Journal of Comparative Physiology A
https://doi.org/10.1007/s00359-017-1238-9

ORIGINAL PAPER

Neural coding of sound envelope structure in songbirds

Santiago Boari[1] · **Ana Amador**[1]

Received: 25 May 2017 / Revised: 23 November 2017 / Accepted: 29 November 2017

Abstract

Songbirds are a well-established animal model to study the neural basis of learning, perception and production of complex vocalizations. In this system, telencephalic neurons in HVC present a state-dependent, highly selective response to auditory presentations of the bird's own song (BOS). This property provides an opportunity to study the neural code behind a complex motor behavior. In this work, we explore whether changes in the temporal structure of the sound envelope can drive changes in the neural responses of highly selective HVC units. We generated an envelope-modified BOS (MOD) by reversing each syllable's envelope but leaving the overall temporal structure of syllable spectra unchanged, which resulted in a subtle modification for each song syllable. We conducted in vivo electrophysiological recordings of HVC neurons in anaesthetized zebra finches (*Taeniopygia guttata*). Units analyzed presented a high BOS selectivity and lower response to MOD, but preserved the profile response shape. These results show that the temporal evolution of the sound envelope is being sensed by the avian song system and suggest that the biomechanical properties of the vocal apparatus could play a role in enhancing subtle sound differences.

Keywords Birdsong · Neural coding · Auditory processing · Zebra finch · Electrophysiology

Abbreviations

BOS	Bird's own song
MOD	Envelope-modified bird's own song
REV	Reverse song
CON	Conspecific song
PSTH	Post-stimulus Time Histogram
MAP	Motif Activity Profile
MRA	Mean Relative Amplitude
RRS	Relative Response Strength
RRV	Relative Response Variability
SC	Spectral Centroid
IRR	Spectral Irregularity
EMG	Electromyography

Introduction

The identity of a complex sound is generally determined by multiple features, which may hinder the study of naturally occurring sounds. Nevertheless, the attack time, defined as the time elapsed between sound onset and the maximum amplitude of the sound envelope, was proven to be of particular importance: changing the attack time alone can be enough to modify the perceived identity of a sound. Compelling examples can be found using musical instruments. For instance, if Bach's Crab Canon is played with a harpsichord (sound produced by plucking a string, and hence has a short attack time) when recorded and played in reverse, the melody is preserved but a sound resembling a pipe organ (wind instrument) is heard instead. This example shows that the attack time could be a meaningful acoustic dimension of timbre, defined as the acoustic attribute that enables the distinction of different sounds of equal pitch, loudness and duration.

A widely used strategy to systematically study timbre is to relate the acoustic differences between sounds to changes perceived by subjects. Previous studies have addressed this by performing perceptual tasks using natural auditory stimuli (e.g., musical instrument discrimination) or by conducting experiments using synthesized sounds to explore

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00359-017-1238-9) contains supplementary material, which is available to authorized users.

✉ Ana Amador
anita@df.uba.ar

[1] Department of Physics, FCEN, University of Buenos Aires and IFIBA, CONICET, Intendente Guiraldes 2160, Pabellon 1, Ciudad Universitaria, Buenos Aires 1428, Argentina

Published online: 12 December 2017

timbre space (Iverson and Krumhansl 1993; McAdams et al. 1995; Caclin et al. 2005). Interestingly, it has been shown that animals can perceive and discriminate acoustic features thought to underlie timbre perception in a similar fashion as humans (for a review see Town and Bizley 2013). This provides new opportunities to study the neurophysiological basis of timbre perception.

Songbirds are a useful animal model for studying the neural basis of complex vocalizations. Zebra finches (*Taeniopygia guttata*) have been widely studied, as they are good laboratory animals that exhibit vocal learning, with vocalizations that could be temporally and spectrally complex (see Fig. 1 as an example). The songs of zebra finches are like an acquired fingerprint: each bird learns its own song during development and maintains it by active physiological processes throughout its life. In this way, each bird produces its own collection of complex "syllables" (denoted by letters in Fig. 1) in a certain order, comprising the bird´s own song (BOS). In zebra finches, the BOS is generally between 0.5 and 1 s long and can be repeated several times within a song bout. Thus, an easy to record and quantifiable behavior is combined with a well-studied neural architecture that mediates song learning and production. This neural architecture

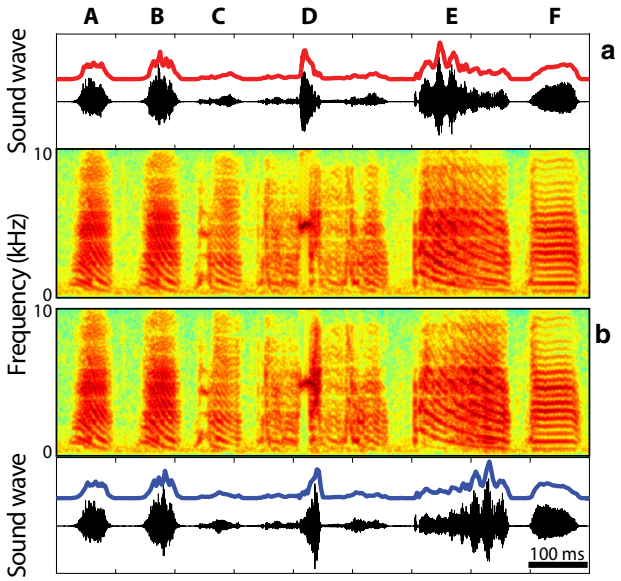

Fig. 1 Example of BOS song and the envelope-modified BOS (MOD). **a** BOS sound signal, its cubic spline-interpolated envelope (red line, vertically displaced, top panel) and song spectrogram (bottom panel). Envelope inspection reveals that some syllables have a mostly symmetric amplitude envelope (syllables A, B and C), while others are asymmetric (syllable D, which is a complex syllable composed of 3 distinct segments and syllables E and F). **b** MOD song spectrogram (top panel), sound signal and envelope (bottom panel). The envelope for each syllable was reversed without affecting the global temporal structure, the syllable amplitude modulations, the spectral centroid and the spectral irregularity found in the original song

is known as the "song system" and is well-preserved in species that learn their vocalizations. Within the song system, the telencephalic nucleus HVC (used as proper name) is a key sensorimotor area, as it receives auditory inputs and is essential for motor control during birdsong production. In this neural nucleus, a complex processing takes place: selectivity to BOS emerges and is propagated downstream to the other neural nuclei of the song system. These "song-selective" neurons respond more strongly to the auditory presentation of BOS than to almost any other sound, including simple stimuli such as pure tones or broadband noise bursts and complex stimuli such as songs of other individuals of the same species (Margoliash 1983, 1986; Margoliash and Fortune 1992; Volman 1996; Lewicki 1996). HVC neurons have also shown a strong sensitivity to slight changes in the acoustic parameters of BOS (Margoliash 1983; Theunissen and Doupe 1998). This selectivity to auditory presentations of BOS presents an unique opportunity to study the neural representation of timbre in songbirds and its relation to sound identity, since these neurons are, in a way, encoding the 'identity' of the stimulus. Moreover, similar neural patterns are observed in HVC during singing and hearing a playback of BOS (Prather et al. 2008), suggesting that in HVC there is a shared neural code for auditory inputs and motor outputs.

Interestingly, natural sounds in zebra finch song have distinguishable envelopes: each syllable has its specific shape, and many of them are highly asymmetric (fast attack times, slow release times). To test the hypothesis that the temporal structure of the sound envelope is a relevant acoustic feature for encoding the identity of the bird's own song, we constructed an envelope-modified BOS (MOD) and studied the responses of HVC neurons to presentations of different auditory stimuli.

Methods

Sound analysis and modifications

Sound envelope modification

To generate the modified BOS (MOD), we took a previously recorded BOS (see section "Stimulus pool") and computed its envelope by detecting positive peaks in the sound signal and performing a cubic spline interpolation among peaks (sampling rate: 20 kHz, detector deadtime 1 ms). Syllable onsets and offsets were detected by applying a relative amplitude threshold on the envelope (2% of envelope maximum) and the envelope for each syllable was reversed. MOD was produced by multiplying the original sound signal by the quotient between the modified and original envelopes, resulting in a sound signal that preserves the temporal structure of

the BOS while having a modified envelope, as is shown in Fig. 1. Syllables that presented discontinuous fundamental frequency traces were subdivided into notes and the same procedure as above was applied for each segment (see Fig. 1, syllable D). This prevented division by near-zero values in the vicinity of these sharp discontinuities and allowed for a smooth transition between BOS and MOD sounds.

Figure 1 summarizes the entire process. In Fig. 1a, the top panel presents the BOS sound signal with its envelope superimposed (red line, vertically displaced for the figure) and the bottom panel shows its spectrogram. Figure 1b shows the resulting MOD sound signal, and its envelope is also superimposed (blue line in bottom panel of Fig. 1b). A direct comparison and visual inspection of the envelopes, sound signal and spectrograms reveals the changes introduced in the stimulus: attack and release times of each syllable have been effectively switched while leaving the overall temporal structure of the spectrum unchanged.

Timbre features of the sound of envelope-modified bird's own song

Given any sound signal, a modification of it could induce changes in either the temporal structure or the spectral properties of the sound. To inspect the spectral differences between BOS and MOD, we measured their spectral centroid and spectral irregularity, as these are standard measurements for timbre characterization (Krimphoff et al. 1994). Spectral analysis was conducted by short-time Fourier transform (STFT) on 10 ms windowed segments (Gaussian window, length = 10 ms, $\sigma = 1$ ms and 90% overlap between segments), such that each analysis segment was separated by 1 ms within each syllable. Spectral centroid and spectral irregularity were computed for each segment.

Spectral centroid was computed as:

$$\mathrm{SC} = \frac{\sum_{n=1}^{N} f(n)x(n)}{\sum_{n=1}^{N} x(n)}.$$

Spectral irregularity was computed as:

$$\mathrm{IRR} = \log_{10}\left(\sum_{n=2}^{N-1}\left|20\log_{10}(x(n)) - \left\{\frac{20\log_{10}(x(n+1)) + 20\log_{10}(x(n)) + 20\log_{10}(x(n-1))}{3}\right\}\right|\right),$$

where $f(n)$ represents the center frequency of frequency bin n and $x(n)$ its magnitudefor the $N = 100$ frequency bins across the range 0–10,000 Hz.

For each BOS–MOD pair, we computed the mean relative differences between spectral centroid (SC) and spectral irregularity (IRR) for each syllable. Additionally, we also inspected the temporal evolution of the MOD-to-BOS relative differences in SC and IRR within each song syllable. To this aim, we computed a windowed average for the spectral quantifiers (SC and IRR, as defined above) using nonoverlapping 5 ms windows. These time traces were used to evaluate the relative difference in SC and IRR in segments within the song syllables (see Online Resources 1 and 2 for the detailed spectral analysis of SC and IRR, respectively).

Electrophysiology experiments

Adult male zebra finches were housed in individual cages in which they had access to food and water ad libitum under a 14/10 h light/dark cycle.

Stimulus pool

For each subject, song recording sessions took place before any surgery was conducted. Recording sessions consisted in placing an adult male zebra finch in its cage inside a sound-attenuation chamber and recording sound using a directional microphone. For each bird, the stereotyped song motif and most common bout length were assessed by inspection of the recordings using Praat (Boersma and Van Heuven 2001). To reduce variability across motif renditions during the experiment, the BOS sound trace was crafted by copying the selected song motif as many times as it was originally found in the most common song bout. This typically led to having a BOS that comprised introductory notes and three copies of the same song motif, but in some cases, only two copies of the motif were necessary to cover bout length.

For each bird, the stimulus pool for the experiment was: (1) BOS; (2) envelope-modified song (MOD); (3) song of an adult male conspecific (CON) and (4) BOS played in reverse (REV).

Stimulus (2) is the target stimulus for which we want to study responses relative to BOS, while stimulus (3) and (4) are control stimuli commonly used to assess the selectivity of the auditory response of HVC neurons. REV is used as it preserves the spectral structure while reversing the temporal structure of the whole song. This is not to be confused with the changes induced in MOD, in which we only reverse each syllable's envelope while leaving the temporal evolution of spectral structure unchanged. Lastly, since REV is not a naturally occurring sound, CON completes the control pool by presenting a natural song of the same species. CON was selected from a set of previously recorded BOS of adult male zebra finches such that its duration was approximately the same as the subject's bout length.

Surgical procedures

During the experiment, a stereotaxic frame was used in which the head was fixed by a stainless-steel post previously implanted on the bird. This procedure ensures that auditory stimuli reach the subject's unobstructed ear canal for auditory processing experiments in HVC.

The stainless-steel post was implanted 2 days prior to experimentation. To this end, the subject was anaesthetized with isoflurane (Baxter Healthcare, USA) and head-fixed in a stereotaxic frame such that the beak formed a 45° angle with respect to the vertical axis of the frame. Lidocaine ointment (2.5% Denver Farma, Argentina) was applied to the scalp, which was then dissected along the midline. The post was fixed at the caudal part of the bird's skull with dental cement and cyanoacrylate.

To conduct in vivo extracellular recordings in HVC, animals were anaesthetized with 20% urethane (typically 60–100 µl total; Sigma, St. Louis, MO) administered in the pectoral muscle in three doses at 30 min intervals. Urethane does not suppress neural activity and has been used for decades in studies of song selectivity in HVC, e.g., (Margoliash 1983; Doupe and Konishi 1991; Mooney 2000). The implanted post was used to head-fix the animal in a stereotaxic frame placed inside a sound-attenuating chamber. HVC location was determined by following stereotaxic coordinates, a small craniotomy was made at the site and the *dura mater* was breached with a 30G syringe needle. Recording electrodes were mounted onto an hydraulic micromanipulator (Narishige MO-10) which, in turn, was mounted to the stereotaxic frame. The frame was used to place the electrode tip near the site of insertion and lower it into the bird's brain by means of the hydraulic micromanipulator. The sound attenuation chamber remained closed during the electrophysiological recordings to avoid auditory contamination by ambient sounds. The subject was monitored during the experiments using a webcam placed inside the chamber.

Electrophysiological recordings

Single-channel, high-impedance tungsten microelectrodes (3–5 MΩ, Microprobes, MD, USA) allow for the recording of multiunit extracellular activity. Recording quality allows for the distinction of the single units (usually 1–3) comprising the multiunit activity by means of a postprocessing *spike sorting* method. For all recordings, sampling frequency was set at 20 kHz and hardware band-pass filtered on acquisition between 300 and 5000 Hz (National Instruments DAQ PCI-6251). Spiking activity was monitored during the experiment using an external audio monitor and an oscilloscope. HVC neurons were identified by stereotaxic coordinates and by spiking features such as firing rates and BOS-selective response.

An experimental protocol consisted of auditory presentations of each stimulus in the pool described above. The auditory stimuli were presented at 70 dB. Stimuli order was randomized and 20 repetitions of each song were presented to the anaesthetized subject, one song presented every 10 s. For each trial, data acquisition onset was synchronized to stimulus presentation onset (delay < 0.03 ms).

Analysis and processing of neural recordings

For each protocol measured during the experiment, spikes from single units were detected and sorted using the software wave_clus. This software utilizes a wavelet transform to extract spike features and automatically assign them to different clusters by an implementation of superparamagnetic clustering (Quiroga et al. 2004). Automatic clustering was manually inspected and corrected when necessary. Spike clusters were determined by spike shape and amplitude, interspike interval (ISI) distributions and cluster separation among projections on wavelet coefficient space. Each recording site yielded activity traces from 1 to 3 different neurons. Recordings in which the signal degraded during the measurement protocol were not included in the analysis.

The neural responses to the auditory stimuli were assessed by computing raster plots of each isolated single unit and the corresponding post-stimulus time histograms (PSTHs, 10 ms average from the raster plots). In these PSTH, each time bin represents the average activity of the neuron during that 10 ms window.

Note that in a selective HVC neuron (as the one shown in Fig. 2), there is a precise temporal pattern of activation and inhibition present in response to BOS (Fig. 2a), but not to CON (Fig. 2c) and REV (Fig. 2d). To quantify this observation, relative response strength (RRS) and relative response variability (RRV) were computed (Boari et al. 2015). RRS and RRV are both global measures of the response, since they consider the mean level and variability of the stimulus response, respectively. RRS and RRV are relative measures of the activity in response to stimulus X with respect to the response that is observed for the BOS, corrected for spontaneous activity:

$$\mathrm{RRS}_{\frac{X}{\mathrm{BOS}}} = \frac{M_{\mathrm{x}} - M_{\mathrm{s}}}{M_{\mathrm{BOS}} - M_{\mathrm{s}}},$$

$$\mathrm{RRV}_{\frac{X}{\mathrm{BOS}}} = \frac{\sigma_{\mathrm{x}} - \sigma_{\mathrm{s}}}{\sigma_{\mathrm{BOS}} - \sigma_{\mathrm{s}}},$$

where X is either MOD, CON or REV, 'M' and 'σ' indicate activity mean and activity variance for each stimulus, respectively, and 'S' indicates spontaneous activity. Note that a value of 1 would indicate the same values are obtained for stimulus X and for BOS, while a value of 0 would indicate an

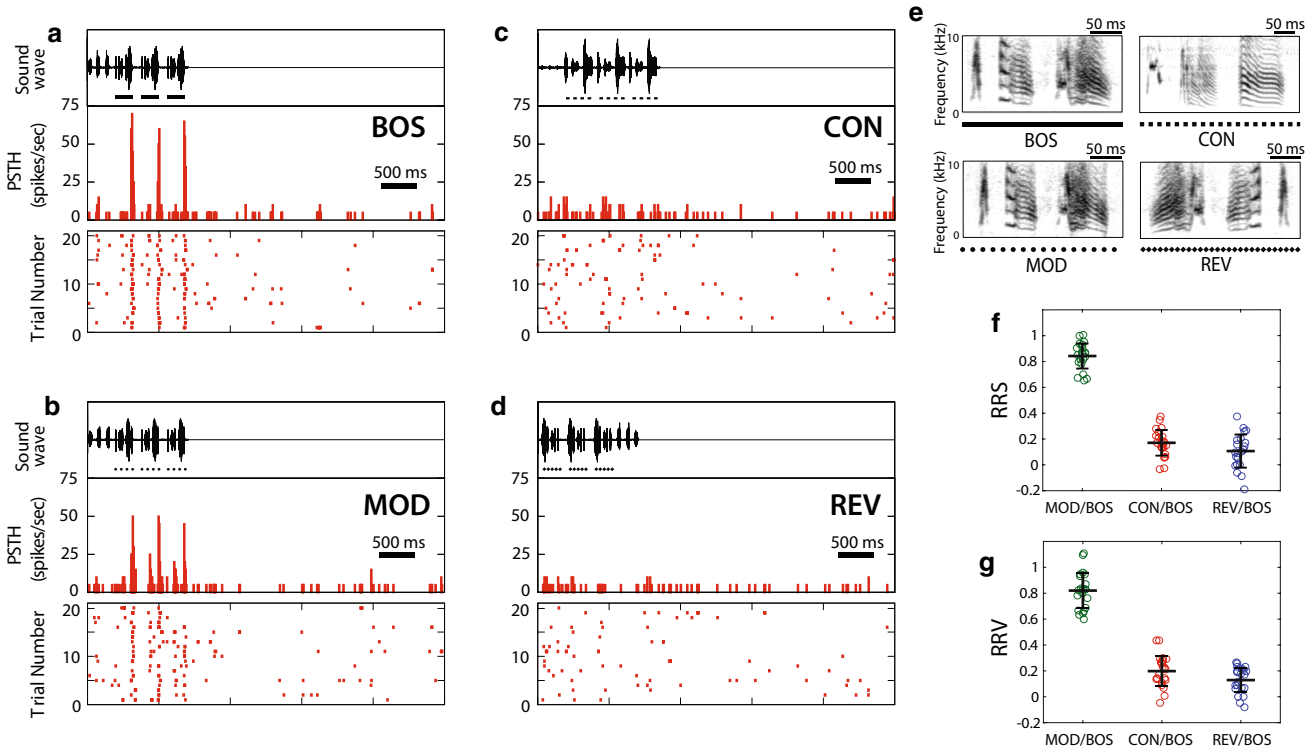

Fig. 2 HVC selective responses to study neural basis of sound envelope processing. (**a**–**d**) Raster plot and PSTHs for a clustered single unit from a protocol from a single bird. 20 trials were conducted and stimuli were randomly presented to the bird. Raster plots show that there is a clear pattern of activation and inhibition in response to BOS (**a**) and to MOD (**b**), but not to CON (**c**) or REV (**d**). The neural response is stronger to BOS, and the pattern is maintained in MOD, albeit with a lower amplitude, as can be seen from the PSTHs. Top panels of each subfigure show the sound signals and pattern-coded bars indicate a song motif. **e** Spectrograms for song motifs of each of the stimuli used in the experiment. Note the similarity between BOS and MOD spectrograms. Global activity measures related to mean activity such as RRS (**f**) and response variability, RRV (**g**) show that for 21 selective HVC neurons, responses to MOD are significantly lower than responses to BOS (t-tests; $RRS_{MOD/BOS} = 0.84 \pm 0.10$, $p = 1.1 \times 10^{-6}$, $RRV_{MOD/BOS} = 0.82 \pm 0.14$, $p = 1.3 \times 10^{-5}$), but significantly larger than those to CON and REV (paired t-tests, $p < 0.01$, $RRS_{CON/BOS} = 0.17 \pm 0.11$, $RRS_{REV/BOS} = 0.11 \pm 0.14$, $RRV_{CON/BOS} = 0.20 \pm 0.12$, and $RRV_{REV/BOS} = 0.13 \pm 0.10$). These results show that sound envelope structure is a quantifiable feature that is taken into account in the discrimination of BOS in highly selective HVC neurons

activity level no different than the spontaneous activity and a negative value would indicate that the stimulus resulted in greater inhibition than would be expected for spontaneous activity.

We conducted further analysis of the data, inspired by the stereotyped response zebra finches present to their BOS motif. Sliding-window PSTHs (10 ms window, sampling rate: 20 kHz) were computed for each stimulus and the motif activity profile (MAP) was obtained by aligning each response to the first motif window and computing the average, which was then smoothed by a 2nd order Savitzky–Golay filter. Since introductory notes had been presented only once per stimulus presentation, they were not considered in this analysis. The MAP allows us to study subtle differences between BOS and MOD song motifs, rather than evaluating on the global scale that RRS and RRV provide.

Peak amplitude was defined as the peak absolute value minus the minimum of the adjacent MAP minima. Prominent peaks for each MAP were defined as peaks that were larger by at least 2 standard deviations from the peak amplitude distribution mean of the peaks from all MAPs.

Then, we defined the Mean Relative Amplitude MOD/BOS ($MRA_{MOD/BOS}$) as the quotient between each peak amplitude found in the MOD response and the corresponding peak in the response to BOS. For phasic firing MAPs, the $MRA_{MOD/BOS}$ is a single value (prominent peak quotient). If the MAP for a given protocol contained more than one significant peak during motif response (as is the case for tonic neurons), the average of all the $MRA_{MOD/BOS}$ values was computed. This was done to give equal weights to the results from all neurons, regardless of whether each individual neuron presented one or several significant

peaks in the response. For grouped data, we present the average and variance from the $MRA_{MOD/BOS}$ distribution.

Given that the MAP of selective HVC neurons to CON and REV do not present prominent peaks, we only studied the difference between MOD and BOS. Nevertheless, we reproduced the measures of RRS and RRV using the activity profiles and found that the two distribution pairs were indistinguishable (paired t-tests, $p > 0.01$; test for $RRS_{MOD/BOS}$, $p = 0.38$ and test for $RRV_{MOD/BOS}$, $p = 0.54$).

Results

To test the hypothesis that the temporal structure of the sound envelope is a relevant feature for encoding the BOS in the song system, we developed a method to modify the temporal evolution of the BOS envelope and then used this modified BOS (MOD) as an auditory stimulus in HVC selectivity experiments. Our goal was to generate subtle changes in the sound, modifying acoustic features that the bird could actively control during singing. In this work, we explored HVC neural sensitivity to changes in the temporal structure of the sound envelope.

Figure 1 shows an example of a modified BOS (MOD), in which the BOS sound envelope was extracted for each syllable (red line in top panel of Fig. 1a) and then reverted (blue line in bottom panel of Fig. 1b). See "Methods" for a detailed description of MOD generation. To verify that the MOD presents modifications in the sound envelope structure while introducing minor variations in the spectral properties, we measured spectral centroid and spectral irregularity for BOS and MOD (see "Methods"). For each BOS–MOD pair, we computed the mean relative differences between spectral centroid (SC) and spectral irregularity (IRR) for each syllable. Analyses were performed on 22 syllables from the songs of 5 birds. BOS and MOD were indistinguishable according to both quantifications. Syllable differences on spectral centroid had no significant deviation from a zero-mean normal distribution (t test, $p = 0.34$), with distribution mean $SC_{MOD/BOS} = (-0.6 \pm 2.9) \times 10^{-3}$, $\pm$ indicates SD. The same resulted for syllable spectral irregularity (t test, $p = 0.85$), with distribution mean $IRR_{MOD/BOS} = (-0.1 \pm 3.6)$ x10^{-3}. Differences taken as absolute value presented distribution means of $SC_{MOD/BOS} = (1.9 \pm 2.3) \times 10^{-3}$ and $IRR_{MOD/BOS} = (2.3 \pm 2.7) \times 10^{-3}$.

As changes in sound amplitude can change the spectral properties when the analysis window is smaller than a syllable, we studied the temporal evolution of SC and IRR within each syllable (see "Methods"). Grouping the data from the 5 birds used in the experiments, we have a total of 530 temporal bins in which we analyzed the relative differences in SC and IRR (for additional detail in the spectral analysis, see Online Resources 1 and 2). We found that only a small number of temporal bins present differences larger than 3% (2% of the temporal bins for SC, 2.8% for IRR) and that only 0.2 and 0.4% present differences larger than 5% for SC and IRR, respectively. Finally, the maximum absolute difference found for any given bin was 5.4% for SC and 5.2% for IRR.

SC and IRR relative difference distributions have a large number of values grouped near zero, yielding mean differences of $(-0.5 \pm 7.4) \times 10^{-3}$ ($\pm$ represents S.D.) for SC and $(-0.1 \pm 1.2) \times 10^{-2}$ for IRR. Statistically, relative difference distributions are indistinguishable from a zero-mean normal distribution (t test, $p = 0.17$ for SC and $p = 0.61$ for IRR).

Alltogether, these results show that the transformation from BOS to MOD does not introduce significant changes to the spectral features of the songs. As revealed by the detailed spectral analysis, small variations from the original BOS can occur in the MOD for some temporal bins within each syllable. However, these changes are mostly within 3% with respect of the original value. On the syllable scale, these changes are small, as is reflected in the mean absolute difference of $SC_{MOD/BOS} = (1.9 \pm 2.3) \times 10^{-3}$ and $IRR_{MOD/BOS} = (2.3 \pm 2.7) \times 10^{-3}$.

During experiments, birds were presented with a pool of auditory stimuli: BOS, MOD, the song of a conspecific adult male (CON) and the reversed BOS (REV). Twenty-one selective HVC neurons were isolated from five birds, arising from 2 to 5 recording sites per bird. Selectivity was determined by stimulus response activity being significantly larger than baseline (paired t test, $p < 0.01$) and presenting ($RRS_{CON,REV/BOS}$ and $RRV_{CON,REV/BOS}$) < 0.5. As this work is focused on assessing whether acoustic modifications in BOS lead to changes in the HVC neural response, units that were non-selective ($RRS_{CON,REV/BOS}$ or $RRV_{CON,REV/BOS}$) ≥ 0.5 were omitted from the following analysis. For the 21 selective neurons, the average mean response CON / BOS was (0.17 ± 0.11), which is comparable to average mean responses CON/BOS reported in previous studies (0.16 ± 0.34), (Margoliash et al. 1994), (0.20 ± 0.22) (Boari et al. 2015), suggesting a shared underlying distribution. Figure 2 shows an example of raster plots and PSTHs for the response of one HVC selective unit to all the stimuli presented (a–d). Song traces have been superimposed as insets on each panel and their respective sonograms are shown in (e). This is the typical paradigm to study neural selectivity to BOS. Interestingly, MOD elicits neuronal responses that preserve the activity profile but the maximum amplitude of the response for MOD tends to be lower than for BOS, suggesting that BOS-selective neurons in HVC may be susceptible to subtle changes of the song envelope. Mean activity and variance were used as quantifiers of stimulus response (see "Methods"). Mean activity was used to assess significance with respect to spontaneous activity. Activity variance, on the other hand, reflects the presence of a pattern formed by activation and inhibition

during the stimulus response. RRS and RRV were used as measurements of the activity relative to BOS, as in a previous study (Boari et al. 2015). Figure 2f, g show the grouped data for all the experiments ($n = 5$ birds, $N = 21$ single units), displaying the calculation for each neuron, the mean value (horizontal solid line) and error bars indicating ± S.D of each distribution. The obtained mean values were $RRS_{MOD/BOS} = (0.84 \pm 0.10)$, $RRS_{CON/BOS} = (0.17 \pm 0.11)$ and $RRS_{REV/BOS} = (0.11 \pm 0.14)$. For RRV we have similar results, $RRV_{MOD/BOS}$ mean value is (0.82 ± 0.14), while the $RRV_{CON/BOS}$ mean value is (0.20 ± 0.12) and $RRV_{REV/BOS}$ (0.13 ± 0.10). These grouped data show that even though there is selectivity to MOD over CON and REV, the response to MOD is significantly lower than the response to BOS. This is represented by distribution means for $RRS_{MOD/BOS}$ and $RRV_{MOD/BOS}$ being lower than 1 (t tests, $p = 1.1 \times 10^{-6}$ and $p = 1.3 \times 10^{-5}$, respectively).

As stated before, although values significantly lower than 1 show that there is a measurable effect, this quantification may not be the most suitable to account for the change in the response. As can be seen in Fig. 2a, HVC neurons display activity patterns of excitation and inhibition. During inhibition periods, firing rate variations are hard to quantify (no release of inhibition was observed). Because RRS and RRV considered the whole interval of stimulus presentation, which includes significant parts of inhibitory activity, this might not be the most appropriate way to quantify subtle changes in neuronal activity. Therefore, we used a complementary measurement.

Upon observation of the data, the emerging effect is that the overall pattern of excitation and inhibition in response to MOD is similar to the one in response to BOS, but that the MOD-related activity reaches a lower level of activation. In other words, the peaks present on the MOD PSTH occur at the same time as those in the BOS PSTH, but with lower amplitude (see Fig. 2a, b). To quantify this observation, we computed the motif activity profile (MAP) by averaging across trials and motif repetitions (see "Methods"). Neural responses have different features depending whether it is elicited by a tonic neuron (Fig. 3a, b) or by a phasic neuron (Fig. 3c, d). By comparing the amplitude of the prominent peaks of the MAP, we considered changes in the neural response pattern that occur at certain times in the motif rather than averaging them across the whole stimulus presentation. Following this idea, we defined the magnitude Mean Relative $Amplitude_{MOD/BOS}$ ($MRA_{MOD/BOS}$) to quantify peak amplitude differences (see "Methods"). The mean $MRA_{MOD/BOS}$ resulted in (0.73 ± 0.16) and was found

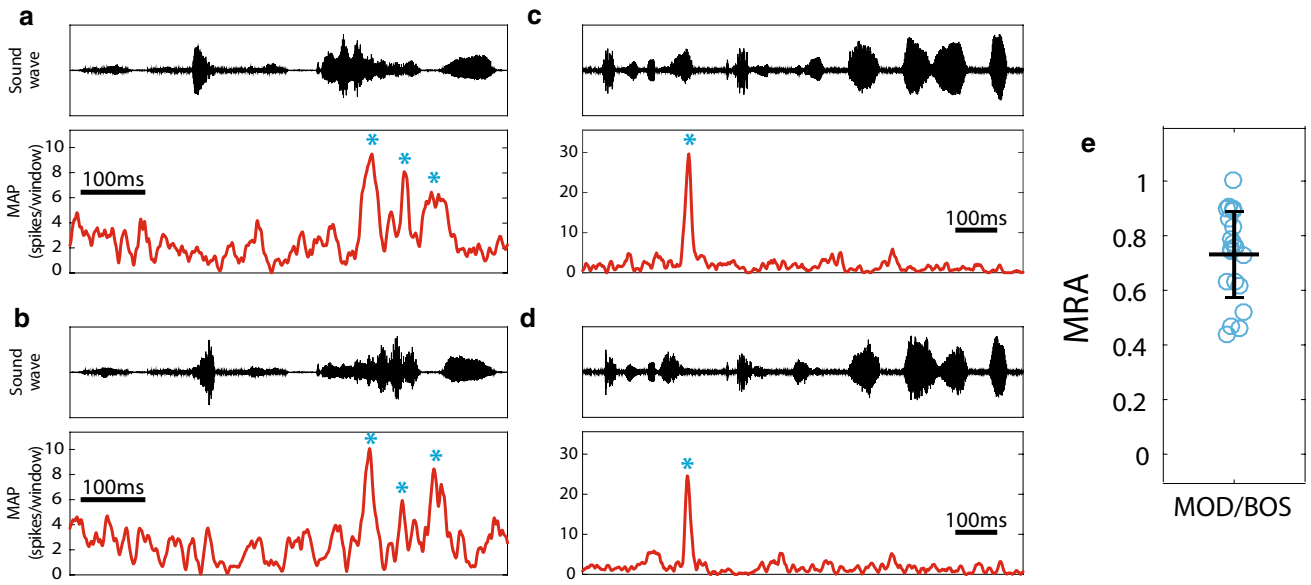

Fig. 3 Motif-averaged neural activity profile (MAP). As a localized measure, complementary to the one provided by RRS and RRV, the neural activity profile was computed by means of a 10 ms sliding window average on the raster plots and averaging responses across motif presentations in the same protocol. The first example **a, b** shows the activity profile for BOS (**a**) and MOD (**b**) for a tonic neuron in HVC. Asterisks (*) indicate prominent peaks, selected from the tail of peak amplitude distribution (greater than 2σ from distribution mean). Relative amplitudes were computed as the quotient between peaks in the MOD MAP and their respective peaks in the BOS MAP. For tonic neurons, the mean relative amplitude was calculated by averaging for all the prominent peaks found in the activity profile. Phasic neurons usually present a single, sharper peak in activity than those found in tonic neurons. The effect described above is also present here, with activity reaching greater amplitude to BOS (**c**) than to MOD (**d**). The distribution of mean relative amplitude values allows for a greater contrast in the quantification of the effect of attack time in auditory recognition of the BOS (**e**), yielding a value of Mean Relative $Amplitude_{MOD/BOS} = (0.73 \pm 0.16)$, which is significantly lower than 1 (t test, $p = 2.5 \times 10^{-7}$). These results show that changes in sound envelope drive changes in the auditory response of HVC units

to be significantly lower than 1 (*t* test, $p = 2.5 \times 10^{-7}$). Figure 3e displays the calculation for each neural response, the mean and the error bars (± S.D). Different neurons in HVC may fire with different profiles to the same acoustic stimuli (e.g., to BOS), with substantial variation in the number of peaks per neuron and in their heights. This could have led to the high variation found in the MRA values but no clear relationship was found between the $MRA_{MOD/BOS}$ values and the acoustic stimuli or the type of neural response. The low values for MRA (i.e., $MRA_{MOD/BOS} < 0.55$) were well distributed along the individuals, and therefore, we found no correlation between this feature and the studied acoustic properties.

Since prominent peaks were calculated from a motif-averaged response across trials, the lower amplitude of the most prominent peaks shows a decrease in firing rates for MOD with respect to BOS, as could already be seen on a global scale by means of RRS and RRV.

Our results show that changes in the temporal features of the BOS sound envelope can elicit a quantifiable effect in the auditory response of HVC neurons. By reverting the sound envelope of each syllable, we kept the overall spectrum properties and the maximum value of the envelope unchanged. In this way, we showed that the envelope shape is a relevant feature encoded in the song system, specifically when the attack and release times are switched.

Discussion

For many years, a collective scientific effort has advanced the knowledge of the production and learning of a complex motor task as is vocal behavior. Intertwined with this program is the understanding of auditory processing of uttered sounds, as auditory feedback plays an important role during learning, and in the maintenance of learned cues that lead to vocal production. The remarkable similarities between songbird and human vocal production and learning (Doupe and Kuhl 1999; Riede and Goller 2010) have led to a synergistic process in which discoveries in one field shed light on the other. For example, it has been shown that adult deafened birds progressively degrade their song and specific brain areas have been identified for this process (Brainard and Doupe 2000). Speech degradation after hearing impairment has also been observed in humans. Also, birds that have been raised in auditory isolation develop an abnormal song (Thorpe 1961; Feher et al. 2009), yet which specific features are relevant in the encoding of sound is still an open question.

Our findings show that the temporal structure of the envelope is a relevant feature for stimuli discrimination in HVC neurons. As the sound envelope of many syllables have an asymmetric shape, reverting the envelope switched the attack and release times. For these syllables, the neurons could be sensing a difference in attack time as well as a mismatch from the expected envelope. It is worth mentioning that sound discrimination is a complex task and could be driven by different cues. A recent behavioral study in European Starlings (*Sturnus vulgaris*) showed that birds rely on the spectral shape of sounds rather than on pitch cues to discriminate auditory stimuli (Bregman et al. 2016). Psychophysical experiments showed that birds are not particularly more sensitive than other vertebrates to slow temporal features of a sound, referred to as envelope characteristics (Dooling et al. 2000). However, Dooling and coworkers have shown that zebra finches seem to be extremely sensitive to the temporal fine structure of the waveform in both synthetic stimuli and natural vocalizations (Dooling et al. 2002; Dooling and Prior 2017). Our experiments tested an intermediate regime in which the envelope was modified reverting the temporal evolution. This generated an attack-time modification for some syllables but in other syllables, the changes were subtle envelope modulations.

Previous works have shown that for zebra finches, there exists a positive correlation between air sac pressure and fundamental frequency (Riede et al. 2010; Ritschard and Brumm 2011). Moreover, it has been shown that changing the air sac pressure during singing resulted in pitch modifications that followed the pressure profile (Amador and Margoliash 2013). These results suggested that respiratory and syringeal muscles were activated in a coordinated fashion to generate a given pitch. The dynamical origin of this effect was studied using a low dimensional biomechanical model for the vocal organ. This biomechanical model had been previously validated through several experiments, including physiological measurements (Mindlin et al. 2003; Perl et al. 2011) and electrophysiological measurements (Amador et al. 2013; Boari et al. 2015). Detailed dynamical analysis of the system describing the sound source, showed that there are constrains for the acoustic output that can arise from the exploration in parameter space that can be related to physiological variables (pressure-tension) (Perl et al. 2011). Moreover, experimental and theoretical tools have been used to test the hypothesis that for birds producing tonal sounds such as domestic canaries (*Serinus canaria*), frequency modulation is determined by both the syringeal tension and the air sac pressure (Alonso et al. 2014). For Bengalese finches (*Lonchura striata domestica*), the relationship between muscle activity and acoustic parameters has also been studied in singing birds, with results showing that EMG activity of single muscles can be correlated to multiple acoustic properties of song syllables (Srivastava et al. 2015). In this way, reverting the envelope in MOD generated some syllables that could be contradicting the biomechanical substrate for singing. As air sac pressure modulations are related to sound amplitude modulations (Boari et al. 2015) and the pressure

patterns during singing are highly stereotyped (Franz and Goller 2002), it could be the case that the envelope modification in MOD is producing mixed cues for HVC neurons in terms of expected amplitude profile, even if the spectral temporal evolution is only slightly modified.

This could be particularly relevant for HVC as this neural nucleus integrates auditory and motor inputs. Moreover, neurons in the song system display a precise auditory–vocal correspondence, resembling mirror neurons (Dave and Margoliash 2000; Prather et al. 2008). Therefore, further understanding of the vocal production mechanisms could also shed light onto which features of song are relevant for the auditory responses to BOS in HVC. The high selectivity in HVC is the result of a hierarchical processing of auditory pathway neurons, which present different levels of BOS selectivity in different nuclei of the system (Theunissen et al. 2004). Synthetic tones or white noise are poor stimuli for understanding auditory processing in higher brain areas of songbirds (Margoliash 1983; Theunissen et al. 2000), and preference for natural sounds and conspecific songs (CON) over synthetic stimuli are part of that processing hierarchy (Grace et al. 2003). Preference for CON song arises at field L, which is one of the auditory inputs to HVC. Therefore, as the auditory processing along the auditory pathway is selective to stimuli ranging from natural sounds to conspecific song, in HVC the selectivity is further tuned to the bird's own song. In performing that discrimination, several features of song are being sensed. As we have shown here, changing the temporal evolution of the amplitude envelope for each syllable while maintaining the temporal structure of song can lead to a decrease in response amplitude, but still remain selective over REV or CON songs.

Unveiling the mechanisms of timbre perception is important beyond the field of birdsong. In human health, perceptual studies have shown that cochlear-implant patients perform poorly with respect to normal-hearing subjects on timbre perception tasks such as musical instrument recognition, but that they can discriminate independent changes in features such as spectral centroid and attack time (Pressnitzer et al. 2005; Drennan and Rubinstein 2008; Kong et al. 2011). Achieving a better understanding of the mechanisms involved in sound perception could provide a valuable tool to drive further advances in cochlear implant technology and bioprosthetic devices.

Acknowledgements We thank Cecilia T. Herbert and Gabriel B. Mindlin for comments and discussions that greatly improved the manuscript. This work was partially supported by Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Argentina), Agencia Nacional de Promoción Científica y Tecnológica (ANCyT, Argentina), Universidad de Buenos Aires (UBA, Argentina) and National Institutes of Health (NIH, USA) through grant R01-DC-012859. Experimentation and surgical procedures were conducted following protocols approved by the Institutional Animal Care and Use Committee (IACUC) of the University of Buenos Aires.

References

Alonso R, Goller F, Mindlin GB (2014) Motor control of sound frequency in birdsong involves the interaction between air sac pressure and labial tension. Phys Rev E 89(3):032706

Amador A, Margoliash D (2013) A mechanism for frequency modulation in songbirds shared with humans. J Neurosci 33(27):11136–11144. https://doi.org/10.1523/JNEUROSCI.5906-12.2013

Amador A, Sanz Perl Y, Mindlin GB, Margoliash D (2013) Elemental gesture dynamics are encoded by song premotor cortical neurons. Nature 495:59–64. https://doi.org/10.1038/nature11967

Boari S, Perl YS, Amador A, Margoliash D, Mindlin GB (2015) Automatic reconstruction of physiological gestures used in a model of birdsong production. J Neurophysiol 114(5):2912–2922

Boersma P, Van Heuven V (2001) Speak and unSpeak with PRAAT. Glot International 5(9–10):341–347

Brainard MS, Doupe AJ (2000) Interruption of a basal ganglia–forebrain circuit prevents plasticity of learned vocalizations. Nature 404(6779):762–766

Bregman MR, Patel AD, Gentner TQ (2016) Songbirds use spectral shape, not pitch, for sound pattern recognition. Proc Natl Acad Sci USA 113(6):1666–1671

Caclin A, McAdams S, Smith BK, Winsberg S (2005) Acoustic correlates of timbre space dimensions: a confirmatory study using synthetic tones. J Acoust Soc Am 118(1):471–482

Dave AS, Margoliash D (2000) Song replay during sleep and computational rules for sensorimotor vocal learning. Science 290(5492):812–816

Dooling RJ, Prior NH (2017) Do we hear what birds hear in birdsong? Anim Behav 124:283–289

Dooling RJ, Lohr B, Dent ML (2000) Hearing in birds and reptiles. In: Dooling RJ, Fay RR (eds) Comparative hearing: birds and reptiles. Springer-Verlag, New York, pp 308–359

Dooling RJ, Leek MR, Gleich O, Dent ML (2002) Auditory temporal resolution in birds: discrimination of harmonic complexes. J Acoust Soc Am 112(2):748–759

Doupe AJ, Konishi M (1991) Song-selective auditory circuits in the vocal control system of the zebra finch. Proc Natl Acad Sci USA 88(24):11339–11343

Doupe AJ, Kuhl PK (1999) Birdsong and human speech: common themes and mechanisms. Annu Rev Neurosci 22:567–631

Drennan WR, Rubinstein JT (2008) Music perception in cochlear implant users and its relationship with psychophysical capabilities. J Rehabil Res Dev 45(5):779

Feher O, Wang HB, Saar S, Mitra PP, Tchernichovski O (2009) De novo establishment of wild-type song culture in the zebra finch. Nature 459(7246):564-U594. https://doi.org/10.1038/nature07994

Franz M, Goller F (2002) Respiratory units of motor production and song imitation in the zebra finch. Dev Neurobiol 51(2):129–141

Grace JA, Amin N, Singh NC, Theunissen FE (2003) Selectivity for conspecific song in the zebra finch auditory forebrain. J Neurophysiol 89(1):472–487

Iverson P, Krumhansl CL (1993) Isolating the dynamic attributes of musical timbre. J Acoust Soc Am 94(5):2595–2603

Kong Y-Y, Mullangi A, Marozeau J, Epstein M (2011) Temporal and spectral cues for musical timbre perception in electric hearing. J Speech Lang Hear Res 54(3):981–994

Krimphoff J, McAdams S, Winsberg S (1994) Caractérisation du timbre des sons complexes. II. Analyses acoustiques et quantification psychophysique. Le Journal de Physique IV 4(C5):C5–C625-C625-628

Lewicki MS (1996) Intracellular characterization of song-specific neurons in the zebra finch auditory forebrain. J Neurosci 16(18):5854–5863

Margoliash D (1983) Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. J Neurosci 3(5):1039–1057

Margoliash D (1986) Preference for autogenous song by auditory neurons in a song system nucleus of the white-crowned sparrow. J Neurosci 6(6):1643–1661

Margoliash D, Fortune ES (1992) Temporal and harmonic combination-sensitive neurons in the zebra finch's HVc. J Neurosci 12(11):4309–4326

Margoliash D, Fortune ES, Sutter ML, Yu AC, Wrenhardin BD, Dave A (1994) Distributed representation in the song system of Oscines: evolutionary implications and functional consequences. Brain Behav Evol 44(4–5):247–264

McAdams S, Winsberg S, Donnadieu S, De Soete G, Krimphoff J (1995) Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. Psychol Res 58(3):177–192

Mindlin GB, Gardner TJ, Goller F, Suthers R (2003) Experimental support for a model of birdsong production. Phys Rev E 68(4):41908. https://doi.org/10.1103/Physreve.68.041908

Mooney R (2000) Different subthreshold mechanisms underlie song selectivity in identified HVc neurons of the zebra finch. J Neurosci 20(14):5420–5436

Perl YS, Arneodo EM, Amador A, Goller F, Mindlin GB (2011) Reconstruction of physiological instructions from zebra finch song. Phys Rev E 84(5):051909. https://doi.org/10.1103/Physreve.84.051909

Prather JF, Peters S, Nowicki S, Mooney R (2008) Precise auditory-vocal mirroring in neurons for learned vocal communication. Nature 451(7176):305-U302. https://doi.org/10.1038/nature06492

Pressnitzer D, Bestel J, Fraysse B (2005) Music to electric ears: pitch and timbre perception by cochlear implant patients. Ann NY Acad Sci 1060(1):343–345

Quiroga RQ, Nadasdy Z, Ben-Shaul Y (2004) Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. Neural Comput 16(8):1661–1687

Riede T, Goller F (2010) Peripheral mechanisms for vocal production in birds—differences and similarities to human speech and singing. Brain Lang 115(1):69–80. https://doi.org/10.1016/j.bandl.2009.11.003

Riede T, Fisher JH, Goller F (2010) Sexual dimorphism of the zebra finch syrinx indicates adaptation for high fundamental frequencies in males. PLoS One 5(6):e11368. https://doi.org/10.1371/journal.pone.0011368

Ritschard M, Brumm H (2011) Effects of vocal learning, phonetics and inheritance on song amplitude in zebra finches. Anim Behav 82(6):1415–1422

Srivastava KH, Elemans CP, Sober SJ (2015) Multifunctional and context-dependent control of vocal acoustics by individual muscles. J Neurosci 35(42):14183–14194

Theunissen FE, Doupe AJ (1998) Temporal and spectral sensitivity of complex auditory neurons in the nucleus HVc of male zebra finches. J Neurosci 18(10):3786–3802

Theunissen FE, Sen K, Doupe AJ (2000) Spectral–temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. J Neurosci 20(6):2315–2331

Theunissen FE, Amin N, Shaevitz SS, Woolley SMN, Fremouw T, Hauber ME (2004) Song selectivity in the song system and in the auditory forebrain. Ann Ny Acad Sci 1016:222–245. https://doi.org/10.1196/Annals.1298.023

Thorpe WH (1961) Bird-song: the biology of vocal communication and expression in birds. Cambridge University Press, Oxford

Town SM, Bizley JK (2013) Neural and behavioral investigations into timbre perception. Front Syst Neurosci 7

Volman S (1996) Quantitative assessment of song-selectivity in the zebra finch "high vocal center". J Comp Physiol A 178(6):849–862