Intrinsic dynamics enhance decodability of neurons in a model of avian auditory cortex

Margot C Bjoring

Department of Psychology,

University of Virginia, Charlottesville, VA

January 5, 2018

Abstract

Birdsong is a complex vocalization that bears important similarities to human speech. Critical to recognizing speech or birdsong is the ability to discriminate between similar sequences of sound that may carry different meanings. The caudal mesopallium (CM) is a secondary area in the auditory system of songbirds that is a potential site for song identification, displaying both between-category selectivity and within-category tolerance to conspecific song. Electrophysiological studies of CM have identified a population of neurons with intrinsically phasic firing patterns in addition to the more typical tonic and fast-spiking neurons. The function of these phasic neurons in processing spectrotemporally complex conspecific vocalizations is not known. We investigated the auditory response properties of phasic and tonic neurons using computational modeling with particular focus on the selectivity and entropy of the simulated responses to birdsong. When biophysical models of phasic and tonic neurons were presented with identical inputs, the phasic models were more selective among syllables and more robust to noise-induced variability, potentially providing an advantage for song identification. Additionally, the overall responsiveness of a model to the stimulus set determined which decoding metric better captured the coding strategy of the model's response. The relationships between measures of decodability found in the model simulations are consistent with extracellular data from zebra finch CM.

Introduction

Auditory Processing

The auditory processing of speech presents a challenging problem that the human auditory system solves with ease. Noisy acoustic environments and speaker-to-speaker variability are just a few of the complications involved in decoding a speech stream. Mammalian models of audition have uncovered key features of auditory cortex such as tonotopic organization [1], feedforward inhibition to sharpen the fine temporal structures of sound [2], and even evidence for harmonic connections across octaves [3]. The ability to extend rodent models to the processing of vocalizations with the temporal and spectral complexity of speech, however, is limited due to the relatively simple and innate vocalizations produced by rodents. In fact, with the exception of cetaceans and bats, mammalian vocalizations do not require auditory experience to produce. The songbird (*Passeriformes*), while a very distant relative of humans and possessing a different vocal apparatus called a syrinx, nevertheless displays many of the vocal traits characteristic of human speech, including complex, learned vocalizations.

Songbird models

Songbirds have generated substantial interest as a model for studying the vocal production and auditory processing of speech. Singing is used to attract mates, strengthen pair bonds, and defend territory [4]. Although

many songbirds inherit a template of their species-appropriate song, which may help juveniles identify suitable tutors, the songs themselves must be learned by memorizing the song of an adult tutor and subsequently practicing vocalizations in an attempt to match the memorized tutor song [5]. In zebra finches (*Taeniopygia guttata*), a popular model for studying language, juveniles deafened prior to song exposure or raised in isolation from a tutor fail to acquire an organized song [6], and juveniles raised with a heterospecific tutor will often attempt to incorporate the content of the tutor's song into their inherent template [7].

Like humans, zebra finches exhibit a critical period for acquiring song, from around 15 days post-hatch (dph) when brainstem auditory responses mature [8] to 60-90 dph [5]. A number of factors can extend the closure of the critical period, including isolation from a suitable tutor [9]. Zebra finches learn a single song, and after the closure of the critical period, this song is crystalized and will not change throughout their life [5]. Other songbirds, like European starlings (*Sturnus vulgaris*), are open-ended learners who can add to their repertoire of songs even in adulthood [10].

The development of song production is the most studied aspect of the critical period, but there is also concomitant development of the auditory system as juveniles learn to hear and identify song. In humans, infants go through well-defined stages of auditory learning including statistical learning of sound patterns leading to categorical perception of language-specific sounds and reduced discrimination of sounds not in their

language [5]. Research in starlings has shown that they are capable of statistical learning of regularities in continuous sound streams [11]. Evidence for categorical perception has been shown for conspecific song notes in zebra finches [12] and for learned vowel sounds in starlings [13]. Auditory experience in development also influences the responses of auditory neurons to song in adulthood [14]. Further research will be necessary to fully explain the developmental stages of the auditory system in juvenile songbirds.

Songbird auditory pathways

The songbird auditory system from the cochlea to the auditory thalamus (nucleus ovoidalis; Ov) is highly consistent with the mammalian auditory pathway [15]. The avian brain lacks a six-layered cortex; the pallium is instead organized into clusters of neurons forming nuclei. The homology of the pallial auditory regions to mammalian auditory cortex has been a matter of debate, although recent studies have identified genetic and functional similarities. Dugas-Ford *et al.* (2012) [16] found conserved cell types among mammals, birds, and reptiles for the layer 4 input and layer 5 output cells of the cortex despite the different architecture of avian and reptilian brains. There is evidence of laminar and columnar organization within the avian auditory forebrain along the dorsorostral-ventrocaudal plane [17]. The avian auditory pallium also shows a marked preference for natural stimuli such as birdsong over artificial stimuli like white noise and

pure tones. The mesencephalicus lateralis dorsalis (MLd), a midbrain auditory nucleus akin to the inferior colliculus in mammals, responds robustly to pure-tone stimulation [18], but at the level of the auditory forebrain the preference for natural sounds or synthetic sounds with statistics that mimic natural sounds emerges [19] [20]. The mammalian auditory system shows a similar emergence of a preference for natural stimuli from midbrain to cortex [21].

Field L2a is the primary thalamorecipient area in the avian auditory forebrain, with downstream areas L1, L3, and L2b. These areas have reciprocal connections with each other and also with the higher-order areas caudomedial nidopallium (NCM) and caudal mesopallium (CM) [22]. Although all of these areas communicate either directly or indirectly with each other, two primary streams emerge from Field L. L3 to NCM is one, and L1 and L2b to CM is the other. More research is needed to determine the functional differences between these two streams of information. NCM and CM are the highest areas in the songbird auditory pathway and may be analogous to supragranular layers of A1 or secondary auditory areas in mammals [23]. Given their position in the auditory hierarchy, it is likely that these areas are responsible for song learning and recognition, and recent research has supported this idea.

NCM is a potential location for the memory of the tutor song that juvenile birds base their own songs on. Immediate early gene expression in NCM when zebra finches are presented with their tutor song is correlated

with the degree of copying between the bird's own song and the tutor song [24]. The strength of song learning is also correlated with the familiarity of the tutor song in NCM as measured by the rate of accommodation of a neural response to auditory stimulation [25]. CM is not involved in the tutor song memory but does play a role in the learning of other conspecific songs. Jeanne *et al.* (2011) [26] showed that learned songs are more effectively encoded by CM neurons than novel songs and that rewarded songs were better encoded than unrewarded songs, indicating not just a bias toward learned songs but toward behaviorally-relevant songs. Meliza and Margoliash (2012) [27] found that the response to within-song variability is an important difference between NCM and CM; NCM shows sensitivity to performance-to-performance differences in a song, while CM is tolerant to these differences.

Current study and its motivation

The tolerance of CM for within-song variability and its preferential response to behaviorally relevant stimuli make it a potential site for the decoding of song identity. In human language, there are meaningful differences between words that can completely change the meaning of an utterance as well as non-meaningful differences in the pronunciation of a single word. The same is true of birdsong: there are variations between performances of a song that a bird must recognize as coding for the same identity, and there are also birds with highly similar songs (e.g., siblings or

a tutor and pupil). Based on its position in the auditory system and its response properties, CM is well positioned to produce this kind of discrimination. The ultimate goal of a birdsong model of language is to explain not only what higher-order areas do but how they do it, and a mechanistic explanation must start at the cell level.

Electrophysiological studies of the broad-spiking, putatively excitatory, cell class within CM by Chen and Meliza (2017) [28] has revealed three distinct cell types within this class based on response properties to current stimulation: tonic, intermediate, and phasic. Tonic neurons are similar to the regular-spiking neurons seen in auditory cortex but show less regularity and higher adaptation rates. Phasic neurons fire only once or a few times regardless of the level and extent of stimulation and are the result of a 4AP-sensitive low-threshold potassium current. This type of firing pattern is not seen in adult mammalian auditory cortex, though it has been observed in juveniles [29] and lower levels of the mammalian auditory system [30]. Intermediate neurons respond tonicly at some levels of stimulation and phasicly at others.

The presence of a phasicly responding neuron in an area of the avian auditory forebrain involved in decoding song identity has interesting implications about the role such neurons might play in addressing some of the complications of auditory processing like noisy acoustic environments and song-to-song variability. In this study, we explore the functional significance of phasic neurons in CM using a modeling approach and test

the hypothesis that phasic neurons may possess an encoding advantage over tonic neurons that make them more informative and less affected by the presence of noise, thereby enhancing the ability of CM to determine the identity of a song stimulus. We then assess the validity of our model's predictions by comparing the results of our model to extracellular data from zebra finch CM. Identifying the functional roles of the cell types of CM is the first step toward understanding the circuit and being able to model the computations required to go from sequences of frequencies to an identifiable, meaningful vocalization.

Methods

Animals

All animal use was performed in accordance with the Institutional Animal Care and Use Committee of the University of Virginia. Adult zebra finches were obtained from the University of Virginia breeding colony. Thirty male zebra finches provided song recordings that were used as stimuli in the simulation experiments. During recording, zebra finches were housed in a soundproof auditory isolation box (Eckel Industries) with *ad libitum* food and water and were kept on a 16:8h light:dark schedule. A mirror was added to the box to stimulate singing. A typical recording session lasted 2-3 days. Birds were returned to the main colony after song recording.

Simulation

Neuron model. The model used in this study is a conductance-based, single-compartment model of CM neurons. The model, based on the ventral cochlear nucleus model of Rothman and Manis (2003) [31], relates the voltage dynamics of a single neuron to currents associated with ion channels. The model used in this study includes 4 voltage-gated potassium and sodium currents, a leak current, and a hyperpolarization activated ion current [28]. The model neuron exhibits a depolarization block to strong currents and a sustained response to weak currents. The model parameter values follow Rothman and Manis (2003) [31] with a few adjustments for resting potential and spike threshold for CM neurons. The calculations presented here used the consensus model parameters from Chen and Meliza (2017) [28] for tonic and phasic cells.

Auditory response simulation. To simulate an auditory response, $I_{stim}(t)$ becomes the convolution of a spectrotemporal receptive field (RF) with a spectrogram of an auditory stimulus. $I_{noise}(t)$ is randomly generated pink noise ($1/f$ distribution) low-pass filtered at 100Hz and scaled relative to the signal to achieve a set signal-to-noise ratio (SNR).

Auditory stimuli are 30 zebra finch songs recorded from our colony. All songs were cut to 2.025s long with 50ms of silence at the beginning to pad the convolution, high-pass filtered at 500Hz with a 4th order Butterworth filter, and scaled to a consistent RMS amplitude. Start and end times of syllables were identified by visual inspection. Repeated

syllables were grouped in the decoding analyses.

RFs were constructed with a Gabor filter based on Woolley *et al.* (2009) [32]:

$$
\begin{aligned}
\mathrm{RF}(t, f) &= H(t) \cdot G(f), \\
H(t) &= e^{-0.5[(t-t_0)/\sigma_t]^2} \cdot \cos(2\pi \cdot \Omega_t(t - t_0) + P_t), \\
G(f) &= e^{-0.5[(f-f_0)/\sigma_f]^2} \cdot \cos(2\pi \cdot \Omega_f(f - f_0)),
\end{aligned}
$$

where H is the temporal dimension of the RF, G is the spectral dimension of the RF, t_0 is the latency, f_0 is the peak frequency, σ_t and σ_f are the temporal and spectral bandwidths, Ω_t and Ω_f are the temporal and spectral modulation frequencies, and P_t is the temporal phase. Parameter values were randomly drawn from distributions set so as to match the modulation transfer function (MTF) of the RF ensemble to the MTF of zebra finch song [33] [32] (Figure 1). The integral of each RF was normalized to one.

In the context of this simulation, a model neuron is a combination of one RF and one model dynamic (phasic or tonic). 60 RFs were generated to produce paired phasic and tonic simulations, and 15 of the RFs were excluded due to MTF values outside the reported distribution of RFs in zebra finch neurons [32] (N = 90 neurons or 45 pairs). The 30 zebra finch songs were presented 10 times each to each neuron with random pink noise producing trial-to-trial variability. Pink noise sets were identical between paired phasic and tonic neurons. The total amplitude of the convolution

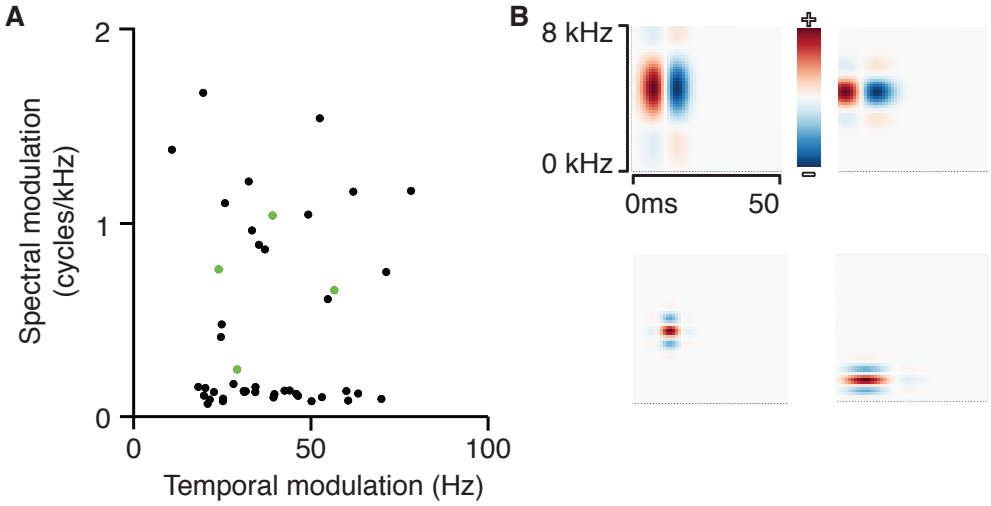

Figure 1: Receptive field parameter distributions. **A**, Combinations of the temporal modulations and spectral modulation parameters used to construct the RFs used in this study. The parameter values were drawn randomly from parameter distributions inferred from experimental data. Values outside the range of reported RFs (temporal modulation > 100Hz or spectral modulation > 2 cycles/kHz) were excluded. The points colored in green are the RFs shown in B. **B**, Examples of 4 of the 45 RFs used in this study.

was normalized by the bandwidth of the RF on the frequency axis (σ_f) to account for the differences in amplitudes between narrowband and broadband RFs. The output of the model was a simulated voltage trace from which spike times were extracted.

Data analysis. Spike times were extracted from the simulated responses. The classification analysis was performed by computing the van Rossum distance [34] (as implemented in neo: http://neo.readthedocs.io/en/0.5.2/) between every pair of spike trains for a model neuron ($n = 300$). We considered multiple time-scales for the τ parameter of the van Rossum distance from 5 to 45ms. A k-means clustering algorithm assigned spike trains to clusters based on their proximity in high-dimensional space. Cluster identity was assigned by a voting scheme as described in Schneider and Woolley (2010) [35] with each spike train casting a vote for its corresponding song. The proportion of correctly clustered spikes for each neuron determined its percent correct value.

We calculated spike rate, $r_{i,j}$, as the number of spikes evoked by syllable i in trial j, divided by the duration of the syllable. Selectivity was quantified using activity fraction [36] [27], a nonparametric index defined as:

$$A = \frac{1 - (\Sigma r_i/N)^2/\Sigma r_i^2/N}{1 - 1/N}$$

where r_i is the rate for syllable i averaged across trials, and N is the total

number of syllables.

Mutual information (MI), response entropy, and noise entropy were calculated following Jeanne et al. (2011) [26]. Response rates were discretized into 15 bins between 0 Hz and the maximum rate of the model. Response (total) entropy was calculated as $H(R) = -\Sigma p(r) \log_2 p(r)$, noise entropy as $H(R|S) = -\Sigma p(s) \Sigma p(r|s) \log_2 p(r|s)$, and mutual information as $I(R;S) = H(R) - H(R|S)$, where r is the rate and s is the syllable. Because of the large number of stimuli and trials, and because we were interested in differences between models presented with exactly the same stimuli, we did not correct entropy or MI for sample size bias.

Extracellular data

Analyses based on extracellular data were performed on the publicly available dataset from Theunissen *et al.* [37] on CRCNS.org. Neural recordings were collected from adult male zebra finches as described in Gill *et al.* [38]. Only cells from CM stimulated with conspecific song were used these analyses ($n = 37$). Selectivity and MI analyses were performed as described above with the exception that 10 response bins were used for MI instead of 15 due to a smaller stimulus set.

Results

To explore the consequences of the intrinsic membrane properties giving rise to phasic and tonic response dynamics in terms of the functional role of the neurons in the auditory processing of song, we use the neuron model described in Chen and Meliza (2017) [28], which replicates the observed phasic and tonic behaviors through the adjustment of the low-threshold potassium current parameter of the model. Auditory response is simulated by setting the current stimulation parameter (I_{stim}) to the normalized convolution of the spectrogram of a zebra finch song and a receptive field constructed from Gabor filters (Figure 2A). Variability in the response is achieved by adding pink noise ($1/f$ spectrum) to the convolution with a signal-to-noise ratio of 4.

Input-matched phasic and tonic neurons produce distinct spiking responses. In general, phasic neurons show reduced variation in spike times and spike numbers to a given syllable of a song (Figure 2B-C). The increased consistency of the responses of phasic neurons indicates an advantage for the decodability of the neural signal. We quantified this effect using several different measures of coding efficiency.

Temporal-based coding. A temporal code uses the pattern of spike times to encode the identity of a signal. An efficient temporal code represents different stimuli with distinguishable patterns of spikes and has high temporal precision across multiple trials of the same stimulus. Because

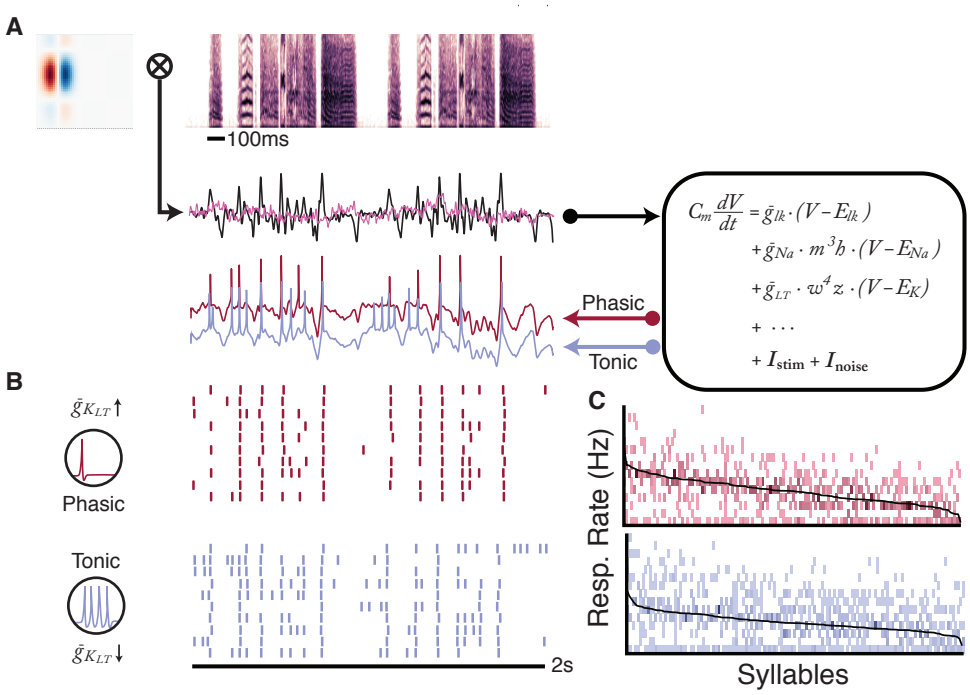

Figure 2: Data simulation and analysis pipeline. **A**, Auditory responses can be simulated through the convolution of a spectrotemporal receptive field (upper left) with a spectrogram (upper right) of an auditory stimulus, in this case a zebra finch song. The resulting convolution (black line) provides the driving current (I_{stim}) of the biophysical model used in this study (right). Low-pass filtered pink noise (pink line) adds variability to the driving current (I_{noise}). The output of the model is a simulated voltage trace (lower left) which can have either phasic (red line) or tonic (blue line) response properties depending on the conductance of a low-threshold potassium channel parameter ($g_{K_{LT}}$: 0 nS or 100 nS for tonic and phasic respectively). **B**, Raster plots of the full simulation for the stimulus-RF pair in A across 10 trials for phasic (red) and tonic (blue) model. The example demonstrates the increased variability in spike number and decreased temporal precision for the tonic model as compared to the phasic model. **C**, Full response distribution for the example neuron. Response rates are calculated per syllable in each song and divided into 15 bins. The black line indicates the average response rate across the syllables and the spread of response rate bins around that line show the trial-to-trial variability of the response rate.

the timescale used in the decoding of a temporal code substantially affects the results, we considered multiple timescales when analyzing the temporal decodability of the simulated neural responses. Figure 3 shows the results of a classification analysis using a k-means clustering approach on the van Rossum distance of each pair of spike trains, calculated at multiple time constants.

Although both groups perform well above chance, the phasic neuron models show clear separation from tonic models in terms of discriminability of temporal codes at all time constants examined, indicating that the neural signal produced by phasic neurons is more temporally precise and distinct than that produced by tonic neurons. Phasic responses are also less sensitive to the time constant used, showing high discriminability at both short and long time constants, in contrast to tonic responses, which show much steeper drop-offs on either side of their ideal time constant.

Rate-based coding. A rate-based code uses the average firing rate across a stimulus to encode identity. The precise timing of spikes matters less than the total excitation of the neuron across a given period of time. Two of the most widely applied rate-based decoding methods in sensory neuroscience are mutual information and selectivity, and these are the metrics we use in this study to assess the decodability of neural simulations. Selectivity measures the tendency of a neuron to respond robustly only to a small subset of all stimuli. Mutual information measures the ability of a neuron to convey information about the identity of multiple stimuli by

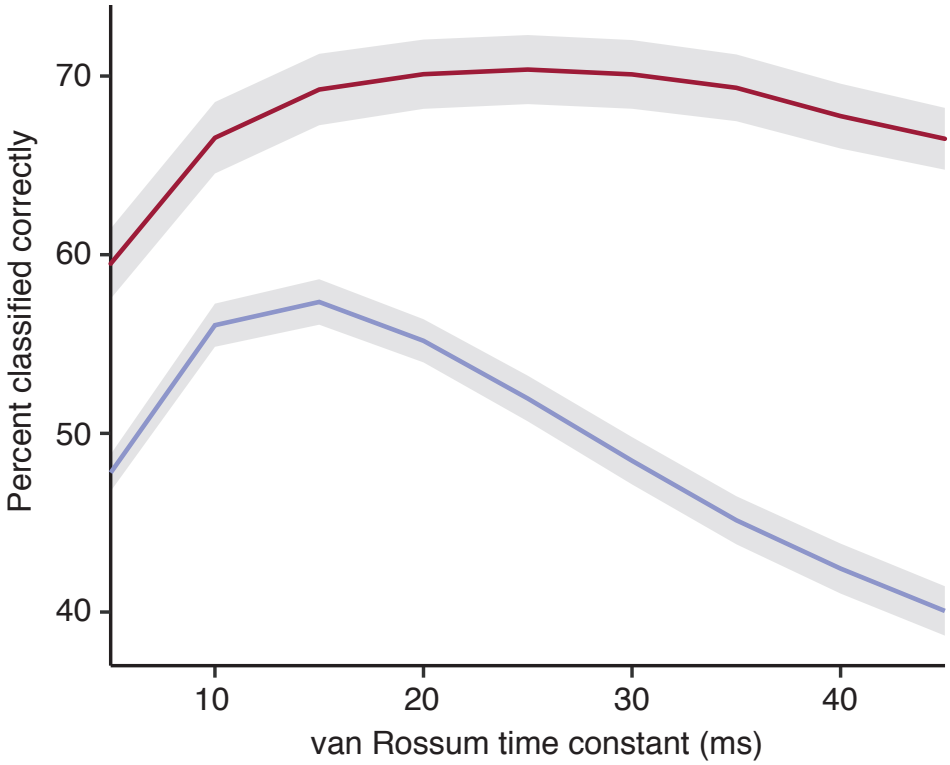

Figure 3: Classification analysis of temporal coding. The classification accuracy of the phasic models (red line) is significantly higher than the tonic models (blue line) at all time constants considered (5-45ms). Classification accuracy is based on a k-means clustering analysis of the van Rossum distances between each simulated spike train of a given neuron model. Gray ribbons show the standard error.

using different firing rates to encode different stimuli. There are two components of mutual information: the response (total) entropy, which represents how much information the neuron can carry based on its range of firing rates, and noise entropy, which represents how much information is lost due to the variability of a neurons firing-rate response within a stimulus. A neuron with high mutual information will have high response entropy and low noise entropy.

In our mutual information (MI) analysis, phasic neuron models showed a higher decodability than their tonic counterparts (paired t-test; $p < 1e-6$). Phasic neurons had a mean MI of 1.636 bits of information, and tonic neurons had a mean MI of 1.414 bits. The difference in MI is due to a reduction in noise entropy in the phasic models relative to the tonic models (phasic: 1.083 bits; tonic: 1.517 bits; paired t-test, $p < 1e-15$). The response entropy is, in fact, slightly higher in the tonic models (tonic: 2.932 bits; phasic: 2.720 bits; paired t-test, $p = 0.0003$), but the large amount of noise entropy in the tonic signal more than cancels out that advantage (Figure 4).

The selectivity analysis shows a similar advantage for phasic model neurons (Figure 5). Phasic models are able to encode song with a higher degree of selectivity than tonic models (tonic: 0.170; phasic: 0.258; paired t-test: $p < 1e-5$) with some phasic models showing very high levels of selectivity (0.60 and 0.78).

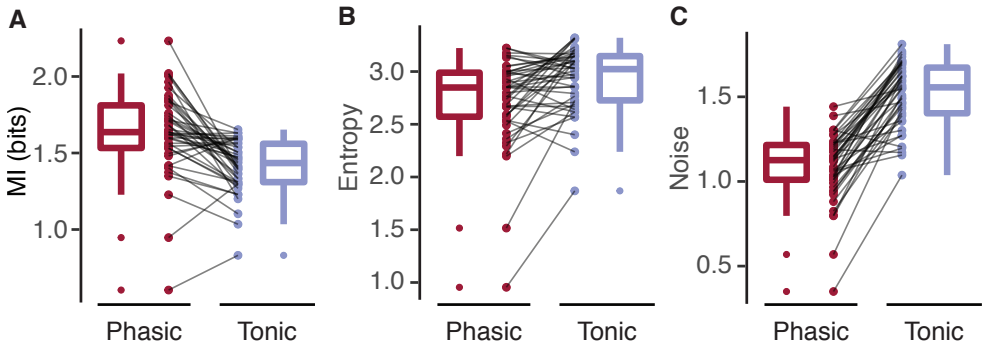

Figure 4: Mutual information analysis. **A**, Phasic models (red) have higher mutual information between firing rate and syllable identity than tonic models (blue) based on a paired t-test ($p < 1e-6$). **B**, One component of mutual information is response (total) entropy which represents the maximum information capacity of the model. Phasic and tonic models have comparable response entropy, though tonic models have a slight advantage ($p = 0.0003$). **C**, The second component of mutual information is noise entropy, which represents variability between repeated trials and decreases the amount of information conveyed from the theoretical maximum. Phasic models have much lower noise entropy than tonic models ($p < 1e-15$) which accounts for their higher mutual information.

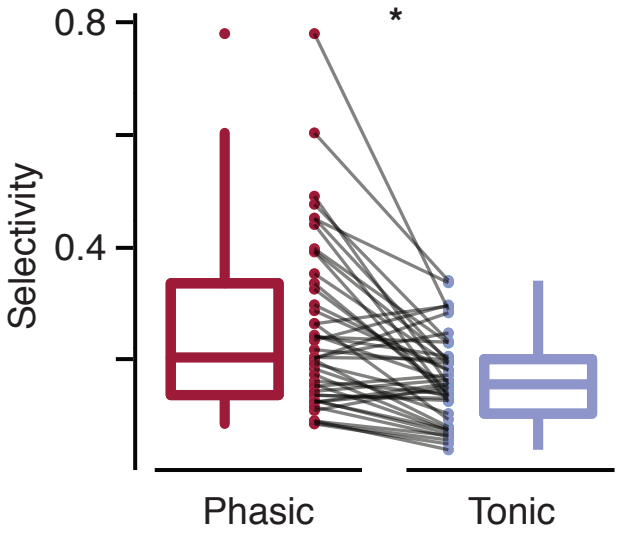

Figure 5: Selectivity analysis. **A**, Selectivity measures the tendency of a neuron to respond robustly only to a small subset of all stimuli. Phasic models (red) have the potential for higher levels of selectivity than tonic models (blue), with some phasic models showing very high levels of selectivity ($p < 1e - 5$).

Discussion

Relationship between decoding measures

Measures of mutual information (MI) and classification accuracy based on the van Rossum distance are positively correlated. This is because these two measures address similar decoding strategies on different timescales; as the time constant of the van Rossum distance increases, the analysis approaches a rate-based analysis.

The relationship between the two rate-based measures used in this study, MI and selectivity, is more complex. There is a general negative correlation (Figure 6A) between the two measures, but there are also models that score low on both measures. The models with low decodability on both measures are overwhelmingly tonic, but there are no models with high decodability on both measures, indicating that these measures are different yet mutually exclusive. This is consistent with extracellular data from zebra finch CM [37] when the same analyses were applied (Figure 6B). This relationship between MI and selectivity has also been previously been shown in starling CM [26].

Overall responsiveness mediates decoding strategy

When considering only the phasic models, the negative correlation between MI and selectivity becomes more pronounced. The overall responsiveness of the model, which we define as the average spiking rate (in Hz) of the model

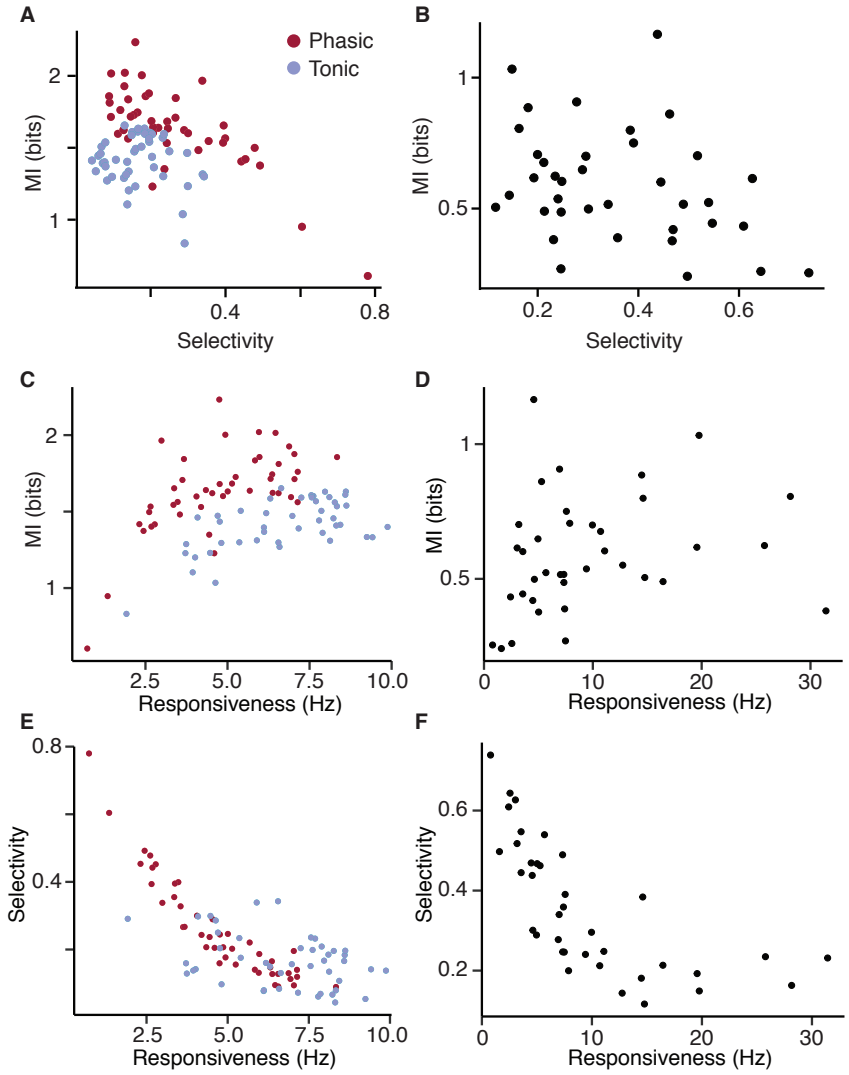

Figure 6: Relationship between MI and selectivity is mediated by responsiveness. **A**, MI and selectivity are inversely related, especially among phasic models (red). Tonic models (blue) tend to rate poorly on both decoding measures. **B**, CM neurons of zebra finches recorded extracellularly show a similar pattern of inverse correlation between MI and selectivity. **C**, Responsiveness is defined as the average response rate of the model to the entire stimulus set in spikes/sec. MI is positively correlated with responsiveness, and the groups of phasic and tonic models are clearly separable along these dimensions. **D**, CM neurons show a similar positive relationship between MI and responsiveness. **E**, Selectivity and responsiveness are negatively correlated in a non-linear fashion. **F**, CM neurons show the same non-linear correlation between selectivity and responsiveness.

over the entire stimulus set, is a strong predictor of whether a model is likely to have high MI or high selectivity. MI is positively correlated with responsiveness, *i.e.* models with higher responsiveness also tend to have higher MI (Figure 6C). Similarly, selectivity is negatively correlated with responsiveness with the most selective models showing very low average firing rates (Figure 6E). The relationships between these measures in the extracellular neural data are very consistent with the predictions of the simulations, indicating that the model is capturing population-level behavior of zebra finch CM (Figure 6D,F).

Figure 7 shows the pairs of phasic and tonic simulations with arrows indicating the phasic part of each pair. Consistent with previous results that show that MI and selectivity are negatively correlated, phasic models tend to increase in decodability relative to the tonic pairs in only one of the two dimensions of MI and selectivity. The direction of increase is determined by the responsiveness of the phasic model. Phasic models with high responsiveness show an increase in MI but not selectivity as compared with the tonic pair; phasic models with low responsiveness show an increase in selectivity but not MI. This relationship is independent of the MI, selectivity, or responsiveness of the tonic model.

Phasicness as slope detection

Because the tonic models are not predictive of whether the phasic models will show increased MI or increased selectivity, we examined the details of

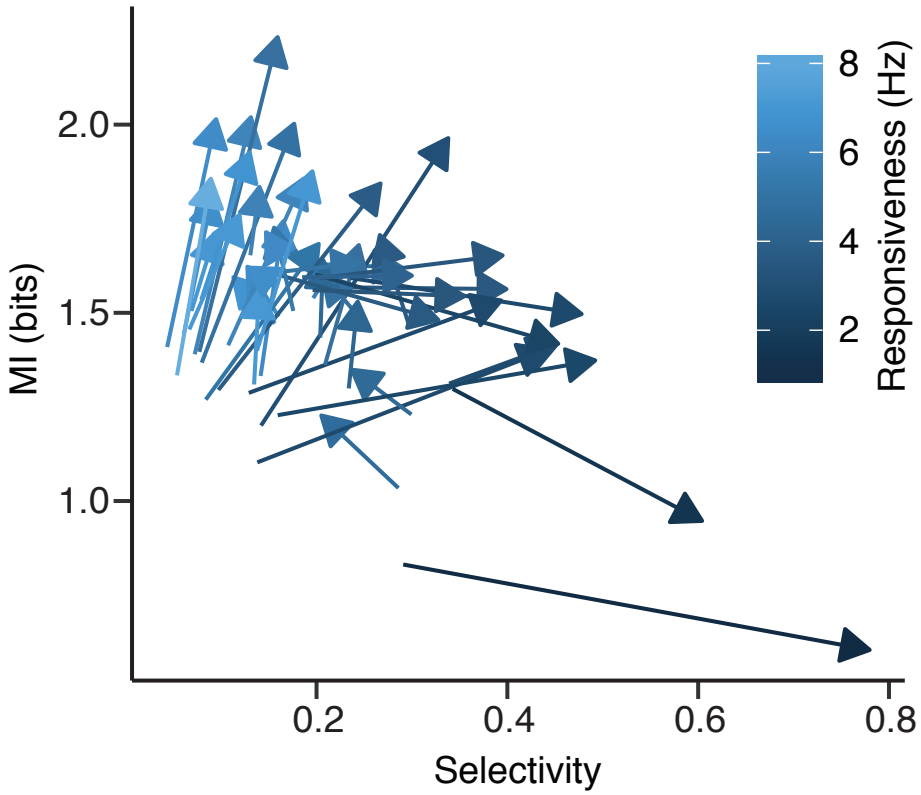

Figure 7: Phasic models increase in either MI or selectivity relative to tonic models. Connecting phasic and tonic pairs (arrows pointing toward the phasic model) shows that the phasic models tend to increase in decodability along only one of the two decoding measures examined here. The location of the tonic model on the measures of MI and selectivity does not seem to determine whether the phasic model will increase in MI or selectivity, but the responsiveness of the phasic model (arrow color) is strongly related. Phasic models with low responsiveness tend to increase in selectivity but not MI relative to tonic models. Phasic models with high responsiveness tend to increase in MI but not selectivity.

the simulations that gave rise to different outcomes. Figure 8 shows two pairs of examples that led to different outcomes. In Figure 8A, the tonic model has MI of 1.60 bits and selectivity of 0.20; the phasic model has similar MI (1.42 bits) but selectivity increases to 0.45. In Figure 8B, the tonic model has MI of 1.39 bits and selectivity of 0.07; the phasic model's selectivity remains similar (0.13) but the MI increases (2.02). The example convolutions in Figure 8 show why this happens.

In Figure 8A, the phasic model responds only to parts of the convolution where the slope increases sharply. This is true not only of the upslope of a peak but also the return to baseline of a negative deflection (black arrow). Because these slope increases are relatively infrequent in this convolution, the phasic model spikes sparsely and therefore shows increased selectivity. The tonic model, on the other hand, responds to the absolute excitation of the signal, treating the sharp peaks and the slower increases of excitation similarly, and this results in broad firing across many of the syllables of the song, reducing the model's selectivity.

In Figure 8B, the convolution contains primarily peaks and not the slow increases in excitation present in Figure 8A. This results in the two models responding similarly to the convolution with the exception of the increased variability of the tonic model as expected from the much higher noise entropy present in the tonic models. In this case, the phasic model acts solely as a noise reducer, thus increasing the MI of its response with only a slight increase in selectivity.

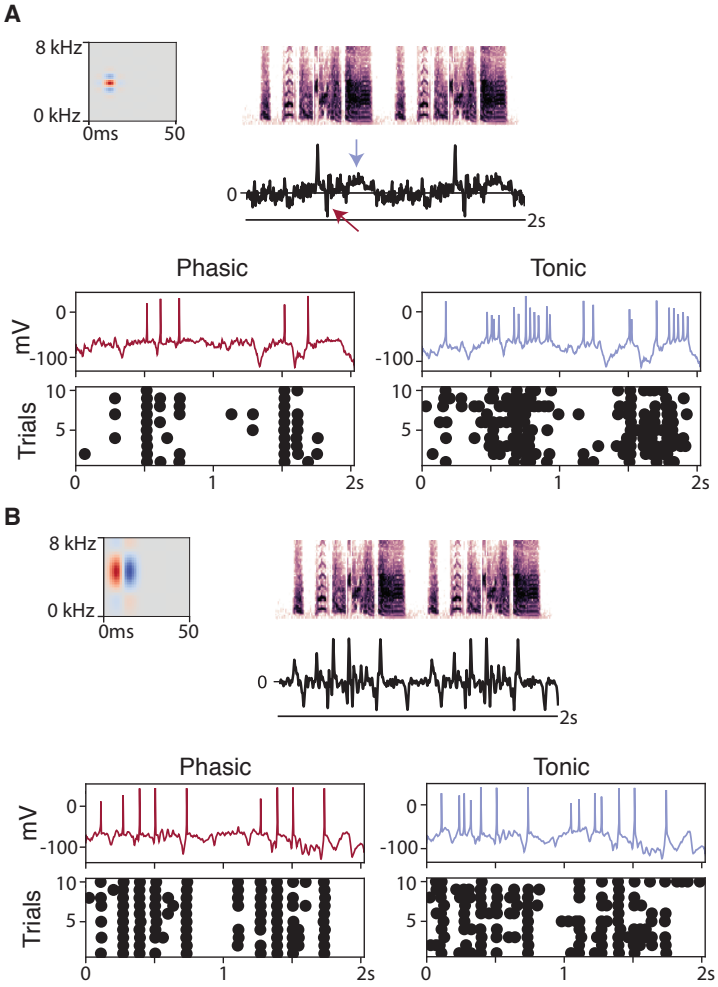

Figure 8: Examples of phasic responses with high selectivity or MI. **A**, A simulated neural response in which the phasic response had higher selectivity (0.45) than the tonic response (0.20). Upper panels show the RF, stimulus spectrogram, and convolution. Middle panels show simulated voltage traces (red: phasic; blue: tonic) and the bottom panels show the spike times across 10 trials of the stimulus. The phasic model responded only to sharp upward deflections of the convolution, including a rebound to baseline from a negative deflection (red arrow). The tonic model responded to all increases in excitation including the slow increases that the phasic model did not respond to (blue arrow). The sparseness of the phasic response boosts selectivity. **B**, A simulated neural response in which the phasic response had higher MI (2.02 bits) than the tonic response (1.39 bits). The phasic and tonic models responded at similar times but the increased temporal precision and decreased variance in spike number increased the MI of the phasic response relative to the tonic.

Ultimately, these simulations point to phasic and tonic neurons responding to fundamentally different features of the signal they receive from upstream neurons. Tonic neurons respond primarily to the level of excitation present in the signal whereas phasic neurons respond to the rate of increase of the excitation. The role of phasic neurons as a slope detector has been shown before, both *in vivo* and *in silico* [39], but these simulations suggest a potential function of that slope-detection property. By responding to the slope rather than the absolute level of excitation, phasic neurons can create selectivity from a signal that is otherwise non-selective, as Figure 8A demonstrates.

Chen and Meliza (2017) [28] found that tonic and phasic neurons differ in their response to high-frequency stimulation as measured by the coherence of their firing to a complex current injection. Phasic neurons were able to follow frequencies up to 30Hz, while tonic neurons had difficulty above 10Hz. They also found that the neuron model used in this simulation produces similar differences in coherence between phasic and tonic models. The ability of phasic neurons to follow higher frequencies may be important to their role in slope detection. Smoothing one of the convolutions used in this simulation with a 10Hz running average filter eliminates the sharpest peaks in the signal, but a 30Hz running average preserves them (Figure 9A). Differencing the 30Hz running average shows that smoothing at that frequency preserves the most important signal deflections (Figure 9B), while the 10Hz running average removes them. In

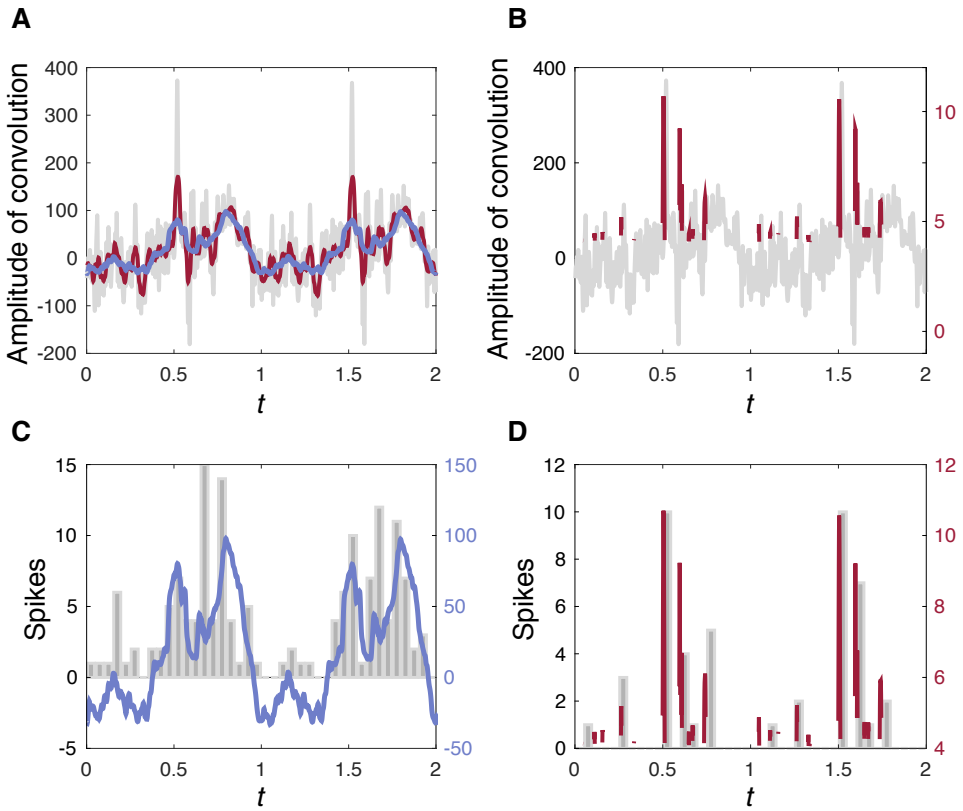

Figure 9: Simple transformations of the convolution predict phasic and tonic responses. **A**, Convolution smoothed with a 10Hz running average (blue) and 30Hz running average (red) based on the frequencies that tonic and phasic neurons are able to follow. 10Hz smooths out the majority of the peaks, but 30Hz preserves the largest ones. **B**, Differenced 30Hz smoothed convolution with a threshold of 1.5 standard deviations highlights the largest upward deflections in the signal. **C**, 10Hz smoothed convolution matches closely the spike-time histogram of the tonic model's response to this convolution (gray bars). **D**, Differenced 30Hz smoothed convolution predicts very accurately the spike-time histogram of the phasic model's response to this convolution (gray bars).

fact, the convolution smoothed with the 10Hz running average fits very well to the spike-time histogram of the tonic model's response to that convolution (Figure 9C), and the differenced 30Hz running average is highly predictive of the spike times of the phasic model (Figure 9D). The higher peak coherence of the phasic neurons may be an important part of their enhanced ability to produce a selective response to song.

Limitations of this model

There are a number of limitations of this model to keep in mind when interpreting these results. The first is that the neuron model used is not specifically a model of a CM neuron but rather a model that reproduces many of the behaviors seen in CM neurons (*e.g.*, response to current steps and coherence to chaotic currents). This model also does not consider a third type of putatively excitatory neuron found in CM, called an intermediate-spiking neuron which shows firing patterns between those of phasic and tonic neurons [28], because we could not arrive at a stable model of this type of neuron using the Rothman-Manis base model.

As described in the methods, the receptive fields used in this analysis were based on a thorough characterization of Field L receptive fields by Woolley *et al.* (2009) [32]. We felt that this was a reasonable approach given that CM is immediately downstream of Field L and that no such comprehensive characterization has been done for CM receptive fields. This is in part due to the fact that receptive fields for CM are difficult to

estimate due to the sparseness of the neurons' firing. We also do not know whether phasic and tonic neurons have a similar distribution of receptive fields. Given the differences in dendritic morphology reported by Chen and Meliza (2017) [28], it is possible that phasic and tonic neurons have systematic differences in their receptive fields. This simulation examined the effect of changing the neural dynamics of a model while keeping the receptive field constant, but that comparison might not completely capture the differences.

This is also a very simple, single-neuron model that lacks lateral connections or feed-forward inhibitory inputs. The auditory system, of course, is much more complex, and there are certainly many additional influences on the behavior of a neuron. It was not our intent to capture all of these complexities in our model, and in fact, the ability of our model to produce selective responses to song syllables despite its simplicity is a strength. There may be other ways to arrive at selectivity, but the fact that selectivity can be created merely by the introduction of phasic neurons into the population may explain, at least in part, the increase in selectivity from Field L to CM [23].

Conclusions

A biophysical neuron model can reproduce the relationship between mutual information and selectivity seen in zebra finch CM. The model predicts that

a decrease in the overall responsiveness of the neuron shifts decoding performance toward selectivity and away from mutual information, and that prediction is supported by evidence from extracellular measurements of CM neurons. The results suggest that phasic neurons represent an advantage for the decoding of stimulus identity and that advantage is due to the precision and selectivity generated by their sensitivity to the rate of increase of excitation. The addition of phasic neurons to the CM population should improve the ability of CM to identify stimuli beyond what tonic neurons could do alone owing to their heightened selectivity and their tolerance to noise.

References

[1] B M Clopton, J A Winfield, and F J Flammino. Tonotopic organization: review and analysis. *Brain research*, 76(1):1–20, August 1974.

[2] Ling-yun Li, Xu-ying Ji, Feixue Liang, Ya-tang Li, Zhongju Xiao, Huizhong W Tao, and Li I Zhang. A feedforward inhibitory circuit mediates lateral refinement of sensory representation in upper layer 2/3 of mouse primary auditory cortex. *The Journal of Neuroscience*, 34(41):13670–13683, October 2014.

[3] Paul V Watkins, Joseph P Y Kao, and Patrick O Kanold. Spatial

pattern of intra-laminar connectivity in supragranular mouse auditory cortex. *Frontiers in neural circuits*, 8(e1002161):15, 2014.

[4] Frédéric E Theunissen and Sarita S Shaevitz. Auditory processing of vocal sounds in birds. *Current opinion in neurobiology*, 16(4):400–407, August 2006.

[5] A J Doupe and P K Kuhl. Birdsong and human speech: common themes and mechanisms. *Annual review of neuroscience*, 22:567–631, 1999.

[6] Olga Fehér, Haibin Wang, Sigal Saar, Partha P Mitra, and Ofer Tchernichovski. De novo establishment of wild-type song culture in the zebra finch. *Nature*, 459(7246):564–568, May 2009.

[7] N S Clayton. The Effects of Cross-Fostering On Selective Song Learning in Estrildid Finches. *Behaviour*, 109(3):163–174, January 1989.

[8] Dan H Sanes and Sarah M N Woolley. A behavioral framework to guide research on central auditory development and plasticity. *Neuron*, 72(6):912–929, December 2011.

[9] A E Jones, C ten Cate, and P J B Slater. Early experience and plasticity of song in adult male zebra finches (Taeniopygia guttata). *Journal of Comparative Psychology*, 1996.

[10] Mountjoy James D and Robert E Lemon. Extended song learning in wild European starlings. *Animal behaviour*, 49(2):357–366, February 1995.

[11] Kai Lu and David S Vicario. Statistical learning of recurring sound patterns encodes auditory objects in songbird forebrain. *Proceedings of the National Academy of Sciences of the United States of America*, 111(40):14553–14558, October 2014.

[12] C B Sturdy, L S Phillmore, J L Price, and R G Weisman. Song-note discriminations in zebra finches (Taeniopygia guttata): Categories and pseudocategories. *Journal of Comparative Psychology*, 1999.

[13] Keith R Kluender, Andrew J Lotto, Lori L Holt, and Suzi L Bloedel. Role of experience for language-specific functional mappings of vowel sounds. *The Journal of the Acoustical Society of America*, 104(6):3568–3582, November 1998.

[14] Sarah M N Woolley, Mark E Hauber, and Frédéric E Theunissen. Developmental experience alters information coding in auditory midbrain and forebrain neurons. *Developmental Neurobiology*, 70(4):235–252, March 2010.

[15] Sarah M N Woolley and Christine V Portfors. Conserved mechanisms of vocalization coding in mammalian and songbird auditory midbrain. *Hearing research*, 305:45–56, November 2013.

[16] Jennifer Dugas-Ford, Joanna J Rowell, and Clifton W Ragsdale. Cell-type homologies and the origins of the neocortex. *Proceedings of the National Academy of Sciences of the United States of America*, 109(42):16974–16979, October 2012.

[17] Yuan Wang, Agnieszka Brzozowska-Prechtl, and Harvey J Karten. Laminar and columnar auditory cortex in avian brain. *Proceedings of the National Academy of Sciences of the United States of America*, 107(28):12676–12681, July 2010.

[18] Sarah M N Woolley and John H Casseday. Response properties of single neurons in the zebra finch auditory midbrain: response patterns, frequency coding, intensity coding, and spike latencies. *Journal of Neurophysiology*, 91(1):136–151, January 2004.

[19] Julie A Grace, Noopur Amin, Nandini C Singh, and Frédéric E Theunissen. Selectivity for conspecific song in the zebra finch auditory forebrain. *Journal of Neurophysiology*, 89(1):472–487, January 2003.

[20] Anne Hsu, Sarah M N Woolley, Thane E Fremouw, and Frédéric E Theunissen. Modulation power and phase spectrum of natural sounds enhance neural encoding performed by single auditory neurons. *The Journal of Neuroscience*, 24(41):9201–9211, October 2004.

[21] Jose A Garcia-Lazaro, Bashir Ahmed, and Jan W H Schnupp. Emergence of tuning to natural stimulus statistics along the central auditory pathway. *PloS one*, 6(8):e22584, 2011.

[22] G E Vates, B M Broome, C V Mello, and F Nottebohm. Auditory pathways of caudal telencephalon and their relation to the song system of adult male zebra finches. *The Journal of comparative neurology*, 366(4):613–642, March 1996.

[23] Ana Calabrese and Sarah M N Woolley. Coding principles of the canonical cortical microcircuit in the avian brain. *Proceedings of the National Academy of Sciences*, 112(11):3517–3522, March 2015.

[24] Nienke J Terpstra, Johan J Bolhuis, and Ardie M den Boer-Visser. An analysis of the neural representation of birdsong memory. *The Journal of Neuroscience*, 24(21):4971–4977, May 2004.

[25] Mimi L Phan, Carolyn L Pytte, and David S Vicario. Early auditory experience generates long-lasting memories that may subserve vocal learning in songbirds. *Proceedings of the National Academy of Sciences*, 103(4):1088–1093, January 2006.

[26] James M Jeanne, Jason V Thompson, Tatyana O Sharpee, and Timothy Q Gentner. Emergence of learned categorical representations within an auditory forebrain circuit. *The Journal of Neuroscience*, 31(7):2595–2606, February 2011.

[27] C Daniel Meliza and Daniel Margoliash. Emergence of selectivity and tolerance in the avian auditory cortex. *The Journal of Neuroscience*, 32(43):15158–15168, October 2012.

[28] Andrew N Chen and C Daniel Meliza. Phasic and Tonic Cell Types in the Zebra Finch Auditory Caudal Mesopallium. *Journal of Neurophysiology*, page jn.00694.2017, December 2017.

[29] Stefan Huggenberger, Marianne Vater, and Rudolf A Deisz. Interlaminar differences of intrinsic properties of pyramidal neurons in the auditory cortex of mice. *Cerebral cortex (New York, N.Y. : 1991)*, 19(5):1008–1018, May 2009.

[30] C E Carr and D Soares. Evolutionary convergence and shared computational principles in the auditory system. *Brain, behavior and evolution*, 59(5-6):294–311, 2002.

[31] Jason S Rothman and Paul B Manis. The roles potassium currents play in regulating the electrical activity of ventral cochlear nucleus neurons. *Journal of Neurophysiology*, 89(6):3097–3113, June 2003.

[32] Sarah M N Woolley, Patrick R Gill, Thane Fremouw, and Frédéric E Theunissen. Functional Groups in the Avian Auditory System. *The Journal of Neuroscience*, 29(9):2780–2793, March 2009.

[33] Nandini C Singh and Frédéric E Theunissen. Modulation spectra of natural sounds and ethological theories of auditory processing. *The*

Journal of the Acoustical Society of America, 114(6 Pt 1):3394–3411, December 2003.

[34] M C van Rossum. A novel spike distance. *Neural computation*, 13(4):751–763, April 2001.

[35] David M Schneider and Sarah M N Woolley. Discrimination of communication vocalizations by single neurons and groups of neurons in the auditory midbrain. *Journal of Neurophysiology*, 103(6):3248–3265, June 2010.

[36] E T Rolls and M J Tovee. Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *Journal of Neurophysiology*, 73(2):713–726, February 1995.

[37] Frederic E Theunissen, Patrick Gill, Amin Noopur, Junli Zhang, Sarah M N Woolley, and Thane Fremouw. Single-unit recordings from multiple auditory areas in male zebra finches. *CRCNS.org*, 2011.

[38] Patrick Gill, Junli Zhang, Sarah M N Woolley, Thane Fremouw, and Frédéric E Theunissen. Sound representation methods for spectro-temporal receptive field estimation. *Journal of computational neuroscience*, 21(1):5–20, August 2006.

[39] Yan Gai, Brent Doiron, Vibhakar Kotak, and John Rinzel. Noise-gated encoding of slow inputs by auditory brain stem neurons with a

low-threshold K+ current. *Journal of Neurophysiology*, 102(6):3447–3460, December 2009.