimulus could have overshadowed the auditory stimulus, even if the females were as sensitive as the males were to songs. If so, then the results could be discussed from the viewpoint of “rich” versus “simple” rather than “biological” or “artificial.” This possibility could be
examined by using transformations of the biological stimuli for the artificial stimuli. Candidates for such transformations would be breaking the visual stimulus apart into pieces or using backwards songs as in acute electrophysiological experiments in songbirds (It has been known that such kind of auditory stimulus evoke fewer neural activities than the bird’s own songs in the song control nuclei of Bengalese finches (Seki & Okanoya, 2006), the same as other songbirds.). Additionally, it would be meaningful to examine whether the results would differ if non-conspecific birds (e.g., zebra finches) were used as stimuli. Another interesting experiment would be to use the tutor song for males and the cage-mate song or the imprinted song for females as auditory stimuli.

These possible explanations based on neural mechanisms, mating system and imprinting, and complicated or simplified stimulus features may or may not relate to the present results. However, our simplified experimental design reveals clear cognitive differences between male and female Bengalese finches in audio-visual discrimination and raises some interesting questions to be tested in future studies.
Figure Captions

*Figure 1a.* Responses to the probe stimuli in the “biological stimulus.” In the training session, the responses to V1A1 were reinforced by food and the responses to V2A2 were followed by blackouts. Error bars represent standard error of mean (SEM). Post-hoc (Tukey’s HSD) test showed there were sex differences in Go-responses to V1 (CI: −63.87 − 0.13, effect size: 1.05) and V2A1 (CI: 13.13 − 76.87, effect size: 1.72). *p < .05; **p < .01.

*Figure 1b.* Responses to the probe stimuli in the “artificial stimulus.” Post-hoc test showed that there were no sex differences in the response to the probe stimuli, in contrast to the “biological stimulus”.
Responses to Biological Stimulus

![Graph showing responses to different stimulus types for males and females.](image)

**Figure 1a**
Responses to Artificial Stimulus

<table>
<thead>
<tr>
<th>Stimulus Types</th>
<th>Go Response (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1A2</td>
<td>Male</td>
</tr>
<tr>
<td>V1</td>
<td>Female</td>
</tr>
<tr>
<td>V2A1</td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td></td>
</tr>
<tr>
<td>V2</td>
<td></td>
</tr>
<tr>
<td>A2</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1b
Section 2:

The effect of sound location in a song discrimination task in Bengalese finches (*Lonchura striata* var. *domestica*)

Abstract

To understand the relationship between song perception and stimulus location, Bengalese finches, a species of songbirds, were trained to discriminate between a normally played conspecific song and the reversed song. Loudspeakers on the opposite side of the response key (experimental condition) and above the subject (control condition) were used to present stimuli. Under the control condition, all subjects learned the task, whereas under the experimental condition some birds did not. In general, birds learned the task more quickly under the control condition than under the experimental condition. Results suggest that for these birds sound location maybe tightly associated with behavioral responses and bioacoustical investigations should be undertaken with these considerations into account.

**Keywords:** operant conditioning, songbird, song discrimination, sound location, touchscreen
Introduction

In many songbird species, only males sing songs, which they learn as juveniles from tutors in social relationships. Males use song as a courtship signal in adulthood (Sossinka & Bohner, 1980). Especially in estrildid finches such as Bengalese finches (Lonchura striata var. domestica) and zebra finches (Taeniopygia guttata guttata), because male courtship behavior includes both song and dance, females listen to males’ song while also being visually stimulated. In the wild and also in captivity, songs sung by different individuals may come from several directions simultaneously within a limited space. Therefore, in these estrildid finches, associating a visual target with the sound from the same direction is important if birds are to perceive the correct stimulus from a signaling bird in a song-learning or courtship situation.

We attempted to clarify how sound source location affects a song discrimination task in Bengalese finches. Because sound waves from various locations can reach the ears of a single listener, we expect birds to be able to discriminate songs regardless of the sound location. However, we previously noted that the location of a loudspeaker seemed to affect the performance of such tasks in Bengalese finches. Thus, we sought to confirm our observation under controlled conditions.
Materials and Methods

Subjects

We used four male and six female Bengalese finches (1-2 years old) that were kept in captivity in the aviary at the Laboratory for Biolinguistics, RIKEN-BSI. These birds had not been used in any experiments previously. The birds were randomly divided into two groups: two males and two females comprised the control group, and two males and four females comprised the experimental group. The temperature and humidity of the aviary were maintained at about 25°C and 60%, respectively, and the light/dark cycle was 13/11 hr. During the experimental period, the subjects were kept together in two steel cages (D 25.5 × W 31.5 × H 40.0 cm) and were fed 3 or 4 hr before daily training; vitamin-enhanced water was available ad libitum.

Apparatus

We used an infrared touchscreen (Unitouch, Touch Panel Systems, Japan) attached to a 15-inch thin film transistor (TFT) liquid crystal display monitor (LCD-A15V, Iodata, Japan; 1024 × 748 pixel resolution). The touchscreen has an acrylic plate that detects the movement of a human finger using scanning lines; however, the beak of a Bengalese finch is so small that the touchscreen could not detect the responses. Therefore, we
removed the panel so that the subject pecked the TFT monitor directly, causing the bird's head to block the scanning lines and allowing the touchscreen to detect the bird's responses. We used a plastic cage designed for small birds (W 15.5×D 30.0×H 22.0 cm) from which one wall was removed and replaced with the touchscreen. Inside the cage, a perch was attached 4.0 cm from the touchscreen and 10.0 cm from the floor. A food cup was located between the perch and the touchscreen, and a feeder (Okubo Sokkoki, Japan) provided one or two grains at a time to the cup as the reward. One loudspeaker (8 Ω, 15 W, 6φ) was placed 6.0 cm above the roof of the cage and another was placed 6.0 cm behind the cage. The system was placed in a ventilated sound-attenuated chamber (inside dimensions: D 40× W 58 × H 37 cm) in which the background noise was 46 dB. The chamber was lit with a miniature bulb (12 V, 0.1 A). The entire system was controlled by a personal computer.

_Simuli_

We used a Bengalese finch song that had been recorded with a sampling frequency of 44.1 kHz at Chiba University; the subjects had not heard the song previously. We converted the song to a Windows PCM file (.wav) and used only a 1.0-sec portion. This was used as the "Go" stimulus, and the reversed song was used as the "NoGo" stimulus.
Avisoft SASLab (Avisoft bioacoustics, Germany) was used to edit the stimuli. The maximum sound amplitude was 74 dB at the location of the bird's head. We presented visual keys (white and black; 20 pixels per circle) on the TFT monitor.

**Procedure**

The experiments were conducted between 10:00 and 13:00. Once the subjects had adapted to the feeder, they were trained to peck the white key. Responses to the key were reinforced by the reward, and the inter-trial interval (ITI) was set at 2 sec. When a subject spontaneously pecked the white key 50 times, the discrimination task was started on the following day.

The experimental group was assigned to the rear-speaker condition (i.e., the experimental condition) and the control group was assigned to the overhead speaker condition (i.e., the control condition).

Birds were trained by Go/NoGo discrimination. One session consisted of 102 trials (Go and NoGo stimuli were randomly presented 51 times each), and one session was performed per day for each subject. In the discrimination task, when the bird pecked the black key (i.e., observation key), the key disappeared and the stimulus song was played immediately. After the song terminated, the white key (i.e., report key) appeared
(limited hold 2 sec). In the Go trial, if the bird pecked the report key (Hit), a reward was provided. In the NoGo trial, if the bird pecked the report key (False Alarm: FA), the room light was turned off for 10 sec. The next trial began after the ITI. If the bird did not peck the key in a Go trial (Miss) or NoGo trial (Correct Rejection: CR), the next trial started following the ITI. The percentage of correct trials was calculated by $(\text{Hits} + \text{CRs}) / (\text{Hits} + \text{CRs} + \text{FAs} + \text{Misses}) \times 100$. When a bird attained $>80\%$ correct responses twice (or $>70\%$ correct in eight sessions for a low-performance bird), we concluded that the bird had acquired the ability to discriminate. To confirm that these birds could maintain performance regardless of speaker location, we then tested these birds in five sessions under the other speaker condition. For subjects that did not reach the performance criterion within 26 sessions, the speaker condition was changed and training was continued for up to 26 more sessions.

Analyses

We used three-way repeated-measures analysis of variance (rm-ANOVA: group $\times$ sex $\times$ session) and two-way rm-ANOVA (group $\times$ session) to analyze our data.
Results

In the experimental group, three of six birds acquired the discrimination ability (in 13, 16, and 25 sessions). However, the other three birds did not meet the performance criterion. These birds had 50-55% correct responses in the 26th session. On the other hand, all birds in the control group reached the performance criterion (in 16, 16, 17, and 23 sessions). One control bird repeatedly had relatively low scores (73-78% correct responses) after it reached 70%. Thus, we adopted the criterion of >70% correct for eight sessions.

Using the session number during which the first bird attained the performance criterion (i.e., the 13th session), we examined the learning curves with a three-way ANOVA. There were significant main effects of session (df = 12, F = 2.512, p < 0.01) and group (df = 1, F = 6.216, p < 0.05; Fig. 1). The effect of sex was not significant (df = 1, F = 0.78, p = 0.38), and there were no significant interactions. These results indicate that when the sound source was located on the opposite side of the visual key, the birds needed more time to learn song discrimination.

After birds acquired the ability to discriminate, they were tested under the other speaker condition. In the first session under the new condition, two birds had about 60% correct responses. However, the average score over five sessions was 73.8-84.8%. There was no significant difference in scores among the five sessions (two-way ANOVA,
group: df = 1, F = 0.83, p = 0.37; session: df = 12, F = 1.27, p = 0.21). Thus, the birds from both groups maintained their performance regardless of speaker location once they had learned the task. This suggests that the sound was equally easy to hear under both conditions.

The birds that did not learn the task in the experimental condition were tested under the control condition after the 26th session. All of these birds attained >80% correct responses within 26 sessions, indicating that they could learn the task if the speaker was located in the control location. Using the session number during which the first bird reached >80% correct (i.e., the tenth session), we tested the learning curves of the two groups with a two-way ANOVA. The main effect of group was significant (df = 8, F = 13.66, p < 0.05), and session was not significant (df = 8, F = 1.73, p = 0.12; Fig. 2). Thus, after the speaker location was changed to the control location, the birds started to learn the discrimination task quickly. It is possible that they already recognized the difference between the stimuli, but could not associate the sound with the Go/NoGo task.

Discussion

When the sound was played from the opposite side of the visual key, half of the subjects could not learn the task within the established period. Under the control condition, all
subjects learned the task. Additionally, the learning curves clearly showed that the task was mastered significantly more quickly under the control condition than under the experimental condition. When the speaker location was changed to the control location, the birds that had been in the experimental group learned the task quickly. Thus, it may be difficult but not impossible for most birds to learn the association between responses to the key and a sound from the direction opposite the visual stimulus. This much said, it was still much easier for the birds to associate the song with the response key when the key was placed close to the sound source.

A previous operant study reported that passerine birds have a performance for discrimination of sound source location (Klump et al., 1986). Thus, in the present study, birds seemed to recognize the sound source direction. However, Lewald (1987) showed the acuity of sound localization became better in a classical conditioning using electrical shock rather than in an operant conditioning using rewards in pigeons. Therefore, a cognitive process of sound location might be affected by the “context”. Because we used not an artificial sound but their song as the discriminative stimulus, it might be affect the results.

In song learning of zebra finches, key pecking (Houx and tenCate, 1999a) and presentation of a stuffed bird as a visual stimulus (Houx and tenCate, 1999b; Bolhuis et
al., 1999) seemed not to affect the learning performance at all. In these studies, birds could copy the played-back tutor song passively in no visual stimulus condition. In the present song discrimination task, therefore, it seemed to be natural if birds can associate the task and played-back song automatically. However, it was difficult for experimental birds to do so. Interestingly, Hultsch et al. (1999) showed auditory-visual pairing enhance the song learning in hand-reared nightingales (*Luscinia megarhynchos*). Thus, effects of association between visual and auditory information might differ between species.

In Bengalese finches, because songs carry the properties of the signaler to the receiver and several songs can come from several directions simultaneously, the receiver must associate the song with the proper visual target. Therefore, it might be difficult for the subjects to associate responses to the visual key with the stimulus song from the opposite direction. Of course, in examining discrimination of sound location in some animals, placing the sound source adjacent to the manipulanda enables rapid learning (Harrison et al., 1971). However, the task in this study was song discrimination, not the discrimination of sound location. Moreover, close proximity of the three terms (i.e., stimulus, response key, and reinforcement place) might not always enhance discrimination learning using touchpanels (Markham et al., 1996).

In many studies of sound discrimination learning in songbirds, the speaker has
been located above and behind the response key or above the subject (e.g., Gentner and Hulse, 1998; Okanoya et al., 2000). Our findings provide a basis for deciding speaker location and may be valuable in improving future studies of song discrimination learning in songbirds.
Figure Legends

Figure 1. Birds under the control condition learned the discrimination task more rapidly than birds under the experimental condition (ANOVA: df = 1, F = 6.216, p < 0.05). In the experimental condition, the sound source was placed on the opposite side of the response key. In the control condition, the sound source was above the subject. Error bars represent the standard error of the mean.

Figure 2. Learning curve of the experimental group after the speaker location was changed to the control location. These birds learned the discrimination task more rapidly than naive birds under the control condition (df = 8, F = 13.66, p < 0.05). Error bars represent the standard error of the mean.
Figure 1
Figure 2
Chapter 2: Production of Acoustic signals and Visual stimulus interruption

Section 1:

Song production pattern of the Bengalese finch (*Lonchura striata* var. *domestica*) is consistent with the statistical “second-order” song unit

Abstract

Bengalese finch (*Lonchura striata* var. *domestica*) songs, unlike those of zebra finches (*Taeniopygia guttata*), do not show an unvarying repetition of motifs; instead, there are variations on partially stereotyped sequences of song syllables. Although these stereotyped sequences consist of multiple syllables, in most cases these syllables occur together. Thus, these stereotyped sequences are song structures and could be called “second-order song units.” To examine whether such structures really exist as a vocal production unit, we subjected singing birds to a light flash and examined when the stimulus stopped the songs. When light interruptions were presented within the stereotyped sequences, the subsequent syllables tended to be produced, whereas interruptions presented during variable sequences tended to cause instantaneous song
termination. This suggests that the associations among the song syllables that compose the stereotyped sequences are orderly than those for the variable sequences, and these relatively robust syllable associations indicate the existence of second-order song units. Additionally, following interruptions, several types of song sequence agitations were observed that had not been previously reported in other studies. These phenomena might be caused by the properties of Bengalese finch songs and the neural systems that control the complicated song structures.

**Keywords:** light flash, song sequence, song unit, songbird, song plasticity

**Introduction**

Birdsongs are sequential vocalizations that are comprised of various sounds, and many songbird species require periods of song learning to sing normal songs (Thorpe 1958). Thus, songbirds provide good models for studying the neuroscience of sequential motor control and learning, and numerous studies have been conducted using zebra finches (*Taeniopygia guttata*). The zebra finch is a closed-ended song learner that has a sensitive phase of song memorization during the juvenile period (Böhner 1990). Once the songs are crystallized, zebra finches produce highly stereotyped songs (Price 1979). The Bengalese finch (*Lonchura striata var. domestica*) is closely related to the zebra
finch and is also a closed-end song learner. However, Bengalese finch songs are complicated and variable compared to zebra finch songs, even after the songs are crystallized.

To facilitate comparison, songs can be visualized as sound spectrograms (Thorpe 1958) and can be divided into “song syllables” separated by silent intervals that appear on the spectrograms. Using this method, Bengalese finch songs appear to consist of units that are composed of several song syllables (Fig. 1). Additionally, while some syllables are always accompanied by other syllables, some syllables do not have such partners. These song structures are not found in zebra finch songs. Zebra finch songs also have hierarchical structures incorporating “syllables,” “motifs,” and “bouts” (Sossinka and Böhner 1980); however, this structure differs from the one of Bengalese finch songs. If a motif is a group of several syllable(s), Bengalese finch songs have motifs as same as zebra finch songs. However, while zebra finch songs consist of repetitions of single motif and all syllables are a member of the motif, Bengalese finch songs include several motifs and consist of various combinations of these motifs, and some syllables are not always a member of these motifs as shown in figure 1. So, in Bengalese finch songs, it is not easy to predict the subsequent motif of a motif in some cases and sometimes it is also not easy to predict the subsequent syllable. Therefore, in
this article, to distinguish syllable groups of Bengalese finch songs from motifs of zebra finch songs, we would like to use the term “second-order” unit, that is, a group of several syllables that always appear together. This structure is a unique feature of Bengalese finch songs.

Some studies reported about the structure of Bengalese finch songs (Immelmann 1969, Clayton 1987, Wooley and Rubel 1997). Then, Honda and Okanoya (1999) reported that Bengalese finches sing relatively complicated songs and suggested the presence of groups of song syllables. Okanoya (2004) analyzed these songs statistically and showed that some song syllables occurred together and could be grouped. Although these studies showed the existence of second-order song structures based on the statistical analyses of the song syllable sequences, that does not mean there is a second-order “motor unit” of song production. Here we used light flashes to interrupt singing behavior following the methods of Cynx (1990) to determine if these higher-order units really exist as motor patterns. Cynx (1990) used this method to successfully identify that the “syllable” is the motor unit of song production in zebra finch. Franz and Goller (2002) also used light flashes to study zebra finch, and similar methods have been used in studies of vocal behavior, e.g., in vocalizing doves (Streptopelia decaocto) (ten Cate and Ballintijn 1996), nightingales (Luscinia
megarhynchos) (Riebel and Todt 1997), and cotton-top tamarins (Saginus oedipus) (Miller et al. 2003).

We hypothesized that in Bengalese finches if a light flash is presented within the second-order unit, the song will not stop immediately and the subsequent syllable will be produced. In contrast, if the stimulus is presented outside of the unit, the song will stop immediately, and the subsequent syllable will not appear.

Materials and Methods

Subjects

The subjects were adult male Bengalese finches (n = 8; 19–26 months old) derived from Chiba University and kept in an aviary at the Laboratory for Biolinguistics at RIKEN-BSI. Because the songs of Bengalese finches crystallize at 120 days post hatching (Okanoya 2004), these birds sang adult songs. The temperature and humidity of the aviary were maintained at approximately 25°C and 60%, respectively, and the light/dark cycle was 13/11 h. Because these subjects had hatched and had been kept in a free crossbreeding environment (D 120 × W 120 × H 200 cm) at Chiba University during their growth and development, they had opportunities to learn complicated songs from multiple tutors. During the experimental period, the subjects were kept together in
a metal cage (D 25.5 × W 31.5 × H 40.0 cm). The subjects had not been used previously for behavioral experiments.

Apparatus

The experimental chamber was a mesh cage (D 35.5 × W 18.5 × H 26.0 cm) equipped with two perches; the chamber was placed in a sound-attenuated room (D 135.7 × W 90.7 × H 183.6 cm, RS3-075, River Field Cosmos, Japan). Eighteen light-emitting diodes (LEDs; SM5H15A-VP, 3.2 V, 850 mcd; Solidlite, Taiwan) were attached to the cage walls as if the birds were besieged, and directed into the cage; the height of the attachment was 7.0 cm above the top of the perches. The lighting arrangement allowed visual stimuli to be presented to the subject from 18 horizontal directions. During the experiment, the chamber was illuminated weakly by an electric bulb (30 m² cd) to enhance the effects of the visual stimulus. Birdsongs were recorded using a microphone (MS957, Sony, Japan) and computer software (Avisoft SASLab Pro, Avisoft Bioacoustics, Germany) at a sampling frequency of 44.1 kHz via an audio mixer (MX-50, Sony, Japan). The same birdcage and sound-attenuating room were used for all trials and song recordings. The recorded data were saved as Windows PCM (.wav) files. The output line from the microphone was connected to one of the input channels of the
audio mixer, whereas an output line of 20-kHz sine waves from an oscillator (AG-204D, Kenwood, Japan) was connected to the other input channel of the audio mixer. The LEDs and the sine wave input line were relayed by a two-channel mechanical relay (G5V-2, Omron, Japan) driven by a mechanical switch. When the switch was pushed down, the LEDs were powered with a regular voltage and the sine wave input was activated. So, the 20-kHz sine waves were not inputted to the system via the microphone but directly to the audio mixer. As a result, although the subjects did not hear the 20-kHz tones at all, when spectrograms were produced, the sine waves appeared as black horizontal lines at 20 kHz within the bird songs and we could detect the visual stimulus onset by the sound spectrograms.

Methods

We recorded normal undirected songs (Sossinka and Böhner 1980) to analyze song structure. After the songs were recorded, the interruption experiments started. The experimenter monitored the singing behavior of the subject using headphones and real time spectrograms on the PC monitor outside of the sound-attenuating room and manually depressed the stimulus switch at the desired moment. The presentation of one light stimulus was counted as one test trial. One session consisted of 10 trials. The
experimental order of the subjects was rotated with an inter-session interval of at least 5 days to reduce habituation to the visual stimulus. For the same reason, the inter-stimulus interval was set at > 3 min, whether or not the subject sang within the interval. Each subject experienced five experimental sessions; consequently, we acquired results for 50 test trials per subject.

Analyses

Normal song transitions  Recordings of the normal undirected songs were visualized as sound spectrograms using the PC software (SASLab). On the software, the frequency resolution was 172 Hz and the temporal resolution was 2.9 milliseconds per one pixel. Intervals without vocalizations that were > 2.0 s long were regarded as inter-bout intervals and the first 30 song bouts were analyzed.

The first songs were normalized by setting of the maximum amplitude within 70% of the dynamic range and were 1.0-kHz high-pass filtered to increase the S/N ratios with SASLab. Then the songs were separated into song syllables by amplitude-based syllable separation in SASLab (While some birds produce syllables that had quite wide range of amplitudes in the song, other birds have relatively fewer variations of syllable amplitude, therefore, we used the criterions 27 - 47 dB SPL below
the maximum recording level, depending on how wide the bird had variations of
syllable amplitudes). In most cases, the syllables resulting from this method were
consistent with our subjective identification of syllables that were separated by silent
periods. When separations were not consistent between the software and the human
observer, we used the software's separation to preserve objectivity. The song syllables
were manually labeled with letters of the alphabet based on the spectrogram. The same
syllable profile was labeled with the same letter in each subject. Bengalese finch songs
sometimes include repetitions of the same song syllables. Some of our subjects had
fixed numbers of repetitions (e.g., two or three times) of the same syllables. In such
cases, the repetitions were labeled with different symbols (i.e., b1, b2), even though
they represented the same patterns. Bout ends were also labeled using alphabetic
symbols. The introductory notes were grouped and labeled using an alphabetic symbol.

Transition matrices were constructed using the alphabetic sequences, and we
calculated the transition probabilities. If a syllable “a” was always followed by one
specific syllable “b”, the maximum transition probability of syllable “a” was 1.00. In
contrast, if a syllable “a” was observed 10 times, and the syllable was followed by a
specific syllable “b” 6 times, but by syllables “c,” “d,” and “e” the remaining times, the
maximum transition probability of syllable “a” was 0.6.
Using the sound spectrograms, we also measured the syllable durations and the silent intervals of song sequence from the first normal song bout of each subject (see Results section). These data were used to establish the criterion for assessment of the effects of the light stimulus (below). From spectrograms of the test trials, we identified the points of the light stimulus onset, then, we assessed the effects of the light stimulus from two angles. The one is "song termination". The other is "subsequent syllable extinction".

**Assessment of song termination**  When a light stimulus onset was between the onset of the ongoing syllable and before the onset of the subsequent syllable, the stimulus presentation was regarded as the stimulus presentation to the on-going syllable, and this on-going syllable was regarded as the target syllable in the trial. Then, when the light stimulus was presented and a vocalization stopped within 300 ms from the offset of the target syllable, this was regarded as song termination induced by the visual stimulus (Fig. 2a). This 300 ms of criterion was decided by the analysis of normal song recordings (see Results section).

**Assessment of subsequent syllable extinction**  We recorded whether the subsequent syllable was produced or not after the presentation of the light stimulus to a target
syllable. The production of the subsequent syllable provided direct evidence; however, in cases of interruption, we used the moment at which subsequent syllable production was expected, instead of the actual production of any subsequent syllable. This allowed us to assess the strength of the association between the ongoing and subsequent syllables. The following cases were regarded as instances of the subsequent syllable extinction: (1) the vocalization was not produced for > 200 ms after the offset of the target syllable; (2) the vocalization stopped, and then the vocalization was resumed within 200 ms, however, any song elements that were anticipated by the song transition matrices were not produced (Fig.2b); (2) some vocalization was produced, but it could not be identified as any part of a specific song syllable, and the song sequence was not produced for > 200 ms after the offset of the target syllable (i.e., this might be a startle vocalization). This 200 ms of criterion was decided based on the normal song recordings (see Results section).

If the visual stimulus was presented during the introductory notes or was presented > 300 ms after the offset of a song syllable and no further vocalizations were produced, the trial was regarded as an irrelevant and was excluded from the data analyses.

There was a possibility that songs would stop without interruptions; therefore,
the visual stimuli were presented within the first half of the average bout length in each subject. Some Bengalese finches, however, tend to produce a typical sequence only once or twice per song bout and then stop singing. In such cases, it is difficult to present the visual stimulus within the first half of the average song bout (as an extremely example, when a song bout consists of A-B-C-D-E-F-end and the bird sings always like this, D, E and F do not include in "the first half of the song bout"). Therefore, if we persist in "the first half", we cannot submit the visual stimulus to these syllables.). Thus, for these birds, sometimes the visual stimuli were presented during the second half of the average song bout, however the results were adjusted to incorporate the probability of spontaneous song stops in ambiguous cases. In such cases, when the song was stopped instantaneously after the visual stimulus but a spontaneous song stop had been observed at the target point in the pre-recordings, we did not count as 1, but rather added the value that reduced the proportion of spontaneous song stops in pre-recordings to the number of times that the subsequent syllable was not produced.

**Statistical analysis** We recorded the number of times that the "song termination" and "subsequent syllable extinction" in each syllable pair these have various transition probabilities. We analyzed the relationship between transition probability and "song
termination” or “subsequent syllable extinction” using GLMM (Generalized Linear Mixed-effect Model). This method was adopted to select a better model from two models for explaining the results; one model has “transition probability” as an independent factor, and the other model does not have it to explain the dependent variable (song was terminated or not; subsequent syllable was produced or not). To test the models, we used AIC scores and the result of likelihood ratio test. The effects of stimulus habituation and differences in stimulus distribution between the first and second halves of the session were tested using the same methods. Additionally, we analyzed these tallies using Fisher’s exact test when these analyses could be used (i.e. the data can be represented as 2 x 2 table).

Results

Normal song recordings

The silent intervals and syllable durations were measured by SASLab from the first song bout of each bird after removing of vague and intermittently vocalizations appeared in the initial song sequence (total number of syllables was 624 from 8 subjects). Based on the frequency distribution, 97% of silent intervals were less than 200ms (average 66ms; S.D. = 52 ms, Fig.3a). Then we measured the lengths from a
syllable offset to the subsequent syllable offset and 99% of the lengths were less than
300 ms (average 131 ms; S.D. = 59 ms, Fig. 3b). Based on these results, we adopted the
criterion for assessment of the song termination and subsequent syllable extinction.

Although the bout length differed among subjects, the total number of song
syllables (excluding the introductory notes) from 30 song bouts amounted to 1142–2622
syllables per subject, with a mean of 1930.9 syllables per subject. Transition matrices
and calculated transition probabilities were created using these data.

Song terminations by light-flash stimulus

After the exclusion of irrelevant trials, we acquired results from 41–46 light-flash trials
per subject, for a total of 348 trials. In many cases, vocalization stopped within several
hundredths of a millisecond after the visual stimulus. Singing sometimes stopped during
the ongoing syllable. At other times the stimulus was followed by one to three song
syllables, but even if singing resumed, vocalizations were not produced until several
hundredths of a millisecond after the song stopped. Finally, in some cases, the songs
were not interrupted by the visual stimulus.

We counted the number of song terminations during the flash trials. For birds 8,
19, 23, and 25, almost all stimulus presentations resulted in a song stop (> 95%). Birds
16, 35, and 52 also stopped singing in most trials (> 80%). Only bird 49 tended to continue singing after the visual stimulus, with a probability of stopping of 56%; however, the song syllables were sometimes collapsed (sounds faded out or unidentified sound was mingled), and larger intervals appeared after the visual stimulus, even in the “no-stop” trials. In all birds, the total number of the song stops was 305 (88%), whereas that of no-stops was 43 (12%). We also counted spontaneous song stops in each target syllable from pre-recordings of each subject, and the total number of the song stops was 137 (1%), whereas that of no-stops was 13122 (99%). Statistical analyses showed a significant difference between the number of song stops and no-stops (Fisher’s exact test: \( p < 0.001 \)), indicating that the visual stimulus could interrupt singing behavior.

However, transition probability did not relate to the song termination based on the result of the model test. To explain the dependent variable “song termination”, there was not different between a model that had “transition probability” as an independent factor and a model that did not has it (AIC 251.35, 249.46; likelihood ratio test, \( \chi^2 = 0.114, \text{ df} = 1, \ p = 0.736 \)).

*Habituation to the visual stimulus*

To assess the effect of habituation to the visual stimulus, we divided the 50 flash trials
into halves and compared the results between the first 25 trials and the second 25 trials within individual birds. After removing irrelevant trials, we statistically analyzed the number of stops and no-stops during the first 25 trials (first half) and the second 25 trials (second half; Table 1). For bird 49, the number of song stops was greater than the number of no-stops during the first 25 trials; however, this relationship was reversed during the second 25 trials. For all other birds, although the numbers of song stops were considerably greater than the numbers of no-stops in both the first and the second halves, the responses to the visual stimulus in some birds decreased in the second 25 trials (birds 49 and 52: one-tailed Fisher’s exact test, \( p < 0.01 \)). A statistical test of pooled data from all subjects showed that the difference between the number of song stops and no-stops was better explained by a model that had a factor of trial (i.e., first or second half of trials) as an independent factor, rather than a model that did not have it (likelihood ratio test, \( \chi^2 = 24.39, \text{df} = 1, p < 0.001 \)). These analyses indicate that some subjects became habituated to the visual stimulus to some extent.

*The subsequent syllable extinction by the visual stimulus*

We arranged the pooled data from all subjects into every 0.1 of transition probability (0.40-0.49: \( n=20 \) from 4 subjects, 0.50-0.59: \( n=38 \) from 6 subjects, 0.60-0.69: \( n=23 \))
from 5 subjects, 0.70-0.79: n=22 from 5 subjects, 0.80-0.89: n=37 from 4 subjects, 0.90-1.00: n=202 from 8 subjects) and calculated percentages of the subsequent syllable extinction in each transition probability. The production rate of the subsequent syllable was greater in 0.90-1.00 than the other ranges (Fig. 4). Therefore, we tested whether the transition probability related to the subsequent syllable extinction or not, using GLMM.

A model that had “transition probability” as an independent factor could explain the dependent variable “subsequent syllable extinction” better than a model that did not have it (AIC 463.12, 468.90; likelihood ratio test, $\chi^2 = 7.777$, df = 1, $p < 0.01$). This result showed a factor “transition probability” affected the result. Then, we adopted that the 90% criterion as the border of “variable” and “stereotyped” (i.e. if a syllable had a high transition probability 0.90 – 1.00, we defined the relation of the syllable and the subsequent syllable as the “stereotyped sequence,” whereas transition probabilities smaller than 0.90 were defined as “variable sequences”). The establishment of this criterion is appropriate, because syllables are thought to be the basic song unit (Cynx 1990) and sometimes transitions from one syllable to other syllables could be broken even within the “second-order song units” in normal condition. However these groups of syllables appear together in highly percentage, so these can be called “stereotyped”. We examined difference in the production probabilities for the subsequent syllable
between the stereotyped and variable sequences. Although percentages of the
subsequent syllable production were varied in each subject, it was higher when the
visual stimulus was submitted within a stereotyped sequence than within a variable
sequence in all subjects (Table 2). A model that contained the independent variable of
"stereotyped or variable" provided a better explanation of the dependent variable (i.e.,
subsequent syllable extinction) than did the other model that did not contained the
independent variable (AIC, 461.01, 468.90; likelihood ratio test, $\chi^2 = 9.887$, df = 1, $p <$
0.01). This indicates that the production probability of the subsequent syllable was
significantly higher in the stereotyped sequences than in the variable sequences.
Additionally, with the presentation of the visual stimulus, the associations of this kind of
"stereotyped" sequences could be broken occasionally. This suggests that the song
syllables, rather than groups of two or more syllables, were the first-order song units as
shown in Cynx (1990).

Because the song bouts of birds 23 and 35 were relatively short, we adjusted
the data of these subjects to remove the probability of spontaneous song stops (see
Materials and Methods, Assessment of subsequent syllable production). However, this
operation did not affect the results.
Distribution of the timing of stimulus presentation

Because habituation was observed, we tested the difference of the distributions of the visual stimulus presentations to stereotyped and variable syllable pairs during the first and second halves of the trials. After removing the irrelevant trials, the visual stimulus was presented 94 times in a stereotyped sequence and 70 times in a variable sequence in the first 25 trials, and was presented 114 times in a stereotyped sequence and 70 times in a variable sequence in the second 25 trials. Statistical test showed the numbers of visual stimulus presentation to stereotyped and variable conditions between the first and second halves of trials were not significantly different (two-tailed Fisher's exact test, \( p = 0.38 \)). Model test also showed there was no significant difference between a model that included “trial” as an independent factor (i.e. the first half or the second half) and a model that did not include it (AIC 39.371, 38.803; likelihood ratio test, \( \chi^2 = 1.433, \text{df} = 1, p=0.23 \)). Therefore, we cannot attribute the differences of the results in stimulation during variable and stereotyped parts of the song to the difference of the stimulus distribution.

Case reports: alterations of song sequences

The light stimulus not only stopped the songs, but also altered the subsequent song
sequences. In our experiment, because the visual stimulus was cued manually, it was
difficult to repeatedly present the stimulus at exactly the same time point in the song
sequence; therefore, we could not acquire results in sufficient numbers to allow
statistical analysis. Thus, we describe these repeated presentations only as case reports,
but alteration phenomena raise interesting questions.

Song sequence alterations included the production of unexpected song syllables.

In a trial of bird 16, light flash evoked an unidentified sound production, then, on
unexpected song sequence followed that sound (Fig. 5a, b). This phenomenon had never
been observed in pre-recordings, although normal sequences that corresponded to this
sequence were appeared 80 times. Bird 35 changed the order of song sequence, such
that a sequence was preceded by a song syllable that would normally appear
subsequently (Fig. 5c, d). With the visual stimulus in the nearly same timing, we
observed the same sequential alteration twice. In pre-recordings (91 observations), this
phenomenon had never been appeared. Another alteration type was the skipping of song
syllables. With visual stimulus, bird 35 skipped one song syllable, but otherwise he
maintained the production timing of song syllables comparable to that of his normal
song (Fig. 6a, b). In pre-recordings (in 91 observations), we had never been observed
this altered sequence. Similarly, bird 25 skipped several song syllables; however, he
resumed the song sequences after leaving an interval and whose length was the same as
when the bird did not skip these song syllables (Fig. 6c, d). This phenomenon had never
been observed in pre-recordings (124 times). Another alteration type was stuttering. In
the case of bird 49, the same syllable seemed to be repeated, although it may be that the
latter half of the original subsequent syllable was omitted (Fig. 7a, b). However, this had
never been observed in pre-recordings (138 times). A more obvious example is bird
52 who tried to re-generate one spoiled syllable over again (Fig. 7c, d). This had also
never been observed in pre-recordings (217 times). Moreover, in bird 49, the intervals
between a specific syllable pair were extended by the light stimulus (Fig. 8a, b). As a
result, the extended intervals were observed 3 times when the light stimuli were
submitted at nearly the same timings. Therefore, we measured the first 10 recordings of
the intervals between the same syllable pair from the normal song recordings, and tested
the difference between them and the 3 altered intervals statistically. The result showed
the intervals between the syllable pair after the visual stimulus were significantly long
(mean normal interval = 0.095 s, mean extended interval = 0.140 s; two-tailed t-test, df
= 11, t = 6.148, p<0.001). These types of disturbances were never reported in zebra
finches (Cynx 1990).
Discussion

When a syllable was always followed by a specific syllable, the association between these two syllables tended to tolerate an interruption such as a light flash. However, even these stereotyped sequences could be broken with a reasonable probability (33%; Table 2). Therefore, we think that these syllable pairs were not the primary song units but the song syllables that are divided by silent periods. The low probability of breakage in syllable pairs within the statistical second-order song units means that song motor patterns of Bengalese finches seem to reflect statistical song structures.

Comparison to other songbird species

The existence of second-order song-production units implies that the Bengalese finch controls sequences for not only each syllable, but also each syllable-group. As a result, songs of Bengalese finches seem to be relatively complicated. Other songbird species also sing complicated songs. For example, some nightingales know several hundred song types and render various combinations of these songs with inter-song intervals of 1–2 s. Nightingale songs have a hierarchical structure that is composed of levels of song class: song types, packages, and bouts (Hultsch and Todt 1989). Riebel and Todt (1997) conducted a light-flash stimulus experiment using the nightingale. They reported the
effects of visual interruption could vary between night and daytime (nightingale songs are more versatile at night than daytime). They also found that an interrupted song was often repeated when the nightingale resumed the bout, however the Bengalese finch did not show such behavior. This might be because of the differences in song structures between the two species; a nightingale song bout is composed of various songs with long silent intervals, whereas the Bengalese finch produces bouts of one idiosyncratic complex song without such intervals. Interestingly, they also reported the number of elements following the light interruption differed among 0 to 3. This suggested that the strength of the association between song elements could be not the same also in nightingales, although their analysis methods differed somewhat from ours.

Researchers are currently interested in the European starling (*Sturnus vulgaris*) because it has complicated songs and may have an ability to recognize acoustic patterns defined by a context-free grammar (Gentner et al. 2006). Starling songs have a hierarchical structure that is composed of syllables, motifs, and song bouts, and a song appears as a sequence of changing motifs (Gentner 2004), although starlings are open-ended song learner and are dissimilar to Bengalese finches in this point. Therefore, if the light-flash experiment were to be conducted in this species, informative results that reflect the song structure may be acquired.
An interpretation of song sequence alterations by visual stimulus

The present results showed the visual stimulus caused song sequence alterations that had not been observed by Cynx (1990) (Fig.5, Fig.6, Fig.7). Although our methods differed from those of Cynx (1990) in that we used LEDs and the stimulus did not blink, the differences in the results might be caused by the substantial difference between the song behaviors of Bengalese finches and zebra finches. To interpret the different results between these species, it may be useful to refer to experiments related to auditory feedback. Although the targeted sensory modality differs, these experiments can be regarded as interruptions of singing behavior. Deafened Bengalese finch stuttered and skipped song syllables soon after the surgery (Okanoya and Yamaguchi 1997; Woolley and Rubel 1997), whereas the songs of zebra finch did not change as quickly (Nordeen and Nordeen 1992). Sakata and Brainard (2006) reported that altered auditory feedback immediately elicited alterations in song sequence and timing in Bengalese finch songs, and these changes were similar to our observations. In their study, although the song alterations were mainly contingent upon the altered auditory feedback, some of these might be attributed to the startle responses caused by the altered auditory feedback, but not by the auditory feedback per se. In the zebra finch study, delayed auditory feedback elicited increases in song interruptions, the repetition of introductory notes, and changes
in the starting places of songs (Cynx and von Rad 2001). These song interruption studies and our results may indicate a substantial difference between the songs of Bengalese finches and zebra finches, that is, relatively higher song sequence plasticity is one of the characteristics of Bengalese finch songs, although the Bengalese finch is a closed-ended learner.

Conconsiderations from physiological studies

Songbirds have song control systems located in the central nervous systems (Nottebohm et al. 1976). According to anatomical findings (Wild 1994), visual inputs reach the nucleus Uvaeformis and there are pathways from the nucleus to the song nucleus HVC and nucleus interface of the nidopallium (NIf). Under anesthesia, the song control nucleus HVC has selective auditory responses to the bird's own songs (Margoliash 1983). Electrophysiological studies showed that light flashes could evoke responses in the HVC of zebra finches (Bischof and Engelage 1985), whereas they did not affect selective auditory responses in Bengalese finches (Seki and Okanoya 2006). Because the HVC is implicated in not syllable production itself but the pattern of syllable sequence (Vu et al. 1994; Yu and Margoliash 1996; Okanoya 2004), these findings were consistent with our observations of the skipping of syllable(s) but maintaining the
pattern of syllable sequence, when the light flash was presented (Fig. 6). In these cases, the upper streams of the motor pathway (e.g., HVC and NIf) might be so robust that they could not be interrupted easily and some kind of fright response or visual input might affect the lower motor control unit such as the robust nucleus of the arcopallium (RA) or the hypoglossal nucleus (Wild 2004). This speculation is consistent with that the examples on figure 6 were similar to the results of an electrical stimulation experiment of RA in zebra finches (Vu et al. 1994). However, because song sequences or timings of syllable production can be affected by visual inputs in some cases (Figs. 5, 7), information processing in the HVC might be affected by light stimuli. These cases were, likewise, similar to the results of electrical HVC stimulation in the same study (Vu et al. 1994). One of the reasons for the combination of these phenomena might be that the Bengalese finch has second-order song units and performs sequence control on several orders. However, the reason why such various song alteration types were caused by the visual interruptions remains an unanswered.

**Figure legends**

*Figure 1.* An example of sound spectrogram of Bengalese finch songs (subject no. 52). The horizontal axis represents time and the vertical axis represents frequency. Song syllables are represented as alphabetic symbols (A, B, C). A number represents a
second-order song unit. In this example, the transition probability from A to B is 1.00; however, that from D to E is 0.80 \( \frac{4 [DE]}{1 [DA] + 4 [DE]} \).

**Figure 2.** (a) Schematic examples of a song termination. (b) A case where the subsequent syllable is not produced.

**Figure 3.** (a). Distribution of the length of silent interval and the syllable duration. (b). Distribution of the silent interval. These data were sampled from the first song bout of the normal song recordings (684 syllables from 8 subjects).

**Figure 4.** Percentage of subsequent syllable production in each transition probability. A model has a factor "transition probability" as an independent variable can explain this result better than a model that does not have the factor (likelihood ratio test, \( \chi^2 = 7.777 \), df = 1, \( p < 0.01 \)). See the text in detail.

**Figure 5.** Examples of unexpected song sequence production as a result of the visual stimulus. These sequences were not observed in normal song behaviors (69 observations for bird 16; 135 observations for bird 35). Normal song sequences (a, c). Song
sequences altered by light flashes (b, d). The horizontal black lines at 20 kHz indicate
the presentation periods of the light stimulus.

Figure 6. Examples of missing song syllable(s) as a result of the visual stimulus. These
types of omissions were not observed in normal song behaviors (91 observations for
bird 35; 124 observations for bird 25). Normal song sequences (a, c). Song sequences
altered by light flashes (b, d). An arrow indicates a missing song syllable (a, b). The
horizontal black lines at 20 kHz indicate the presentation periods of the light stimulus.
Dotted lines show the interval of missing syllables (d) and normal song (c).

Figure 7. Examples of stutter as a result of the visual stimulus. These types of stutter
were not observed in normal song behaviors (91 observations for bird 35; 124
observations for bird 25). Normal song sequences (a, c). Arrows show normal
production of song syllables corresponding to each song in the lower panel. Song
sequences altered by light flashes (b, d). Arrows show stoppers. The horizontal black
t lines at 20 kHz indicate the presentation periods of the light stimulus.

Figure 8. An example of extended silent interval as a result of the visual stimulus. This
extension of silent interval was observed 3 times in the same syllable pair when the
stimuli were submitted at the nearly same point and the songs were not terminated. The
lengths of these intervals were significantly longer than the intervals of corresponding
syllable pair in the normal songs. See the text in detail.
Table 1 Percentages of song stops as a result of visual stimulation or habituation: comparison of observations from the first 25 (first half) and second 25 (second half) trials.

<table>
<thead>
<tr>
<th>Subject No.</th>
<th>Stop</th>
<th>%</th>
<th>Stop</th>
<th>%</th>
<th>No-stop</th>
<th>Stop</th>
<th>%</th>
<th>No-stop</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>100</td>
<td>18</td>
<td>0</td>
<td>96</td>
<td>23</td>
<td>0</td>
<td>80</td>
<td>149</td>
<td>35</td>
</tr>
<tr>
<td>16</td>
<td>100</td>
<td>17</td>
<td>2</td>
<td>74</td>
<td>17</td>
<td>6</td>
<td>80</td>
<td>149</td>
<td>35</td>
</tr>
<tr>
<td>19</td>
<td>100</td>
<td>23</td>
<td>0</td>
<td>100</td>
<td>23</td>
<td>0</td>
<td>80</td>
<td>149</td>
<td>35</td>
</tr>
<tr>
<td>23</td>
<td>100</td>
<td>23</td>
<td>0</td>
<td>91</td>
<td>21</td>
<td>2</td>
<td>80</td>
<td>149</td>
<td>35</td>
</tr>
<tr>
<td>25</td>
<td>100</td>
<td>22</td>
<td>0</td>
<td>91</td>
<td>21</td>
<td>2</td>
<td>80</td>
<td>149</td>
<td>35</td>
</tr>
<tr>
<td>35</td>
<td>86</td>
<td>19</td>
<td>3</td>
<td>90</td>
<td>19</td>
<td>2</td>
<td>80</td>
<td>149</td>
<td>35</td>
</tr>
<tr>
<td>49</td>
<td>83</td>
<td>15</td>
<td>3</td>
<td>35</td>
<td>8</td>
<td>15</td>
<td>80</td>
<td>149</td>
<td>35</td>
</tr>
<tr>
<td>52</td>
<td>100</td>
<td>19</td>
<td>0</td>
<td>68</td>
<td>17</td>
<td>8</td>
<td>80</td>
<td>149</td>
<td>35</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>96</td>
<td>156</td>
<td>8</td>
<td>80</td>
<td>149</td>
<td>35</td>
<td>80</td>
<td>149</td>
<td>35</td>
</tr>
</tbody>
</table>
Table 2 Relationship between the subsequent sound production and the strength of the syllable association.

<table>
<thead>
<tr>
<th>Subject no.</th>
<th>Variable sequence</th>
<th>Stereotyped sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Production</td>
<td>Production Extinction</td>
</tr>
<tr>
<td>8</td>
<td>39%</td>
<td>12%</td>
</tr>
<tr>
<td>16</td>
<td>60%</td>
<td>3%</td>
</tr>
<tr>
<td>19</td>
<td>44%</td>
<td>7%</td>
</tr>
<tr>
<td>23</td>
<td>25%</td>
<td>3%</td>
</tr>
<tr>
<td>25</td>
<td>67%</td>
<td>8%</td>
</tr>
<tr>
<td>35</td>
<td>31%</td>
<td>5%</td>
</tr>
<tr>
<td>49</td>
<td>35%</td>
<td>7%</td>
</tr>
<tr>
<td>52</td>
<td>50%</td>
<td>14%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>42%</td>
<td>59%</td>
</tr>
</tbody>
</table>

Production: Number of times the subsequent syllable was produced.
Extinction: Number of times the subsequent syllable was not produced.
Figure 1
a  
**SONG TERMINATION**

Stim. Onset \(\rightarrow\) stop \(< 300\text{ms}\) 

Song elements \(\rightarrow\) 

target \(\rightarrow\) time

b  
**SUBSEQUENT SYLLABLE EXTINCTION**

Stim. Onset \(\rightarrow\) \(> 200\text{ms}\) 

Song elements 

target \(\rightarrow\) time

Figure 2
Figure 3
Chapter 2: Section 1

SONG PRODUCTION PATTERN OF BENGALESE FINCHES

Figure 4
Figure 5
Figure 6
Figure 7
Figure 8
Chapter 3: Looking for Neural visual pathway into Song production system

Section 1:

Effects of visual stimulation on the auditory responses of the HVC song control nucleus in anesthetized Bengalese finches

Abstract

In songbirds, the auditory neurons of the telencephalic song control nuclei, especially those in the high vocal center (HVC), respond to the bird’s own song (BOS) selectively. Since songs are elicited by the sight of conspecific females and interrupted by intense visual stimulation, such as strobe lights, visual input might modulate this auditory selectivity. This study used acute electrophysiological experiments using Bengalese Finches (*Lonchura striata* var. *domestica*) to examine whether strobe lights affect this auditory response. The results showed that visual inputs did not affect the neural activities in response to the BOS. When the visual stimulus was presented alone, we did not record comparable neural activities to auditory stimuli, although vague, weak electrical potential fluctuations were observed. This means that direct visual inputs do not reach all HVC neurons that have BOS selectivity, and the effects of visual information might be very limited in the song control system. Although visual
information should have some relationship to singing behaviors, such effects might be mediated by indirect connections from the visual system via unidentified emotional modules.

**Keywords**: Bengalese Finches (*Lonchura striata* var. *domestica*), Electrophysiology, HVC, Visual stimulus

**Introduction**

Songbirds, including Bengalese Finches (*Lonchura striata* var. *domestica*), produce complex temporal sound patterns called "songs". Songs are classified into two types by their context and the way they are produced. Males sometimes sing alone without much motion and these songs are called "undirected song". By contrast, when a male faces an attractive female, the male starts singing, erects its plumage, and dances. This type of song is called "directed song" (Sossinka & Bohner 1980). Therefore, directed songs are governed partially by visual input (Takahasi et al. 2005).

For many songbird species, songs have been analyzed by listening directly or to acoustical recordings, and by transforming them into spectrograms for visual inspection. Using this technique, a song can be divided into "elements" or "notes". In this way, song structures can be studied more objectively, and songs are known to have temporal
sequences consisting of several song elements. For example, Okanoya discovered that the songs of Bengalese Finches consist of "chunks", which are groups of song elements forming a second-order song unit (Honda & Okanoya 1999).

Songbirds have also been the subjects of neuro-ecological studies for over 30 years. Nottebohm et al. (1976) discovered that songbirds have "a song control system" in the brain. This system consists of several telencephalic nuclei. Under anesthesia, neurons in some of these nuclei fire only if the bird's own song (BOS) is presented as auditory stimuli. The high vocal center (HVC) has especially strong BOS selectivity (Katz & Gurney 1981; Margoliash & Konishi 1985). The responses of HVC neurons seem to relate to the temporal pattern of song sequences. An electrical stimulation study in Zebra Finches, *Taeniopygia guttata*, (Vu et al. 1994) and lesion studies in Bengalese Finches (Uno & Okanoya 1998; Okanoya et al. 2000) support this idea.

One of the behavioral approaches used to understand song structures is to use a visual stimulus. In these experiments, singing birds were exposed to strobe lights to observe when the birds stopped singing. This method has been used to identify the unit of song production in Zebra Finches (Cynx 1990; Franz & Goller 2002). Some studies using Nightingales, *Luscinia megarhynchos* (Riebel & Todt 1997) and vocalizing doves, *Streptopelia* sp. (ten Cate & Ballintijn 1996), also identified the song or coo unit by
observing when the birds stopped vocalizing in response to light flashes at various intervals.

Very few studies have examined relationships between the song control system and visual information. One study showed anatomical evidence that the song control system could receive inputs from the visual system, via the optic tectum to the Uva, then to the HVC and nucleus interfacialis (NIf) (Wild 1994). Furthermore, a physiological study showed that light flashes triggered an evoked electrical potential in the HVC in anesthetized Zebra Finches; the amplitudes of the visually evoked potential waves were smaller than the auditory responses, and the waveforms were not always stable (Bischof & Engelage 1985).

Based on these studies, we conducted acute electrophysiological experiments in Bengalese Finches to examine whether visual input affects the HVC while it is processing auditory information. This might clarify the hierarchical structure of the song in this species. We subjected anesthetized subjects to light flashes while recording neural activities in response to the BOS in the HVC. If visual input reaches the HVC directly, it could affect auditory processing there. We hypothesized that the effects of light flashes during the strongly bonded song elements (inside a chunk) would be relatively small, whereas the effects would be larger when the visual stimuli were
presented outside a song chunk. Such results could clarify the existence of the representation of second-order unit of songs in neurons.

In addition, because behavioral observations show that some visual information affects singing behaviors (e.g., directed songs are elicited on seeing other individuals, strobe lights can interrupt singing behaviors, and social interaction is important for song learning (Bolhuis et al. 1999)), such inputs should reach the song control system. This study might provide insights into the neural substrates that cause these phenomena.

Materials and Methods

1) Subjects

The subjects were six adult male Bengalese Finches kept at Chiba University.

2) Apparatus

To record neural activity, we used a CED1401Micro and application software Spike2 (Cambridge Electronic Design, Cambridge, UK), an extra-cellular amplifier (Neurodata ER98, Cygnus Technology, PA, USA), a stereotaxic device, a micro-positioner (Model 640, David Kopf Instruments, CA, USA), and tungsten parylene-coated electrodes (5M #57340, 0 2M #573200, A-M Systems, WA, USA) or
coaxial electrodes (IMB-6502 Inter Medical, Tokyo, Japan). Recordings were made at frequencies from 100 Hz to 10 kHz using a filter on the ER98. All neural activity recordings were made in a sound-attenuated, electromagnetically shielded box (SC-2, Music Cabin, Kawasaki, Japan). During the experiment, the room light was turned off and the background noise was about 22dB.

The visual stimuli were produced using a super-bright light-emitting diode (LED White NSPW500BS 3.6V 9200mcd Nichia, Tokushima, Japan) and an optic fiber (Eska CK60E Mitsubishi Rayon, Tokyo, Japan). A voltage of 3.6 generated by the CED1401Micro was sent to the LED and the resulting light was transmitted along five bound optical fibers. The end of the fibers was fixed 1 cm from the eye contralateral to the recording hemisphere in the subjects.

Before the physiological experiments, songs were recorded with a microphone and personal computer with application software (Avisoft SASLab pro, Berlin, Germany) at a sampling frequency of 44.1kHz in a sound-attenuated room. These songs were saved as .wav files, and then used for song analyses and as electrophysiological stimuli.
3) Song analyses

Songs were visualized as sound spectrograms using SASLab. From the spectrograms, each sound separated by silent periods was identified as a song element. These song elements were labeled with letters of the alphabet manually. For each bird, if two elements were regarded as identical, based on their form, they were assigned the same letter. In this way, songs could be represented as strings of letters.

From these strings, element transition probabilities were calculated. Whenever an element was followed by one specific element in the strings, the transition probability between these two elements was 1.00. Conversely, if the following element was not fixed, the element transition probability was the appearance rate of that combination versus all other following elements in all strings.

4) Stimuli

Songs were transformed from .wav files into the file format used in Spike2. For each subject, seven types of stimuli were prepared. Two of these were auditory stimuli: the bird’s own song (BOS) and the BOS reversed (REV). Three were “BOS + Visual Stim”, which were combinations of the BOS as the auditory stimulus and a strobe flash as a visual stimulus: a 50-ms-long strobe flash was presented between song elements.
while playing the BOS. The timings of the visual stimuli were based on the different element transition probabilities from the song analyses. The other two stimuli were the strobe light alone and the control (nothing presented). All sound modifications were made using SASLab. All acoustical stimuli were adjusted to a maximum sound pressure, to 60-70 dB at the bird’s ear.

Stimuli were presented in the same order repeatedly to all subjects (1. BOS + VisualStim1; 2. BOS + VisualStim2; 3. REV; 4. control; 5. BOS + VisualStim3; 6. BOS only; 7. light flash only). One trial consisted of these seven types of stimuli. Twenty trials were conducted for each recording site. The inter-stimulus intervals were randomized to be within 1.0-1.8 ms.

5) Surgery

The subjects were deeply anesthetized with 0.2 ml of 10% urethane (injected as several doses of 30-50 μl at approximately 20-min intervals). The birds were then fixed on the stereotaxic device using ear-bars and a beak-holder that held the beak tip at an angle of 45° down from the horizontal plane. Then, we put Xylocaine gel on the subject’s head and removed the feathers and skin. A custom-made 3-point fixation device (Narishige, Tokyo, Japan) was attached to the rostral part of the skull surface.
using dental cement. Small holes were then opened in the skull just above the HVC (2.0 mm lateral to and 0.5 mm anterior from the bifurcation of the mid-sagittal sinus), the dura was removed, and an electrode was inserted. Before making the physiological recordings, the ear-bars were removed.

After making recordings in the HVC, to ensure that the visual stimuli reached the central nervous system, the neural activities in response to visual stimuli were recorded in the nucleus rotundus (Rt), which is part of the visual system (Hodos & Karten 1966; Schmidt & Bischof 2001), in three birds. The surgical area was 1.8 mm lateral to and 2.0 mm anterior from the bifurcation of the mid-sagittal sinus, and the operations were performed in the same manner as for the HVC.

After recording the neural activities, the subjects were sacrificed with 0.12 ml of Nembutal and were perfused with saline and paraformaldehyde to prepare brain slice specimens.

6) Analysis

Multi-unit neural activities were recorded as waveforms with Spike2. After the experiments, the acquired data were analyzed using the same software. When a potential difference was generated, and it exceeded the threshold setup for detecting auditory
responses adequately, it was counted as neural activity. Then, PSTHs (peri-stimulus
time histograms) and raster data (8-ms bins) were made from these counts for each
recording site. To evaluate BOS selectivity versus REV, we calculated \(d'\) as a
psychophysical measure (Green & Swets 1966) as follows:

\[
d'_{A-B} = \frac{2(RS_A - RS_B)}{\sqrt{\sigma_A^2 + \sigma_B^2}}
\]

The value of \(d'\) was used as index of the signal detection theory. The theory is useful to
examine the degree of selectivity between signal and noise by subjects, in this case,
neurons. The index is a better way to consider BOS selectivity of the song control nuclei.
Here, \(RS\) is the "Response Strength", which is the difference between the firing rate of
spikes with a stimulus and the spontaneous firing rate and \(\sigma^2\) is the variance of each
mean \(RS\) (Solis & Doupe 1997). \(d_{BOS \ vs \ REV}\) was calculated from the raster data, and if
the value exceeded 1.00, we considered significant BOS selectivity to exist at the
recording site (Theunissen & Doupe 1998); it was identified as BOS-selective neural
activity inside the nucleus HVC.

The effects of visual stimuli on auditory responses were also examined using \(d'\)
for each condition versus BOS for 200 ms from the visual stimuli onsets (BOS +
VisualStim1 vs. BOS; BOS + VisualStim2 vs. BOS; BOS + VisualStim3 vs. BOS).
Results

1) Song recording

A sufficient number of songs were recorded from all subjects (each element appeared at least 21 times in the recorded songs: average 53.6 times), and the insert timings of the visual stimulus to the BOS were decided from these records. Because 3 types of “BOS + VisualStim” were prepared for each 6 birds, the total number of auditory-visual combination stimuli became 18 ($6 \times 3$). As the result, for each subject, the transition probabilities between song elements to which light flashes were inserted were following: $Bird1$: 1.00, 0.48, 0.95. $Bird2$: 0.71, 1.00, 0.67. $Bird3$: 1.00, 1.00, 0.61. $Bird4$: 0.42, 1.00, 0.75. $Bird5$: 1.00, 0.61, 1.00. $Bird6$: 1.00, 0.17, 0.33.

2) Neural activities in the visual system (in the nucleus rotundus)

There were very strong responses to light flashes in the nucleus rotundus in three subjects (Fig. 1). This ensured that the experimental system worked well and the visual inputs reached the central nervous system, although the strobe lights were presented over the eyelids to anesthetized animals.
3) **BOS selectivity in the HVC**

Based on the criterion $d'_{BOS\ vs.\ REV} > 1.00$, 53 recording sites in the HVC from six Bengalese Finches were regarded as BOS selective (maximum 10.43, minimum 1.02, average 2.43, Fig. 2 upper and middle). The records from these sites were used as data for all the following analyses.

4) **Effects of visual stimuli on auditory responses to the BOS**

The effects of a visual stimulus on the auditory response to the BOS were considered for three types of “BOS + VisualStim” (Fig. 3). Since the stimuli were presented in the same order in all sessions, order effects were observed in the response strength to the part of the three types of stimuli containing the BOS itself. Statistical analysis of $d'_{BOS+\ VisualStim\ vs.\ BOS}$ for 200 ms before the visual stimulus onset showed that the difference was significant ($F(2)=4.88$, $P<0.01$ one-way ANOVA), although the stimuli for these parts were exactly the same. Therefore, to examine the effects of visual stimuli while processing auditory information, we compared the values of $d'_{BOS+\ VisualStim\ vs.\ BOS}$ before stimulus onset with the 200-ms period after the onset. If no effects derived from visual stimuli exist, there should be no difference between before and after the stimuli. The statistical analysis showed that the difference was not significant (Fig. 4.
BOS + VisualStim1: t=1.22, P=0.23, BOS + VisualStim2: t=0.13, P=0.90, BOS + VisualStim3: t=0.16, P=0.87, two-tailed t-test).

Then, we examined the effects of the visual stimulus timing, according to the element transition probability. For convenience, the transition probabilities were categorized into three levels (1.00; 0.95-0.65; <0.65). Then, $d'_{BOS + VisualStim vs. BOS}$ from all the subjects was analyzed using two-way ANOVA; these probability levels and the stimulus submission order were set as factors. This analysis found no relationship between the transition probability and visual stimulus effects ($F(2,2)=0.79$, $P=0.46$), although the submission order effect was significant ($F(2,2)=6.44$, $P<0.01$). Therefore, the visual stimulus had no effect, regardless of stimulus timing.

In addition, after removing the order effects from the values of $d'_{BOS + VisualStim vs. BOS}$, a test of the population mean was attempted. The null hypothesis that the values of $d'_{BOS + VisualStim vs. BOS}$ were zero was not rejected ($t=1.14$, $P=0.25$, two-tailed t-test). Therefore, the visual stimuli did not modify the BOS-selective neural activity.

5) Responses to visual stimuli in the HVC

For the recording sites, the visual stimulus alone did not seem to elicit electrical fluctuations over the threshold level necessary to detect auditory responses adequately.
(Fig. 2 lower). This result implied that there was no visual input in the HVC, at least at a level comparable to the auditory responses to the BOS.

However, when the threshold level was reduced, some vague, weak visual neural activity was observed in the HVC. When the threshold was lowered markedly, minor subtle visual responses were found in the PSTHs (Fig. 5). The statistical analyses of $d_{VisualStim \text{ vs. control}}$ from these data showed that there were significant differences between the visual stimuli and control conditions in some sessions (17/53 sessions from 5/6 subjects, $P<0.05$, t-test). In the PSTHs, the peak latency was not always the same and the peaks were not clear. These results might be consistent with Bischof and Engelage (1985), who showed that the visual evoked potential was smaller than the auditory one, and that these waveforms were not always the same.

**Discussion**

Our results clearly confirmed that HVC neurons respond to the BOS. Auditory inputs to the anesthetized birds strongly activated neurons in the HVC. In addition, we confirmed that the visual system (the nucleus rotundus) responded to the LED light. However, in this study, no "strobe light effect", in which the strobe light disrupted or interrupted on-going singing, appeared in the neural activities during auditory information
processing in the HVC. HVC neurons were not affected by visual interruption, whereas they were activated by auditory input. Previous experiments clearly showed that strobe-stimulated birds stop singing (Cynx 1990; ten Cate & Ballintijn 1996; Riebel & Todt 1997; Franz & Goller 2002). In these cases, it is possible that some changes occurred in the song control system. Therefore, in conscious birds, neurons in the song control system might receive startle signals from some part of the nervous system, while in anesthetized birds, as in this study, the neurons might not receive such information because the anesthetic turns off the "emotional" modules. Of course, these light flash experiments revealed song production, not auditory processing, and the auditory response cannot be regarded as song production.

It has been known that HVC neurons could be classified into several types, X projection, RA projection, and interneurons, by their various properties (Dutar et al. 1998). We did not confirm whether the neurons we had recorded be classified into which type. Although each type of neurons responded to BOS as auditory stimulus (Mooney 2000), RA projection neurons contribute song production directly. Considering behavioral experiments as described, possibly, visual response might reach at the motor side of the nucleus more effective. Even if that may be the case, we can be sure that neurons with strong BOS selectivity were not affected by light flashes in the
present experimental condition. What happens in the nervous system when singing birds are frightened will be revealed in electrophysiological recordings with freely moving, conscious animals.

This speculation suggests, however, that visual stimuli reach the song control system indirectly. This is consistent with an anatomical study of Zebra Finches, which showed that the song control system has indirect projections from the medial preoptic nucleus (POM) that might regulate sexually motivated song (Riters & Alger 2004).

Although the POM is not part of the song control system, it projects to the GCt (griseum centrale) of the tectum, the VTA (ventral tegmental area), and the LoC (locus coeruleus), which connect to several song nuclei. Since male songbirds start singing when they see attractive females, visual information clearly has an important role in relation to song initiation. Therefore, it is important to understand the connection between the visual system and the song control system.

In this study, some responses to visual stimuli were observed in the HVC, although they were very weak and vague. This might result from connections from the Uva, confirmed previously (Wild 1994). However, the effects of strobe lights were not so large as to interfere with activities in response to the BOS in HVC neurons.
We also found that the auditory responses of HVC neurons could not be disturbed very easily. Once the process starts, it might be unstoppable as long as auditory input continues, especially in the case of important input for individuals like the BOS. Auditory feedback is very important for normal song production in adult Bengalese Finches (Okanoya & Yamaguchi 1997; Woolley & Rubel 1997, 2002).

Consistent with this fact, this robustness and sensitivity for auditory input in the HVC might support the importance of auditory input in song control.

The relationships between the song control system and auditory inputs have been investigated in detail (Vates et al. 1996), while those with other sensory modalities have not. An anatomical study reported that somatosensory inputs had a number of connections to the song control system (Wild 1999). Therefore, understanding the functional roles of these pathways remains an important question. In the future, this will be investigated using neuro-ecological approaches, especially with combined behavioral and electrophysiological experiments.
Figure Legends

Fig. 1. Visual responses in the nucleus rotundus (bin size = 5 ms). A 50-ms-long strobe flash was presented at 1.00 s. Bursts of neural activity were observed.

Fig. 2. Upper: Raster data and PSTH (bin size = 8ms) of typical responses to the BOS in the HVC. Middle: Responses to the reversed song. Auditory stimuli were presented from 1.00 to 4.22 s (the bars represent these stimulus periods). Comparing the upper and middle graphs, BOS selectivity is clearly shown. Lower: Responses to light flashes, with 50-ms-long strobe flashes presented at 1.00 s (the triangle represents the visual stimulus onset). The threshold level is the same as in the upper two analyses. No responses to visual stimuli were observed.

Fig. 3. Responses to the BOS and strobe light (bin size = 8 ms) with 50-ms-long light flashes at 1.60 s (upper), 2.00 s (middle), and 2.59 s (lower). No effects of the strobe lights were observed.
Fig. 4. Differences in neural activity before and after the strobe flash. These are the values of $d'$ for each BOS + VisualStim condition versus BOS condition. No effects of the strobe light on the neural response to the BOS were observed (BOS + VisualStim1: $t=1.22$, $P=0.23$, BOS + VisualStim2: $t=0.13$, $P=0.90$, BOS + VisualStim3: $t=0.16$, $P=0.87$, two-tailed t-test). Only the order effect of stimulus presentation was significant.

Fig. 5. An example of visual responses in the HVC (bin width = 8 ms). In some cases, the threshold level was lowered markedly to count electrical fluctuations, and vague responses were detected in the HVC. Stimuli were presented at 1.00 s.
Figure 1
Figure 2
Figure 3
Figure 4
Figure 5
Section 2:

Functional evidence for internal feedback in the songbird brain nucleus HVC

Abstract

The song control system of songbirds consists mainly of the "motor pathway" and "anterior forebrain pathway". The medial magnocellular nucleus of the anterior nidopallium (mMAN) projects to the song control nucleus HVC which is the point of divergence of the two pathways. We made simultaneous multi-unit electrophysiological recordings from the mMAN and HVC in anesthetized Bengalese finches. We confirmed that mMAN neurons responded selectively to the bird's own song and found temporal correlations between song-related activities of the mMAN and HVC neurons. The temporal relationship between the neural activation of the HVC and mMAN suggests that these nuclei are parts of a closed loop, which could provide internal feedback to the HVC for sequential syllable control.

Keywords: songbirds, birdsong, mMAN, cross-correlation
Introduction

In many songbird species including Bengalese finches (*Lonchura striata* var. *domestica*), male birds learn songs as juveniles from tutors in social relationships. Song learning and production are controlled by the song control system (Nottebohm et al. 1976). This system consists of two main pathways: the "motor pathway" and "anterior forebrain pathway." The HVC (used as a proper name) is an important song nucleus because it is located upstream in the system and is the divergence point of the two pathways. The medial magnocellular nucleus of the anterior nidopallium (mMAN) is also considered to be a song nucleus but this nucleus is not included in the two main pathways. However, the mMAN apparently has important roles in song learning, based on the finding that mMAN lesions impaired normal song learning (Foster and Bottjer, 2001). Moreover, anatomical studies have shown mMAN–HVC and robust nucleus of arcopallium (RA)–nucleus dorsomedialis posterior thalami (DMP)–mMAN projections (Nottebohm et al 1982, Bottjer et al, 1989, Fortune et al. 1995, Vates et al. 1997, Foster et al. 1997, Foster and Bottjer 1998). The DMP–mMAN projection might relate to bilateral coordination (Schmidt et al. 2004). In addition, a recent study suggested that Area X projects to the mMAN via the DMP and that the neural signals from Area X are indirectly transmitted to the HVC (Kubikova et al. 2007).
To investigate the functional role of the mMAN, we attempted to record multi-unit neural activity in the mMAN of anesthetized male Bengalese finches and examined the auditory responses to the bird's own song (BOS). In addition, we examined the temporal correlation between the neural activities in the mMAN and HVC with simultaneous recordings in these nuclei.

Materials and Methods

Subjects

The subjects were 20 adult Bengalese finch males (ages: 1-2 years). These birds were kept in an aviary at the Laboratory for Biolinguistics, RIKEN-BSI. The aviary was maintained at a temperature of about 26°C and a humidity of about 60%. Birds could take grain and water with added vitamins *ad libitum*.

Stimuli

Songs from each bird were recorded at 44.1 kHz in a sound-attenuated chamber using a microphone (Sony, ECM-MS907), personal computer, and software program (Avisoft Bioacoustics, SASLab Pro) within 1 week before the experiments. The song was then cut into a typical portion (0.8-2.5 s) and converted to a Windows PCM file (.wav
format; 20 kHz). The song was used as an auditory stimulus (BOS) and the reversed song was used as another stimulus (REV) for each bird. Stimuli were submitted using CED1401 micro and Spike2 version 4 (Cambridge Electronic Design Limited) after band-path filtering at 100-10 kHz (NF Electronic Instruments, 3611; Frequency Devices, 900), and the stimuli were monitored using spectrograms on Spike2. One session consisted of 3 (BOS, REV, no stimulus) × 40 trials. Each stimulus was presented in random order, and the inter-stimulus intervals were randomized to be 1.0-1.8 s. The maximum sound amplitude of the stimuli was about 70 dB SPL at the birds’ head.

_Surgery_

The subjects were deeply anesthetized with 10% Urethan (2.0 ml) and fixed on a stereotaxic device (David Kopf Instruments, Model 955). The head was fixed with ear bars, and the beak tip was lowered to 45° from the horizontal plane and fixed. The head skin was removed, and the skull was opened just above the mMAN (5.1 mm anterior to the midsagittal sinus bifurcation and 0.3-0.5 mm lateral to the midsagittal sinus) and HVC (0.5 mm anterior to the midsagittal sinus bifurcation and 1.8 mm lateral to the midsagittal sinus). The bird’s head was fixed at that point with a three-point fixed device (Narishige) using dental cement, and the ear bars were removed. Then
parylene-C-insulated tungsten electrodes (A-M system, #573200) were lowered, and selective auditory responses of the nuclei were sought.

**Electrophysiological recording**

The signals were amplified and 100- to 10-kHz band-path filtered (Cygnus Technology, Model ER-91, ER-98; WPI, DAM80; NF Electronic Instruments, 3315), and CED1401micro and Spike2 were used for data acquisition and processing. An adequate threshold was established in each recording, and when the electrical fractionation of the waveforms exceeded the threshold, it was recorded as a wavemark with Spike2. These wavemarks were processed using original scripts and transformed to raster data and peri-stimulus time histograms (PSTHs), and the temporal cross-correlation of the neural activities of the mMAN and the HVC were examined using Spike2.

After recording, the recording sites were marked by small lesions made with a 10-μA current for 20 s (Nihon Kohden, SEN-3301 and SS-104J; WPI DS8000 and DLS-100). After the experiments, all of the subjects were deeply anesthetized with 20% Nembutal (0.8 ml) and perfused with 40 ml of 0.1 M phosphate-buffered 0.9% saline and 40 ml of 4% paraformaldehyde (PFA).
Histology

After fixation with PFA, the brain was cut into frontal sections (50 μm) and stained for acetylcholinesterase following the method of Bolam (1992). This stain enabled us to clearly assess the borders of the mMAN and HVC (Sadananda 2004). We confirmed the markers and electrode traces.

Data analysis

Based on the raster data, the BOS selectivity was quantified using $d'$, which has been used in signal detection theory, and the values were calculated referring to Solis and Doupe (1997). When $d' > 1.0$, we considered the units as having BOS-selective activity.

The temporal correlation was quantified using a cross-correlation coefficient (Tomita and Eggermont 2005). The bin size was 5 ms, and the lag time was adjusted from −500 to 500 ms. When a peak of the coefficient values was found within 25 ms of the zero-lag point and the value was more than three standard deviations from the mean, the peak was considered a significant correlation.
Results

Auditory selective response

In six birds, we recorded auditory responses from mMAN neurons only. We acquired strong BOS-selective responses \((d' > 1.00)\) in the left hemisphere in four birds (1.07-2.32) and in the right hemisphere in all six birds (1.29-2.94). These results clearly showed that mMAN neurons have BOS-selective auditory responses (Fig. 1).

Temporal correlation between the mMAN and HVC

We attempted to acquire simultaneous recordings from the ipsilateral mMAN and HVC in 16 birds. Both nuclei showed auditory responses in 29 recordings. Eleven of these recordings lacked significant peaks based on the cross-correlation coefficient. Eighteen recordings had a significant peak. In three of the 18 recordings, because the peaks were observed at the exact zero-lag point, an artifact might have caused the correlations. In six recordings, the peaks preceded the zero-lag point (−10 to −15 ms in 5 recordings; −20 to −25 ms in 1 recording; Fig. 2, upper panel). These time delays seemed to correspond to the results of a previous antidromic stimulation study, in which such evoked potentials were referred as “late onset activity” (Foster et al. 1997). These
results showed that the neural responses of the mMAN preceded those of the HVC in some cases. In nine recordings, the peaks followed the zero-lag point (Fig. 3, upper panel), and the distribution of these peaks varied in distance from the zero-lag point (2 to 5 ms in 2 recordings; 10 to 15 ms in 2 recordings; 15 to 20 ms in 3 recordings; 20 to 25 ms in 1 recording). In these cases, the neural responses of the mMAN followed those of the HVC.

To examine the possibility that these correlations were caused by a simple response to the BOS in each nucleus, and not by the neural transmissions between the two nuclei, we shifted one trial for the recordings of the mMAN only and calculated the correlation between trial $i$ of the HVC and trial $i + 1$ of the mMAN (i.e., the HVC recording in the first trial and the mMAN recording in the second trial, ..., the HVC recording in the 40th trial and the mMAN recording in the first trial). This extinguished all of the correlations (Figs. 2, 3; lower panels). Therefore, we could not conclude that the correlations were caused by a direct stimulus effect.

*Difference in the firing rates before and after the zero-lag point*

In many cases, the mMAN neurons maintained a higher firing rate for longer than 100 ms after the zero-lag point (Fig. 4), although some of these recordings did not have a
significant peak in terms of the cross-correlation coefficient. Therefore, we tested the difference in the firing rate between 50 ms before and after the zero-lag point (i.e., from -100 to -50 ms and from 50 to 100 ms) in the 29 simultaneous recordings. In 18 recordings, the firing rate of the mMAN neurons after HVC activation was significantly higher than that before (two-tailed paired t-test, $p < 0.05$). Conversely, only two recordings showed a significantly higher firing rate before the zero-lag point. These results mean that the mMAN neurons maintained a high firing rate for a long time after activation of the HVC.

Discussion

We found BOS-selective responses in the mMAN and temporal correlations between the neural activities of mMAN neurons and HVC neurons in anesthetized Bengalese finches. Moreover, some of the neural activities of the mMAN preceded those of the HVC. These results suggest that a mMAN–HVC pathway actually transmits song-related neural information. Since many recordings showed that mMAN neurons were activated for a period of time after HVC activation, the mMAN responses might originate in the auditory responses of the HVC (Vates et al. 1997) via several pathways. However, the HVC seemed to be reactivated by some of these responses of the mMAN.
This means that these pathways might form a closed functional loop that provides internal feedback to the HVC.

Some studies have shown that the HVC is involved in syllable sequence control (Vu et al. 1994, Yu and Margoliash 1996, Okanoya 2004). If the HVC were to control the syllable sequence, it would be very important to refer to stored information on several preceding syllables. Given that the duration of each syllable is several tens of milliseconds, the song control system must maintain the information for several tens or a few hundreds of milliseconds. Our data showed that mMAN neurons remained activated for 100 ms after HVC activation. Therefore, the HVC might receive such information from internal feedback via the mMAN, if such long activation were to relate to the information on preceding syllable production or perception.

Area X lesions affect the syllable sequence in Bengalese finches (Kobayasi et al. 2001). Considering only the two main pathways, it is difficult to interpret this phenomenon. If signals from Area X were to be able to reach the HVC, it would be reasonable to conclude that the song sequence was altered in birds with Area X lesions. The present study provides insight for understanding the reason for this. Our data suggest that the mMAN probably serves as a relay nucleus for the Area X–HVC projection.
However, the time from the zero-lag point of the peaks with a cross-correlation coefficient varied among recordings. These results might originate in a local circuit in the HVC and multiple inter-nuclei pathways. Thus far, several possible pathways for signal transmission from the HVC to the mMAN have been reported (i.e., HVC–Area X–DMP–mMAN, HVC–RA–DMP–mMAN, and coordination of the two hemispheres; Kubikova et al. 2007, Nottebohm et al. 1982, Bottjer et al. 1989, Fortune et al. 1995, Vates et al. 1997, Foster et al. 1997, Foster and Bottjer 1998). In addition, the time by which the peaks preceded the zero-lag point was too long (10-25 ms) for signal transmission via only one synaptic connection from the mMAN to the HVC. Foster et al. (1997) suggested that such activity is attributable to axon collaterals of HVC-projecting neurons acting locally within the mMAN.

We recorded multi-unit activities to identify neural activities related to song perception and the temporal correlation between the HVC and mMAN, and thus we did not identify the type of HVC neuron (Mooney, 2000). These points can be improved in future studies. To date, mMAN neurons are known to receive inputs from only the DMP in song related circuits; it would be interesting to test whether DMP lesions affect mMAN activity. In addition, this study showed that the neural activity occurred during song perception and not production. Therefore, direct evidence that the mMAN–HVC
pathway is related to song production and the role of sequence modification should be investigated in further studies.
Figure legends

**Figure 1.** Examples peri-stimulus time histograms for BOS-selective mMAN neurons.

When the BOS is submitted as an auditory stimulus, neurons show strong responses (upper panel). The reversed BOS does not cause such a response (lower panel). One session consisted of 40 trials. Bin size = 8 ms. Bars represent stimulus periods.

**Figure 2.** Example of a cross-correlogram between the activities of the mMAN and the HVC for 40 BOS trials. In this case, a peak in the coefficient appears before the zero-lag point because mMAN neurons fired before HVC neurons (upper panel). However, when the recordings of the mMAN are shifted for one trial, this peak disappears (lower panel). Therefore, this correlation could not be attributed simply to stimulus effects.

**Figure 3.** Example of a cross-correlogram between the activities of the mMAN and the HVC for 40 BOS trials. In this case, a peak of the coefficient appears after the zero-lag point because HVC neurons fired before mMAN neurons (upper panel). However, when the recordings of the mMAN are shifted for one trial, this peak disappears (lower panel). Therefore, this correlation could not be attributed simply to stimulus effects.
Figure 4. Example of a cross-correlogram between the activities of the mMAN and the HVC for 40 BOS trials. In many cases, the activities of mMAN neurons after the zero-lag point lasted longer than 100 ms. In 18 of 29 recordings, the activations after the zero-lag point were significantly greater than the activations before the point (two-tailed paired $t$-test, $p < 0.05$).
TEMPORAL CORRELATION OF NEURAL ACTIVITY BETWEEN HVC AND MMAN

Figure 1
Figure 2
Figure 3
Chapter 3: Section 2

TEMPORAL CORRELATION OF NEURAL ACTIVITY BETWEEN HVC AND MMAN

Figure 4
General Discussion

In this dissertation, I found sex differences in audiovisual information processing and that a lack of correspondence between visual attention and sound source location affected acoustic discrimination learning. Moreover, I discovered hierarchical structures of the vocal pattern, and showed that auditory-selective neural activity of the song nucleus HVC was not affected by light flashes in anesthetized Bengalese finches. In addition, I presented a possible neural pathway of visual input to the song control system in songbirds.

Sex differences in sensory information processing

As argued in the first chapter, sex differences might originate in the neural system, although I did not conduct physiological experiments and do not have data to support that conclusion. In humans, some sex differences have been reported in sensory information processing and they originate in physiological differences (reviewed in Cahill 2006). Interestingly, a study that investigated auditory information processing using the McGurk effect (McGurk & MacDonald 1976) showed that women were more affected by visual information than men (Irwin et al. 2006). As shown in the first
chapter, in an audiovisual discrimination task, female Bengalese finches always depended on the visual cue, but males did not. Future studies will clarify whether this kind of sex difference is common across species.

**Effects of a lack of correspondence between the target of visual attention and sound source**

As reported in the second section of the first chapter, sound source location affected the performance of a song discrimination task in Bengalese finches. In humans, when subjects participate in an acoustic psychological experiment, they receive instructions and use headphones. Thus, I could not compare songbirds and humans. Moreover, little is known about this effect in other species, although many studies have examined the discrimination of sound locations.

However, this kind of experiment provides certain suggestions for neuroethological studies. As described in the section, once the birds learned the discrimination task, they could discriminate the stimulus even when the sound source was located on the opposite side of the response key. However, the learning speed of the experimental group (i.e., the sound source was placed on the opposite side of the response key) was significantly slower than that of the control group (i.e., the sound
source was located above the subject). Some subjects in the experimental group could not learn the song discrimination task. These results might be related to more rigid innate restriction in the audiovisual integration of songbirds. If so, this kind of effect in other species would be weaker than that in songbirds because associating a visual target with a sound is less important for other species relative to songbirds.

Hierarchical vocal pattern of Bengalese finches and humans

In the second chapter, I examined the song motor pattern of Bengalese finches. When the syllable sequence of recorded songs was analyzed statistically, the songs consisted of regular syllable pairs and irregular pairs, and the songs had a hierarchical structure. Using light flashes, I showed that the strength of the association between song syllables reflected the statistical structure. This result has not been reported in studies of other species (Cynx 1990; ten Cate & Ballintijn 1996; Riebel & Todt 1997; Franz & Goller 2002). Thus, this result is the basis for the argument that the Bengalese finch is a good animal model for studying human language. Human language also has a hierarchical structure (i.e., phoneme, word, and sentence). Therefore, it might be possible to test the vocal pattern of human language with a similar experiment using transcranial magnetic stimulation (Barker et al. 1985).
Visual input to the song control system of songbirds

In the third chapter, I showed electrophysiological results regarding the song control system of Bengalese finches. As shown in the results, neurons in the song nucleus HVC showed selective auditory responses to a bird’s own song (BOS); however, the neurons did not respond to visual inputs that rivaled the auditory responses. Some neurons in the nucleus uvaformis (Uva) have visual response (Coleman et al. 2007) and Uva projects into the HVC. Therefore, visual inputs might arrive in the song control system via the Uva. However, because the basal ganglia (AreaX) is related to singing motivation (i.e., neural activity differs when singing to females and when singing alone; Hessler & Doupe 1999), I hypothesized that the basal ganglia circuit [i.e., anterior forebrain pathway (AFP) in songbirds] might also receive visual inputs. Therefore, I attempted electrophysiological recording in the nucleus mMAN (medial magnocellular nucleus of anterior nidopallium) because this nucleus is projected from AreaX indirectly (Kubikova et al. 2007) and projects to the HVC. As shown in the results, mMAN neurons had a BOS-selective auditory response, and spikes of the auditory responding neurons temporally correlated with spikes of some HVC neurons. Thus, mMAN neurons project to the HVC and transmit some singing-related activities and the AFP
might project to the HVC via the mMAN. I speculate that some activity through this pathway carries visual information.

The significance of this dissertation and comparisons with human communication

When we see an excellent ventriloquist, we feel as if the doll were talking although the sound source is the mouth of the ventriloquist. This “ventriloquism effect” is well known to psychologists (Jack et al. 1973); visual information affects the signal receiver so that the receiver cannot correctly locate the sound source. This is an example of how humans use visual information more predominantly than auditory information in spatial audiovisual information processing. In the first chapter, I showed not only that Bengalese finches had sex differences but also that they used predominantly visual cues in an audiovisual discrimination task. This is similar to audiovisual perception in humans.

I must point out another similarity between communication of Bengalese finches (or some Estrildidae) and that of humans. Many species use vocalization as a communication signal in situations in which the receiver cannot be seen (e.g., over a long distance or in darkness). For example, chimpanzees prefer to use gestures when the receiver is adjacent and visible (Pollick & de Waal 2007). However, humans
communicate with adjacent parties by vocalization. Similarly, many Estrildidae sing "songs" only toward adjacent targets because the sound amplitude is so low. They have another type of vocal signal called the "long distance call"; this type of vocalization is used to communicate with invisible targets. However, the calls are substantially different from songs (e.g., a song consists of many sounds, whereas a call consists of only one sound). Additionally, songs are learned vocalizations, as are human languages.

Few animal species have vocal-learning ability. Other than birds and humans, some whales and bats are vocal learners (reviewed in Brainard & Doupe 2002). It is reasonable to conclude that whales evolved this ability because they must communicate over long distances; their songs can be detected several miles away. Similarly, because bats fly in the dark and visual information cannot be used, it is reasonable to conclude that bats evolved acoustic communication for this reason. However, Bengalese finches and humans use complex and learned vocalizations to communicate with an adjacent and visible party, although both species predominantly use visual information.

This dissertation provides not only knowledge enhancing our understanding of acoustic communication by Bengalese finches, but also suggestions for the evolution of universal acoustic communication. I also demonstrated the importance of visual information in acoustic communication. As research progresses, I hope we will
understand how and why systems (e.g., human communities) evolve complex acoustic communication even when such communication is not always necessary.
References


References


magnocellular nucleus of the anterior neostriatum in zebra finches. J. Comp. Neurol., 382, 364–381.


References


Leitner, S., Catchpole, C. K. (2002). Female canaries that respond and discriminate more between male songs of different quality have a larger song control nucleus (HVC) in the brain. J. Neurobiol., 52, 294–301.


References


Riebel, K., Todt, D. (1997). Light flash stimulation alters the nightingale’s singing style:
implications for song control mechanisms. Behaviour, 134, 789-808.


Schmidt, A., Bischof, H.J. (2001). Integration of information from both eyes by single


auditory neurons in the nucleus HVC of male zebra finches. J. Neurosci., 18, 3786-3802.


