Local structure sensitivity in auditory information processing in avian song nuclei

Takuya Koumura[a], Yoshimasa Seki[a,b,c] and Kazuo Okanoya[a,b,c]

Birdsong is an excellent research model for sound sequences consisting of complex structures. Neural and behavioral experiments have shown that auditory feedback is necessary for songbirds, especially Bengalese finches, to maintain the quality of the songs and that the nucleus HVC (used as proper name) and the anterior forebrain pathway (AFP) in the nervous system play key roles in this maintenance process. Neurons in the HVC and AFP exhibit higher spike rate to the bird's own song (BOS) than to other sound stimuli, such as temporally reversed song. To systematically evaluate what aspects of the BOS are captured by the different types of neural activities, both average spike rate and trial-to-trial spike timing variability in the BOS-selective neurons in the HVC and Area X (used as proper name), a gateway to the AFP from the HVC, were investigated following the presentation of auditory stimuli consisting of the BOS with systematic temporal inversion. Within-subjects analysis of the average spike rate and spike timing revealed that neural activity in the HVC and Area X is more sensitive to the local sound modulation of songs than to the global amplitude modulation. In addition, neurons in the HVC exhibit greater consistency of spike timing than neurons in Area X. *NeuroReport* 25:562–568 © 2014 Wolters Kluwer Health | Lippincott Williams & Wilkins.

NeuroReport 2014, **25**:562–568

Keywords: anterior forebrain pathway, Area X, Bengalese finches, HVC, neurophysiology, songbirds, temporal structure

[a]Department of Life Sciences, Graduate School of Arts and Sciences, University of Tokyo, [b]ERATO, Japan Science and Technology Agency, Tokyo and [c]RIKEN Brain Science Institute, Wako, Japan

Correspondence to Kazuo Okanoya, PhD, Graduate School of Arts and Sciences, University of Tokyo, Komaba 3-8-1, Meguro-ku, Tokyo 153-8902, Japan
Tel: +81 354 546 266; fax: +81 354 546 725;
e-mail: kazuookanoya@gmail.com

Received 28 November 2013 **accepted** 29 January 2014

Introduction

Birdsong has a complex temporal structure. For example, songs in Bengalese finches and zebra finches last a few tens of seconds and are divided into syllables, which are discrete sound elements with a complex spectrotemporal structure of a few tens of milliseconds separated by intervals of silence (Fig. 1a) [1]. Within a song, syllables appear in a certain order, and thus the temporal structure of a song is dissected into the local structure within syllables, represented by the sound modulation in a fine time scale, and the global structure, represented by the order of syllables [2,3]. Both levels of temporal structure are actively maintained by utilizing the auditory feedback of the bird's own song (BOS) [4–6]. Songs in Bengalese finches deteriorate relatively fast after deafening, which indicates the importance of auditory feedback in this species [4]. As a result, Bengalese finches are a suitable model with which to investigate the neural mechanisms of auditory feedback processing.

In the avian nervous system, the HVC (used as proper name) and the anterior forebrain pathway [AFP; consisting of Area X (used as proper name), the medial nucleus of the dorsolateral thalamus, and the lateral magnocellular nucleus of the anterior neostriatum] each play a key role in the maintenance of song [7,8]. Auditory information from both primary and higher auditory brain regions is sent to the HVC [9–11], which, in turn, sends it to the AFP through the connections with the HVC and Area X [12,13]. Neurons in the HVC and the AFP show much higher spike rate for a BOS over other auditory stimuli such as temporally reversed song, in which both local sound modulation and global amplitude modulation of BOS are disrupted [13–15]. Previous physiological studies in zebra finches and Bengalese finches have shown that BOS-selective neurons exhibit high spike rate to the stimuli with the same local sound modulation as the BOS [14–16]. As songs of Bengalese finches vary greatly among individuals, however, to understand the universal mechanisms underlying auditory information processing, it is important to compare the neural activity in the HVC and AFP in the same individuals. In addition, it is important to examine the neural activities other than the spike rate, such as trial-to-trial variability of spike timing. Therefore, in the present study, both average spike rate and trial-to-trial spike timing variability in the BOS-selective neurons in the HVC and Area X, a gateway to the AFP from the HVC, were investigated using a within-subjects design. To systematically evaluate what aspects of a BOS (local sound modulation or global amplitude modulation) are captured by these neurons, a BOS with syllable inversion and/or syllable order inversion was used as the auditory stimulus.

Methods

Subjects

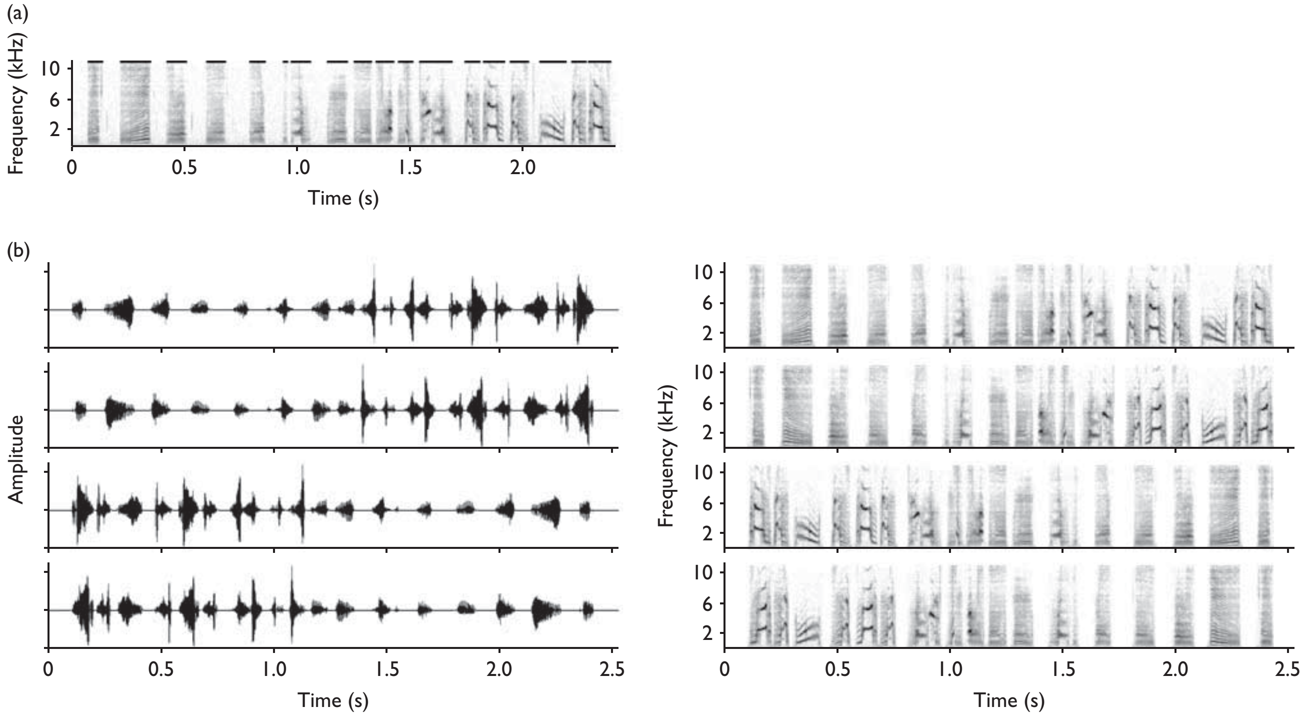
Subjects included 10 adult male Bengalese finches (> 120 days posthatch) obtained from local pet suppliers or born

---

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (*www.neuroreport.com*).

0959-4965 © 2014 Wolters Kluwer Health | Lippincott Williams & Wilkins DOI: 10.1097/WNR.0000000000000136

Fig. 1



(a) Spectrogram of a song in the Bengalese finch. Songs in Bengalese finches have a complex spectrotemporal structure, which is divided into syllables (upper black bars), or elements of a few tens of milliseconds, separated by intervals of silence. Thus, a song consists of two levels of temporal structure: the lower level of which is the sound modulation of each syllable and the upper level of which is the order of syllables. Note that this figure shows only a part of the song, which has duration of a few tens of seconds. (b) An example of four stimuli constructed from the song in (a) (bird's own song, local reverse, order reverse, and whole reverse from the top). Sound amplitudes are shown in the left column and spectrograms in the right.

in our laboratory. All participants were kept in an aviary in a controlled environment (temperature of ~25°C and humidity of ~50%). All experiments were performed in accordance with the guidelines of the Animal Experiment Committee of the University of Tokyo.

Auditory stimuli

Songs were recorded in a sound attenuation box with a microphone (ECM-MS907; Sony Corporation, Tokyo, Japan), an analog-to-digital converter (SE-U77; Onkyo Corporation, Osaka, Japan), and a computer software (SASLab Pro; Avisoft Bioacoustics, Glienicke, Germany) with a sampling rate of 22.05 kHz. From the recorded song of each bird, a segment with a length shorter than 3 s was chosen as a stimulus. Syllables are determined by visual inspection using a 256-point sound spectrogram and extracted with fade in and out in 3 ms at both ends. Four stimuli were constructed from the original song (Fig. 1b): BOS, local reverse (LR), order reverse (OR), and whole reverse (WR). BOS was constructed by rearranging the extracted syllables with the same timing as the original song. The other three stimuli were constructed by reversing the syllables and/or syllable order of the BOS: WR was the temporally reversed sound of the BOS; LR and OR were constructed by arranging the temporally reversed syllables with the same timing as BOS and WR, respectively. In other words, both local and global amplitude modulations of the original song were preserved in BOS, local modulation was disrupted and global modulation was preserved in LR, local modulation was preserved and global modulation was disrupted in OR, and both were disrupted in WR (Table 1). The preservation of global amplitude modulation in BOS and LR was confirmed by the correlation coefficients of the amplitude envelopes (figure in Supplemental digital content 1, *http://links.lww.com/WNR/A278*).

Surgery

Birds were anesthetized with 0.2 ml 10% ethyl carbamate and then fixed on a stereotaxic device (Model 900; David Kopf Instruments, Tujunga, California, USA). The angle of the line connecting the ears and the beak tip was fixed to 45° below the horizontal plane. After applying an anesthesia product (Xylocaine; Astrazeneca K.K., Osaka, Japan) to the head skin, it was removed along with the skull and the dura above the HVC and Area X. The HVC in Bengalese finches is located ~2.0 mm lateral and 0.5 mm anterior to the bifurcation of the midsagittal sinus

Table 1 Auditory stimuli

	Local sound modulation	Global amplitude modulation
BOS	Preserved	Preserved
LR	Disrupted	Preserved
OR	Preserved	Disrupted
WR	Disrupted	Disrupted

BOS, bird's own song; LR, local reverse; OR, order reverse; WR, whole reverse.

and Area X is located ~1.7 mm lateral and 4.7 mm anterior to the bifurcation.

Electrophysiological recording

To conduct electrophysiological recordings, birds were placed onto the stereotaxic device within an electromagnetically shielded sound attenuation box. Neural activity was sequentially recorded from the HVC and Area X from the left hemisphere because of the left-side dominance of these song nuclei in the Bengalese finch [17]. The signals were recorded with a tungsten electrode with an impedance of 2 MΩ (#573200; A-M Systems Inc., Carlsborg, Washington, USA), an amplifier (DAM80; World Precision Instruments, Sarasota, Florida, USA), a band-pass filter with a band width from 100 Hz to 25 kHz (Multichannel SR Filter 3315; NF Corporation, Yokohama, Japan), an analog-to-digital converter (MICRO1401; Cambridge Electronic Design Ltd, Cambridge, UK), and a computer software (Spike2; Cambridge Electronic Design Ltd) with a sampling rate of 25 kHz. In each recording, auditory stimuli were presented 30 times in random order for each of the four stimuli. Intervals between trials were randomly varied between 3 and 5 s. The stimuli were presented with the computer software (Spike2), digital-to-analog converter (MICRO1401), a low-pass filter with the cutoff frequency of 10 kHz (900C9L8B; Frequency Devices Inc., Ottawa, Illinois, USA), and a speaker with an amplifier (SRS-Z1PC; Sony Corporation). The amplitude of the stimuli at the location of participants' ears was ~70 dB in sound pressure level.

Histology

Following electrophysiological recording, electric lesions were made by applying a 20 μA current for 20 s using a stimulator and an isolator (SEN-3301; Nihon Kohden Corporation, Tokyo, Japan). Then the bird was deeply anesthetized with an overdose of pentobarbital before being perfused with 4% paraformaldehyde for the fixation of the brain. The brain was cut into 50 μm slices and stained with cresyl violet. The stained slices were inspected under a microscope to confirm the recording sites (figure in Supplemental digital content 2, *http://links.lww.com/WNR/A278*).

Analysis of average spike rate

Multiunit spike activity was defined as extracellular waveforms with minimum peaks smaller than the threshold (mean 3 SD). The average spike rate was calculated from stimulus onset to 0.1 s after the offset of auditory stimuli because the effect of the stimuli on neural activity often persists subsequent to stimuli offset [13]. Spontaneous spike rate was calculated from two sections each with the length half of a stimulus: immediately before the stimulus onset and 1.6 s after the stimulus offset. It is assumed that, using visual inspection, 1.6 s is long enough for the effect of the stimulus on neural activity to disappear completely. The effect of the stimuli on the increase in the average spike rate was measured by $d'(\text{stimulus}_A/\text{stimulus}_B)$, a psychophysical measure representing the selectivity of neural activity on stimulus_A over stimulus_B [18].

As strong selectivity in the HVC and Area X on BOS over WR has been previously reported [13,15], in this study, neural activity was analyzed if it exhibited a greater spike rate compared with the spontaneous spike rate during BOS presentation ($P < 0.05$ in paired t-test) and a stronger response to BOS than WR [$d'(\text{BOS/WR}) > 0.5$], and other data were discarded [14]. In all recorded birds, at least one recording in each nucleus reached these criteria, enabling the within-subjects comparison of the HVC and Area X. Selectivity on BOS over LR and OR was measured by $d'(\text{BOS/LR})$ and $d'(\text{BOS/OR})$, respectively. If multiple recordings were obtained from the same nucleus in a single bird, d's in the same nucleus were averaged. Then the d's were statistically analyzed using a within-subjects two-way analysis of variance (ANOVA) with factors of stimuli (LR, OR, or WR) and nuclei (HVC or Area X). If the main effect of the stimuli was significant, the effect of the stimuli was compared using a pairwise paired t-test with Bonferroni correction.

Analysis of spike timing

The effect of auditory stimuli on spike timing was measured using the z-scores of the mean correlation coefficient between the instantaneous spike rate for all pairs of trials with the same stimuli. First, the instantaneous spike rate $r(t)$ was estimated by smoothing the spike trains with a Gaussian kernel with a half-width of 20 ms. Next, the mean correlation coefficient between the instantaneous spike rate for all pairs of trials with the same stimuli was calculated [19].

$$\mathrm{CC} = \frac{1}{N_{\text{pairs}}} \sum_{i<j} \mathrm{CC}_{ij},$$

$$\mathrm{CC}_{ij} = \frac{\langle r_i(t) r_j(t) \rangle_t}{\left\langle r_i(t)^2 \right\rangle_t \left\langle r_j(t)^2 \right\rangle_t},$$

where N_{pairs} is the number of trial pairs, CC_{ij} is the correlation coefficient between the instantaneous spike rate of trial i and j, $r_{\mathrm{i}}(t)$ is the instantaneous spike rate of trial i, and $\langle\ \rangle_t$ represents averaging over t. The correlation coefficient itself depends on the average spike rate because, even if the timing of spikes is random, the higher the average spike rate, the more likely it is that spikes between trials will occur with the same timing. To obtain a measure that depends only on the spike timing, the z-scores of the mean correlation coefficient were calculated on the basis of the null distribution of spike trains with independent timing and the same average spike rate as the original data obtained by randomly shifting the original spike trains [20].

$$z = \frac{\mathrm{CC} - (\text{mean of } \mathrm{CC}_{\mathrm{null}})}{(\text{SD of } \mathrm{CC}_{\mathrm{null}})},$$

where $\mathrm{CC}_{\mathrm{null}}$ is the mean correlation coefficient of the spike trains with independent timing. A thousand patterns of $\mathrm{CC}_{\mathrm{null}}$ were randomly generated to calculate the null distribution. If multiple recordings were obtained from the same nucleus in a single bird, the z-scores in the same nucleus were averaged. Then the z-scores were statistically analyzed using a within-subjects two-way ANOVA with factors of stimuli (BOS, LR, OR, or WR) and nuclei (HVC or Area X). If the main effect of the stimuli was significant, the effect of the stimuli was compared using a pairwise paired t-test with Bonferroni correction.

Results

To confirm that global amplitude modulation of BOS was indeed preserved in LR, but not in OR or WR, correlation coefficients between the amplitude envelope of BOS and those of reversed stimuli were calculated (figure in Supplemental digital content 1, *http://links.lww.com/WNR/A278*). As amplitude envelope extracts the global amplitude modulation of sounds, large correlation coefficient indicates the similarity of the global amplitude modulation between two sounds. Indeed, the correlation coefficients between the amplitude envelope of BOS and LR were large (mean ± SD 0.63 ± 0.16), whereas those between BOS and OR (mean ± SD 0.022 ± 0.14) and between BOS and WR (mean ± SD 0.0022 ± 0.19) were small, which indicate that global amplitude modulation is preserved in LR, but not in OR or WR. Note that it is obvious that local sound modulation is preserved in OR because OR was constructed by rearranging the same syllables as BOS.

Multiunit spike activity was recorded from BOS-selective neurons in both the HVC and Area X in 10 Bengalese finches; the raster plots and the mean spike rate over trials in response to auditory stimuli in the HVC and Area X from a single bird are shown (Fig. 2). In both the HVC and Area X, the spike rate increased sharply during BOS and OR presentation but not during LR and WR presentation. There appeared to be more large peaks in the HVC than in Area X, suggesting that the spike timing was more consistent in the HVC.

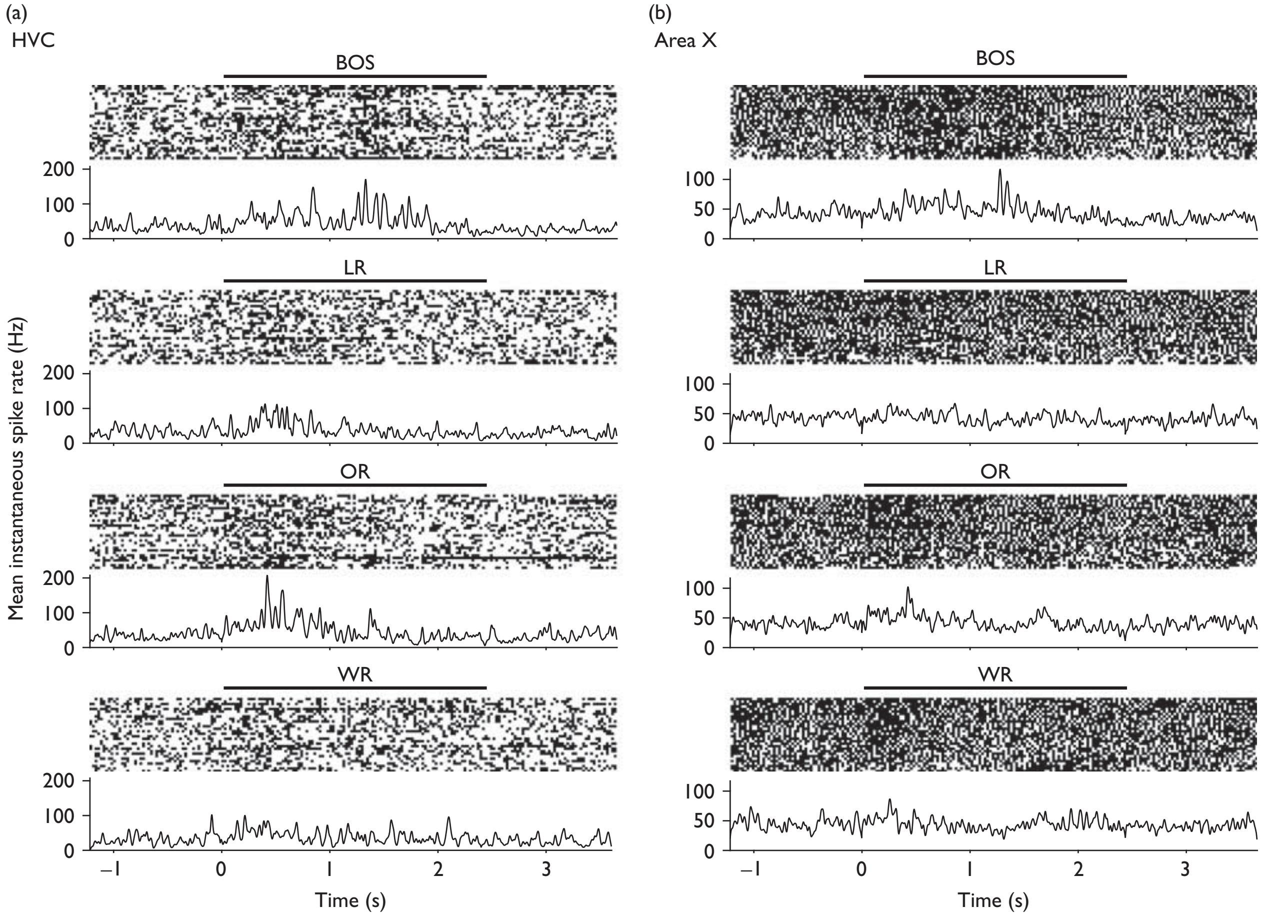
To quantify the effect of auditory stimuli on average spike rate and spike timing, the d's of the increase in the average spike rate and the z-scores of the mean correlation coefficients between the pairs of the instantaneous spike rate, respectively, were calculated. In both the HVC and Area X, the d's were significantly different among the three reversed stimuli (within-subjects two-way ANOVA, $F = 24.95$, $P = 6.47 \times 10^{-6}$), and subsequent pairwise comparison revealed that d'(BOS/WR) was the largest, followed in order by d'(BOS/LR) and d'(BOS/OR) [paired t-test: $t = 4.14$, $P = 5.57 \times 10^{-4}$ for d'(BOS/WR) and d'(BOS/LR); $t = 8.31$, $P = 9.41 \times 10^{-8}$ for d'(BOS/WR) and d'(BOS/OR); $t = 3.94$, $P = 8.74 \times 10^{-4}$ for d'(BOS/LR) and d'(BOS/OR)] (Fig. 3a). Note that the significance level for the pairwise comparison was set to $\alpha = 0.017$ according to the Bonferroni correction. These findings indicate that the inversion of the syllables is more responsible for the BOS selectivity than the inversion of syllable order. The differences between the HVC and Area X as well as the interactions between stimuli and nuclei were not significant ($F = 1.10$, $P = 0.32$ for the main effect of the nuclei; $F = 2.77$, $P = 0.090$ for the interaction). Note that the large d'(BOS/WR) is trivial because neurons with d'(BOS/WR) of 0.5 or less are excluded from the analysis in the first place to focus on the BOS-selective neurons.

In both the HVC and Area X, the z-scores of the mean correlation coefficients between the pairs of instantaneous spike rates were significantly different among the four stimuli (within-subjects two-way ANOVA, $F = 6.37$, $P = 0.0021$), and subsequent pairwise comparison revealed that the z-scores during the presentation of BOS and OR were larger than that of LR and WR (paired t-test: $t = 2.98$, $P = 0.0076$ for BOS and LR; $t = 3.07$, $P = 0.0063$ for BOS and WR; $t = 3.58$, $P = 0.0020$ for OR and LR; $t = 3.55$, $P = 0.0021$ for OR and WR) (Fig. 3b). These findings indicate that spike timing was more consistent for BOS and OR, stimuli with original local sound modulation, than for LR and WR, stimuli with disrupted local modulation. The differences in the z-scores between other pairs of stimuli were not significant ($t = 2.07$, $P = 0.052$ for BOS and OR; $t = 0.54$, $P = 0.59$ for LR and WR). Note that the significance level for the pairwise comparison was set to $\alpha = 0.0083$ according to the Bonferroni correction. The z-scores in the HVC were larger than those in Area X ($F = 9.94$, $P = 0.012$), indicating that spike timing was more consistent in the HVC than in Area X. The interactions between stimuli and nuclei were not significant ($F = 2.42$, $P = 0.088$).

Discussion

The average spike rate and trial-to-trial spike timing variability in BOS-selective neurons in the HVC and Area X are investigated with four stimuli: BOS, LR, OR, and WR (Table 1). BOS-selective neurons are defined as the neurons that exhibit higher average spike rate for BOS than WR, which differs from BOS in the following two aspects: local sound modulation and global amplitude

Fig. 2



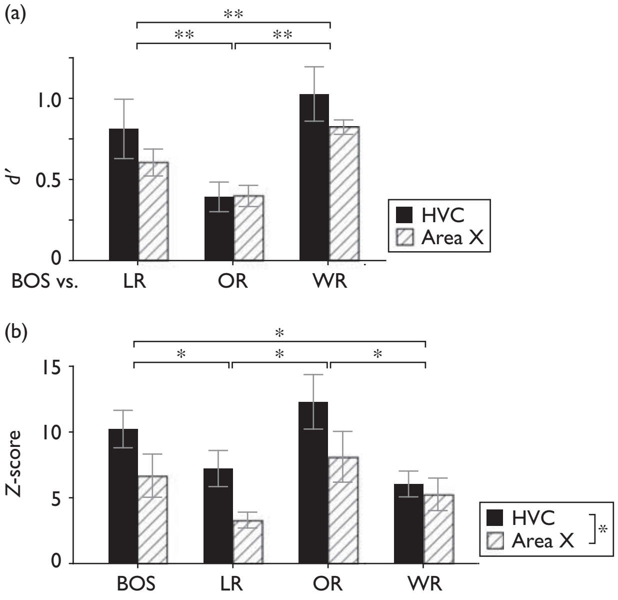
Raster plots and mean instantaneous spike rate over the 30 trials in response to auditory stimuli in the HVC (a) and Area X (b) in a single bird. Upper bars show the time of the stimuli presentation [bird's own song (BOS), local reverse (LR), order reverse (OR), and whole reverse (WR) from the top]. In both the HVC and Area X, the mean instantaneous spike rate increases sharply in response to the presentation of BOS and OR. The number of large peaks in the instantaneous spike rate appears to be higher in the HVC than in Area X. Time zero is set at the beginning of each stimulus.

modulation. The results demonstrated that d'(BOS/LR) was larger than d'(BOS/OR). The d'(BOS/reversed stimulus) is the index of the selectivity of the average spike rate on the BOS over the reversed stimulus. Such selectivity was stronger for the LR than OR, suggesting that the average spike rate was more influenced by the inversion of the syllables than by the inversion of the syllable order. In other words, the average spike rate in the HVC and Area X was sensitive to the local sound modulation of the BOS.

In addition, the z-scores of the mean correlation coefficients between the pairs of the instantaneous spike rate during the presentation of BOS and OR were larger than those of LR and WR, and those in the HVC were larger than those in Area X. The z-score is an index of the spike timing consistency. Thus, these results indicate that spike timing was more consistent for the stimuli with original local sound modulation than those with disrupted local modulation. In other words, spike timing in the HVC and Area X was more sensitive to the local sound modulation of the BOS. In addition, spike timing was more consistent in the HVC than in Area X.

The analyses of average spike rate and spike timing both illustrate that the neural activity in the HVC and Area X is more sensitive to the local structure of syllables rather than the order of syllables. This suggests that neurons in the HVC and Area X code more information regarding the local sound modulation of the BOS than the global amplitude modulation. Interestingly, sensitivity for local structure in the BOS-selective neurons contrasts with the conclusion of our previous study on the sensitivity of behavior in Bengalese finches [21]; in that study, following an operant conditioning procedure, birds perceived their own song globally and the other subjects' song locally. A possible reason for this difference in the local or global sensitivity is that anesthesia might deprive the birds of their sensitivity to the global amplitude modulation of the BOS. Another possibility is that in the

Fig. 3



(a) The d' of the increase in the average spike rate between bird's own song (BOS) and reversed stimuli in the HVC (black) and Area X (slanted lines). The larger the d', the more sensitive the average spike rate is to the temporal structure reversed in the stimuli (sound modulation of each syllable and/or the order of syllables). (b) The z-scores of the mean correlation coefficients between the pairs of the instantaneous spike rate in the HVC (black) and Area X (slanted lines). The larger the z-score the more consistent is the spike timing for the temporal structure preserved in the stimuli. *$P<0.05$, **$P<0.01$, in the statistical analysis described in the Methods section. Error bars are SEM ($n=10$ birds). LR, local reverse; OR, order reverse; WR, whole reverse.

avian nervous system there might be regions other than the HVC and Area X that are sensitive to the global amplitude modulation of one's own song that take control of behavior during operant conditioning tasks. Higher auditory regions are a possible candidate for this mechanism [22]. Both the HVC and higher auditory regions receive inputs from primary auditory regions [9,23], which respond not only to BOS, but also to auditory stimuli with more general spectrotemporal structures [10,11]. Therefore, information processed in the primary auditory regions might be transmitted to two independent pathways: one is the HVC and AFP for song maintenance where sensitivity to the local sound modulation arises and the other is the pathway to higher auditory regions associated with behavioral tasks where sensitivity to the global amplitude modulation arises. Although the average spike rate was more sensitive to the local sound modulation, it was to some extent sensitive to the order of syllables [d'(BOS/OR) $= 0.39 \pm 0.09$ in the HVC and 0.40 ± 0.07 in Area X (mean ± SEM), $n = 10$ birds]. Such sensitivity to the syllable order of average spike rate is consistent with the conclusions of the previous studies [13,15].

The current findings show that neurons in the HVC exhibit a greater consistency with spike timing than neurons in Area X. During singing, spike timing in the HVC is precisely correlated with the sound structure of the song within an error of ~20 ms [24]. During the process of song maintenance, the HVC is considered to have the role of telling the timing of the song to Area X, which conducts another calculation using several inputs, including the inputs from the HVC [25]. One possible explanation for the difference in spike timing consistency is that neurons in Area X might calculate signals for song maintenance by temporally integrating inputs from the HVC, which exhibit precise timing information, resulting in the lower consistency of spike timing in Area X than in the HVC.

Acknowledgements

This study was supported by Grant-in-Aid for Scientific Research, Basic Research A (#23240033) from MEXT, Japan to K.O.

Conflicts of interest

There are no conflicts of interest.

References

1 Berwick RC, Okanoya K, Beckers GJ, Bolhuis JJ. Songs to syntax: the linguistics of birdsong. *Trends Cogn Sci* 2011; **15**:113–121.
2 Cynx J. Experimental determination of a unit of song production in the zebra finch (*Taeniopygia guttata*). *J Comp Psychol* 1990; **104**:3–10.
3 Glaze CM, Troyer TW. Temporal structure in zebra finch song: implications for motor coding. *J Neurosci* 2006; **26**:991–1005.
4 Okanoya K, Yamaguchi A. Adult Bengalese finches (*Lonchura striata* var. domestica) require real-time auditory feedback to produce normal song syntax. *J Neurobiol* 1997; **33**:343–356.
5 Woolley SMN, Rubel EW. Bengalese finches *Lonchura striata* domestica depend upon auditory feedback for the maintenance of adult song. *J Neurosci* 1997; **17**:6380–6390.
6 Sakata JT, Brainard MS. Real-time contributions of auditory feedback to avian vocal motor control. *J Neurosci* 2006; **26**:9619–9628.
7 Kao MH, Doupe AJ, Brainard MS. Contributions of an avian basal ganglia–forebrain circuit to real-time modulation of song. *Nature* 2005; **433**: 638–643.
8 Sakata JT, Brainard MS. Online contributions of auditory feedback to neural activity in avian song control circuitry. *J Neurosci* 2008; **28**:11378–11390.
9 Kelley DB, Nottebohm F. Projections of a telencephalic auditory nucleus – field L – in the canary. *J Comp Neurol* 1979; **183**:455–469.
10 Lewicki MS, Arthur BJ. Hierarchical organization of auditory temporal context sensitivity. *J Neurosci* 1996; **16**:6987–6998.
11 Janata P, Margoliash D. Gradual emergence of song selectivity in sensorimotor structures of the male zebra finch song system. *J Neurosci* 1999; **19**:5108–5118.
12 Nottebohm F, Stokes TM, Leonard CM. Central control of song in the canary, *Serinus canarius. J Comp Neurol* 1976; **165**:457–486.
13 Doupe AJ. Song- and order-selective neurons in the songbird anterior forebrain and their emergence during vocal development. *J Neurosci* 1997; **17**:1147–1167.
14 Theunissen FE, Doupe AJ. Temporal and spectral sensitivity of complex auditory neurons in the nucleus HVc of male zebra finches. *J Neurosci* 1998; **18**:3786–3802.
15 Nakamura KZ, Okanoya K. Neural correlates of song complexity in Bengalese finch high vocal center. *Neuroreport* 2004; **15**:1359–1363.
16 Kojima S, Doupe AJ. Neural encoding of auditory temporal context in a songbird basal ganglia nucleus, and its independence of birds' song experience. *Eur J Neurosci* 2008; **27**:1231–1244.
17 Okanoya K, Ikebuchi M, Uno H, Watanabe S. Left-side dominance for song discrimination in Bengalese finches (*Lonchura striata* var. domestica). *Anim Cogn* 2001; **4**:241–245.
18 Green DM, Swets JA. *Signal detection theory and psychophysics*. New York: Wiley; 1966.

19 Goldberg JH, Adler A, Bergman H, Fee MS. Singing-related neural activity distinguishes two putative pallidal cell types in the songbird basal ganglia: comparison to the primate internal and external pallidal segments. *J Neurosci* 2010; **30**:7088–7098.

20 Olveczky BP, Andalman AS, Fee MS. Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *PLoS Biol* 2005; **3**:e153.

21 Okanoya K, Tsumaki S, Honda E. Perception of temporal properties in self-generated songs by Bengalese finches (*Lonchura striata* var. domestica). *J Comp Psychol* 2000; **114**:239–245.

22 Gentner TQ, Margoliash D. Neuronal populations and single cells representing learned auditory objects. *Nature* 2003; **424**:669–674.

23 Bolhuis JJ, Okanoya K, Scharff C. Twitter evolution: converging mechanisms in birdsong and human speech. *Nat Rev Neurosci* 2010; **11**:747–759.

24 Amador A, Perl YS, Mindlin GB, Margoliash D. Elemental gesture dynamics are encoded by song premotor cortical neurons. *Nature* 2013; **495**:59–64.

25 Fee MS, Goldberg JH. A hypothesis for basal ganglia-dependent reinforcement learning in the songbird. *Neuroscience* 2011; **198**:152–170.