



HAL
open science

Neuronal encoding in a high level auditory area: from sequential order of elements to grammatical structure

Aurore Cazala, Nicolas Giret, Jean-Marc Edeline, Catherine del Negro

► To cite this version:

Aurore Cazala, Nicolas Giret, Jean-Marc Edeline, Catherine del Negro. Neuronal encoding in a high level auditory area: from sequential order of elements to grammatical structure. *Journal of Neuroscience*, 2019, 39 (31), pp.6150 - 6161. 10.1523/JNEUROSCI.2767-18.2019 . hal-02191908

HAL Id: hal-02191908

<https://hal.science/hal-02191908>

Submitted on 18 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Neuronal Encoding in a High-Level Auditory Area: From Sequential Order of Elements to Grammatical Structure

Aurore Cazala, Nicolas Giret, Jean-Marc Edeline, and Catherine Del Negro

Paris-Saclay Institute of Neuroscience, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 9197, Paris-Saclay University, 91405 Orsay, France

Sensitivity to the sequential structure of communication sounds is fundamental not only for language comprehension in humans but also for song recognition in songbirds. By quantifying single-unit responses, we first assessed whether the sequential order of song elements, called syllables, in conspecific songs is encoded in a secondary auditory cortex-like region of the zebra finch brain. Based on a habituation/dishabituation paradigm, we show that, after multiple repetitions of the same conspecific song, rearranging syllable order reinstated strong responses. A large proportion of neurons showed sensitivity to song context in which syllables occurred providing support for the nonlinear processing of syllable sequences. Sensitivity to the temporal order of items within a sequence should enable learning its underlying structure, an ability considered a core mechanism of the human language faculty. We show that repetitions of songs that were ordered according to a specific grammatical structure (i.e., ABAB or AABB structures; A and B denoting song syllables) led to different responses in both anesthetized and awake birds. Once responses were decreased due to song repetitions, the transition from one structure to the other could affect the firing rates and/or the spike patterns. Our results suggest that detection was based on local differences rather than encoding of the global song structure as a whole. Our study demonstrates that a high-level auditory region provides neuronal mechanisms to help discriminate stimuli that differ in their sequential structure.

Key words: artificial grammar; auditory perception; multielectrode; sequence; songbirds

Significance Statement

Sequence processing has been proposed as a potential precursor of language syntax. As a sequencing operation, the encoding of the temporal order of items within a sequence may help in recognition of relationships between adjacent items and in learning the underlying structure. Taking advantage of the stimulus-specific adaptation phenomenon observed in a high-level auditory region of the zebra finch brain, we addressed this question at the neuronal level. Reordering elements within conspecific songs reinstated robust responses. Neurons also detected changes in the structure of artificial songs, and this detection depended on local transitions between adjacent or nonadjacent syllables. These findings establish the songbird as a model system for deciphering the mechanisms underlying sequence processing at the single-cell level.

Introduction

Learning to recognize temporal sequences is fundamental to sensory perception and social communication. Like speech and language, the songs of many species of songbirds consist of several distinct vocal elements called syllables, organized into sequences.

The sequential ordering of syllables results from a learning process (Lipkind et al., 2013). Modifying the temporal arrangement of vocal items may influence receiver's responses (Holland et al., 2000; Clucas et al., 2004; Gentner, 2008; Comins and Gentner, 2010; Briefer et al., 2013; Suzuki et al., 2016, 2017). Songbirds therefore have perceptual capacities for assessing the temporal ordering of vocal sounds.

Several basic processes for the coding of temporal sequences of items in the brain have recently been proposed (Dehaene et al., 2015). The encoding of transitions between two adjacent items, the merging of successive distinct elements into a single auditory object and other hypothetical mechanisms have been proposed (Kiggins et al., 2012; Dehaene et al., 2015). Songbirds are well-suited species for the investigation of how neuronal circuits encode the temporal sequences of vocal items as their brain contains

Received Oct. 25, 2018; revised April 24, 2019; accepted April 24, 2019.

Author contributions: A.C., N.G., and C.D.N. performed research; A.C., N.G., and C.D.N. analyzed data; N.G. and C.D.N. designed research; N.G. and J.-M.E. edited the paper; C.D.N. wrote the paper.

This work was supported by the Centre National de la Recherche Scientifique, the Idex Neuro-Saclay, and the University of Paris Sud. N.G. was supported by Idex Neuro Saclay Postdoctoral Fellowship. A.C. was supported by French Minister of Research and Technology. We thank Christophe Pallier and Stanislas Dehaene for advice in the conception of the experimental design.

The authors declare no competing financial interests.

Correspondence should be addressed to Catherine Del Negro at catherine.del-negro@u-psud.fr.

<https://doi.org/10.1523/JNEUROSCI.2767-18.2019>

Copyright © 2019 the authors

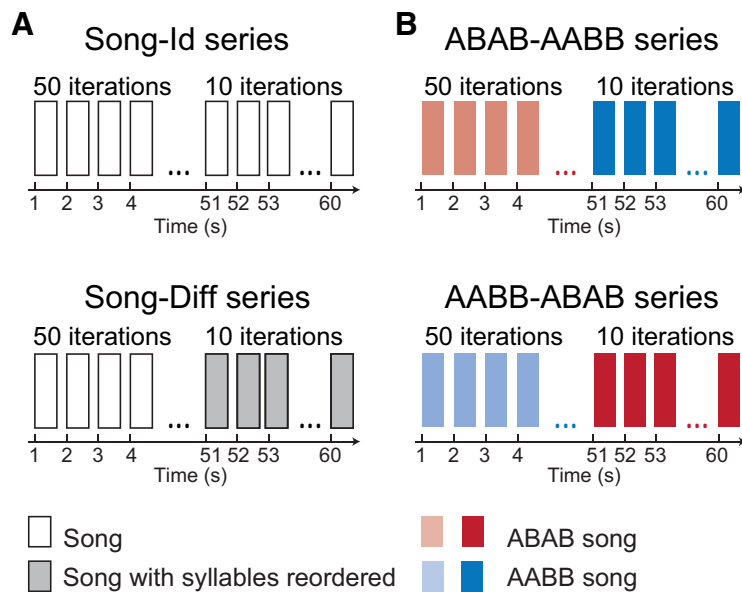


Figure 1. Schematic diagram of the series of songs used as stimuli. Except the Song-Id series, which consisted of 60 song iterations, the three other series consisted of 50 song iterations followed by 10 iterations of the same song in which syllables were reordered (Song-Diff series) or organized into a different structure (AABB-ABAB and ABAB-AABB series). Sets of two series were used: **A**, a Song-Id and its respective Song-Diff series built from the same syllables; **B**, an AABB-ABAB and an ABAB-AABB series built from distinct syllables.

many interconnected neural networks specialized in perception, learning, and production of vocal sounds, like humans. Sensitivity to temporal combinations of syllables has already been investigated in the set of brain nuclei specialized for singing and song learning (Margoliash, 1983; Doupe, 1997; Kojima and Doupe, 2008). For example, neurons in the nucleus HVC (used as a proper name) are sensitive to changes affecting the temporal structure of the bird's own song (Margoliash and Fortune, 1992; Lewicki and Konishi, 1995).

Much less is known about the encoding of temporal order of syllables within conspecific songs and the role of telencephalic auditory regions in sequence processing (Lu and Vicario, 2014; Ono et al., 2016). One of these regions, the caudomedial nidopallium (NCM), receives input from thalamo-recipient Field L and is a likely candidate area for discriminating songs according to syllable ordering. This is mainly because NCM neurons display nonlinear response properties. Indeed, neuronal responses to a multinote syllable cannot be explained by responses to the various notes presented separately (Ribeiro et al., 1998). Given these findings, it may be assumed that NCM neurons extend their nonlinear processing to sequences of contiguous syllables. In addition, NCM neurons show stimulus-specific adaptation where responses decrease with repeated presentation of the same song and exposure to a novel song reinstates responses (Chew et al., 1995; Mello et al., 1995; Beckers and Gahr, 2010; Menardy et al., 2012). This phenomenon has been interpreted as reflecting memory formation of the previously heard song. To date, most studies have only used songs produced by distinct birds as stimuli. In the present study, we provide evidence that a rearrangement in the temporal order of syllables is sufficient to reset neuronal responses after repeated presentations of a conspecific song.

Sensitivity to the order of items in a sequence may depend on the encoding of local transitions between adjacent or nonadjacent items or on its global structure (i.e., the items being grouped together as a “chunk”). Short sequences may present an oppor-

tunity to examine sensitivity to the local or global structure (Bekinschtein et al., 2009). The ability to learn the structure of sequences, considered as a core mechanism of human language faculty, has been investigated in both humans or nonhuman animals, including songbirds. In experiments, individuals were exposed to sound stimuli organized according to a structure based on artificial grammar, for example, (AB) n or a AnB n sequence (A and B denoting vocal sounds, with $n = 2$ or 3) (Fitch and Hauser, 2004; Friederici et al., 2006; Gentner et al., 2006; van Heijningen et al., 2009; Wilson et al., 2017; Wakita, 2019). Here, based on extracellular recordings in anesthetized and awake zebra finches, we show that NCM neurons responded differently to song stimuli organized into either an AABB or an ABAB structure.

Materials and Methods

Subjects and housing conditions. The subjects were 19 adult male zebra finches (*Taeniopygia guttata*), reared socially in the breeding colony of the Paris-Sud University. Birds were kept under a 12:12 light-dark cycle, with food and water *ad libitum*, and an ambient temperature

of 22°C–25°C. Experimental procedures were performed in compliance with national (JO 887–848) and European (86/609/EEC) legislation on animal experimentation, and following the guidelines used by the animal facilities of Paris-Sud University (Orsay, France), approved by the national directorate of veterinary services (#D91-429).

Auditory stimuli. As auditory stimuli, we first used natural zebra finch songs and variants of these songs that were built by manipulating syllable order. The set of songs included four distinct original zebra finch songs and three variants of these songs. Original songs came from a collection of songs previously recorded (sampling rate: 32 kHz) from adult male zebra finches that had lived in the laboratory's aviary years before the experiment. Subjects had never been exposed to these songs before the electrophysiological experiment. The four original songs were approximately the same length (mean \pm SD: 0.75 \pm 0.02 s). They consisted of five distinct syllables called A, B, C, D, and E. Both the delivery order of both syllables and silent gaps between syllables were manipulated to create song variants of the same length as the original ones. We made three variants of ABCDE songs by changing three or four of the four transitions between adjacent syllables: the BCEAD, EDCBA, or AECBD variants. Original songs and their respective variants were used to build sets of two series types (Fig. 1A). Each set included a series that consisted of 60 repeats of a given song (Song-Id series) and another series in which the same song was repeated 50 times followed by 10 presentations of a variant (Song-Diff series).

We also used four-element songs as stimuli. These artificial songs consisted of two element types, As and Bs, that were ordered in either the ABAB or the AABB structure. To build ABAB and AABB stimuli, we first selected 81 song syllables produced by 12 male zebra finches from our song database. The As and Bs were chosen to build ABAB or AABB stimuli of 0.70 \pm 0.30 s duration with 30–50 ms as intersyllable intervals, as typically found in zebra finch songs. Syllable duration ranged from 50 to 150 ms. All ABAB- or AABB-structured song stimuli started with the same introductory note. We created two series types: the ABAB-AABB or the AABB-ABAB series (Fig. 1B). They consisted of 50 repeats of either an ABAB or an AABB song followed by 10 presentations of the same song in which the serial positioning of the two central syllables A and B changed. In other words, the AABB and ABAB song stimuli of a given series were constructed from the same syllables A and B. We also built a third type of

series that consisted of only 60 repeats of an ABAB song. From the 81 syllables, we created a large set of ABAB-AABB or the AABB-ABAB series.

When a series was played back, the song stimulus was delivered at a rate of one per second.

Electrophysiological recordings. Neuronal activity in NCM was first recorded in anesthetized males while presenting the sets of songs that included either Song-Id and Song-Diff series or AABB-ABAB and ABAB-AABB series. Because NCM contains at least two populations of neurons that can be distinguished on the basis of their action potential (AP) width and their firing rate (FR) (Schneider and Woolley, 2013; Ono et al., 2016), we only analyzed the activity of very well isolated NCM single neurons (Fig. 2). Subsequently, we performed multiunit recordings in awake zebra finch males ($n = 5$) while presenting AABB-ABAB and ABAB-AABB series.

Acute recordings. Birds were anesthetized with isoflurane gas (in oxygen; induction: 3%, maintenance: 1.5%) that flowed through a small mask over the bird's beak. The bird was immobilized in a custom-made stereotaxic holder that allowed the head to be tilted at 45° and placed in a sound attenuation chamber. Lidocaine cream was applied to the skin. A window was opened in the inner skull layer, and small incisions were made in the dura. A multielectrode array of 8 or 16 tungsten electrodes (1–2 M Ω impedance at 1 kHz; Alpha Omega Engineering), which consisted of two rows of four (or eight) electrodes separated by 100 μ m apart, with 100 μ m between electrodes of the same row, was lowered to record extracellular activity. The array was positioned 0.3–0.5 mm lateral and 0.7–0.9 mm rostral to the bifurcation of the sagittal sinus in either the left or the right hemisphere, with a micromanipulator, as in previous studies (Stripling et al., 1997; Menardy et al., 2012, 2014). The probe was lowered very slowly until electrode tips reached 1200 μ m below the brain surface. From 1200 to 1900 μ m below the brain surface, auditory stimuli were delivered when the amplitude of AP waveforms recorded with at least one of the eight wires was clearly distinct from background noise. Recording sites were at least 100 μ m apart to minimize the possibility that the neural activity recorded from two successive sites originated from the same single units. Electrode signals were amplified and filtered (gain 10,000; bandpass: 0.3–10 kHz; AlphaLab SnR, AlphaOmega Engineering) to extract multiunit activity. During recordings, voltage traces and APs were monitored in real time using the AlphaLab SnR software. Auditory stimuli were concomitantly recorded and digitized to precisely determine the onset of NCM responses with respect to the sound stimulus. While spiking activity was recorded, auditory stimuli were broadcasted through a speaker situated 30 cm from the bird. In the first experiment, the set of the two series of song stimuli, which included the Song-Id and Song-Diff series, was presented. The set of stimuli that consisted of both the ABAB-AABB and ABAB-AABB series was used in the second experiment. From one recording site to the following one, we changed the set of series used as auditory stimuli and the order of series. All stimuli had been normalized to achieve maximal amplitude of 70 dB (Audacity software) at the level of the bird's head. Spike sorting of neuronal activity was done offline.

Chronic recordings. Surgical procedures were similar as described above. To perform chronic recordings, we used a custom-built screw microdrive that allows a microelectrode array to be repositioned. We used arrays of eight electrodes (two rows of four electrodes separated by 100 μ m apart; with a ground silver wire and a reference wire; 1–2 M Ω impedance at 1 kHz; AlphaOmega Engineering). Once the array was lowered into the brain to a depth of 1200 μ m, the reference wire was inserted between the outer and the inner skull layers. The microdrive was secured to the skull using dental cement. Subjects were allowed to recover for a few days. In the sound-attenuation chamber, the implanted micro-

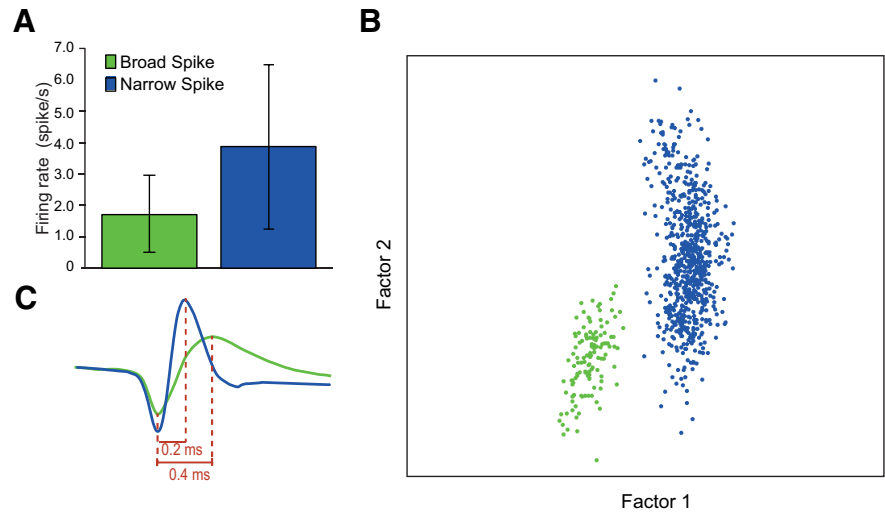


Figure 2. Two distinct cell types in NCM: BS and NS cells. **A**, Spontaneous FR (spike/s) of the populations of BS and NS neurons. Error bar indicates mean \pm SD. **B**, Representative example of spike clustering that distinguishes a BS (green) from an NS neuron (blue). **C**, Example of a spike waveform of these two neurons.

drive was connected through a commercial tether and head stage (AlphaOmega Engineering) to a mercury commutator located on the roof of the cage (Dragonfly Systems). An elastic thread built into the tether helped to support the weight of the implant. Subjects remained tethered during the experiment. The screw drive held the electrode array. Each full turn of the screw advanced the array by 200 μ m. Before a recording session, we rotated the screw by half-turn to advance the microelectrode array in step as \sim 100 μ m. Birds were not freely moving during the recording session. They were restrained with a jacket around their bodies. At least 24 h separated two recording sessions. As stimuli, we used one set of two series (ABAB-AABB or AABB-ABAB series) and also a third series that consisted of 60 repeats of an ABAB song (ABAB-ABAB series). From one recording session to the following one, we changed the set of series used as auditory stimuli and the order of series.

Data processing and analysis. In anesthetized birds, spike sorting was performed using the template-matching algorithm of the Spike2 software (version 8.0, Cambridge Electronic Design). In awake birds, spikes were detected by amplitude thresholding. Responses to stimuli were quantified by calculating average evoked FRs and Z scores. Evoked responses correspond to the FR during stimuli presentations. Z scores were measured as follows:

$$Z \text{ score} = \frac{\text{Evoked FR} - \text{Baseline FR}}{SD_{\text{Baseline FR}}}$$

The baseline FR was calculated for the last 200 ms period that preceded stimulus presentation. We calculated Z score values per block of 10 presentations, giving us 6 values per series (one per block of 10 iterations of the stimulus). The temporal pattern of responses evoked by both AABB and ABAB songs was quantified calculating the following: (1) the average FRs evoked by individual syllables contained in these songs; and (2) the spike-timing reliability coefficient (CorrCoef), which was used to quantify the iteration-to-iteration reliability of responses. It was computed per block of 10 stimulus iterations: it corresponds to the normalized covariance between each pair of AP trains and was calculated as follows:

$$\text{CorrCoef} = \frac{1}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{\sigma x_i x_j}{\sigma x_i \sigma x_j}$$

where N is the number of iterations and $\sigma x_i x_j$ is the normalized covariance at zero lag between spike trains x_i and x_j , where i and j are the iteration numbers. Spike trains x_i and x_j were previously convolved with a 10-ms-width Gaussian window. The CorrCoef was used because this index was not influenced by fluctuations of FR (Gaucher et al., 2013).

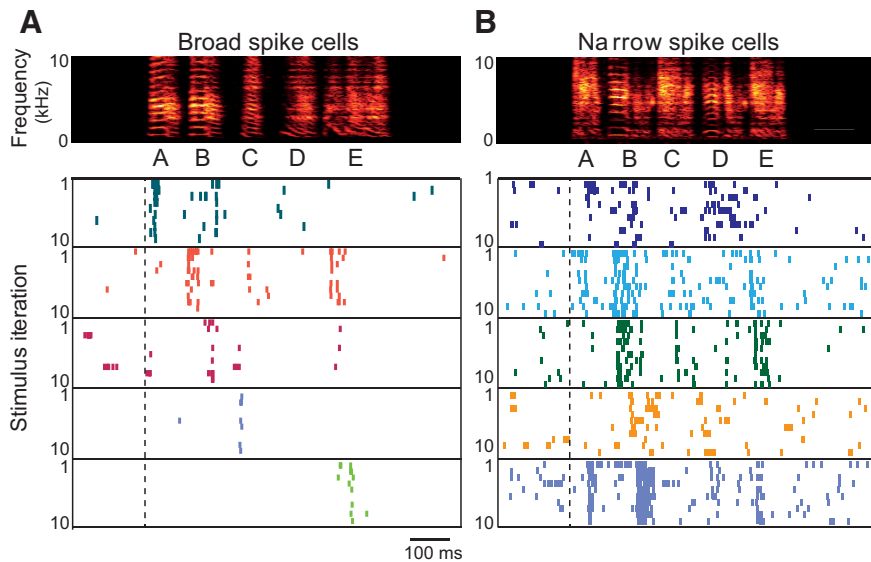


Figure 3. From one cell to another, song-driven spiking events can occur at different times during the presentation of a song stimulus. Neuronal responses of 5 BS and 5 NS cells are shown as raster plots (first 10 iterations). Top, The spectrogram of the song. BS neurons (**A**) produced more precise spike trains than NS neurons (**B**). Songs consisted of five distinct syllables (called from A to E). All the syllables of a given song can potentially elicit changes in spiking activity.

Statistical computations were performed in Statistica (version 8.0; StatSoft). FRs and Z score values were analyzed using repeated-measures ANOVA in a GLM. Several cofactors were included in the model, depending on the analysis: the block effect ($n = 6$), the series type (Song-Id vs Song-Diff; ABAB-AABB vs AABB-ABAB; ABAB-AABB vs ABAB-ABAB), the cell type (broad spike cells vs narrow spike cells), and the serial position of the syllable within the song ($n = 4$). We used planned contrast and Tukey HSD *post hoc* tests for assessing pairwise differences.

Histology. At the end of each experiment, the animal was killed with a lethal dose of pentobarbital and the brain quickly removed from the skull and placed in a fixative solution (4% PFA). Sections (100 μm) were cut on a vibratome to examine the location of multielectrode array penetration tracks.

Results

Experiment 1: repeated playbacks of a conspecific song followed by a change in the sequential order of syllables

Auditory responses in NCM decrease with song or call repetition, but a novel song or call stimulus reinstates responses (Chew et al., 1995; Stripling et al., 1997; Beckers and Gahr, 2010; Menardy et al., 2012). Various studies have exploited this property to characterize what is being encoded by the NCM neurons. Here, we assessed whether NCM neurons are sensitive to the auditory temporal context in which syllables occur by manipulating the temporal order of syllables. Such manipulation preserves the local and spectral structure within each syllable but alters what precedes each syllable. Transitions between adjacent syllables changed in all but one case (the transition B to C when ABCDE was reordered to BCEAD). We hypothesized that, if NCM neurons are only sensitive to the acoustic features of syllables, their responses should continue to decrease after changing the sequential order of syllables. On the contrary, if NCM neurons are also sensitive to what precedes syllables (i.e., depends on the auditory context), changing the temporal order of syllables within songs should reinstate robust neuronal responses.

Extracellular recordings from the NCM were performed in eight isoflurane-anesthetized adult males. Only well-isolated single units (Fig. 2), which responded to song playback with increased FRs relative to baseline activity, were selected. The

spiking activity of 205 auditory NCM neurons was analyzed. These units were from the dorsorostral portion (maximal depth 2000 μm) to the dorsocaudal portion as described by Menardy et al. (2012). They were driven by the playback of two series: the Song-Id and the Song-Diff series.

As described in previous studies (Meliza and Margoliash, 2012; Schneider and Woolley, 2013; Ono et al., 2016; Yanagihara and Yazaki-Sugiyama, 2016), responsive NCM neurons can be distinguished by the width of their AP and their spontaneous FR as follows: broad-spiking (BS) neurons (AP width: 0.50 ± 0.07 ms; FR: 1.81 ± 1.34 spikes/s; mean \pm SD, $n = 134$) and narrow-spiking (NS) neurons (AP width: 0.21 ± 0.05 ms, 4.00 ± 2.77 spikes/s; $n = 71$; Fig. 2). The pattern of song-evoked responses differed between BS and NS neurons. The responses of five BS cells and five NS cells to the first 10 iterations of a given song are displayed in Figure 3. The BS units showed phasic responses, firing only at precise restricted

points during the song stimulus. The NS units responded to the conspecific song with less accurate timing in their spike trains and continued to fire more or less during the entire song. As has been previously reported (Schneider and Woolley, 2013) and is illustrated in Figure 3, spiking events from one cell to another occurred at different time points in the song. All syllables can potentially drive spiking events from both BS and NS neurons, which supports the idea that neural coding of songs in the NCM relies on distributed population coding (Schneider and Woolley, 2013).

To assess NCM neuron sensitivity to repeated song exposure and to changes in syllable order, we calculated normalized Z score values per block of 10 iterations and performed a repeated-measures ANOVA on Z score values. Three cofactors, including block, series type, and cell type, were used. The time course of changes differed between the two series (interaction block \times series type, $F_{(5,1000)} = 2.40$, $p = 0.038$; Fig. 4A). Repetition of the same song 50 times decreased the response magnitude, but this did not differ between the two series (main effect of block factor, $F_{(5,1000)} = 9.01$, $p < 0.001$; from block 1 to block 5, *post hoc* HSD Tukey tests, all $p > 0.5$). Changing the syllable order within the song or extending its repetition differentially affected response magnitude (planned contrast of series type on the two last blocks, $F_{(1,200)} = 8.36$, $p = 0.004$; Song-Id series vs Song-Diff series on block 6, $p = 0.003$; Fig. 4A). Rearranging syllable order reinstated strong responses (Song-Diff series, block 5 vs block 6, $p = 0.003$).

The time course of responses and the sensitivity to syllable order rearrangement differed between the two cell types (interaction between the three factors: $F_{(5,1000)} = 2.91$, $p = 0.01$). NS cells did not show any modulation in their responses with stimulus repetition (Fig. 4B). A transient change in responses, however, occurred: from the first to the second presentation, FRs sharply declined (series Song-Id, first iteration mean \pm SD = 5.43 ± 0.61 spikes/s; second iteration mean \pm SD = 3.57 ± 0.63 spikes/s; paired *t* test, $t_{(70)} = 8.62$, $p < 0.001$; series Song-Diff: first iteration: 4.93 ± 0.51 spikes/s; second iteration: 3.2 ± 0.39 spikes/s; paired *t* test: $t_{(70)} = 11.73$, $p < 0.001$). FRs subsequently remained stable despite song repetition and syllable order rear-

repetition. In contrast, BS cells showed a gradual decrease in response magnitude when the same song was repeated (Fig. 4C; $F_{(5,500)} = 6.89$, $p < 0.0001$; *post hoc* tests for both series, block 1 vs 4 or 5, $p < 0.0001$). Reordering the syllables reset the FR to a high level ($F_{(1,100)} = 28.58$, $p < 0.0001$; *post hoc* test, block 5 vs 6, $p < 0.0001$); no change was observed when the order was kept the same (block 5 vs 6, $p = 0.99$). During the last block of song presentation, responses also differed between the two series ($p < 0.0001$). Of the 134 BS cells, 98 (73.1%) showed stronger responses when syllables were reordered. These results provide evidence that the majority of BS cells are sensitive to manipulation of the sequential ordering of syllables within songs. The Figure 4D displays the spiking activity of one representative BS neuron.

Interestingly, the mean FR of other BS cells did not increase when syllables were reordered, but responses were clearly affected by syllable reordering. Three such examples are presented in Figure 5; these also illustrate context dependency of responses. In the first example, the neuron showed peak activity at song onset during the repeated song (Fig. 5A, syllable A). Following syllable reordering (i.e., when syllable A was preceded by syllable E in the BCEAD song), the neuron no longer responded to this syllable (Fig. 5A, Song-Diff series). A peak in spiking activity was still observed when the syllable remained at the start of the ABCDE song (Fig. 5A, Song-Id series). Figure 5B shows the responses of another representative neuron. Rearranging the syllable order suppressed responses to syllable D (Fig. 5B, blue arrows, syllable D). Interestingly, these two neurons displayed in Figure 5A and B showed a greater response to another syllable when syllables were reordered (red arrows, syllable B in Fig. 5A; syllable E in Fig. 5B). Thus, the loss of response driven by a particular syllable could be accompanied by the reinstatement of robust responses to another syllable. The third example, which is displayed in Figure 5C, showed complete suppression of responses to syllable C after syllable reordering, even though this syllable remained at the same position within the song. A total of 12 cells displayed similar changes in discharge patterns. This suggests that a large fraction of BS neurons (110 of 134, 82%) encoded the auditory context in which syllables occurred through song repetitions either by increases in FR and/or changes in discharge patterns. Together, these results provide evidence that responses elicited by a multiple-syllable song cannot be reduced to responses evoked by individual syllables. This illustrates the sensitivity of NCM neurons to the auditory context in which syllables are presented.

Experiment 2: repeated playbacks of either an ABAB or AABB-structured song stimulus followed by a change in structure

To further examine the context dependency of NCM neurons, we used artificially AABB- or ABAB-structured song stimuli that

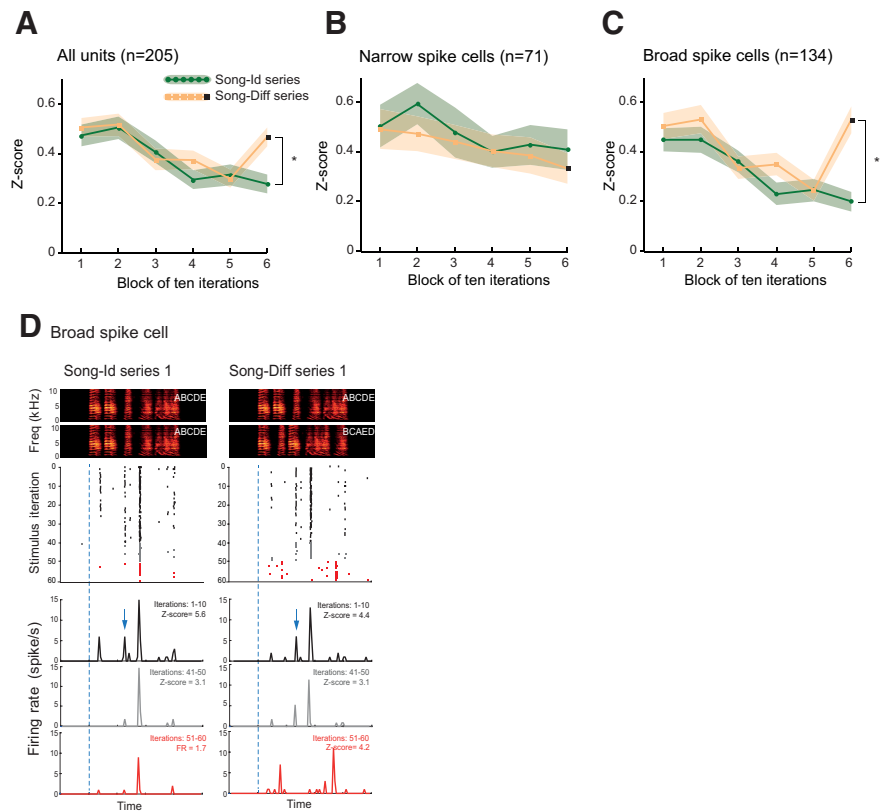


Figure 4. Reordering the syllables within a conspecific song affects the modulation of single-unit responses. Responses (Z score values) of the whole population of cells (**A**), the subset of NS cells (**B**), and the subset of BS cells (**C**) to repeated songs of either the Song-Diff (orange lines) or the Song-Id (green lines) series. Z score values were calculated per block of 10 song iterations. Changes in syllable order occurred at the beginning of the sixth block of the Song-Diff series. Thick line indicates mean. Shaded area represents SEM. * $p < 0.05$. **D**, Responses of a representative BS neuron to 50 presentations of a song consisting of five distinct syllables (ABCDE) followed by 10 presentations of the same song in which syllables are in the same order (Song-Id series) or arranged according to a new one (Song-Diff series; here BCAED). Neuronal responses are shown as raster plots (middle, 60 iterations) and peristimulus histograms (bottom) that are time-aligned with song spectrograms (top, the song repeated 50 times; bottom, the same song or the song with reordered syllables). Dotted lines indicate song onset.

differed in the serial positioning of the two central syllables A and B. The first and the fourth syllable remained at the same position. We aimed to examine how neurons responded to multiple repetitions of AABB and ABAB songs. Zebra finches are behaviorally able to discriminate between both songs (van Heijningen et al., 2009) and, based on previous studies, the NCM could provide a neural substrate for this ability. As suggested by the first experiment, if NCM neurons encode relationships between syllables, once responses are habituated, permuting the two central syllables A and B should reinstate larger responses.

At first, we performed neuronal recordings in anesthetized birds ($n = 7$). A total of 160 single units showed auditory responses to song stimuli. Both the AABB and ABAB songs induced a strong increase in FRs. From one neuron to another, spiking events occurred at different times in the songs (Fig. 6), and both syllables types could drive spiking activities. As shown in Figure 7A, responses to both song types were greatest during the first block of 10 iterations and showed a gradual decrease across successive blocks (main effect of block repetition, $F_{(5,790)} = 24.13$, $p < 0.001$; *post hoc* tests; block 1 vs 3, 4, and 5, for both series, $p < 0.01$ in all cases). However, lower responses were elicited by AABB compared with ABAB songs from the second block, suggesting that sequential structure of songs had an impact on responses (main effect of series type, $F_{(1,158)} = 9.25$, $p = 0.003$; blocks 2 and 5, $p < 0.002$). Neither the transition from AABB to

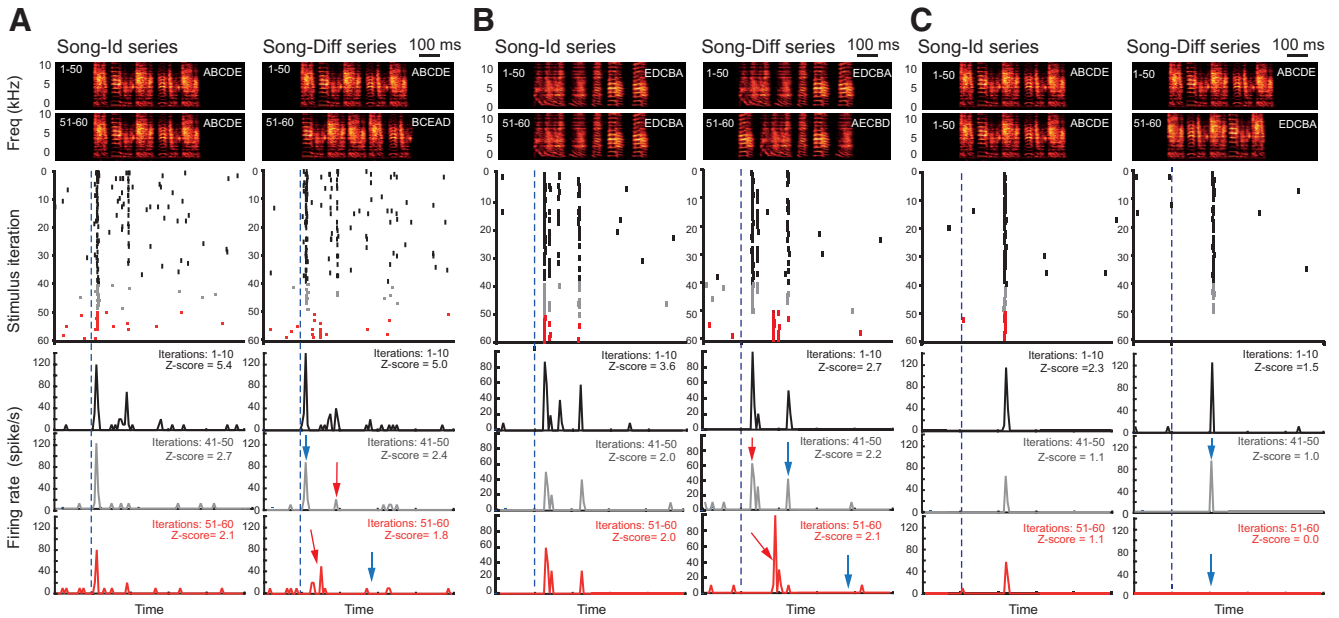


Figure 5. Reordering song syllables affects responses of three example BS cells. **A**, Spiking activity was suppressed after moving the syllable A from the first to the fourth position (ABCDE: initial order of syllables; BCEAD: new order in Song-Diff series; iterations: 51–60; blue arrows). There is reinstatement of responses to the syllable B after syllable reordering (red arrows). **B**, Reordering song syllables suppressed responses to the syllable D (EDCBA: initial order; AECBD: new order; blue arrows) and reinstated responses to syllable E (red arrows). **C**, Suppression of responses to the syllable C that remained to the same position after syllable reordering (ABCDE: initial order; EDCBA: new order; blue arrows). The preceding syllables were different. Neuronal responses are shown as raster plots (middle, 60 iterations) and peristimulus histograms (bottom) that are time-aligned with the song spectrograms. Spectrograms of songs used as stimuli: Top, During the first 50 iterations. Bottom, During the 10 last ones. Dotted blue lines indicate song onset.

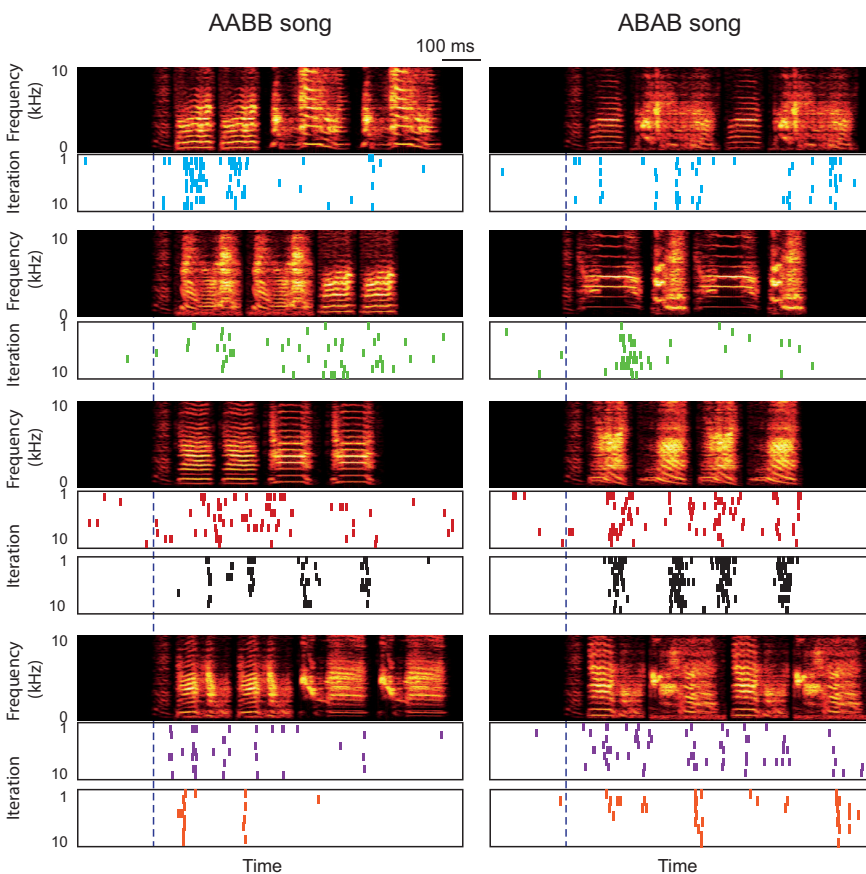


Figure 6. Responses of six example neurons to the playback of ABAB and AABB song stimuli. Neuronal responses are shown as raster plots (the first 10 iterations) that are time-aligned with the song spectrograms. Song-driven spiking events can occur at different times during both ABAB and AABB songs. Dotted blue lines indicate song onset. All the syllables of a given song can elicit changes in spiking activity.

ABAB nor vice versa was detected by the population of NCM cells as a whole.

Responses of both NS ($N = 68$) and BS cells ($N = 92$) decreased in magnitude with song repetition ($F_{(5,790)} = 24.13, p < 0.001$); however, a difference in the amount of decrease between AABB and ABAB songs was seen ($F_{(1,158)} = 9.25, p = 0.003; p < 0.002$). NS cells were not sensitive to changes in song structure once responses were habituated (Fig. 7B). In contrast, BS cells detected the transition from AABB to ABAB structure (“cell type” factor, $F_{(1,158)} = 6.12, p = 0.014$; *post hoc* test, block 5 vs 6; $p = 0.038$; Fig. 7C). Figure 7D illustrates a neuron showing this detection in its responses.

To assess whether responses could depend on behavioral state, we performed multiunit recordings in awake birds (recording sites $n = 52$ in 4 birds). On average, three (range 2–5) recording sessions (4.4 electrodes per recording session, range 2–8) per bird were performed with, on average, 5 d between two successive recording sessions (range 3–10). Most results are consistent with those obtained from anesthetized birds. Responses (Z score values) were the highest during the first block of song iterations and decreased during subsequent blocks (main effect of block repetition, $p < 0.001$). The AABB songs induced lower responses than ABAB songs (main effect of the song

type: $F_{(1,51)} = 11.9$, $p = 0.001$; from block 2 to block 5, *post hoc* tests: $p < 0.001$; Fig. 8A). After 50 song iterations, changing the song structure did not reinstate stronger responses (Fig. 8A). For example, changing from ABAB to AABB did not induce greater responses than continuing to repeat ABAB up to 60 times (Fig. 8A). The spiking activity of 18 single units from our recordings were further analyzed. As illustrated by the example presented in Figure 8B, BS cells showed an increase in FR when AABB songs were changed to ABAB. Thus, we confirm the results obtained from anesthetized birds that BS cells can detect the change in song structure from AABB to ABAB.

The difference in response habituation between AABB and ABAB songs and the differential effect of changing the song structure on responses led us to analyze how neurons respond to individual syllables contained in songs. We aimed to assess whether syllable-evoked responses express a context dependency and, if so, whether this context dependency is limited to local transitions between adjacent syllables. The first syllable B occurred in the following two different auditory contexts: (1) it was preceded by two iterations of the syllable A in AABB songs; and (2) it was preceded by only one in ABAB songs. The syllable pair AB, which included the first syllable B, was preceded by single syllable A in AABB songs while it occurred in silence in ABAB songs. If the auditory context does not affect syllable-evoked responses, similar responses to the first syllable B and the syllable pair AB (from syllable A to B) should be observed when AABB and ABAB songs are played back. On the contrary, differences between AABB and ABAB songs should be observed if the auditory context has an impact on responses.

Variations in syllable-evoked FRs differed between AABB and ABAB songs from the first to the fourth song syllable in both anesthetized and awake birds (“syllable position” effect: $F_{(3,465)} = 36.55$, $p < 0.001$ and $F_{(3,177)} = 8.52$, $p < 0.001$; interaction between “syllable position” and “song type” factors: $F_{(3,465)} = 24.72$, $p < 0.001$ and $F_{(3,177)} = 8.97$, $p < 0.001$; Fig. 9). As shown in Figure 9A and B, there was a gradual decrease in FR when syllables were organized according to the AABB structure (*post hoc* tests, from block 1 to 5, serial position 1 vs 3 and 4; 2 vs 3 and 4, all $p < 0.01$). In contrast, when syllables were alternated in ABAB songs, the FR first increased before it decreased and then increased again (Fig. 9A,B). Thus, FRs from one syllable to the next thus strongly differed between the two songs.

Responses driven by the first syllable B were weaker when AABB songs rather than ABAB songs were played back (*post hoc* test, $p < 0.01$). Changes in FR from syllable A to B differed. When syllable B followed two syllables A in AABB songs, the FR decreased (Fig. 9A,B, blocks 1 and 5). When syllable B followed a single A in ABAB songs, the FR increased (Fig. 9A,B, blocks 1 and 5). These results indicate that responses to the first syllable B

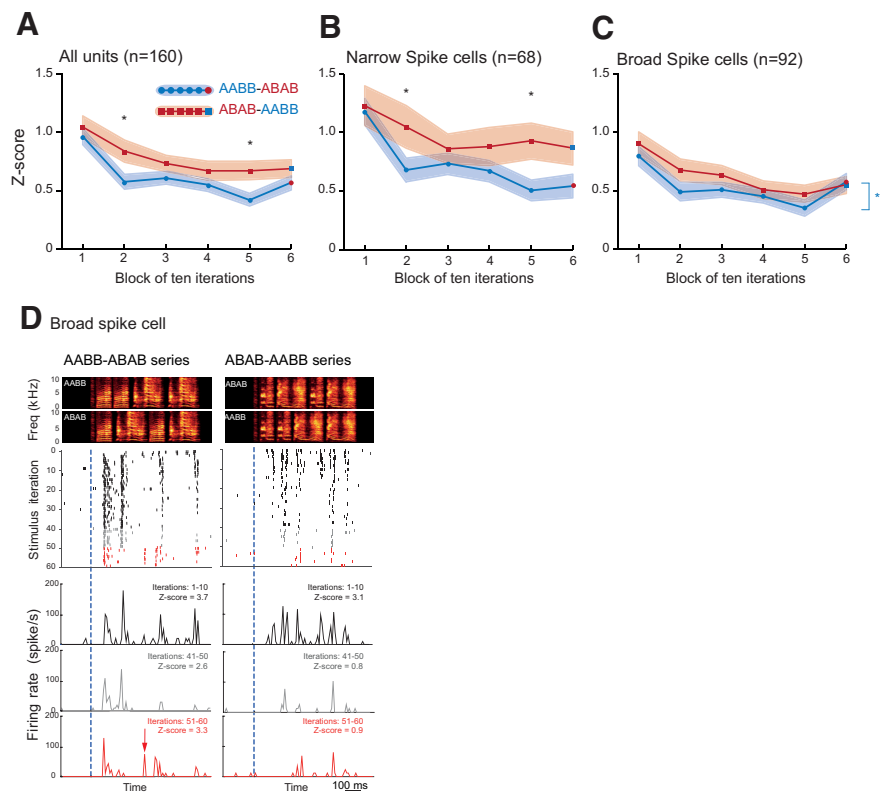


Figure 7. Responses driven by ABAB and AABB song stimuli in anesthetized birds. Responses (Z score values) of the whole population of cells (A), the subset of NS cells (B), and the subset of BS cells (C) across the successive blocks of ABB-ABAB and ABAB-AABB series. Z score values were calculated per block of 10 song iterations. BS cells detected the transition from ABB to ABAB structure. Thick line indicates mean. Shaded area represents SEM. $*p < 0.05$. D, Responses of a representative BS neuron to the playback of ABB and ABAB songs. Neuronal responses are shown as raster plots (middle, 60 iterations) and peristimulus histograms (bottom) that are time-aligned with song spectrograms. Spectrograms of songs used as stimuli: Top, During the first 50 iterations. Bottom, During the 10 last ones. Dotted blue lines indicate song onset. There is an increase in the response to the first syllable B (red arrow) when song structure changed from ABB to ABAB.

depended on the number of As that preceded it. Also, differences in responses between syllable A and the following syllable B depended on the auditory context in which the syllable pair AB occurred (silence vs preceded by the presentation of a syllable A).

A puzzling result, observed in both anesthetized and awake birds, was the large difference in response habituation between ABB and ABAB songs (Figs. 7A; 8A). The first syllable of ABB songs should have induced weaker responses than the first syllable of ABAB songs if the larger decrease in Z score values when ABB songs were played back resulted from fatigue. Instead, the first syllable A of ABB songs drove stronger responses than the first syllable A of ABAB songs in both anesthetized and awake birds (block 1 to 5, $p < 0.01$; in all cases). In contrast, the last syllable of ABB songs evoked much less spiking activity than the last syllable of ABAB songs both in anesthetized and awake birds (blocks 1 to 5, $p < 0.01$). These differences (larger responses to the first syllable and lower responses to the last syllable) between ABB and ABAB songs clearly illustrate the drastic decrease in FR during the ABB song. In addition, from the last syllable of a given ABB song iteration to the first syllable of the following one, the FRs partially recovered high values (Fig. 9A,B, blue dotted lines). This highlights another important difference between ABB and ABAB songs.

The impact of transition from one structure to the other one on syllable-evoked FRs of NS and BS cells in anesthetized conditions was subsequently assessed. In Figure 9C, responses elicited by central syllables during block 5 were artificially permuted to

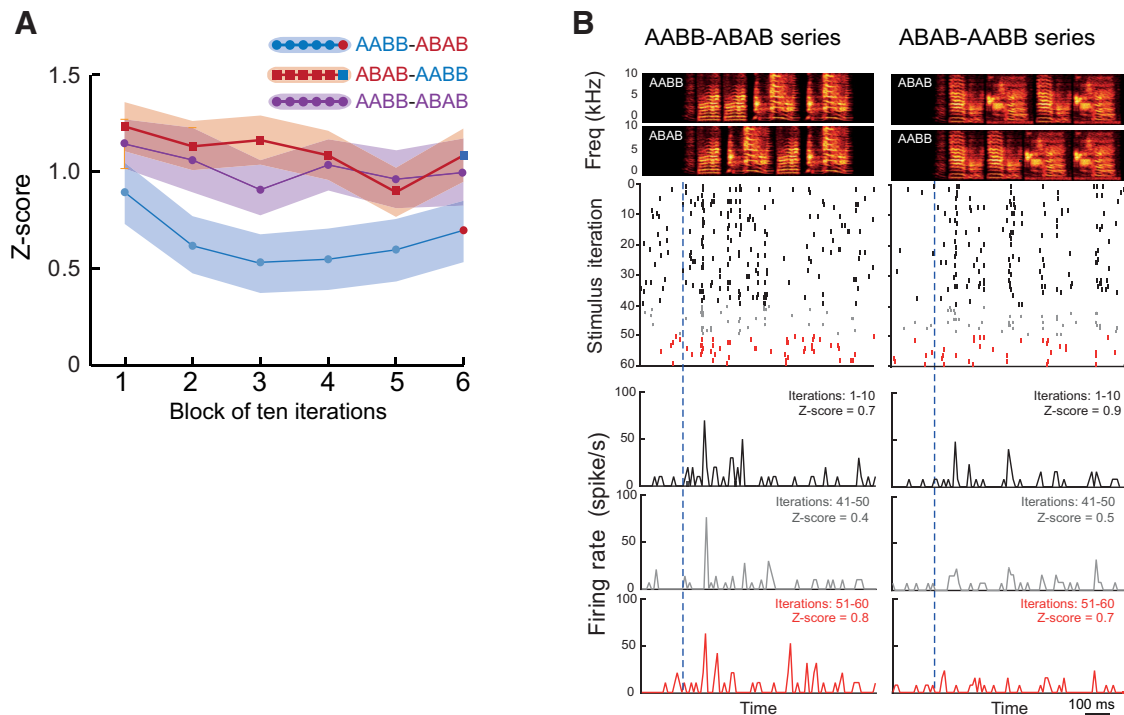


Figure 8. Responses driven by ABAB and AABB song stimuli in awake birds. **A**, Changes in multiunit activity across the successive blocks of AABB-ABAB (blue line) and ABAB-AABB (red line) series. A third series, the ABAB-ABAB series (purple line), was also played back. Thick line indicates mean. Shaded area represents SEM. **B**, Responses of a representative single unit (BS neuron) to the two structured songs. There is a reset in responses when the song structure changed from AABB to ABAB. Neuronal responses are shown as raster plots (middle, 60 iterations) and peristimulus histograms (bottom) that are time-aligned with song spectrograms (top). Spectrograms of songs used as stimuli: Top, During the first 50 iterations. Bottom, During the 10 last ones. Dotted blue lines indicate song onset. FR was averaged over the stimulus duration.

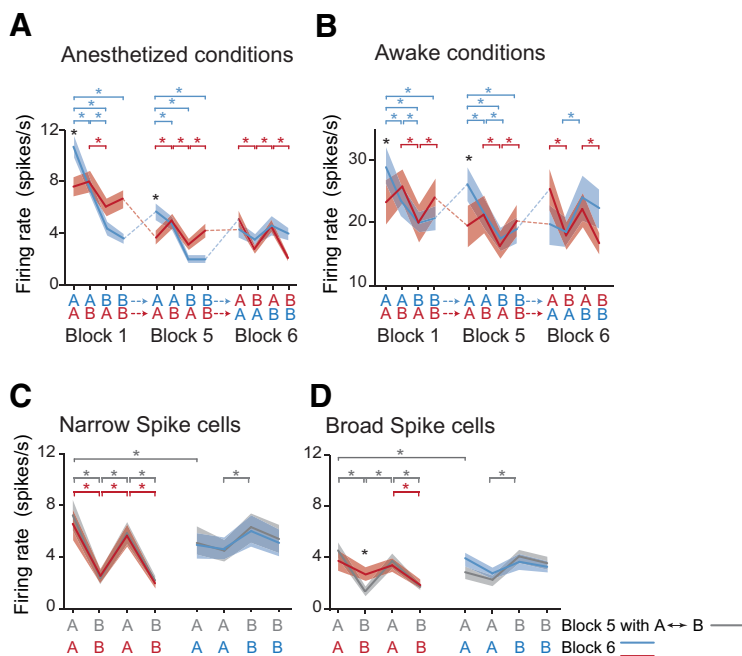


Figure 9. Variations in syllable-evoked FRs during song playbacks depend on song structure. In both anesthetized (**A**) and awake conditions (**B**), responses of the whole population of cells to the four song syllables of AABB (blue line) and ABAB (red line) songs. Thick line indicates mean. Shaded area represents SEM. Data collected before (blocks 1 and 5) and after (block 6) changes in song structure. There is a significant difference in response to the first syllable A between AABB and ABAB songs during blocks 1 and 5 (black $p < 0.001$) and the lack of this difference during block 6. Responses of NS (**C**) and BS (**D**) cells during blocks 5 and 6. Responses to the central syllables A and B during block 5 (in gray) are artificially permuted. Responses to AABB and ABAB songs are represented by blue and red, respectively. $*p < 0.05$, for the corresponding song structure (blue and red $*$ for AABB and ABAB songs, respectively).

show syllable-evoked responses according to the song structure used after the transition. Responses of NS cells during block 6 could be predicted by those elicited before permuting the central syllables (Fig. 9C). In contrast, BS cells showed a marked difference in their FR between blocks 5 and 6. More robust responses to the first syllable B after the transition from AABB to ABAB were evident (interaction between “block” and “song type” factors: $F_{(3,273)} = 3.11, p = 0.02$; block 5 vs 6; $p = 0.001$; Fig. 9D). In terms of the syllable pair AB that included the first syllable B, the FR of BS cells no longer differed between syllable A and B after transitions (Fig. 9D). These results clearly demonstrate that responses to song syllables did not strictly depend on acoustic identity of syllables but rather were influenced by the auditory context in which they occurred.

Responses of BS cells to the first song syllable A did not differ between AABB and ABAB songs after transitions (interaction between “block 5 and 6” and “song type” factors: $F_{(1,90)} = 7.8, p = 0.006$; block 5: $p < 0.001$; block 6: $p = 0.8$). In contrast, responses to the fourth syllable still differed between the two songs ($p < 0.001$). Responses to the first syllable A appeared to be influenced by the song structure as the

acoustic identity of both the first and the fourth syllables remained the same.

To evaluate to what extent the pattern of neuronal responses was affected by changes in song structure, we compared the precise timing of spike trains between two blocks of song presentations (block 4 vs 5 and block 5 vs 6) in awake birds. We computed the CorrCoef index (see Materials and Methods) from spike trains driven by each song syllable (Fig. 10A). CorrCoef values of ~ 0.2 were observed when spike trains recorded during blocks 4 and 5, before changing the song structure, were compared (Fig. 10B). These CorrCoef values represent a range usually seen for cortical neurons (Gaucher et al., 2013; Gaucher and Edeline, 2015). The CorrCoef values obtained for the second syllable reached a value close to 0 after altering the song structure (block 5 vs 6; Fig. 10C). This indicated a strong reduction in spike-timing reliability. Following transition from ABAB to AABB song, the second syllable of AABB songs (A) was followed by B in ABAB songs while the second syllable of ABAB songs (B) was preceded by two As in AABB songs. Therefore, changing the auditory context and particularly increasing the number of presentations of the syllable A before the syllable B had a strong impact on temporal reliability of responses to this syllable.

We also computed the CorrCoef index from spike trains driven by the syllable pair AB that included the first syllable B. This allowed us to examine whether the precise timing of APs elicited during this syllable pair was impacted by a change in the auditory context in which it occurred. We obtained CorrCoef values ~ 0.2 for the two songs before song transitions (block 4 vs 5; Fig. 10D). Following transition from one structure to the other (block 5 vs 6 and block 1 vs 6; Fig. 10D), CorrCoef values differed between the two transitions. The spike timing was less reliable after the transition from AABB to ABAB than from ABAB to AABB (block 1 vs 6, paired t test, $t_{(51)} = 2.70$, $p = 0.008$; block 5 vs 6, $t_{(51)} = 4.73$, $p < 0.001$; Fig. 10D). This, in addition to the results based on FR, indicate that the temporal reliability of neuronal discharges was markedly impacted by the transition from AABB to ABAB.

Discussion

Our results provide clear evidence that the neuronal responses of NCM neurons to playback of conspecific songs are sensitive to the song context in which syllables occur, and this sensitivity is expressed by changes in FRs and/or discharges patterns. The hypothetical outcomes of our study are summarized in Figure 11.

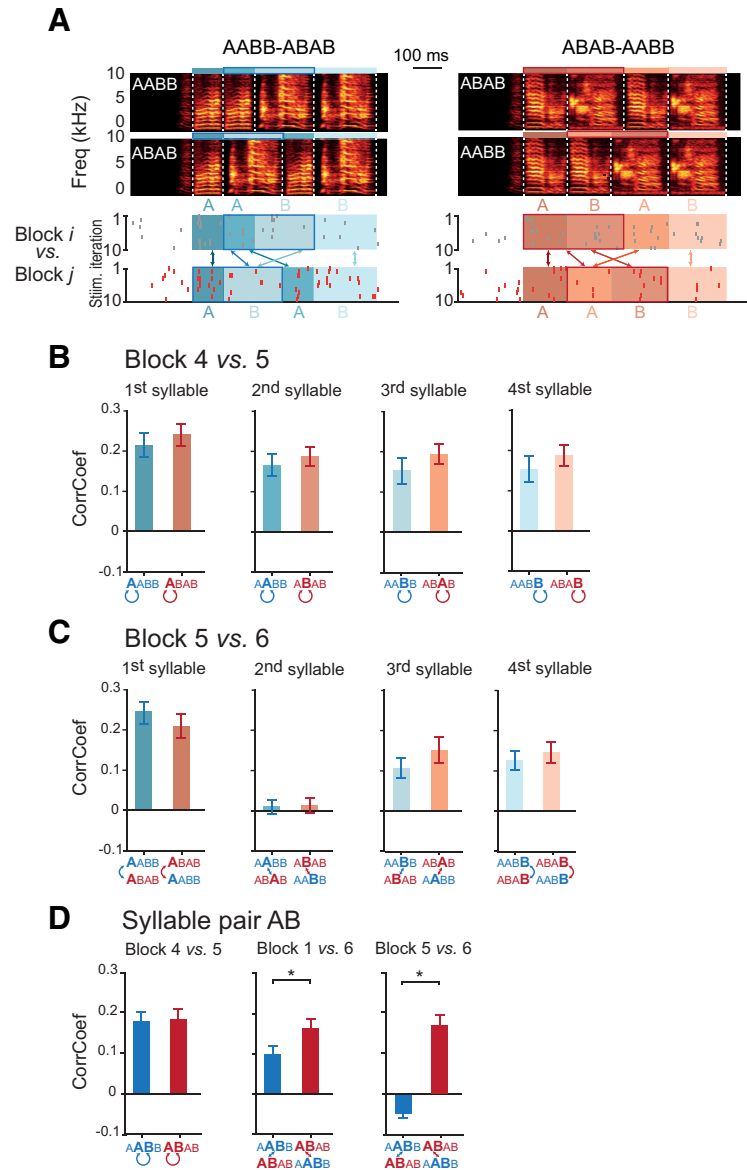


Figure 10. Effects of changes in song structure on temporal reliability of spike trains. **A**, Neuronal responses during two blocks of 10 song iterations are shown as raster plots that are time-aligned with song spectrograms. The CorrCoef index was computed by comparing the precise timing of spike trains between two blocks (*i* vs *j*) of song presentations (e.g., between iterations 1 to 10 of block *i* vs iterations 1 to 10 of block *j*). Computations were performed on spike trains elicited by individual syllables (AABB-ABAB and ABAB-AABB series: blue and red shaded areas on raster plots, respectively) or by the syllable pair AB that included the first syllable B (blue and red rectangles on raster plots). White dotted lines on spectrograms indicate individual syllables. Arrows above and below raster plots indicate the serial position of syllables before and after song transition, respectively. **B**, CorrCoef values (mean \pm SEM) for spike trains elicited by each of the four song syllables (in bold below histograms) of the AABB (in blue) and the ABAB (in red) songs during blocks 4 and 5. **C**, CorrCoef values for spike trains elicited by the same syllable (in bold below histograms) before (block 5) and after (block 6) changes in song structure. Arrows below histograms indicate the serial position of the syllable within the song before and after song transition. Moving the syllable from the second to the third position affected the temporal reliability of responses driven by this syllable. **D**, CorrCoef values for spike trains elicited by the syllable pair AB (in bold below histograms) during two blocks (4 and 5; 5 and 6; 1 and 6). * $p < 0.05$.

Sensitivity to the temporal order of syllables in conspecific songs

Our results show the repetition-induced decrease in responses, described by numerous studies both in mammals (Malmierca et al., 2014; Khouri and Nelken, 2015) and songbirds (Chew et al., 1995; Stripling et al., 1997; Phan et al., 2006; Beckers and Gahr, 2012; Smulders and Jarvis, 2013; Lu and Vicario, 2014). Once neuronal responses were decreased, the presentation of a novel song, generally sung by a distinct individual, resets the responses

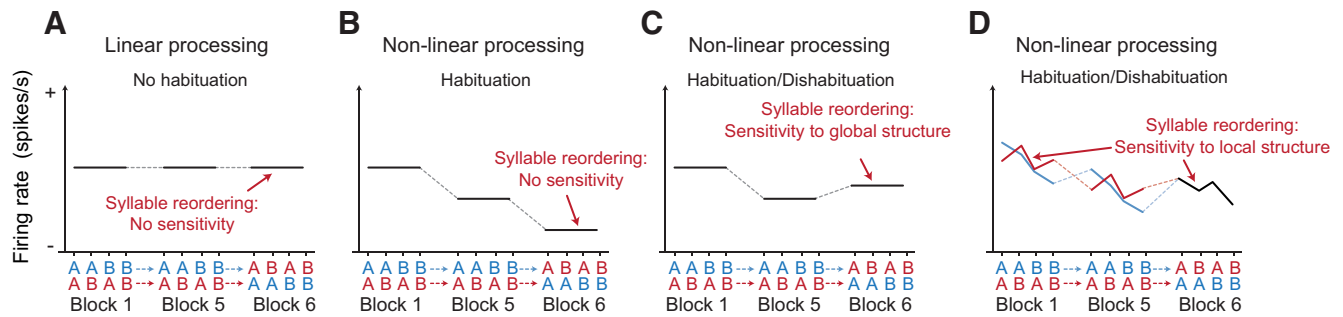


Figure 11. Hypothetical outcomes of our study. The different scenarios represent the responses of the whole population of cells to the presentation of AABB-ABAB and ABAB-AABB series during blocks 1, 5, and after changes in song structure (i.e., during block 6). **A**, Linearity of song responses. Neurons are only sensitive to certain acoustic features, and their responses are neither modulated by repetitions nor by changes in song structure (e.g., neurons of the songbird midbrain). **B–D**, Nonlinearity of song responses: repetition induces a decrease in responses. **B**, Neurons do not encode the temporal order of syllables within songs. **C**, Neurons encode the global structure of the song, syllables being grouped together as a “chunk.” Changing the song structure by permuting the two central syllables A and B reinstates stronger responses to all song syllables. **D**, Neurons encode the transitions between syllables. Changing the song structure by permuting the two central syllables A and B affects responses to these syllables. The sensitivity to local song structure is also detected in the initial responses during block 1. Responses to the first syllable B depend on the number of syllable A that precedes it. These responses are based on a set of results obtained for BS cells. NS cells did not show any changes in responses to song syllables after changes in song structure.

(Chew et al., 1995, 1996; Stripling et al., 1997; Smulders and Jarvis, 2013). This change in response magnitude represents a methodological approach to examine whether two stimuli that differ in certain aspects are encoded as similar or distinct auditory objects (Lu and Vicario, 2011; Moorman et al., 2011; Smulders and Jarvis, 2013). The reinstatement of strong responses when syllables were rearranged indicates that the temporal order of syllables within conspecific songs contributes to song identity.

To the best of our knowledge, no previous study has demonstrated that NCM neurons encode the sequential ordering of syllables within natural songs. Neurons in NCM have been previously found to be sensitive to the temporal order of sounds, for example, harmonic stacks in short or long streams (Lu and Vicario, 2014) and song elements in certain sequences of three elements (Ono et al., 2016). It was crucial to use conspecific songs as the activation of NCM neurons is largest when birds are exposed to this song type, as opposed to heterospecific songs or artificial stimuli (Mello and Clayton, 1994; Chew et al., 1996; Stripling et al., 1997; Van Meir et al., 2005). Behavioral studies have shown that the type of sounds used to build auditory stimuli and the way strings are segmented both influence the ability to detect changes in temporal sequences (Spierings et al., 2015; Neiworth et al., 2017).

Dependency on song context

Once the song stimulus has been repeated, larger responses to the novel arrangement of syllables are shown by BS cells rather than NS cells. Based on the width of the APs, these two cell types could correspond to excitatory (broad) and inhibitory (narrow) neurons (see the discussion in Schneider and Woolley, 2013). Our results therefore provide additional evidence that the two cell types differ in song processing (Meliza and Margoliash, 2012; Schneider and Woolley, 2013; Ono et al., 2016; Yanagihara and Yazaki-Sugiyama, 2016). Approximately three-fourths of BS cells showed increased responses when syllables were reordered; another subset of BS neurons ceased to fire at a given syllable when syllable order differed. A previous study reported that BS cells may fail to respond to a particular syllable included in a song but responded to this same syllable played in silence (Schneider and Woolley, 2013). By manipulating the temporal order of syllables in songs, we provide the first evidence that the responses to a given syllable depend on which syllables immediately precede it. As song responses in the NCM are nonlinear compared with song

responses in upstream auditory regions (Woolley, 2013), we assume that the extension of the nonlinear processing to contiguous syllables represents an emergent property of the NCM.

Syllables in songs were rearranged to alter transitions between adjacent syllables. In most cases, syllables were assigned to a different position than before (but see the example on Fig. 5C). Consequently, our song stimuli did not allow us to disentangle the following two possible interpretations for the response reinstatement: an encoding of either the transition between adjacent and nonadjacent syllables or the ordinal position of syllables in songs. However, both encodings require long-term integration of the sequential information that extends beyond individual syllables.

Sensitivity to grammatically structured songs

In both anesthetized and awake birds, NCM neurons (as a population) were sensitive to differences in song stimuli where two syllables were repeated twice or alternated. When the syllable B followed two As in AABB songs, the FR decreased. When the syllable B followed a single A in ABAB songs, the FR increased. Songbirds are behaviorally able to discriminate between differently structured sequences (Gentner et al., 2006; van Heijningen et al., 2009, 2013; Abe and Watanabe, 2011; Comins and Gentner, 2013, 2014). Our findings suggest that NCM may provide neural mechanisms to help discriminate between stimuli that differ in their sequential structure. We assume that the gradual decrease in FRs seen over the four syllables of AABB songs is a direct consequence of repetition of the first syllable. A study using XXY or YXX sequences (X and Y denoting song elements) as stimuli to examine responses in NCM of male zebra finches (Ono et al., 2016) found a gradual decrease over the three elements of XXY sequences (Ono et al., 2016, their Fig. 9).

Grammatically structured sequences have been used in many behavioral studies to determine whether nonhuman animals are able to learn the underlying structure of sequence sets (Berwick et al., 2011; Petkov and Wilson, 2012; Zaccarella and Friederici, 2017). This ability, which remains a matter of great debate (Beckers et al., 2017; Ghirlanda et al., 2017), was assessed by determining whether subjects detected the transition from one structure pattern to another or generalized the discrimination between stimuli to novel stimuli: in nonhuman primates (Fitch and Hauser, 2004; Uhrig et al., 2014; Wilson et al., 2015; Neiworth et al., 2017) and in songbirds (for review, see ten Cate, 2017). In our

study, we assessed detection of changes in grammatical structure by addressing this question at the neuronal level. Both the increase in responses of NCM neurons seen when AABB songs were rearranged into ABAB ones and the changes in temporal reliability of spike patterns seen when both transitions occurred provided support that encoding of relationships between syllables has occurred. However, most of the results highlight changes that are related to the first syllable B or the pair AB (which includes B), suggesting a detection of transitions based on local regularities/irregularities of songs rather than encoding of the global song structure (AABB vs ABAB). Behaviorally, zebra finches are able to generalize the discrimination between AABB and ABAB stimuli to novel song stimuli (van Heijningen et al., 2009). However, investigations have demonstrated that individuals do not learn the whole structure of the stimuli. Rather, they discriminate based on simpler differences, such as differences in the first (AA vs AB) or final two syllables (van Heijningen et al., 2013; Spierings and ten Cate, 2016), potentially by focusing on local transitions between two syllables (van Heijningen et al., 2009, 2013; Seki et al., 2013; Chen and ten Cate, 2015; Ghirlanda et al., 2017). Therefore, our experiments provide insights into the neuronal underpinnings of the behavioral ability to distinguish between AABB and ABAB structures, which has hitherto been unexplored at a cellular level in the avian and mammalian auditory system.

Despite the limited sensitivity of neurons to local song structure, NCM neurons could extend their encoding beyond contiguous syllables, as previously suggested (Lu and Vicario, 2014). The difference in BS cell responses to the first syllable A in AABB and ABAB songs (Fig. 9A,B), and the loss of response with a change in song structure supports this assumption.

Based on evidence provided by both experiments, our study demonstrates that an avian brain auditory region shows sensitivity to the sequential organization of sound elements in songs. Until recently, such sensitivity has been found mostly in the sensorimotor nucleus HVC, using the bird's own song as stimulus (Margoliash, 1986; Margoliash and Fortune, 1992; Lewicki and Konishi, 1995). The primary auditory area known as Field L shows less sensitivity than HVC neurons to manipulations in the order of syllables (Lewicki and Arthur, 1996). Sensitivity to the sequential order in which sounds occur could be a functional property that characterizes high-level auditory regions. However, very few studies have examined neuronal responses to changes in the ordering relationships of elements within sequences by recording single-unit activity in the auditory cortex of mammalian species (Weinberger and McKenna, 1988; McKenna et al., 1989; Kikuchi et al., 2017).

In conclusion, the present study provides new insights into properties of high-order auditory neurons by showing sensitivity to sequential organization of natural communication signals. Our findings establish the songbird as a model system for deciphering basic neural processes by which, not only ordering relationships, but also the global structure can be encoded.

References

- Abe K, Watanabe D (2011) Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nat Neurosci* 14:1067–1074.
- Beckers GJ, Gahr M (2010) Neural processing of short-term recurrence in songbird vocal communication. *PLoS One* 5:e11129.
- Beckers GJ, Gahr M (2012) Large-scale synchronized activity during vocal deviance detection in the zebra finch auditory forebrain. *J Neurosci* 32:10594–10608.
- Beckers GJ, Berwick RC, Okanoya K, Bolhuis JJ (2017) What do animals learn in artificial grammar studies? *Neurosci Biobehav Rev* 81:238–246.
- Bekinschtein TA, Dehaene S, Rohaut B, Tadel F, Cohen L, Naccache L (2009) Neural signature of the conscious processing of auditory regularities. *Proc Natl Acad Sci U S A* 106:1672–1677.
- Berwick RC, Okanoya K, Beckers GJ, Bolhuis JJ (2011) Songs to syntax: the linguistics of birdsong. *Trends Cogn Sci* 15:113–121.
- Briefer EF, Rybak F, Aubin T (2013) Does true syntax or simple auditory object support the role of skylark song dialect? *Anim Behav* 86:1131–1137.
- Chen J, ten Cate C (2015) Zebra finches can use positional and transitional cues to distinguish vocal element strings. *Behav Process* 117:29–34.
- Chew SJ, Mello C, Nottebohm F, Jarvis E, Vicario DS (1995) Decrements in auditory responses to a repeated conspecific song are long-lasting and require two periods of protein synthesis in the songbird forebrain. *Proc Natl Acad Sci U S A* 92:3406–3410.
- Chew SJ, Vicario DS, Nottebohm F (1996) A large-capacity memory system that recognizes the calls and songs of individual birds. *Proc Natl Acad Sci U S A* 93:1950–1955.
- Clucas BA, Freeberg TM, Lucas JR (2004) Chick-a-dee call syntax, social context, and season affect vocal responses of Carolina chickadees (*Poecile carolinensis*). *Behav Ecol Sociobiol* 57:187–196.
- Comins JA, Gentner TQ (2010) Working memory for patterned sequences of auditory objects in a songbird. *Cognition* 117:38–53.
- Comins JA, Gentner TQ (2013) Perceptual categories enable pattern generalization in songbirds. *Cognition* 128:113–118.
- Comins JA, Gentner TQ (2014) Auditory temporal pattern learning by songbirds using maximal stimulus diversity and minimal repetition. *Anim Cogn* 17:1023–1030.
- Dehaene S, Meyniel F, Wacongne C, Wang L, Pallier C (2015) The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron* 88:2–19.
- Doupe AJ (1997) Song- and order-selective neurons in the songbird anterior forebrain and their emergence during vocal development. *J Neurosci* 17:1147–1167.
- Fitch WT, Hauser MD (2004) Computational constraints on syntactic processing in a nonhuman primate. *Science* 303:377–380.
- Friederici AD, Bahlmann J, Heim S, Schubotz RI, Anwander A (2006) The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proc Natl Acad Sci U S A* 103:2458–2463.
- Gaucher Q, Edeline JM (2015) Stimulus-specific effects of noradrenaline in auditory cortex: implications for the discrimination of communication sounds. *J Physiol* 593:1003–1020.
- Gaucher Q, Huetz C, Gourévitch B, Edeline JM (2013) Cortical inhibition reduces information redundancy at presentation of communication sounds in the primary auditory cortex. *J Neurosci* 33:10713–10728.
- Gentner TQ (2008) Temporal scales of auditory objects underlying bird-song vocal recognition. *J Acoust Soc Am* 124:1350–1359.
- Gentner TQ, Fenn KM, Margoliash D, Nusbaum HC (2006) Recursive syntactic pattern learning by songbirds. *Nature* 440:1204–1207.
- Ghirlanda S, Lind J, Enquist M (2017) Memory for stimulus sequences: a divide between humans and other animals? *R Soc Open Sci* 4:161011.
- Holland J, Dabelsteen T, Paris AL (2000) Coding in the song of the wren: importance of rhythmicity, syntax and element structure. *Anim Behav* 60:463–470.
- Khouri L, Nelken I (2015) Detecting the unexpected. *Curr Opin Neurobiol* 35:142–147.
- Kiggins JT, Comins JA, Gentner TQ (2012) Targets for a comparative neurobiology of language. *Front Evol Neurosci* 4:6.
- Kikuchi Y, Attaheri A, Wilson B, Rhone AE, Nourski KV, Gander PE, Kovach CK, Kawasaki H, Griffiths TD, Howard MA 3rd, Petkov CI (2017) Sequence learning modulates neural responses and oscillatory coupling in human and monkey auditory cortex. *PLoS Biol* 15:e2000219.
- Kojima S, Doupe AJ (2008) Neural encoding of auditory temporal context in a songbird basal ganglia nucleus, and its independence of birds' song experience. *Eur J Neurosci* 27:1231–1244.
- Lewicki MS, Arthur BJ (1996) Hierarchical organization of auditory temporal context sensitivity. *J Neurosci* 16:6987–6998.
- Lewicki MS, Konishi M (1995) Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order. *Proc Natl Acad Sci U S A* 92:5582–5586.
- Lipkind D, Marcus GF, Bemis DK, Sasahara K, Jacoby N, Takahashi M, Suzuki K, Feher O, Ravbar P, Okanoya K, Tchernichovski O (2013) Stepwise

- acquisition of vocal combinatorial capacity in songbirds and human infants. *Nature* 498:104–108.
- Lu K, Vicario DS (2011) Toward a neurobiology of auditory object perception: what can we learn from the songbird forebrain? *Curr Zool* 57:671–683.
- Lu K, Vicario DS (2014) Statistical learning of recurring sound patterns encodes auditory objects in songbird forebrain. *Proc Natl Acad Sci U S A* 111:14553–14558.
- Malmierca MS, Sanchez-Vives MV, Escera C, Bendixen A (2014) Neuronal adaptation, novelty detection and regularity encoding in audition. *Front Syst Neurosci* 8:111.
- Margoliash D (1983) Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. *J Neurosci* 3:1039–1057.
- Margoliash D (1986) Preference for autogenous song by auditory neurons in a song system nucleus of the white-crowned sparrow. *J Neurosci* 6:1643–1661.
- Margoliash D, Fortune ES (1992) Temporal and harmonic combination-sensitive neurons in the zebra finch's HVC. *J Neurosci* 12:4309–4326.
- McKenna TM, Weinberger NM, Diamond DM (1989) Responses of single auditory cortical neurons to tone sequences. *Brain Res* 481:142–153.
- Meliza CD, Margoliash D (2012) Emergence of selectivity and tolerance in the avian auditory cortex. *J Neurosci* 32:15158–15168.
- Mello CV, Clayton DF (1994) Song-induced ZENK gene expression in auditory pathways of songbird brain and its relation to the song control system. *J Neurosci* 14:6652–6666.
- Mello C, Nottebohm F, Clayton D (1995) Repeated exposure to one song leads to a rapid and persistent decline in an immediate early gene's response to that song in zebra finch telencephalon. *J Neurosci* 15:6919–6925.
- Menardy F, Touiki K, Dutrieux G, Bozon B, Vignal C, Mathevon N, Del Negro C (2012) Social experience affects neuronal responses to male calls in adult female zebra finches. *Eur J Neurosci* 35:1322–1336.
- Menardy F, Giret N, Del Negro C (2014) The presence of an audience modulates responses to familiar call stimuli in the male zebra finch forebrain. *Eur J Neurosci* 40:3338–3350.
- Moorman S, Mello CV, Bolhuis JJ (2011) From songs to synapses: molecular mechanisms of birdsong memory. *BioEssays* 33:377–385.
- Neiworth JJ, London JM, Flynn MJ, Rupert DD, Alldritt O, Hyde C (2017) Artificial grammar learning in tamarins (*Saguinus oedipus*) in varying stimulus contexts. *J Comp Psychol* 131:128–138.
- Ono S, Okanoya K, Seki Y (2016) Hierarchical emergence of sequence sensitivity in the songbird auditory forebrain. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 202:163–183.
- Petkov CI, Wilson B (2012) On the pursuit of the brain network for proto-syntactic learning in non-human primates: conceptual issues and neurobiological hypotheses. *Philos Trans R Soc Lond B Biol Sci* 367:2077–2088.
- Phan ML, Pytte CL, Vicario DS (2006) Early auditory experience generates long-lasting memories that may subserve vocal learning in songbirds. *Proc Natl Acad Sci U S A* 103:1088–1093.
- Ribeiro S, Cecchi GA, Magnasco MO, Mello CV (1998) Toward a song code: evidence for a syllabic representation in the canary brain. *Neuron* 21:359–371.
- Schneider DM, Woolley SM (2013) Sparse and background-invariant coding of vocalizations in auditory scenes. *Neuron* 79:141–152.
- Seki Y, Suzuki K, Osawa AM, Okanoya K (2013) Songbirds and humans apply different strategies in a sound sequence discrimination task. *Front Psychol* 4:447.
- Smulders TV, Jarvis ED (2013) Different mechanisms are responsible for dishabituation of electrophysiological auditory responses to a change in acoustic identity than to a change in stimulus location. *Neurobiol Learn Mem* 106:163–176.
- Spierings MJ, ten Cate C (2016) Zebra finches as a model species to understand the roots of rhythm. *Front Neurosci* 10:345.
- Spierings M, de Weger A, ten Cate C (2015) Pauses enhance chunk recognition in song element strings by zebra finches. *Anim Cogn* 18:867–874.
- Stripling R, Volman SF, Clayton DF (1997) Response modulation in the zebra finch neostriatum: relationship to nuclear gene regulation. *J Neurosci* 17:3883–3893.
- Suzuki TN, Wheatcroft D, Griesser M (2016) Experimental evidence for compositional syntax in bird calls. *Nat Commun* 7:10986.
- Suzuki TN, Wheatcroft D, Griesser M (2017) Wild birds use an ordering rule to decode novel call sequences. *Curr Biol* 27:2331–2336.e3.
- ten Cate C (2017) Assessing the uniqueness of language: animal grammatical abilities take center stage. *Psychon Bull Rev* 24:91–96.
- Uhrig L, Dehaene S, Jarraya B (2014) A hierarchy of responses to auditory regularities in the macaque brain. *J Neurosci* 34:1127–1132.
- van Heijningen CA, de Visser J, Zuidema W, ten Cate C (2009) Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proc Natl Acad Sci U S A* 106:20538–20543.
- van Heijningen CA, Chen J, van Laatum I, van der Hulst B, ten Cate C (2013) Rule learning by zebra finches in an artificial grammar learning task: which rule? *Anim Cogn* 16:165–175.
- Van Meir V, Boumans T, De Groof G, Van Audekerke J, Smolders A, Scheunders P, Sijbers J, Verhoye M, Balthazart J, Van der Linden A (2005) Spatiotemporal properties of the BOLD response in the songbirds' auditory circuit during a variety of listening tasks. *Neuroimage* 25:1242–1255.
- Wakita M (2019) Auditory sequence perception in common marmosets (*Callithrix jacchus*). *Behav Process* 162:55–63.
- Weinberger NM, McKenna TM (1988) Sensitivity of single neurons in auditory cortex to contour: toward a neurophysiology of music perception. *Music Perception* 5:355–389.
- Wilson B, Kikuchi Y, Sun L, Hunter D, Dick F, Smith K, Thiele A, Griffiths TD, Marslen-Wilson WD, Petkov CI (2015) Auditory sequence processing reveals evolutionarily conserved regions of frontal cortex in macaques and humans. *Nat Commun* 6:8901.
- Wilson B, Marslen-Wilson WD, Petkov CI (2017) Conserved sequence processing in primate frontal cortex. *Trends Neurosci* 40:72–82.
- Woolley SM (2013) The songbird auditory system. In: *Animal models of speech and language disorders* (Helekar SA, ed), pp 61–88. New York: Springer.
- Yanagihara S, Yazaki-Sugiyama Y (2016) Auditory experience-dependent cortical circuit shaping for memory formation in bird song learning. *Nat Commun* 7:11946.
- Zaccarella E, Friederici AD (2017) The neurobiological nature of syntactic hierarchies. *Neurosci Biobehav Rev* 81:205–212.