

Invariant neural responses for sensory categories revealed by the time-varying information for communication calls

Julie E. Elie^{1,2*} and Frédéric E. Theunissen^{1,3}.

1. Helen Wills Neuroscience Institute. University of California, Berkeley.

2. Department of Bioengineering. University of California, Berkeley.

3. Department of Psychology. University of California, Berkeley.

*Corresponding Author:

julie.elie@berkeley.edu

Short Title: Time-varying information for auditory objects.

Abstract.

Although information theoretic approaches have been used extensively in the analysis of the neural code, they have yet to be used to describe how information is accumulated in time while sensory systems are categorizing dynamic sensory stimuli such as speech sounds or visual objects. Here, we present a novel method to estimate the cumulative information for stimuli or categories. We further define a time-varying categorical information index that, by comparing the information obtained for stimuli versus categories of these same stimuli, quantifies invariant neural representations. We use these methods to investigate the dynamic properties of avian cortical auditory neurons recorded in zebra finches that were listening to a large set of call stimuli sampled from the complete vocal repertoire of this species. We found that the time-varying rates carry 5 times more information than the mean firing rates even in the first 100 *ms*. We also found that cumulative information has slow time constants (100-600 *ms*) relative to the typical integration time of single neurons, reflecting the fact that the behaviorally informative features of auditory objects are time-varying sound patterns. When we correlated firing rates and information values, we found that average information correlates with average firing rate but that higher-rates found at the onset response yielded similar information values as the lower-rates found in the sustained response: the onset and sustained response of avian cortical auditory neurons provide similar levels of independent information about call identity and call type. Finally, our information measures allowed us to rigorously define categorical neurons; these categorical neurons show a high degree of invariance for vocalizations within a call-type. Surprisingly, call-type invariant neurons were found in both primary and secondary avian auditory areas.

Author Summary

Just as the recognition of faces requires neural representations that are invariant to scale and rotation, the recognition of behaviorally relevant auditory objects, such as spoken words, requires neural representations that are invariant to the speaker uttering the word and to his or her location. Here, we used information theory to investigate the time course of the neural representation of bird communication calls and of behaviorally relevant categories of these same calls: the call-types of the bird's repertoire. We found that neurons in both the primary and secondary avian auditory cortex exhibit invariant responses to call renditions within a call-type, suggestive of a potential role for extracting the meaning of these communication calls. We also found that time plays an important role: first, neural responses carry significant more information when represented by temporal patterns calculated at the small time scale of 10 ms than when measured as average rates and, second, this information accumulates in a non-redundant fashion up to long integration times of 600 ms. This rich temporal neural representation is matched to the temporal richness found in the communication calls of this species.

Introduction

Information theoretic analyses are well suited to the study of neural representation since this mathematical framework was developed to quantify and optimize the encoding of informative signals in communication channels [1]. In sensory systems, Information Theory (IT) has been applied extensively as a complimentary approach to the estimation of stimulus-response functions such as tuning curves, spatio-temporal or spectro-temporal receptive fields or other higher-level encoding models [2]. Information theoretic approaches have been particularly powerful in explorations of the nature of the neural code and its redundancy or efficiency [3-6]. For example, IT was used in early studies in the visual system to

demonstrate that spike patterns contain information beyond average rate both for static images [7] and dynamic visual stimuli [8]. IT was also used to show that spike doublets can contain synergistic information that cannot be explained by an analysis of successive single spikes [9] and that, although information can only decrease in a signal processing chain, the neural coding efficiency increases as one moves to higher levels of sensory processing [10]. Finally, IT investigations also revealed that neural efficiency is higher when sensory systems process natural stimuli versus synthetic stimuli [11-13], in support of ethological theories of optimal sensory processing [14].

In sensory systems, the mutual information between a stimulus and the neural responses has often been estimated in a stimulus reconstruction framework and for continuous dynamic stimuli in stationary conditions, where time averages can be performed. In the stimulus reconstruction framework, one attempts to estimate the information about all aspects of the stimulus; for example, in audition, the stimulus would be represented by its exact sound pressure waveform. As long as the stimulus set is rich (i.e. has very large entropy), the mutual information can be an estimate of the maximum information that can be transmitted by a neural communication channel, also known as the channel capacity [3, 4]. For instance, one can obtain the mutual information of an adapted auditory neuron processing white noise or colored noise sounds [11]. As long as the stationary assumption is valid, using continuous stimuli is also beneficial as it provides large data sets that are needed to estimate the joint probability of stimuli and neural responses, both of which can have high dimensions. Even in these conditions, it is noteworthy that a direct estimation of information is only possible when many repeats of the same stimulus can be obtained [15] or when simplifying assumptions are made [9]. Ultimately, the calculation of information based on stimulus reconstruction gives a single number corresponding to the information transmitted by a single neuron or an ensemble of neurons for a particular stimulus ensemble. By repeating the

calculation for different stimulus ensembles, one can investigate how the channel capacity of particular neurons or neural ensembles might depend on the stimulus statistics (e.g. for natural vs synthetic stimuli). Furthermore, by repeating the calculation using different symbols to represent the response, the potential nature of the neural code (e.g. time patterns vs. rate) can be revealed.

Here we are using IT in a different sensory encoding context: the accumulation of information in a recognition task, such as face recognition in the visual system [16] or word recognition in the auditory system [17]. Recognition or identification is one of the key computations performed by higher sensory areas as opposed to the task of efficient stimulus representations that is performed in lower sensory areas and that might therefore be well quantified by information values based on stimulus reconstruction. In the recognition tasks, each stimulus is described by a simple label, such as the word corresponding to a given speech sound or that is used to label a given visual object. The relevant value of information in that task is then the information about these discrete labels and the information capacity of the system in its ability to identify the stimulus as a whole. In the recognition framework, one can ask how the information about the stimulus identity or label changes as a function of time relative to the stimulus onset and to what extent that time-varying information is redundant and, thus, how it accumulates over time. For example, one could ask at what time after stimulus onset does the performance of single neurons or ensemble of neurons match a behavioral performance of word recognition. Such an IT analysis has been performed in the primate visual system using a delay-matching to sample paradigm, and using spike counts, estimated in progressively longer windows, as the neural symbols [18].

In information studies based on continuous stimulus reconstruction, the neural code can be investigated in terms of its temporal resolution (i.e. letter size) and its integration time (i.e. word length). While the same properties of the neural code can be deciphered in the

recognition framework, one can also examine the relationship between spikes at different points in time and time-varying information. This analysis is meaningful because a time zero corresponding to stimulus onset can be clearly defined and is behaviorally relevant. Moreover, stimulus-response functions for such discrete stimuli are not time-invariant. Responses in sensory neurons, in vision [19, 20] and in audition [21, 22], are often characterized by an onset response (or on-response) and a sustained response, where both the precision of spikes and the information coded might be different [23, 24]. For example, a first spike latency code has been proposed as a fast encoding scheme in vision [25], audition [26] and somato-sensation [27]. Rolls et al. tested this hypothesis, by quantifying the fraction of information that is present in the first spike relative the on-going response [28].

Finally, in the recognition framework, one can also compare the information values obtained when different labelling schemes are used for identifying the stimuli as objects. For example, speech sounds could be labeled hierarchically as unique utterances, as phonemes, as syllables, as words, etc.. One can then compare time-varying information about each of the levels in such hierarchical labelling scheme and gain insight on the neuro processing involved in object categorization. Although such a hierarchical representation of stimulus features has been used in encoding models for studying human processing [29], it has not yet been used in an IT analysis.

In this study, we developed a new approach for estimating time-varying information and cumulative information for sensory object identification task. Our approach assumes that time-varying neural responses can be modeled as inhomogeneous Poisson processes and generalizes well to large number of stimulus categories and to long integration times relative to the dynamics of the time-varying response. Our motivation for developing this methodology was to gain additional understanding on the neural representation of communication signals in high level auditory areas. Animal communication calls, just as

speech sounds in humans, are categorized into behaviorally meaningful units. Significant progress has been made in identifying brain regions involved in categorizing sounds, in particular in the primate brain, where neural responses that are correlated with progressively more abstract concepts are found in primary auditory cortex, the lateral belt of the auditory cortex and the prefrontal cortex [30]. However, the neural computations involved in generating categorical responses remain poorly described [31] and only a small number of studies have examined the neural categorization of natural communication calls in non-human species [32-36]. We and others have been developing an avian model system to study the neural processing of relatively large and complex vocal repertoires [37, 38]. Our prior studies include a detailed bioacoustical analysis of the features that define each call-type of the complete vocal repertoire of the zebra finch [39] and the first characterization of neural responses to the calls from that large repertoire in primary and secondary avian auditory cortical areas [40]. In that study, we found that approximately 45% of auditory neurons encode information about call-type categories. Among those, a minority show strong selectivity for single call-type categories and invariance for calls within that category. Here, we investigated the processing in time that could lead to those observed categorical responses by comparing the time-varying information for stimuli labelled as individual utterances to the time-varying information for the same stimuli labelled by their call-type category. With that analysis, we were able to obtain values of temporal integration for stimulus identification and call-type category identification. We also analyzed the relationship between the time-varying firing rate and the time-varying information and, in particular, examined differences in selectivity in the onset versus sustained response. Finally, we used anatomical data to examine the distribution of neurons in primary and secondary avian auditory cortical areas with distinct responses properties as revealed by this IT analysis.

Results

We studied the time-varying information in a population of neurons recorded from primary and secondary regions in the avian auditory cortex of head-fixed urethane anesthetized zebra finches listening to a large set of natural communication calls. Zebra finches emit various calls in different behavioral contexts and have a complete vocal repertoire composed of 11 call-types. The acoustical characteristics of each call-type have previously been described in detail [39]. Here we focused on the neural representation of 9 call-types: 1) 3 pro-social calls emitted for pair bonding and social cohesion: the Distance Call (DC), the Tet Call (Te), the Nest Call (Ne); 2) the Song (So) that is emitted as a sexual display in males; 3) 2 calls emitted in aggressive encounters: the aggressive Wsst Call (Ws) and the Distress Call (Di); 4) 1 alarm call, the Thuk Call (Th) and 5) 2 calls emitted by juveniles: the Begging Call (Be), used by young to request food, and the Long Tonal Call (LT), a contact call that is a precursor of the adult DC. A stimulus set was composed of approximately 10 different exemplars of calls, train of calls or song produced by different vocalizers for each of the 9 call-types. The call stimuli were randomly sampled from a large annotated data base of calls and songs from the complete repertoire of the male and female zebra finch. Neural responses were recorded using electrode arrays implanted in both hemispheres of 4 male and 2 female adult zebra finches. We recorded from a total of 914 single auditory units in both primary (Field L) and secondary (CML, CMM and NCM) avian auditory areas. In previous analysis, we showed using a decoding approach that information about call-types was found in 404 (44%) of these units [40]. Here, we further restricted our population analysis of information to neurons, from that same data set, that reached a significant level of information about the stimulus (see methods) and analyzed the time-varying information of 337 neurons during the first 600 ms of their response. Note that many calls are shorter than 600 ms, but also that they are often produced in succession with short

inter-call intervals. Thus, this analysis window could contain two or more calls or the beginning of a longer song motif comprised of multiple syllables. We only analyzed the response in the first 600 ms because the estimation of the cumulative information for longer time windows became unreliable, as we will explain below. Additional details on the neurophysiological recordings can be found in the methods section and in Elie and Theunissen [40].

In the *Results*, we first describe the approach we developed to estimate instantaneous and cumulative time varying information. We illustrate these calculations with specific examples of model and actual neurons in the avian auditory cortex. We then analyze the time-varying coding properties of the avian cortical auditory neurons with an emphasis on: the relationship between spike rate and instantaneous information, the time constants observed for the cumulative information and the relative fraction of stimulus cumulative information that is used for extracting the behaviorally relevant categories corresponding to distinct call-types.

Estimation of the time-varying information

At a given time t , the *instantaneous* mutual information between the stimulus S and the response Y_t can be written as a difference in Shannon entropies:

$$I_t = H(Y_t) - H(Y_t|S)$$

Here, $H(Y_t)$ is the response entropy for a window at time t , while $H(Y_t|S)$ corresponds to entropy of the response given the stimulus or the conditional response entropy. $H(Y_t|S)$ can also be called the neural noise since it represents the variability in the neural response to the same stimulus. For spiking neurons, y_t represents the number of spikes in the window at time t (Note: In our notation, capitals are used for random variables and lower case for a sample from that random variable).

Similarly, the cumulative mutual information in neural responses that are discretized into time intervals is given by:

$$CI_t = H(Y_t, Y_{t-1}, Y_{t-2}, \dots, Y_0) - H(Y_t, Y_{t-1}, Y_{t-2}, \dots, Y_0 | S)$$

The entropies now include the time course of the neural responses starting at $t=0$ and up to time t . The reference time $t=0$ is set to the stimulus onset in our analyses but could be any arbitrary reference point.

The conditional response entropy and the response entropy are obtained from the distribution of the conditional probability of neural responses given the stimulus, $p(y_t|s)$, and the distribution of probability of each stimulus $p(s_i)$:

$$H(Y_t|S) = \sum_i p(s_i) \sum_{y_t=0}^{R_{Max}} -p(y_t|s_i) \log_2 p(y_t|s_i)$$

$$H(Y_t) = \sum_{y_t=0}^{R_{Max}} -p(y_t) \log_2 p(y_t)$$

$$\text{with } p(y_t) = \sum_i p(s_i) p(y_t|s_i)$$

y_t , the neural response at time t , is measured in spike counts and takes values from zero to a maximum rate value, R_{Max} (for example, as dictated by the neuron's refractory period or numerically as $p \log p$ becomes infinitely small for high spike counts that have very small probability of occurring). The probability of the stimulus $p(s_i)$ is usually taken as $1/n_s$, where n_s is the number of stimuli, unless the study incorporates natural stimulus statistics.

The probability of spike counts at time t given a stimulus, $p(y_t|s_i)$, could be estimated empirically by recording hundreds or thousands of responses of the same neuron to the same stimulus. Although this approach has been shown to be possible in certain preparations [15], it severely limits the number of stimuli that can be investigated in most neurophysiological experiments. Here we propose a parametric approach where we model the distribution of

neural responses to a given stimulus s_i as an inhomogeneous Poisson process. The conditional probability of response (spike count) given the stimulus is then:

$$p(y_t | s_i) = \frac{\mu_{s_i}(t)^{y_t}}{y_t!} e^{-\mu_{s_i}(t)}$$

where $\mu_{s_i}(t)$ is the mean response at time t for stimulus s_i . This mean rate was estimated empirically using the time varying kernel density estimation (KDE) proposed by Shimazaki and Shinomoto [41]. The instantaneous information estimated in this fashion is relatively straightforward, as long as the Poisson assumption is valid and a sufficient number of trials is obtained to estimate $\mu_{s_i}(t)$ (see methods).

The estimation of the cumulative information then extends this approach to joint probabilities of responses in successive time windows, $(y_t, y_{t-1}, y_{t-2}, \dots)$. Due to what has been labelled as the “curse of dimensionality”, the numerical estimation of the unconditional probability becomes exponentially more expensive as the integration time increases. We evaluated multiple approaches based on different assumptions and found that Monte Carlo with importance sampling yielded the best results (see methods and Sup. Fig. 3 for comparison to alternative approaches).

Given our Poisson assumption, the conditional probability of response at t is independent of the conditional response at previous times.

$$p(y_t, y_{t-1}, y_{t-2}, \dots | s_i) = p(y_t | s_i) p(y_{t-1} | s_i) p(y_{t-2} | s_i) \dots$$

Because of this probabilistic independence, it can be shown that the joint conditional response entropy is simply the sum of the conditional response entropies at each time point (see methods):

$$H(Y_t, Y_{t-1}, Y_{t-2}, \dots | S) = H(Y_t | S) + H(Y_{t-1} | S) + H(Y_{t-2} | S) + \dots$$

Thus, the estimation of the conditional response entropy is straightforward and not affected by the integration time. The problem of dimensionality arises in the estimation of the unconditional probabilities of response and the corresponding response entropy.

The probability of the time varying response is the joint probability of observing $(y_t, y_{t-1}, y_{t-2}, \dots)$. This joint probability cannot be expressed as the product of the probabilities at different times because these are not independent. More intuitively observing a particular y_{t-1} will affect the probability of observing y_t . This is true because the time varying means of the Poisson distributions $\mu_{s_i}(t)$ are correlated in time; for example, if $\mu_{s_i}(t-1)$ is high we might expect $\mu_{s_i}(t)$ to also have high values. These high values could be true for one particular stimulus s_i but not for the other stimuli. Then observing a high value of y_{t-1} would predict a higher value than expected for y_t (and an increase in probability that it was caused by s_i). The joint unconditional probability distribution is:

$$p(y_t, y_{t-1}, y_{t-2}, \dots) = \sum_{s_i} p(s_i) [p(y_t | s_i) p(y_{t-1} | s_i) p(y_{t-2} | s_i) \dots] \neq p(y_t) p(y_{t-1}) p(y_{t-2}) \dots$$

Given the lack of independence, the response entropy must then be calculated from the joint probability distribution:

$$H(Y_t, Y_{t-1}, Y_{t-2}, \dots) = - \sum_{y_t=0}^{R_{Max}} \sum_{y_{t-1}=0}^{R_{Max}} \sum_{y_{t-2}=0}^{R_{Max}} \dots p(y_t, y_{t-1}, y_{t-2}, \dots) \log_2 p(y_t, y_{t-1}, y_{t-2}, \dots)$$

The estimation of this entropy was performed using Monte Carlo with importance sampling. In Monte Carlo, random samples of a vector $(y_t, y_{t-1}, y_{t-2}, \dots)$ are drawn from a proposal distribution $q(y_t, y_{t-1}, y_{t-2}, \dots)$ and used to estimate the expected value of $\log_2 p(y_t, y_{t-1}, y_{t-2}, \dots)$ by an algebraic average weighted by the likelihood ratio of $p(y_t, y_{t-1}, y_{t-2}, \dots)/q(y_t, y_{t-1}, y_{t-2}, \dots)$. The sampling stops when entropy estimations reach an equilibrium. Information estimations are also known to suffer from positive bias [42]. Here, biased-corrected estimates and errors on information values were obtained from

Jackknifing the estimation of the time-varying rates and bootstrapping the Monte Carlo samples.

For these calculations, one has also to determine the size of the time window used for estimating the instantaneous information and correspondingly the steps for the cumulative information. This time window is used to estimate spike counts and the average rate $\mu_{s_i}(t)$ and depends both the dynamics of the stimulus and on the response properties of the neurons. By performing a coherence analysis on spike trains and a power spectral analysis on time varying rate in response to natural stimuli, we found that 10 ms (or 50 Hz) captured between 97% and 99% of the dynamics in our system (see methods and Sup. Fig. 1).

Finally, we also estimated the information values for stimulus categories by performing the weighted sum of probabilities for stimuli belonging to each category. The information about categories at time t :

$$I_t = H(Y_t) - H(Y_t|C).$$

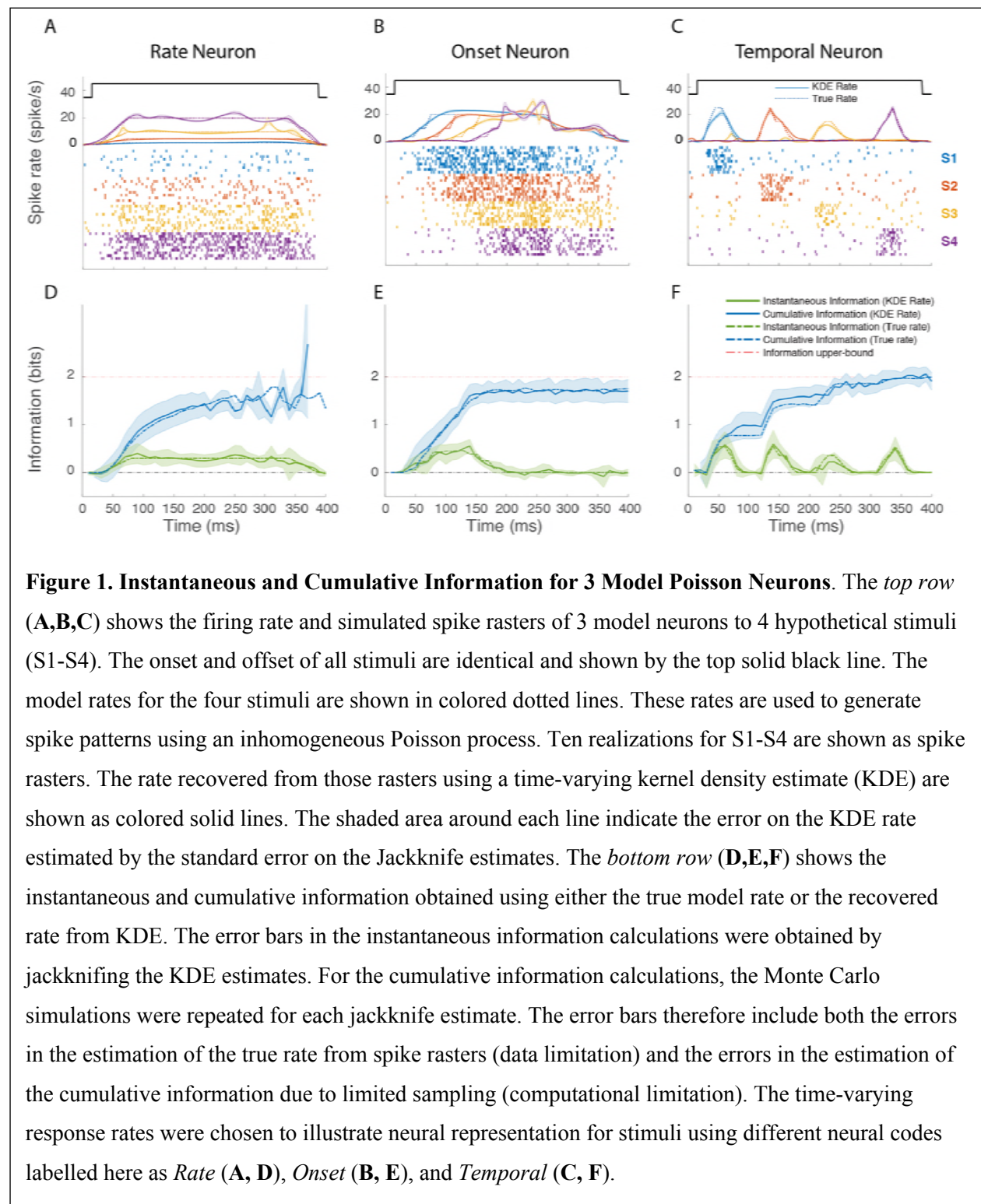
is obtained from the conditional probability of response given the category c_k , which is in turn calculated as the average conditional probability of response for the stimuli belonging to that category:

$$p(y_t|c_k) = \sum_{s_k} p(s_k)p(y_t|s_k)$$

Here, $p(s_k)$ is the probability of occurrence of stimulus s_k within the category c_k . In controlled playback experiments (as here), $p(s_k)$ is $1/n_{s_k}$ where n_{s_k} is the number of stimuli used to sample category c_k . Similarly, $p(c_k)$, the probability of occurrence of vocalizations in category c_k was taken as $1/n_c$ where n_c is the number of categories. In our system, the stimuli are individual renditions of vocal communication calls that fall into 9 call categories of the zebra finch vocal repertoire. We will contrast stimulus information to categorical information, both instantaneous and cumulative for these behaviorally relevant categories of

call-types.

Time-varying information for model neurons.



To validate our approach and to illustrate the behavior of time-varying information

values, we calculated instantaneous and cumulative information for model neurons with

simple and stereotyped response properties. Figure 1 shows the firing rates, raster plots and information values for 3 model neurons in response to 4 stimuli (S1-S4). One model neuron responds to the 4 stimuli with different mean firing rates that are constant in time (Rate Neuron). A second model neuron responds to the four stimuli with the same fixed firing rate but with different latencies (Onset Neuron). The third model neuron also responds with equal average firing rates to the four stimuli but the response occurs at different times (Temporal Neuron).

These simulations illustrate some very basic principles of neural coding. First, many different response profiles can lead to very similar rates of information: in all three simulations, the cumulative information approaches the maximum possible value (2 bits). Second, the coding capacities of neurons are a function of both the range of firing rates that can be achieved (as in the rate neuron) and the modulations in time of this neural activity. Third, the estimation of the instantaneous information gives an incomplete picture of the neural coding of a neuron as it does not incorporate the redundancy or independence of the neural representation over time. For example, on the one hand, comparing the instantaneous information in the Rate neuron to that of the Onset neuron, one might erroneously conclude that the Rate neuron has more information while, in fact, the cumulative information shows that the Onset neuron is more informative at short time scales. On the other hand, one can also observe that the cumulative information in the Rate neuron continues to increase while the firing rate is constant; additional time points allows for a better assessment of that firing rate by time-averaging out neural noise.

These simulations also allowed us to validate our methods. The KDE for the empirical estimation of the time-varying rate based on the generated spike rasters (solid line) gave very good predictions of the actual model rates (dashed lines): over all stimuli and model neurons, the average error was less than 0.02 spikes/ms. Not surprisingly then, using

333 the actual rate versus the estimated rate yielded practically identical results in the information
 334 calculations (dashed vs solid lines). We also checked that the bias corrected estimates were
 335 accurate: the instantaneous information was indeed centered at zero when the response to the
 336 4 stimuli was identical. We verified that the actual values of instantaneous and cumulative
 337 information were correct. For example, in the Temporal neuron a peak instantaneous
 338 information of 0.5 bits is expected as 1 out 4 stimuli will be almost perfectly discriminated.
 339 Finally, we also assessed the limitations of Monte Carlo with importance sampling for the
 340 estimation of the cumulative information. For neurons, with continuously high firing rates,
 341 this estimation can become unreliable at longer integration times as illustrated by the
 342 calculation for the Rate neuron. However, in those cases, the estimate of the standard error
 343 also increased drastically and allowed us to define end points for the calculation of the
 344 cumulative information.

Avian auditory neurons have Poisson statistics.

Since our estimation of the time-varying information values was based on a Poisson parametrization for the distribution of spike counts, we first assessed the validity of this assumption. Given the small number of trials, we could only assess whether the first and second moments (mean and variance) obeyed Poisson statistics. Figure 2 shows the Fano Factor estimated at consecutive time points after stimulus onset and both for individual stimuli (onset) and average across stimuli. The mean Fano factor for the population was 1.04 with SEM of 0.04.

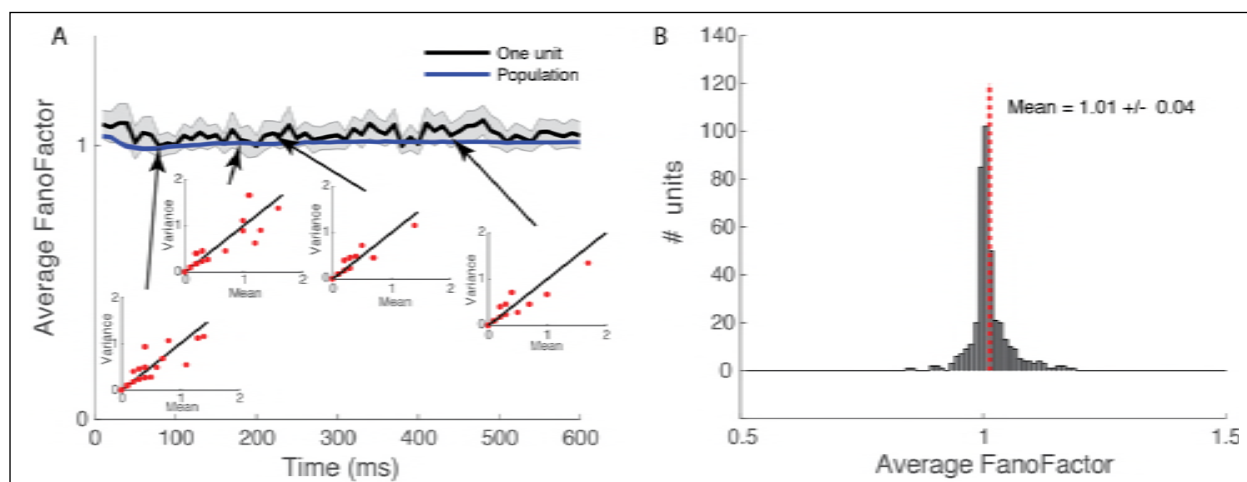


Figure 2. Avian Auditory Neurons have Poisson Statistics. **A.** Time-varying Fano Factor (Variance/Mean) obtained from the empirical distribution of spike counts estimated in successive 10 ms windows. Spike count distributions are obtained from 10 trials and the average Fano Factor from repeating the estimation of mean and variance calculation for all stimuli presented (min=54, max=104). The Fano Factor is shown for one example neuron (solid black line) and also averaged across the entire population (n=404). Error bars are $\pm$ two SEMs. The insets show the mean and variance spike count for the example neuron at a given time shown by the arrows. On those plots, each red dot corresponds to one stimulus (many points overlap). **B.** Distribution of time averaged Fano Factors for the population of neurons (n=404). Fano Factors are not significantly different from 1.

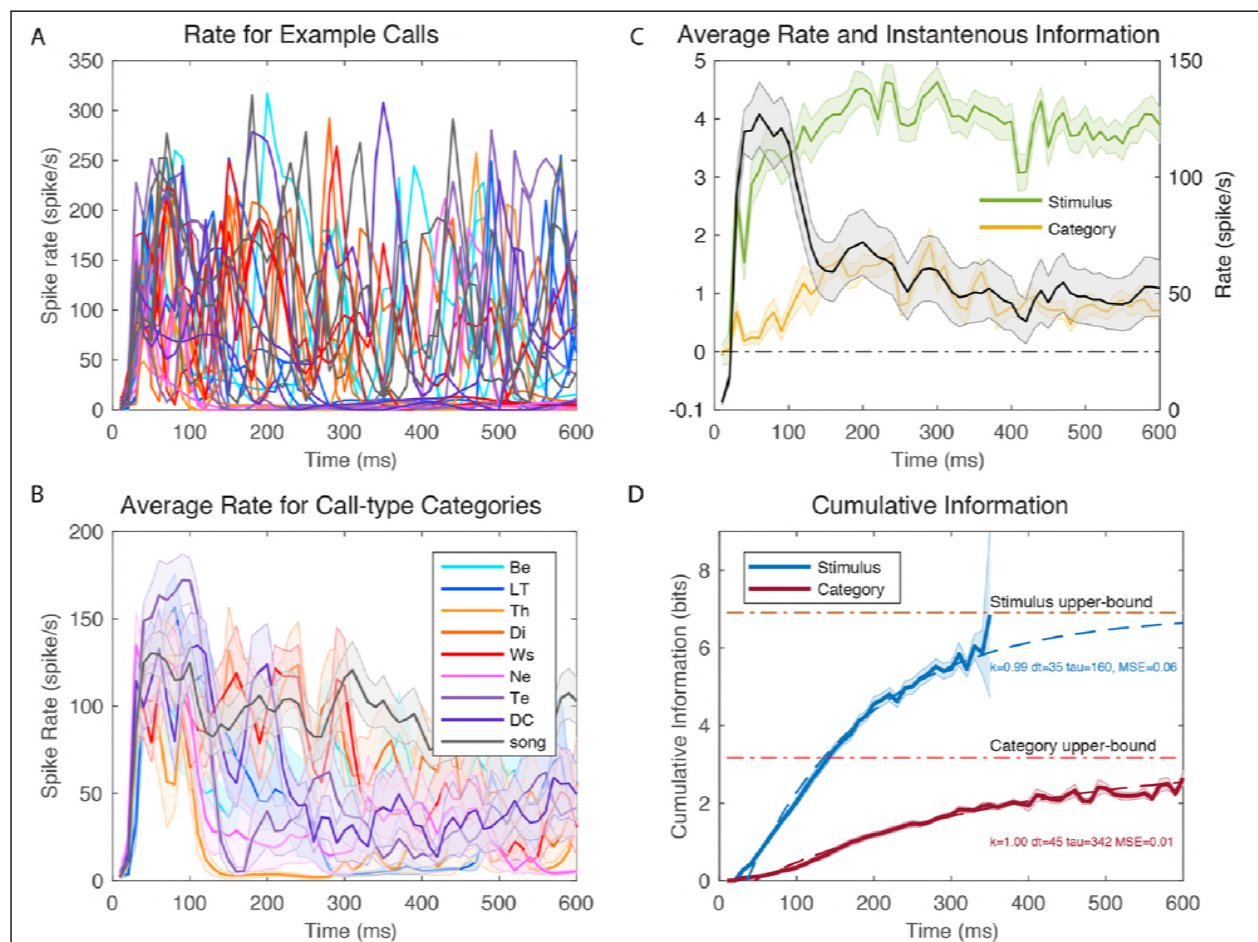
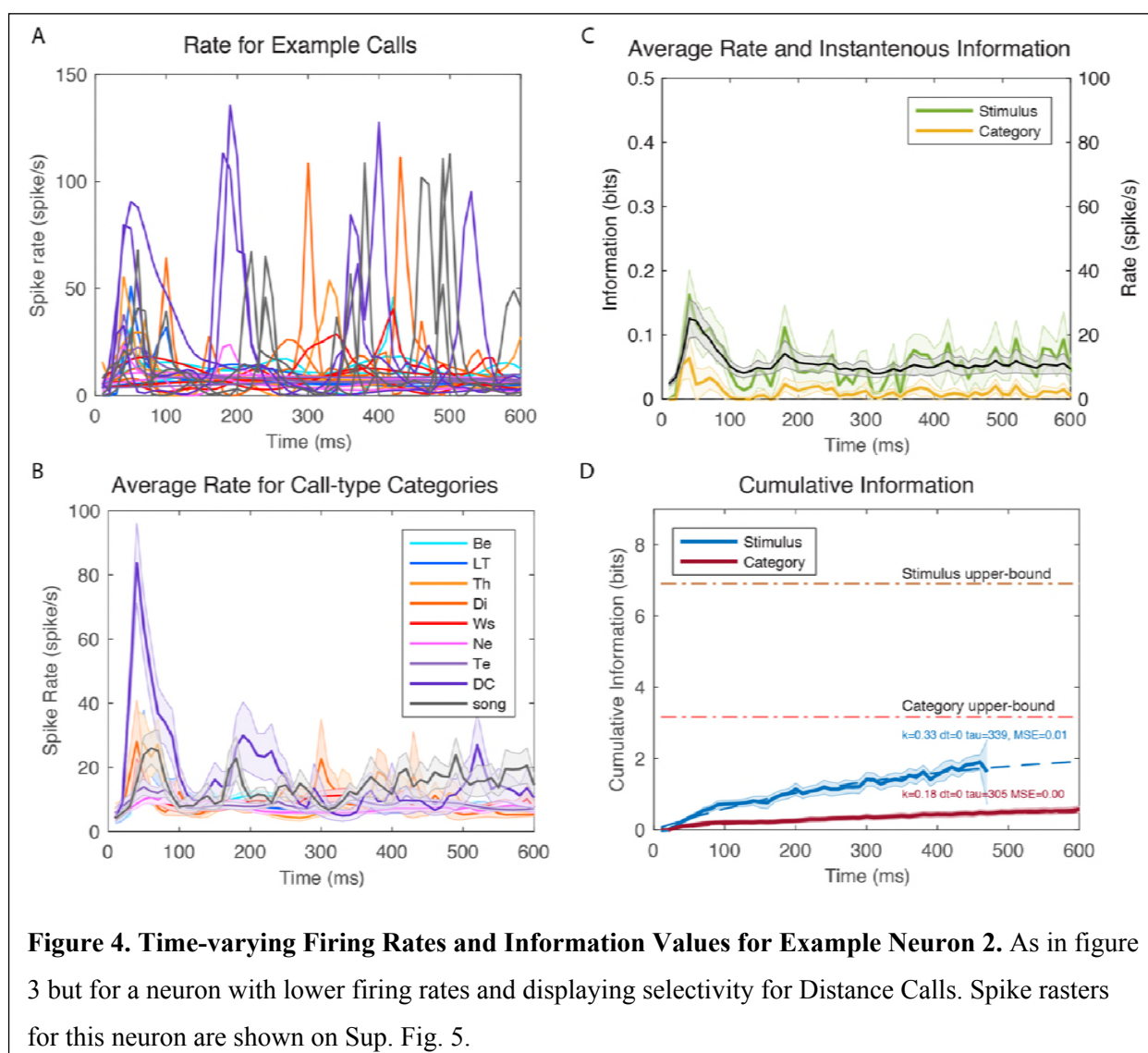


Figure 3. Time-varying Firing Rates and Information Values for Example Neuron 1. **A.** Time-varying firing rate obtained from kernel density estimation (KDE) of ten repetitions of different example stimuli. The plot shows the responses to two examples of calls from each call-type: 18 time series color coded by call-type using the color scheme shown in the legend of B. 0 ms corresponds to stimulus onset. **B.** Averaged time-varying firing rate obtained for each call-type category. These average responses are obtained by averaging the KDE estimates of firing rates for each stimulus belonging to the same call-type (~ 10 example stimuli per call-type). Shaded error bars show $\pm$ one SEM. **C.** The overall (averaged over all stimuli) time-varying firing rate for the same example neuron is shown by the solid black line (right y-axis) with $\pm$ SEM as shaded error bars. The green line and yellow line (left y-axis) correspond to the instantaneous information for stimuli and call-type categories calculated in successive 10 ms windows. Shaded error bars were obtained by a jackknife procedure. **D.** Cumulative information for stimuli and call-type categories for the same example neuron. The shaded error bars are $\pm$ SE obtained by jackknifing and resampling (see Fig. 1 and Methods). Each cumulative information curve is also fitted using an exponential function (dashed lines) characterized by three parameters: a latency (dt, in ms), an exponential rate (tau, in ms) and the infinite time limit value expressed as a fraction of the maximum information achievable (k). The MSE is the mean square error of the fit in bits. This example neuron had a very high stimulus evoked mean firing rate with rapid and reliable time-varying dynamics leading to high information rates. Spontaneous firing was very low (~ 0 spikes/s). Spike rasters for this neuron are shown on Sup. Fig. 4.

354 Time-varying Information for 3 example neurons.

355 In Figures 3-5, we show the time-varying firing rates, the time-varying instantaneous

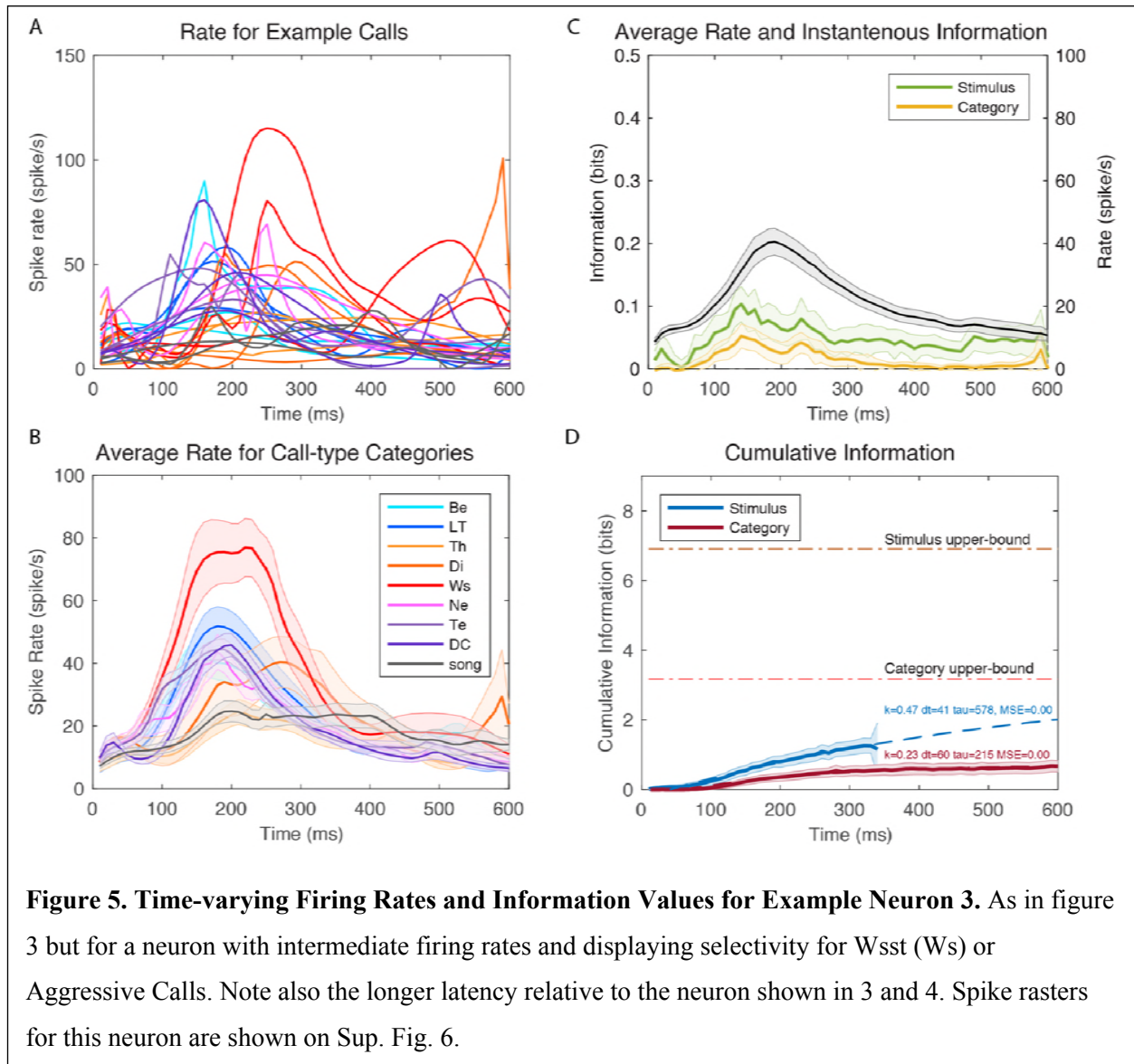


356 information and the cumulative information for 3 zebra finch auditory neurons with distinct

357 response properties. The neuron in Figure 3 responded robustly to all communication calls

358 with high and reliable firing rates. It also responded in a time locked fashion sometimes at

multiple time points for single calls. Although this neuron was not selective for a particular stimulus or call-type category, by combining rate and temporal codes it reached very high levels of instantaneous information. Moreover, this instantaneous information showed little redundancy yielding very high cumulative information. One can also observe, that for this particular neuron, the average time-varying firing rate is not correlated with instantaneous



information. This neuron shows a strong onset response in its firing rate while the instantaneous information is almost constant and even slightly lower during the onset response. The neurons in Figures 4 and 5 have much lower firing rates and exhibit selectivity for a call-type, the Distance Call (DC) and the Wsst Call (Ws) respectively. The neuron in

Figure 4 exhibits both an onset and sustained response both of which are selective for DC. The neuron in Figure 5 has a much longer latency for response with correlated peaks in firing rate and instantaneous information found between 100 and 300 ms after stimulus onset (Example of spike rasters from single trials for these three neurons are shown in Sup. Figs. 4-6).

The cumulative information curves were fitted with an exponential function that allows us to quantify the time constant of information accumulation (τ), the saturation level (k) relative to the maximum information that could be achieved (I_{Max}) and the latency (dt):

$$CI_{mod}(t) = kI_{Max}(1 - e^{-(t-dt)/\tau})$$

Most of the calls in the zebra finch repertoire have durations that are shorter than 300 ms [39]. In order to investigate the cumulative information accumulated at this behaviorally relevant fixed point in time, we estimated the relative cumulative information as the value of the model at 300 ms relative to I_{Max} :

$$k(300) = CI_{mod}(300)/I_{Max}.$$

The results of the fit are shown in dashed lines in the cumulative information plot for each neuron. One can observe that the neuron in Figure 3 has exceptionally high values of saturation: with sufficiently long integrations time, single spike trials could be used to perfectly assess what stimulus (out of those used in the experiment) was heard and, thus, which call-type category it belonged to. The selective neurons in Figures 4 and 5 have much lower saturation levels as expected since they mostly respond to stimuli from a single call-type category. The neuron in Figure 5 has longer latency in its categorical cumulative information but this is not the case for the neuron in Figure 4 that has a rapid yet selective onset response. The high firing neuron (Figure 3) also has a faster time constant τ for the stimulus information than the selective neurons shown in Figures 4-5. All three neurons have

similar and relatively slow time constants for the categorical information in the 200-300 ms range.

Rate and Time-Varying Information: Population Analysis

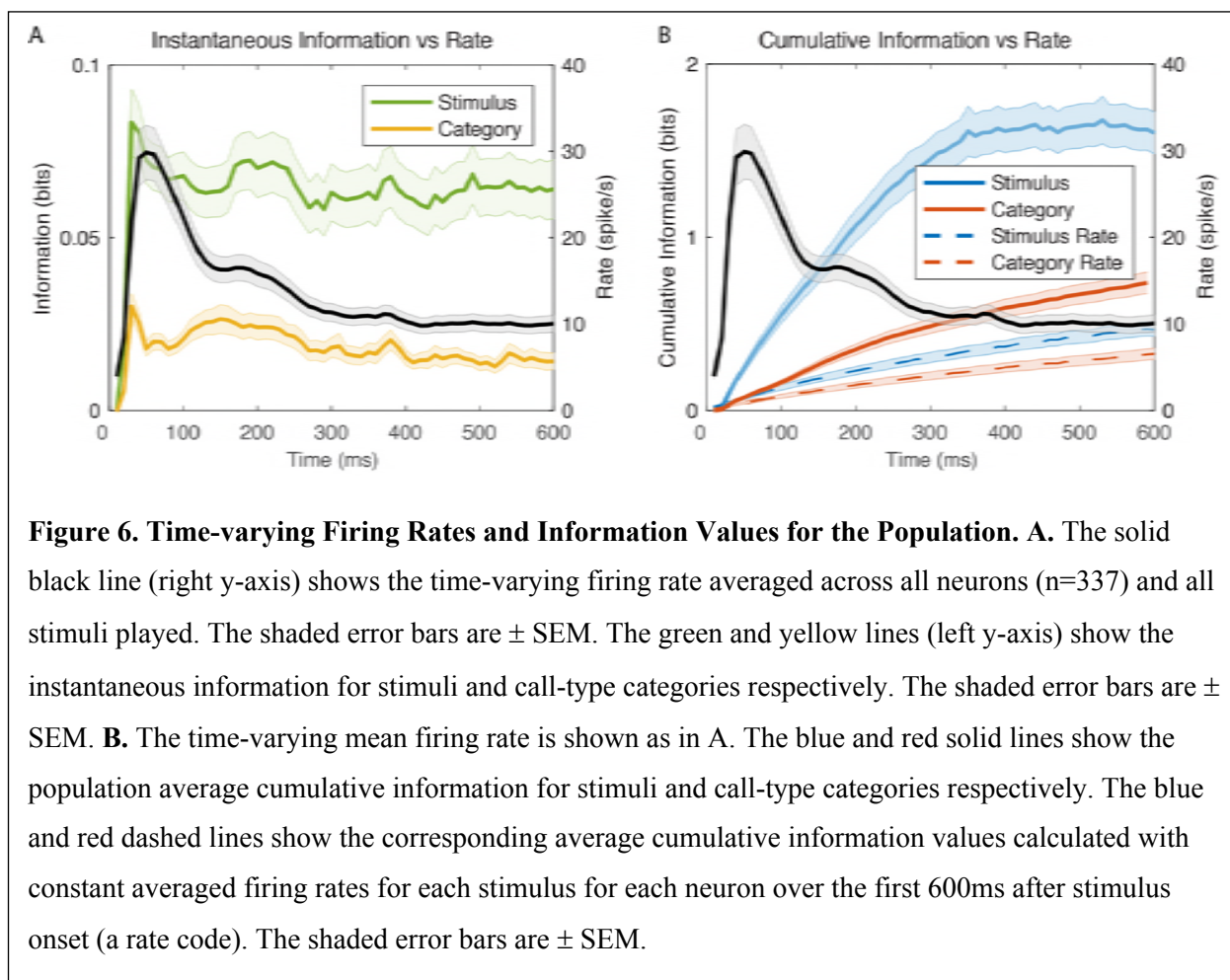


Figure 6A shows the time course of the firing rate averaged across all stimuli and its relationship with the instantaneous information. On average, avian cortical auditory neurons show an onset response followed by a sustained response as observed at many stages of auditory processing and in many sensory systems. The instantaneous information, however, remains almost constant during the entire time. This is true both for the information about individual stimuli or the information about categories. Thus, the onset response is only less selective/informative than the sustained response in bits/spike but not in bits/s; when processing natural vocalizations, the onset and sustained response are equally informative.

Moreover, the information in the sustained response continues to provide new information as reflected by the continuous and relatively fast increase in cumulative information shown in Figure 6B. That rapid rate of increase should be compared to the one observed when the mean firing rate for each stimulus estimated across the entire time-window was used in the calculation (dashed lines in Fig. 6B). As seen in those curves, the increases in the cumulative information for the rate code is much smaller (the information still increases because of noise averaging as explained above). The time-varying rates observed in neural responses (as illustrated in Fig. 3A) provide additional information. How much more? For the coding of individual stimuli (comparing the dashed blue line to the solid blue line in Fig. 6B), a fixed rate code (or assumption) captures only 24% of the information at 100 ms and 21% at 300 ms. The effect for categorical information is smaller because some of the coding dynamics in time-varying responses to stimuli belonging to the same category effectively become neural noise: for categorical information, a fixed rate code captures 50% of the information at 100 ms and 41 % at 300 ms.

The distributions of time constants, τ for the cumulative information for stimuli and call-type categories are shown on Figure 9. The distributions of relative cumulative information at 300ms ($k(300)$) are shown on Figure 10.. The range of time constants observed across the population of neurons was large (100 ms-600 ms) with average time constants of 459 ms for cumulative stimulus information and 372 ms for cumulative categorical information. This difference in means of time constants is statistically significant (Paired t-test $t(214)=3.49$, $p=0.00058$); the ongoing time-varying rate changes continue to provide more information for decoding stimuli and less so for decoding categories of stimuli. There is also a wide distribution of relative cumulative information values ($k(300)$) ranging from close to zero to 0.6 . On average across neurons, the $k(300)$ was 0.2 (or 20%) for stimuli and 0.15 (or 15%) for categories. These differences in relative information values are highly

significant suggesting that, single neurons, capture more variability in stimuli than in categories. Note however that measures of relative information depend on the number of stimuli or categories. Therefore, a direct statistical comparison is not warranted. The comparison between stimulus representation and category representation requires estimations of expected values of categorical information given stimulus information which is performed below. Relative cumulative information values can, however, be used for stimulus and categories independently to assess other coding properties: although average time-varying rates and instantaneous time varying information are not well correlated within single neurons (as shown in Figs. 3 and 6), the relative cumulative information is correlated with average firing rates. For cumulative stimulus information, one finds an increase in $k(300)$ of 1% per spike/s (Adj $R^2=0.36$, $F(1,213)=120$, $p=1.82 \cdot 10^{-22}$) and, for cumulative categorical information, an increase in k_{300} of 0.8% per spike/s (Adj $R^2=0.34$, $F(1,213) = 112$, $p = 2.26 \cdot 10^{-21}$).

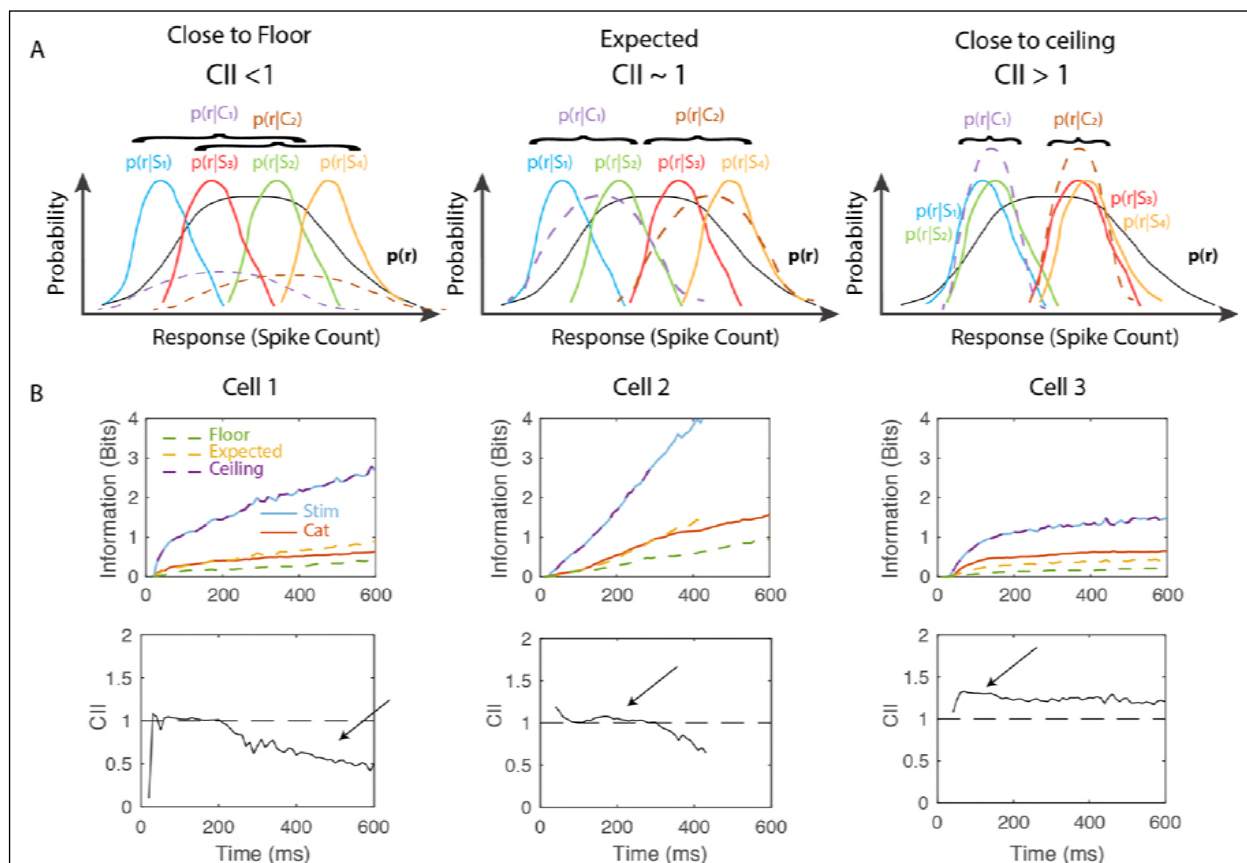
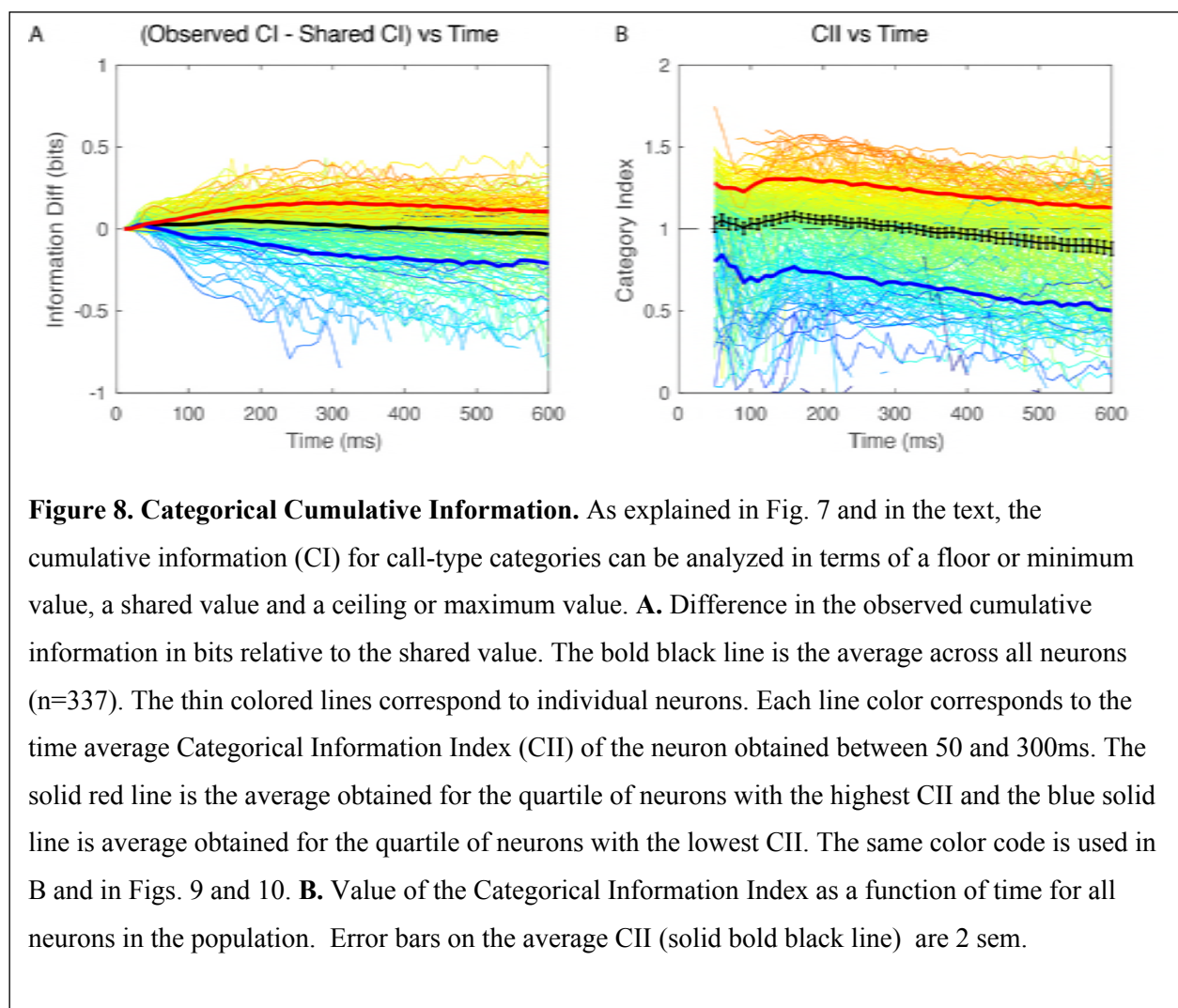


Figure 7. Cumulative Information and the Categorical Information Index. **A.** Schematic of hypothetical response probability distributions that yield different values of the Categorical Information Index (CII). CII is defined by comparing the cumulative categorical information to a floor ($CII = 0$), an expected value predicted from the stimulus information ($CII=1$) and a ceiling value ($CII = 2$) as described in the text and methods. Each plot represents the distributions of hypothetical neural responses for different stimuli (S_1 to S_4) that yield different values of CII. When stimuli are grouped randomly into categories (S_1 and S_2 belong to C_1 , while S_3 and S_4 belong to C_2), CII is close to 0 (left panel). When stimulus categories preserve the order of the stimulus conditioned response distributions and these distributions are equally separated, effectively coding for both stimuli and categories, CII is equal to 1 (middle panel). When stimuli belonging to the same category have identical responses, the stimulus information is equal to the category information, the response being effectively invariant within categories and CII is equal to 2 (right panel). Note that the actual response probability distributions are multidimensional (one dimension for each time lag) with Poisson marginals. **B.** Example of the cumulative information for stimuli and call-type categories and CII for 3 different representative neurons. The blue and red lines are the cumulative information for stimuli and categories respectively. Estimates of the floor (dashed-green), expected (dashed-orange) and ceiling (dashed-purple, here overlapping with blue) values for the categorical cumulative information as described in A are given as well. The CII is shown as a solid line in the second row for each example neuron. Cell 1 has long periods of time with $CII < 1$, Cell 2 has long periods of time with $CII \sim 1$ and Cell 3 has long periods of time with $CII > 1$ as shown by the arrows.

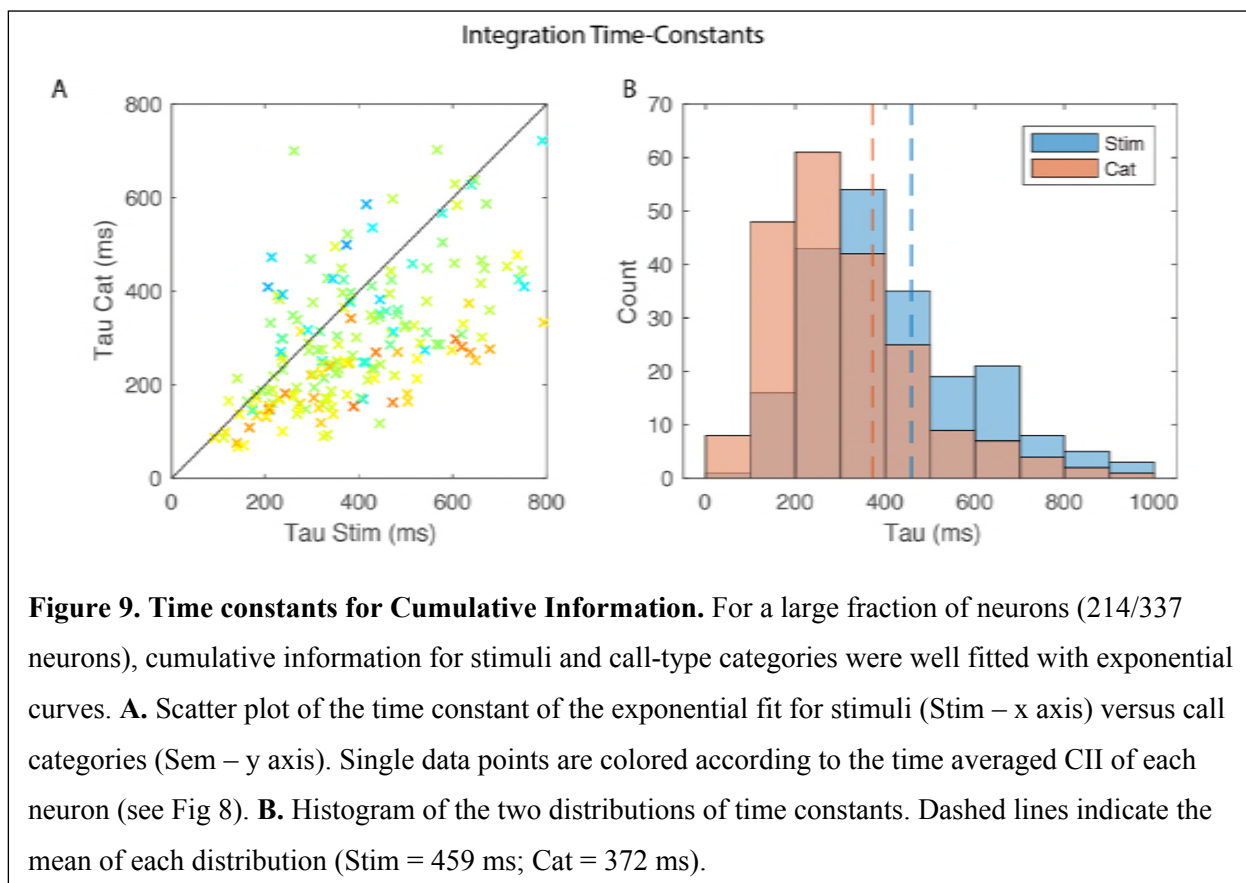
Analysis of the Categorical Information

We are interested in identifying neurons that could play an important role in categorizing vocalizations. We had previously identified example neurons that were highly selective for particular call-types and showed a high degree of invariance such as those shown in Figures 4 and 5 [40]. Here, we attempted to quantify the neural invariance for call renditions within call-type categories along time. For this purpose, we computed a Categorical Information Index (CII). The CII compares the actual categorical cumulative information to three potential values: a floor or minimum value (set at 0), an expected value for shared information between stimuli and categories (set at 1) and a ceiling or maximum value (set at 2). The floor is the categorical information that one would obtain if stimuli are randomly grouped. The shared-information value is the information that one would obtain if the information about stimuli is equally shared across all stimuli and the neural responses for stimuli are perfectly sorted for each natural call-type category; for example, at a given point in time the 10 renditions of DC would elicit the 10 highest rates, the 10 renditions of LT call the next 10 higher firing rates and so forth. This shared-information value does therefore assume that, for a given neuronal signal to noise ratio, neural responses segregate categories maximally while also preserving the maximum discrimination between stimuli within categories. The ceiling value is the categorical information value that one would obtain if a maximum amount of information about stimuli was used for the discrimination of categories and a minimum for discriminating stimuli within categories; it assumes maximum invariance to variations within a category. The ceiling value is equal to the stimulus information until it reaches $\log_2(n_c)$, where n_c is the number of categories, corresponding here to the 9 call-types.

Figure 7B shows the time-varying floor (dashed-green), shared (dashed-orange) and ceiling (dashed-red) values of cumulative categorical information along with the actual stimulus (solid blue) and categorical (solid red) cumulative information for 3 neurons chosen to illustrate CII values that are below 1, around 1 and above 1. Figure 7A shows cartoon probability distributions of neural responses for particular stimuli and particular categories that correspond to the floor, shared and ceiling values.



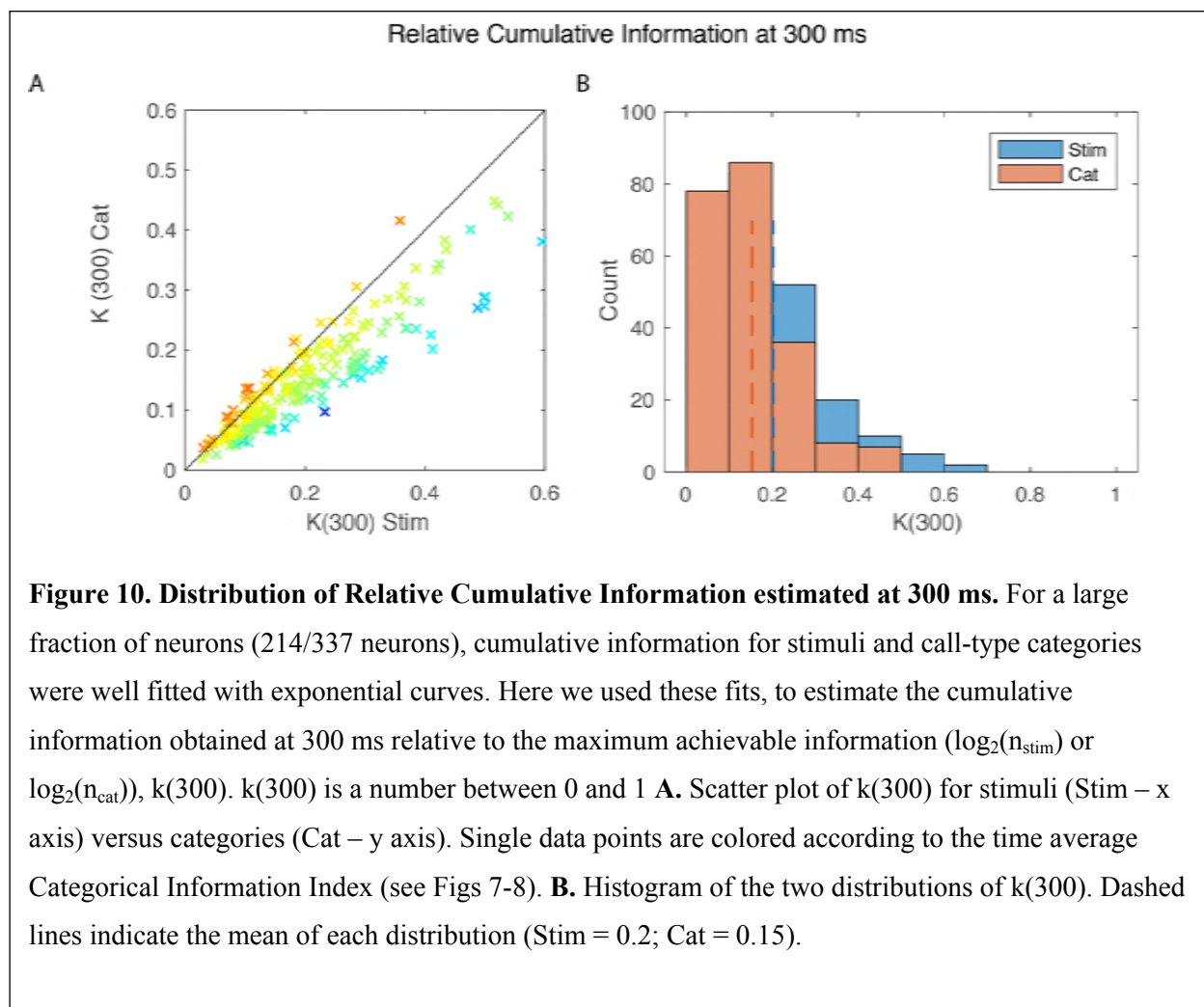
The plots in Figure 8 show the results of this analysis for the population both in absolute information units (left panel) and in the normalized units of CII. The thin colored lines correspond to CII curves for single neurons and they are colored according to the time average CII. The average CII over neurons (bold line on right panel) is very close to the shared value of 1 as one might expect if acoustical differences across stimuli drive neural



responses in a linear fashion along some acoustical feature *and* call categories segregate perfectly along that same acoustical feature. The average CII is slightly (and significantly) above 1 between 120 and 260 ms showing some small degree of average invariance for call renditions within a call-type category during those times. Focusing, on the 25% of neurons with the highest CII, we found a peak at 175 ms. More significantly, perhaps, it is clear that there is a wide distribution of CII around the shared value of 1. However, this distribution includes many neurons that exhibit a high degree of invariance for call renditions within call-type categories as shown by the average CII for the top quartile (red solid line). For that top 25%, the absolute value of additional categorial cumulative information relative to the expected shared information value reaches a maximum of 0.16 bits at 320 ms (Figure 8A).

Do these high CII neurons exhibit other characteristic response properties? In the scatter plots of Figures 9 and 10, we examined the relationship between CII (color coded) and, respectively, time constants and relative level of the cumulative information. It can be

503 seen from Figure 9, that neurons with high CII have relatively long stimulus time constants in
504 comparison to their corresponding time constant observed for categorical information. This
505 relationship is also significant for the entire population of neurons (Linear Regression
506 explaining CII from τ_{cat}/τ_{stim} : Adj $R^2=0.11$, $F(1,213) = 28.5$, $p=2.37 \cdot 10^{-7}$). As shown in
507 Figure 10, neurons with high CII also have higher relative levels of categorical information in
508 comparison to their relative values for stimulus information although this result is expected
509 given our definition of CII (Linear Regression explaining CII from $k_{cat}(300)-k_{stim}(300)$: Adj
510 $R^2=0.76$, $F(1,213) = 675$, $p=5.5 \cdot 10^{-68}$). Finally, one can also notice that neurons with high CII
511 have low values of relative cumulative information (Linear Regression explaining CII from
512 $k_{stim}(300)$: Coef = -0.84 Adj $R^2=0.28$, $F(1,213) = 84$, $p=3.96 \cdot 10^{-17}$). This effect is caused by
513 the correlation between invariance and selectivity as we have shown previously[40]: neurons
514 that show the highest degrees of invariance also tend to respond to a small number of call-
515 type categories and thus a small number of stimuli. As such, high invariance and high
516 selectivity goes hand in hand with lower values of information.



517

518 Anatomical Organization

519 We examined whether neurons were organized in the avian auditory cortex based on
 520 their CII, cumulative information time constants (τ) and saturation levels (k). Most of our
 521 recording sites were identified histologically and could be assigned to avian cortical areas
 522 that had been segregated into the thalamic recipient area, L2; intermediate primary auditory
 523 regions, L1, L3, CML and L; and secondary areas, CMM and NCM [43-46]. We also
 524 obtained spatial x,y,z coordinates of the recording sites relative to the midline, the position
 525 along the rostral-caudal axis where the lamina pallio-subpallialis (LPS) is the most dorsal,
 526 and the top of the brain (Fig. 11).

In all regions, we found a range of CII and thus only weak anatomical trends across areas. Using the L2 dorsal-ventral oblique axis as reference point, CII was slightly higher as one moved rostrally or caudally to higher regions of auditory processing (Linear Regression: Adj $R^2 = 0.03$, $F(3,194) = 4.19$, $p = 0.016$). However, an ANOVA also suggested that both regions NCM and L2 had slightly higher mean CII (Adj $R^2 = 0.027$, $F(4,185) = 2.77$, $p = 0.042$). We also observed an increase in the time constant (τ) for the stimulus cumulative information that parallel the increase in CII as one moved away from the L2 axis (Linear Regression: Adj $R^2 = 0.03$, $F(3,194) = 4.79$, $p = 0.0093$). These significant anatomical trends were characterized by very small effect sizes. We also did not find any anatomical organization of saturation constants (k) for the cumulative information; neurons that had high levels of information could be found next to neurons with much lower levels and similar levels of average information were found in all regions.

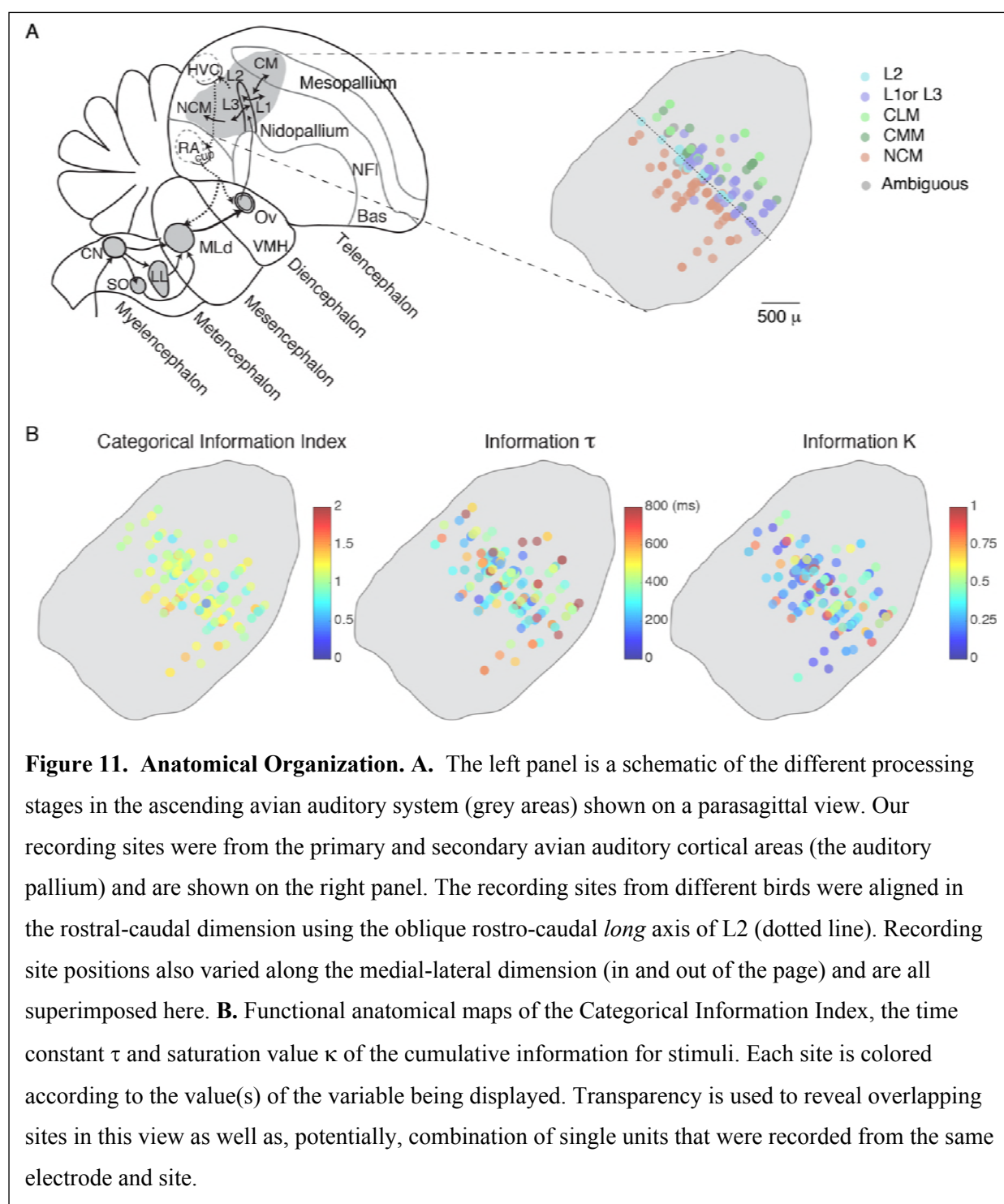


Figure 11. Anatomical Organization. A. The left panel is a schematic of the different processing stages in the ascending avian auditory system (grey areas) shown on a parasagittal view. Our recording sites were from the primary and secondary avian auditory cortical areas (the auditory pallium) and are shown on the right panel. The recording sites from different birds were aligned in the rostral-caudal dimension using the oblique rostro-caudal *long* axis of L2 (dotted line). Recording site positions also varied along the medial-lateral dimension (in and out of the page) and are all superimposed here. **B.** Functional anatomical maps of the Categorical Information Index, the time constant τ and saturation value κ of the cumulative information for stimuli. Each site is colored according to the value(s) of the variable being displayed. Transparency is used to reveal overlapping sites in this view as well as, potentially, combination of single units that were recorded from the same electrode and site.

Discussion

We modeled responses that are observed in the auditory system as inhomogeneous Poisson processes in order to estimate the time-varying instantaneous and cumulative information for vocalizations used in communications. We showed that using Kernel Density Estimation for the time-varying firing rate and Monte Carlo with importance sampling for estimating probabilities, we were able to obtain accurate and bias-free estimates of these time-varying information values. This parametric approach is powerful because a relative small number of trials can be used to estimate the time-varying response and thus information values can be estimated in response to a relative large set of stimuli (here over 100 distinct vocalizations) given typical recording times. Poisson statistics were observed in the auditory cortical neurons recorded here when they were stimulated with short natural communication calls. More generally, the same procedures can also be used with spike trains statistics that can be parametrized with probability functions that depend only on a time varying rate such as inhomogeneous gamma or inhomogeneous inverse Gaussian [12, 47] and could also be extended to ensemble of neurons. Although Poisson statistics are often observed in neural data and will correctly fit any data set obtained from pooling responses across a large number of trials [48], it makes the strong assumption that the firing rate for a particular trial for a particular neuron depends only on time. Refractory period in spiking neurons and other correlations in single neurons or in an ensemble of neurons that are not phased locked to the stimulus (i.e. noise correlations) are common violations of this assumption. Taking noise correlations into account can increase stimulus decoding accuracies [23]. In cases where noise correlations are informative, the method proposed here would only yield lower bound estimates of information theoretic values. Alternatively, one should try the use of spike metrics measures [49] in combination with stimulus decoding approaches to then obtain measures of information from the confusion matrix of predicted versus actual stimuli [13, 40,

50, 51]. If spike metrics can be estimated accurately (with limited data), repeating the stimulus decoding procedure for progressively longer time windows would yield cumulative information values such as those calculated here but which could consider noise correlations.

Beyond our methodological contribution, the principal goal of this analysis was to characterize the neural code in higher auditory areas for communication calls. We found that, on average, auditory cortical neurons responded to these natural stimuli with time-varying firing rates that exhibited an onset and a sustained component of the response. Although in some situations the auditory cortex appears to respond only transiently [26, 52], our data supports findings from mammalian species that have showed strong sustained responses when neurons are driven by preferred stimuli [21]. Similarly, in the human superior temporal gyrus, the sustained responses to speech have been shown to be more informative for decoding speech phonemes [53]. Natural sounds in general have also been shown to be particularly efficient at driving auditory areas [12, 54-59] and thus the presence of informative sustained responses is not surprising. Indeed, in some of our most selective neurons, such as the neuron in Figure 5, the onset response is missing and only the sustained response is observed. More generally, and on average, we found that both the transient and sustained response had information about the stimulus identity *and* that the information in these two response phases was not redundant: the stimulus space of natural vocalizations is very large and although the initial response provides some clue as to the nature of the vocalization, new and additional information is observed in the sustained response. On the one hand, one might argue that, in natural vocalizations, the sound itself is changing with time. These stimulus changes occur both within calls that are made of composite notes or in call bouts and song motifs made of multiple syllables. Sustained responses in such cases could be interpreted as multiple onset responses (but with decreasing amplitude). On the other hand, from a behavioral perspective, these communication calls correspond to a single

auditory object: a message from a particular individual about a particular state. From either perspective, these observations and analyses illustrate the importance of using behaviorally relevant stimuli when analyzing the nature of the neural code.

The information about stimulus identity was shown to be approximately equal in bits/s in the onset and sustained response. Given that the average onset response (in spikes/s) is greater than the sustained response, one might conclude that the onset coding (in bit/spikes) is not as efficient. Although, this is true from the point of view of a single neuron, it is almost certainly not true when ensemble of neurons is considered; the relative timing of the first spike has been shown both in audition [26] and in other sensory modalities [25, 27] to be highly informative. Moreover, in addition to stimulus identity other stimulus features are also encoded in neural responses; the timing of the stimulus is clearly marked by the onset response or transient response [22, 53, 60, 61] but other stimulus attributes such as the location, fundamental and loudness/distance of the sound source are also processed in auditory cortex [62-64]. It is therefore very likely that the onset response contains information about stimulus features other than stimulus identity or category and, potentially to a greater extent than in the sustained response (e.g. relative timing of onset). We note, however, that the opposite was found for the neural discrimination of two vowel sounds in the ferret auditory cortex where the onset response or early response was the most informative for decoding vowel identity relative to other attributes [64].

Our IT analysis also revealed the importance of stimulus locked spike patterns even in relatively short neural responses: in just 100 ms, the mean rate captured only a quarter of the information present when time-varying firing rates are considered. Thus, our analyses provide additional evidence that spiking patterns carry a significant amount of stimulus information and should therefore not be ignored in the analysis of neural responses [65]. The spiking precision analyzed here was relatively coarse (10 ms windows) and matched to the

time scales of the relevant dynamics in the stimulus: although zebra finch calls are much longer than 10 ms, they are complex sounds with fast spectro-temporal features [39]. Therefore, although the neural code observed here uses fast varying spike rates, it cannot be labelled as a temporal code. In a rigorous definition of a temporal code, temporal information in spike patterns must code stimulus attributes other than the stimulus dynamics [66].

The cumulative information for stimulus identity or for categories increased for sustained periods of time before saturating. These saturating curves were well fitted with exponentials and yielded relatively long-time constants of approximately 460 ms for stimuli identity and 370 ms for call categories. These *information* time constants are long in comparison to the integration times that are usually found for auditory cortical neurons; the spectro-temporal receptive fields of avian auditory neurons rarely extend beyond 50 ms [67-70] although adaptive responses on longer time scales have also been described [71-73]. Information time constants depend on the integration and adaptation time constants of neurons but *also* on the stimulus dynamics: although stimulus identity and stimulus category are fixed in time, the sound itself has time varying features that can provide additional information as time goes on. It is the triple combination of the dynamics of the natural stimulus statistics, the neuronal integration time and the neuronal signal to noise levels that are going to affect the information time constant. Natural sounds and in particular communication calls are informative objects not only because of their spectral structure but also their rich dynamical structure. The importance of time in the neural code used in the auditory system has been emphasized multiple times [74-78] and our cumulative information analysis further stresses the importance of using natural sounds or synthetic sounds that carefully match natural dynamics when probing auditory neural encoding. Ultimately, it is an information time constant that includes stimulus dynamics that should be compared to behavioral responses [18]. A behavioral assay of reaction time for all the calls in the zebra

finch repertoire has not been performed but the values of a few hundred of ms correspond to the shortest time intervals between call and call back in anti-phonal calling in paired zebra finches [79]. Although, faster times could be obtained in reaction to any sounds (e.g. startle or orientation), the processing of stimulus identity to extract the information on who is calling and what is being said might require these longer processing times.

Finally, we quantified the fraction of the information about stimulus identity that could be used for extracting the call-type category. We used that analysis to characterize neurons not only in terms of their absolute coding capacities for call-type categories but also in terms of their invariance in their response to different stimuli belonging to the same call-type category. We found that, on average, information for categories is very close to what one would expect if neural representation for stimulus identity is also segregated along call-type categories. However, we also found neurons that had a high degree of invariance and could therefore be classified as categorical. On the one hand, these categorical and invariant neurons had even longer time constants for cumulative stimulus information suggesting that they could be higher in the auditory processing stream. On the other hand, we did not find a separate population of invariant neurons across all of our recordings of neurons informative for categories: the distribution of our categorical index was unimodal. Moreover, we found only weak anatomical correlations with very small increases in the Categorical Information Index along an anatomical axis corresponding to lower vs higher auditory areas. Higher avian auditory cortical areas have been associated with more complex spectro-temporal receptive fields [67, 70, 80], increase robustness to noise [81-83] and more specificity for processing natural sounds [12, 84]; but, it is also striking to see that in all those analyses as well as those that focused on single avian auditory cortical areas [85] there is a high degree of heterogeneity in the neural responses within each area. Here, we also found that neurons with high Categorical Information Indices could be found anatomically next to neurons with low

Categorical Information Indices. Some of this functional heterogeneity could be associated with different cell types [69, 84] and a better understanding of the micro-circuitry in the avian auditory cortical areas is needed [30]. It is interesting to note, however, that this mixing of low level and high-level response properties is not unique to the avian auditory system as similar heterogeneity has been found in the mouse [86] and ferret auditory cortex [87, 88]. Contrary to the visual system, the auditory system might preserve a higher mixture of low-level and high-level sensory responses properties at multiple stages of processing including the higher auditory areas involved in auditory object recognition. If this is true and universally found in vertebrates, it might be a necessary property of the computations needed for auditory object recognition, potentially related to the fact that complexity in auditory signals is in time-varying spectral patterns that quickly disappear; the fleeting nature of sounds could prevent higher processing stages from subsequently accessing lower-level representations for additional information.

Methods

Animals and Stimuli

Four male and two female adult zebra finches (*Taeniopygia guttata*) were used for the electrophysiological experiments. The birds were bred and raised in family cages until they reached adulthood, and then maintained in uni-sex groups. Although birds could only freely interact with their cage-mates, all cages were in the same room allowing for visual and acoustical interactions between all birds in the colony. All birds were given seeds, water, grid and nest material ad libitum and were supplemented with eggs, lettuce and bath once a week. All animal procedures were approved by the Animal Care and Use Committee of the University of California Berkeley and were in accordance with the NIH guidelines regarding the care and use of animals for experimental procedures.

Vocalizations used as stimuli during neurophysiological experiments were recorded from 15 adult birds and 8 chicks (20-30 days old). The vocalization bank obtained contains 486 vocalizations that included for each bird most of the calls in the Zebra finch repertoire: 7 call-types in adults and 2 in chicks. The adult calls included the following affiliative calls: Song (So), Distance Call (DC), Tet call (Te) and Nest Call (Ne); and the following non-affiliative calls: Wsst or aggressive call (Ws), the Distress Call (Di) and one of the two alarm calls, the Thuk (Th). The juvenile calls included the Begging call (Be) and the Long Tonal Call (LT). Additional information about these stimuli and their behavioral meanings can be found in [39, 40, 89].

For the neurophysiological experiments, a new subset of the vocalization bank was used at each electrophysiological recording site. This subset was made from a representative number of vocalizations from the repertoire of individuals: three adult females, three adult males, two female chicks and two male chicks. From each individual caller, we randomly chose 3 call bouts from each category or fewer if fewer than 3 call bouts were obtained for

that particular call-type and individual. The maximum number of stimuli that could be selected in that procedure was therefore $3 \times 7 \times 3$ (adult males x repertoire x renditions) + $3 \times 6 \times 3$ (adult females x repertoire x renditions) + $4 \times 2 \times 3$ (juveniles x repertoire x renditions) = 141. Fewer stimuli were used when we had fewer renditions than 3 for a particular bird or when the signal from a single unit was lost before the end of the recording. The average number of stimuli played for each single unit was 114 (sd = 22, min = 34, max = 123). Ten trials were acquired for each stimulus with a few exceptions (min=9, max = 11). Sounds were broadcasted in a random order using an RX8 processor (TDT System III, sample frequency 24414.0625 Hz) connected to a speaker (PCxt352, Blaupunkt, IL, USA) facing the bird at approximately 40cm. The sound level was calibrated on song stimuli to obtain playbacks at 75dB SPL measured at the bird's location using a sound meter (Digital Sound Level Meter, RadioShack).

Neurophysiological and Histological Procedures.

Extra-cellular electrophysiological recordings were performed in 6 urethane anesthetized adult zebra finches. The birds were placed in a sound-attenuated chamber (Acoustic Systems, MSR West, Louisville, CO, USA) and sound presentation and neural recording were performed using custom code written in TDT software language and TDT hardware (TDT System III). Sounds were broadcasted in a random order as described above. Neural responses were recorded using the signal of two (5 subjects) or one (1 subject) 16-tungsten electrode arrays, band-pass filtered between 300Hz and 5kHz and collected by an RZ5-2 processor (TDT System III, sample frequency 24414.0625 Hz). The electrode arrays consisted of two rows of 8 electrodes with row separation of 500 mm and inter-electrode separation within row of 250 mm. Electrode impedances were approximately 2 MOhms. When two electrode arrays were used, they were placed each in one hemisphere. Spike arrival times and spike shapes of multiple units were obtained by voltage threshold. The level

of the threshold was set automatically by the TDT software using the variance of the voltage trace in absence of any stimuli. Electrodes were progressively lowered and neural responses were collected as soon as auditory responses to song, white noise, Distance call or limited modulation noise could be identified on half of the electrodes in each hemisphere (the stimuli used to identify auditory neurons were different from the stimuli used in the analysis).

Several recording sites were randomly selected by progressively deepening the penetration of the electrodes and ensuring at least 100 μm between two sites. On average 4.2 ± 2 sites (mean $\pm$ sd) were recorded per bird and per hemisphere at a depth ranging from 400 μm to 2550 μm .

After the last recording site, the subject was euthanized by overdose of isoflurane and transcardially perfused. Coronal slices of 20 μm obtained with a cryostat were then alternatively stained with Nissl staining or simply mounted in Fluoroshield medium (F-6057, Fluoroshield with DAPI, Sigma-Aldrich). While Fluoroshield slices were used to localize electrode tracks, Nissl stained slices were used to identify the position of the 6 auditory areas investigated here: the three regions of Field L (L1, L2 and L3), 2 regions of Mesopallium Caudale (CM): Mesopallium Caudomediale (CMM) and Mesopallium Caudolaterale (CLM); and Nidopallium Caudomediale (NCM). By aligning pictures, we were able to anatomically localize most of the recording sites and calculate the approximate coordinates of these sites. Since we could not localize the Y-sinus on slices, we used the position of the Lamina Pallio-Subpallialis (LPS) peak as the reference point for the rostro-caudal axis in all subjects. The surface of the brain and the midline were the reference for respectively the dorsal-ventral axis and the medial-lateral axis. The approximate coordinates of units were used to build 3-D reconstructions of all single unit positions in an hypothetic brain.

Single unit isolation was performed off-line using custom software that used a combination of supervised and unsupervised clustering algorithms. These clustering algorithms used the spike-snippets shape as described by a PCA. Sorted units were declared

to be single units based on spike shape reliability across snippets. The spike shape reliability measure was a signal to noise ratio (SNR) where the signal was the difference between the maximum and the minimum of the average snippet and the noise was the standard deviation of this measure across all snippets. Single units in our data set have an $\text{SNR} > 5$. Additional details on these experimental procedures can be found in [40].

Data Analysis: Time-varying Spike Rate Estimation

The data of neural responses from 404 out of 914 isolated single units were used in this study. The 404 were selected based on a prior analysis that showed that this subset of units were not only auditory but also contained information about call-types, in the sense that call-types could be decoded above chance level from neural responses (see [40]). Here, we analyzed the neural response in the first 600ms after stimulus onset.

For each stimulus, the 9 to 11 raw spike patterns of 600ms, sampled at 10kHz, were combined to obtain the corresponding time varying spike rate (sample frequency set at 1kHz) by applying a locally adaptive kernel bandwidth optimization method [41]. In cases where the neuron did not respond to any of the presentations of the stimulus or responded only once over all presentations, the rate was estimated as being constant for the 600ms duration of the neural response. For those two unresponsive cases, the rate was set to be $1/(2 \cdot N_{\text{trials}} \cdot N_{\text{times}})$ in the absence of any spike or $1/(N_{\text{trials}} \cdot N_{\text{times}})$ in case of one spike, with N_{trials} the number of stimulus presentations (9-11) and N_{times} the number of time points at which the rate was estimated (here 600, for a 600ms neural response section with a sampling rate set at 1kHz).

Calculations of cumulative information requires the estimation of very large distributions which sizes grow exponentially with the number of time points investigated. To investigate cumulative information values up to 600ms after stimulus onset, time-varying

rates were sampled at 10ms (Nyquist limit frequency of 50Hz). To estimate, the amount of information potentially lost by this low-pass filtering, we estimated an information value based on coherence analysis of the signal to noise ratio in the raw spike train. The coherence between a single spike train (R) and the actual time-varying mean response (A) γ_{AR}^2 can be derived from the coherence between the peristimulus time histogram (PSTH) obtained from half of the trials and the PSTH obtained from the other half [90].

$$\gamma_{AR}^2 = \left[1 - \frac{M}{2} \times \left(1 - \sqrt{\frac{1}{\gamma_{R_1, M/2}^2 \gamma_{R_2, M/2}^2}} \right) \right]$$

where M the total number of trials (presentations of the stimuli) and $\gamma_{R_1, M/2}^2 \gamma_{R_2, M/2}^2$ the coherence between the two PSTHs calculated on half of the trials. The coherence between two responses is a function of frequency (ω). An estimate of the mutual information (in bits per second) between R and A responses can then be obtained by integrating over all frequencies [4, 90]:

$$I_{AR} = - \int_0^\infty \log_2 [1 - \gamma_{AR}^2(\omega)] d\omega$$

For each unit, we estimated the percentage of information preserved as the ratio between I_{AR} calculated up to 50Hz and I_{AR} calculated over all frequencies. Over all units, 96.7% $\pm$ 6.9% (mean $\pm$ SD) of information was conserved by a lowpass filtering at 50Hz and only 88 out of 404 cells had information losses greater than 5%. For each unit, we also calculated the proportion of cumulative power across frequencies in the time varying spike rate estimation obtained with the KDE before low-pass filtering and down-sampling (averaged periodogram in overlapping 200 ms Hanning windows and 1 kHz sampling rate). The cumulative sum of the power was calculated across frequencies and normalized by the maximum power value to obtain the proportion of cumulative power. On average across

units, the cumulative power reached $98.8\% \pm 1.6\%$ (mean $\pm$ SD) at 50Hz, further validating our choice of the temporal resolution (Supp Fig. 1).

Information Theoretic Calculations.

As described in the results, the time-varying *instantaneous* mutual information between the stimulus S and the response Y_t can be written as a difference in Shannon entropies:

$$I_t = H(Y_t) - H(Y_t|S)$$

and the *cumulative* mutual information for neural responses that are discretized into time intervals is given by: $CI_t = H(Y_t, Y_{t-1}, Y_{t-2}, \dots, Y_0) - H(Y_t, Y_{t-1}, Y_{t-2}, \dots, Y_0 | S)$

In the present paper we calculated 4 different types of information: the stimulus instantaneous information, the categorical instantaneous information, the stimulus cumulative information and the categorical cumulative information. Stimulus instantaneous and cumulative information were calculated for all 404 units, while categorical instantaneous and cumulative information were calculated on a restricted set of 337 neurons that presented at least one time point with a significant value of stimulus cumulative information (significant threshold set as 3 times the local error, see below for error calculations). While we verified our assumptions (minimal information loss with spike rate binning, Poisson distributions of spike counts and maximum value of spike counts) on the full set of 404 units, the population analysis of time-varying information presented in the results section only include the relevant dataset of 337 neurons.

The custom Matlab code used to calculate time-varying information values is available at <https://github.com/julieelie/PoissonTimeVaryingInfo> along with a tutorial on how to use the core functions.

828 The Instantaneous Information

829 The conditional response entropy and the response entropy are obtained from the distribution
830 of the conditional probability of neural responses given the stimulus, $p(y_t|s)$, and the
831 distribution of probability of each stimulus $p(s_i)$:

$$832 \quad H(Y_t|S) = \sum_i p(s_i) \sum_{y_t=0}^{R_{Max}} -p(y_t|s_i) \log_2 p(y_t|s_i)$$

$$833 \quad H(Y_t) = \sum_{y_t=0}^{R_{Max}} -p(y_t) \log_2 p(y_t)$$

$$834 \quad \text{with } p(y_t) = \sum_i p(s_i) p(y_t|s_i)$$

835

836 We modeled the distribution of neural responses to a given stimulus s_i as an inhomogeneous
837 Poisson process. The conditional probability of response (spike count) given the stimulus is
838 then:

$$839 \quad p(y_t|s_i) = \frac{\mu_{s_i}(t)^{y_t}}{y_t!} e^{-\mu_{s_i}(t)}$$

840 And the local entropy is:

$$841 \quad H(Y_t|s_i) = \mu_{s_i}(t) [1 - \log_2 (\mu_{s_i}(t))] + e^{-\mu_{s_i}(t)} \sum_{y_t=0}^{R_{Max}} \frac{\mu_{s_i}(t)^{y_t} \log_2 (y_t!)}{y_t!}$$

842 Because, in our data, the probability of response is very small for high values of y_t ,
843 calculations of entropies were bounded for y_t between zero and R_{Max} . Here we set $R_{Max} = 20$
844 which corresponds to a rate of 2 spike/ms in the 10 ms analysis windows. The maximum rate
845 observed in all of our neurons across all time bins was 0.8 spike/ms (Sup. Fig. 2). Note that
846 this very high firing rate (800 Hz) was only observed only once; that is in one 10 ms time
847 windows across all neurons (404) and all stimuli ($114 \times 60 = 6840$) or with a $p = 1/2763360$.
848 This is clearly the very upper limit of a distribution with a long tail. Our time-varying rates

were well below that upper bound but we verified that the cumulative probability up to R_{Max}
 $= 20$ was numerically identical to 1 before estimating entropies. As described above (see
time-varying spike rate estimation), we also enforced a lower bound for $\mu_{s_i}(t)$ of $1/20$.

The total conditional entropy at time t is:

$$H(Y_t | S) = \sum_{i=1}^{n_s} \sum_{y_t=0}^{R_{Max}} -p(s_i)(p(y_t|s_i)\log_2 p(y_t|s_i))$$

where $p(s_i)$ is the probability of observing stimulus s_i and n_s is the number of stimuli
sampled: it is the average of the local entropies obtained for the conditional probability of
response for each stimulus s_i . Assuming that our sample is representative of stimuli
encountered, each stimulus is equally probable, $p(s_i) = \frac{1}{n_s}$ or:

$$H(Y_t | S) = \frac{1}{n_s} \sum_{i=1}^{n_s} H(Y_t | s_i)$$

Alternatively, one can assume that each call category is equally probable. If k_i
is the number of stimuli in the particular call category c to which s_i belongs and n_c is the
number of categories, then:

$$p(s_i) = \left(\frac{1}{k_i}\right)\left(\frac{1}{n_c}\right)$$

for $s_i \in c$. Then:

$$H(Y_t | S) = \frac{1}{n_c} \sum_{i=1}^{n_s} \frac{1}{k_i} H(Y_t | s_i)$$

The unconditional probability (i.e. across all stimuli) of a response at time t is not Poisson
but is given by a weighted sum of Poisson with different mean rates:

$$p(y_t) = \sum_{i=1}^{n_s} p(s_i) p(y_t | s_i)$$

$$p(y_t) = \frac{1}{n_c} \sum_{i=1}^{n_s} \frac{1}{k_i} \left[\frac{\mu_{s_i}(t)^{y_t}}{y_t!} e^{-\mu_{s_i}(t)} \right]$$

869

870 The response entropy is obtained from these unconditional probabilities:

$$H(Y_t) = - \sum_{y_t=0}^{R_{Max}} p(y_t) \log_2(p(y_t))$$

$$H(Y_t) = - \sum_{y_t=0}^{R_{Max}} \left[\frac{1}{n_c} \sum_{i=1}^{n_s} \frac{1}{k_i} \left[\frac{\mu_{s_i}(t)^{y_t}}{y_t!} e^{-\mu_{s_i}(t)} \right] \right] \log_2 \left[\frac{1}{n_c} \sum_{i=1}^{n_s} \frac{1}{k_i} \left[\frac{\mu_{s_i}(t)^{y_t}}{y_t!} e^{-\mu_{s_i}(t)} \right] \right]$$

873

874 The Cumulative Information.

875 The conditional probability of a time varying response is the joint probability of
876 observing $(y_t, y_{t-1}, y_{t-2}, \dots)$ given s_i . Given our Poisson assumption, the conditional
877 probability of response at t is independent of the conditional response at previous times.

$$p(y_t, y_{t-1}, y_{t-2}, \dots | s_i) = p(y_t | s_i) p(y_{t-1} | s_i) p(y_{t-2} | s_i) \dots$$

879 We can show that the conditional entropy of the joint responses is the sum of the individual
880 entropies:

$$\begin{aligned} & H(Y_t, Y_{t-1}, Y_{t-2}, \dots | S) \\ &= \frac{1}{n_c} \sum_{i=1}^{n_s} \frac{1}{k_i} \sum_{y_t=0}^{\infty} \sum_{y_{t-1}=0}^{\infty} \sum_{y_{t-2}=0}^{\infty} \dots (p(y_t | s_i) p(y_{t-1} | s_i) p(y_{t-2} | s_i) \dots) \\ & \log_2 \{p(y_t | s_i) p(y_{t-1} | s_i) p(y_{t-2} | s_i) \dots\} \end{aligned}$$

882

$$\begin{aligned} & H(Y_t, Y_{t-1}, Y_{t-2}, \dots | S) \\ &= \frac{1}{n_c} \sum_{i=1}^{n_s} \frac{1}{k_i} \left\{ \sum_{y_t=0}^{\infty} (p(y_t | s_i) \log_2 p(y_t | s_i)) + \sum_{y_{t-1}=0}^{\infty} \right. \\ & \left. (p(y_{t-1} | s_i) \log_2 p(y_{t-1} | s_i)) + \sum_{y_{t-2}=0}^{\infty} (p(y_{t-2} | s_i) \log_2 p(y_{t-2} | s_i)) + \dots \right\} \end{aligned}$$

$$H(Y_t, Y_{t-1}, Y_{t-2}, \dots | S) = H(Y_t | S) + H(Y_{t-1} | S) + H(Y_{t-2} | S) + \dots$$

885

886 The probability of the time varying response is the joint probability of observing

887 $(y_t, y_{t-1}, y_{t-2}, \dots)$. This joint probability cannot be expressed as the product of the

888 probabilities at different times because these are not independent. The joint unconditional

889 probability distribution is:

$$890 \quad p(y_t, y_{t-1}, y_{t-2}, \dots) = \frac{1}{n_c} \sum_{i=1}^{n_s} \frac{1}{k_i} [p(y_t | s_i) p(y_{t-1} | s_i) p(y_{t-2} | s_i) \dots] \neq p(y_t) p(y_{t-1}) p(y_{t-2}) \dots$$

891 This joint probability distribution could be expressed as a product of probabilities by

892 assuming that most of the interdependence can be calculated from the previous time point, as

893 in the Markov chain assumption at a beginning of the time series at $t=0$. In all cases, the joint

894 probability distribution can be written as:

$$895 \quad p(y_t, y_{t-1}, y_{t-2}, \dots) = p(y_t | y_{t-1}, y_{t-2}, \dots) p(y_{t-1} | y_{t-2}, y_{t-3}, \dots) \dots p(y_0)$$

896 or when it is approximated by a first-order Markov chain:

$$897 \quad p(y_t, y_{t-1}, y_{t-2}, \dots) \cong p(y_t | y_{t-1}) p(y_{t-1} | y_{t-2}) \dots p(y_0)$$

898 The conditional probability of y_t given y_{t-1} is:

$$899 \quad p(y_t | y_{t-1}) = \frac{p(y_t, y_{t-1})}{p(y_{t-1})}$$

$$900 \quad p(y_t | y_{t-1}) = \frac{\frac{1}{n_c} \sum_{i=1}^{n_s} \frac{1}{k_i} [p(y_t | s_i) p(y_{t-1} | s_i)]}{\frac{1}{n_c} \sum_{i=1}^{n_s} \frac{1}{k_c} [p(y_{t-1} | s_i)]}$$

901

902 Note that this joint probability distribution can have very high dimensions. Assuming the

903 number of spikes $y_t \in [0, 19]$, the number of probabilities that must be estimated is 20^{nt} where

904 nt is the number of windows in time. For example, calculating all the probability of all the

905 outcomes for 10 windows (100ms) requires $20^{10} \sim 10^{13}$ calculations. On the other hand, the

Markov chain approximation only requires the estimation of all pair-wise joint probability distributions: for 10 windows and 20 outcomes, the number of calculations is $(10)(20^2) \sim 4000$.

The response entropy is then calculated from the joint probability distribution:

$$H(Y_t, Y_{t-1}, Y_{t-2}, \dots) = - \sum_{y_t=0}^{\infty} \sum_{y_{t-1}=0}^{\infty} \sum_{y_{t-2}=0}^{\infty} \dots p(y_t, y_{t-1}, y_{t-2}, \dots) \log p(y_t, y_{t-1}, y_{t-2}, \dots)$$

To estimate this response entropy, we investigated various approaches: a time-running cumulative information, the Markov chain approximation and Monte Carlo with importance sampling. Monte Carlo with importance sampling gave the best results and was therefore used in our analyses. We briefly describe the three approaches. The time-running cumulative information consisted in calculating the full cumulative information (using all possible spike events) but only for a fixed number of successive time windows. We estimated that we could easily calculate all possible probabilities for 4 time-windows, corresponding to a 40 ms history. This approach gave the best estimate of the information in 40 ms windows but, in our system, grossly underestimated the cumulative information: some of the information in successive 40 ms is clearly independent (Sup. Fig. 3).

The second approximation was based on Markov chain of variable orders up to 4 (also 40 ms). With this approximation, we overestimated the cumulative information: the correlation time of the time-varying rates for different stimuli is clearly also greater than 40 ms. Using the first order Markov chain approximation:

$$H(Y_t, Y_{t-1}, Y_{t-2}, \dots) = - \sum_{y_t} \sum_{y_{t-1}} \dots \sum_{y_0} \{ p(y_t | y_{t-1}) p(y_{t-1} | y_{t-2}) \dots p(y_0) \log (p(y_t | y_{t-1}) p(y_{t-1} | y_{t-2}) \dots p(y_0)) \}$$

$$= - \sum_{y_t} \sum_{y_{t-1}} \dots \sum_{y_0} \{ (p(y_t | y_{t-1}) p(y_{t-1} | y_{t-2}) \dots p(y_0)) (\log(p(y_t | y_{t-1})) + \log(p(y_{t-1} | y_{t-2})) + \dots + \log(p(y_0))) \}$$

928

929 Expanding that sum, the last term is:

$$\begin{aligned} L.T. &= - \sum_{y_t} \sum_{y_{t-1}} \dots \sum_{y_0} \{ (p(y_t | y_{t-1}) p(y_{t-1} | y_{t-2}) \dots p(y_0)) \log(p(y_0)) \} \\ &= - \sum_{y_0} \sum_{y_1} \dots \left\{ \sum_{y_t} p(y_t | y_{t-1}) \right\} (p(y_{t-1} | y_{t-2}) \dots p(y_0)) \log(p(y_0)) \\ &= - \sum_{y_0} \sum_{y_1} \dots \left\{ \sum_{y_{t-1}} p(y_{t-1} | y_{t-2}) \right\} (p(y_{t-2} | y_{t-3}) \dots p(y_0)) \log(p(y_0)) \end{aligned}$$

933

...

$$= - \sum_{y_0} \left\{ \sum_{y_1} p(y_1 | y_0) \right\} p(y_0) \log(p(y_0))$$

$$= - \sum_{y_0} p(y_0) \log(p(y_0))$$

$$= H(Y_0)$$

937

938 The second to last term is:

$$\begin{aligned} &= - \sum_{y_0} \sum_{y_1} \dots \left\{ \sum_{y_t} p(y_t | y_{t-1}) \right\} (p(y_{t-1} | y_{t-2}) \dots p(y_0)) \log(p(y_1 | y_0)) \\ &= - \sum_{y_0} p(y_0) \sum_{y_1} p(y_1 | y_0) \log(p(y_1 | y_0)) \\ &= H(Y_1 | Y_0) \end{aligned}$$

942

943 The third to last term is:

$$= - \sum_{y_0} \sum_{y_1} \dots \left\{ \sum_{y_t} p(y_t | y_{t-1}) \right\} \{ (p(y_{t-1} | y_{t-2}) \dots p(y_0)) (\log(p(y_2 | y_1))) \}$$

$$\begin{aligned}
 &= - \sum_{y_0} \sum_{y_1} \sum_{y_2} \{ (p(y_2 | y_1) p(y_1 | y_0) p(y_0) (\log (p(y_2 | y_1))) \} \\
 &= - \sum_{y_1} \sum_{y_2} \left\{ \left(p(y_2 | y_1) \sum_{y_0} \{ p(y_1 | y_0) p(y_0) \} \right) \log (p(y_2 | y_1)) \right\} \\
 &= - \sum_{y_1} \sum_{y_2} (p(y_2 | y_1) p(y_1) \log (p(y_2 | y_1)) \\
 &= - \sum_{y_1} p(y_1) \sum_{y_2} p(y_2 | y_1) \log (p(y_2 | y_1)) \\
 &= H(Y_2 | Y_1)
 \end{aligned}$$

and, similarly, for all the other terms.

Thus, the response entropy using the Markov chain approximation is:

$$H(Y_t, Y_{t-1}, Y_{t-2}, \dots) = H(Y_t | Y_{t-1}) + H(Y_{t-1} | Y_{t-2}) + \dots + H(Y_0)$$

$$\text{with } H(Y_t | Y_{t-1}) = - \sum_{y_{t-1}} [p(y_{t-1}) \sum_{y_t} p(y_t | y_{t-1}) \log_2 (p(y_t | y_{t-1}))]$$

This approximation can be extended using estimates at two previous time points, *etc.*:

$$\begin{aligned}
 H(Y_t, Y_{t-1}, Y_{t-2}, \dots) &= H(Y_t | Y_{t-1}, Y_{t-2}) + H(Y_{t-1} | Y_{t-2}, Y_{t-3}) + \dots + H(Y_2 | Y_1, Y_0) + H(Y_1 | Y_0) \\
 &+ H(Y_0)
 \end{aligned}$$

where the conditional probability based on two prior measurements is:

$$\begin{aligned}
 p(y_t | y_{t-1}, y_{t-2}) &= \frac{p(y_t, y_{t-1}, y_{t-2})}{p(y_{t-1}, y_{t-2})} \\
 p(y_t | y_{t-1}, y_{t-2}) &= \frac{\frac{1}{n_c} \sum_{i=1}^{n_s} \frac{1}{i_c} [p(y_t | s_i) p(y_{t-1} | s_i) p(y_{t-2} | s_i)]}{\frac{1}{n_c} \sum_{i=1}^{n_s} \frac{1}{i_c} [p(y_{t-1} | s_i) p(y_{t-2} | s_i)]}
 \end{aligned}$$

962 where $H(Y_t | Y_{t-1}, Y_{t-2}) = - \sum_{y_{t-1}}$

963 $\sum_{y_{t-2}} [p(y_{t-1}, y_{t-2}) \sum_{y_t} p(y_t | y_{t-1}, y_{t-2}) \log_2 (p(y_t | y_{t-1}, y_{t-2}))]$

964

965 Finally, we estimated cumulative information using Monte Carlo with importance

966 sampling. In Monte Carlo in conjunction with importance sampling, N_i samples or, here,

967 time varying responses $\vec{y}_i$, are taken from a proposal distribution, $q(\vec{y})$. The actual

968 probability, $p(\vec{y}_i)$, is calculated exactly at those samples and an estimate of the expected

969 value of the measure of interest (here $f(\vec{y}) = \log (p(\vec{y}))$) is obtained by the average of $f(\vec{y})$

970 at the sample points, $\vec{y}_i$, weighted by the likelihood ratio p/q :

971
$$E[f(\vec{y})] = \frac{1}{N_i} \sum_i (p(\vec{y}_i)/q(\vec{y}_i)) f(\vec{y}_i)$$

972 Our proposal distribution was based on the distribution at each time point and assuming

973 independence across time:

974
$$q(y_t, y_{t-1}, y_{t-2}, \dots) = p(y_t) p(y_{t-1}) p(y_{t-2}) \dots$$

975 For each sample obtained from the proposal distribution q we calculated the actual

976 probability value using:

977
$$p(y_t, y_{t-1}, y_{t-2}, \dots) = \frac{1}{n_c} \sum_{i=1}^{n_s} \frac{1}{k_i} [p(y_t | s_i) p(y_{t-1} | s_i) p(y_{t-2} | s_i) \dots]$$

978 and used that probability value in the estimation of the entropy. Monte Carlo samples were

979 chunked in groups of 100,000 samples and at each additional sample chunk, a

980 bootstrapped/jackknife bias corrected mean and standard error were estimated (see below).

981 Sampling stopped when the standard error was below 0.2 bits or at a max number of samples

982 set to 5,000,000 samples. If the error at the maximum number of samples was greater than

983 0.6 bits, the estimation was deemed unreliable and the calculation was not performed for any

984 successive time points.

Bias Correction and Standard Error for Information Calculations.

The small number of trials used to estimate spike rates is the source of bias and uncertainty in our calculation of information: a small number of stimulus presentations increase the probability of obtaining by chance estimated spike rate that are different between stimuli and so yields a positive bias on information calculations. We used a Jackknife procedure on the estimation of spike rate for each stimulus to correct for this positive bias. Jackknife kernel density estimate of the rate were obtained by applying the locally adaptive kernel bandwidth optimization method on the N_{trials} possible sets of $N_{trials}-1$ spike patterns of each stimulus (N_{trials} being the number of stimulus presentation of a given stimulus). Moreover, uncertainty about information values comes from the sampling errors on spike rate and on the Monte Carlo estimation of joint spike rate probability distributions (cumulative information only). These errors on information calculations were estimated by bootstrapping the jackknife procedure:

$$Error = \sqrt{\frac{\sum_{NB} var(I_{JN})}{NB}}$$

with NB the number of bootstrap ($NB = 20$), I_{JN} the bias-corrected

estimation of information obtained from the Jackknife procedure.

Calculation of the Expected Value of Categorical Information Given the Stimulus Information

We computed a Categorical Information Index (CII) that compared the empirical categorical cumulative information for call-type categories, CCI, to three hypothetical values: a floor (CCI_{Floor}), an expected value (CCI_{Exp}) and a ceiling value (CCI_{Ceil}). The floor value is the categorical cumulative information obtained from random categories. The expected value is the categorical cumulative information that would be achieved if the stimulus information

was 1) equally distributed for each stimulus and 2) could be used for classifying stimuli into groups. Note that the second assumption is not necessarily true in the actual data because the categorical information is based on averaging the probabilities for stimulus from the same category and thus effectively averaging time-varying rates. If time-varying rates are not grouped by categories, then it is possible that two stimuli from two different categories are distinguishable based on their time-varying rate, but that, the average time-varying rates for the two categories are not distinguishable, or less so than expected from the average pair-wise distances. The ceiling value corresponds to the case where all the cumulative information about stimuli is used for the categorization and none to discriminate stimuli belonging to the same category: CCI_{Ceil} . The CII is a number between 0 and 2 that is then calculated as:

$$CII = \frac{CCI - CCI_{Floor}}{CCI_{Exp} - CCI_{Floor}} \text{ if } CCI < CCI_{Exp}$$

$$CII = 1 + \frac{CCI - CCI_{Exp}}{CCI_{Ceil} - CCI_{Exp}} \text{ if } CCI \geq CCI_{Exp}$$

The following three steps were taken to calculate the expected categorical information (CCI_{Exp}) from the stimulus cumulative information. First, the stimulus mutual information, mi , was expressed as the conditional probability of correct stimulus decoding, p , for any given stimulus (and assumed to be equal for all stimuli). Given a confusion matrix of size $n \times n$ obtained from a decoder for n stimuli, with p as the conditional probability given a stimulus of correct decoding (diagonal terms) and thus $(1-p)/(n-1)$ as the conditional probability of error (off-diagonal terms), the mutual information is equal to:

$$mi = p \left[2 \log p - \log \frac{p}{n} \right] + (1-p) \left[2 \log \frac{1-p}{n-1} - \log \frac{1-p}{(n-1)n} \right]$$

The above equation was inverted numerically to solve for p given mi . Second, a new matrix was generated by grouping rows and columns of joint probabilities (and not conditional) to form a confusion matrix for categories. The number of stimulus in each

category was matched to the actual values in the neurophysiological data on a unit per unit basis. Third, the expected mutual information for categories was then estimated from this new confusion matrix by subtracting the total entropy obtain from the joint probabilities, from the sum of the entropies of the marginal distributions for the rows and columns:

$$mi_{cat} = H_{Row} + H_{Col} - H_{Tot}$$

Data availability

The custom Matlab code used to calculate time varying information values is available at <https://github.com/julieelie/PoissonTimeVaryingInfo>, along with a tutorial on modeled data.

Acknowledgement

This research used the Savio computational cluster resource provided by the Berkeley Research Computing program at the University of California, Berkeley (supported by the UC Berkeley Chancellor, Vice Chancellor for Research, and Chief Information Officer).

1047

1048 References

- 1049 1. Shannon CE, Weaver W. The mathematical theory of communication. Chicago:
1050 University of Illinois Press; 1963.
- 1051 2. Wu MCK, David SV, Gallant JL. Complete functional characterization of sensory
1052 neurons by system identification. Annual Review of Neuroscience. Annual Review of
1053 Neuroscience. 292006. p. 477-505.
- 1054 3. Rieke F, Warland D, de Ruyter van Steveninck R, Bialek W. Spikes: Exploring the
1055 Neural Code. Cambridge, MA: MIT Press; 1997.
- 1056 4. Borst A, Theunissen FE. Information theory and neural coding. Nat Neurosci.
1057 1999;2(11):947-57. doi: 10.1038/14731. PubMed PMID: 10526332.
- 1058 5. Rolls ET, Treves A. The neuronal encoding of information in the brain. Progress in
1059 Neurobiology. 2011;95(3):448-90. doi: 10.1016/j.pneurobio.2011.08.002. PubMed PMID:
1060 WOS:000297614000009.
- 1061 6. Stone JV. Principles of Neural Information Theory: Sebtel Press; 2018. 197 p.
- 1062 7. Optican LM, Richmond BJ. Temporal encoding of two-dimensional patterns by single
1063 units in primate inferior temporal cortex. III. Information theoretic analysis. Journal of
1064 Neurophysiology. 1987;57:162-78.
- 1065 8. Reinagel P, Reid RC. Temporal coding of visual information in the thalamus. J
1066 Neurosci. 2000;20(14):5392-400. PubMed PMID: 10884324.
- 1067 9. Brenner N, Strong SP, Koberle R, Bialek W, de Ruyter van Steveninck RR. Synergy
1068 in a neural code. Neural Comput. 2000;12(7):1531-52. PubMed PMID: 10935917.
- 1069 10. Chechik G, Anderson M, Baryosef O, Young E, Tishby N, Nelken I. Reduction of
1070 Information Redundancy in the Ascending Auditory Pathway. Neuron. 2006;51(3):359-68.
1071 doi: 10.1016/j.neuron.2006.06.030.
- 1072 11. Rieke F, Bodnar DA, Bialek W. Naturalistic stimuli increase the rate and efficiency of
1073 information transmission by primary auditory afferents. Proc R Soc Lond B Biol Sci.
1074 1995;262(1365):259-65.
- 1075 12. Hsu A, Woolley SM, Fremouw TE, Theunissen FE. Modulation power and phase
1076 spectrum of natural sounds enhance neural encoding performed by single auditory neurons. J
1077 Neurosci. 2004;24(41):9201-11. Epub 2004/10/16. doi: 10.1523/JNEUROSCI.2449-04.2004.
1078 PubMed PMID: 15483139.
- 1079 13. Amin N, Gastpar M, Theunissen FE. Selective and efficient neural coding of
1080 communication signals depends on early acoustic and social environment. PLoS One.
1081 2013;8(4):e61417. Epub 2013/05/01. doi: 10.1371/journal.pone.0061417. PubMed PMID:
1082 23630587; PubMed Central PMCID: PMC3632581.
- 1083 14. Atick J. Could information theory provide an ecological theory of sensory
1084 processing? Network. 1992;3:213-51.
- 1085 15. Strong SP, Koberle R, de Ruyter van Steveninck R, Bialek W. Entropy and
1086 information in neural spike trains. Phys Rev Letters. 1998;80(1):197-200.
- 1087 16. Kanwisher N, Yovel G. The fusiform face area: a cortical region specialized for the
1088 perception of faces. Philosophical Transactions of the Royal Society B-Biological Sciences.
1089 2006;361(1476):2109-28. doi: 10.1098/rstb.2006.1934. PubMed PMID:
1090 WOS:000242685500005.
- 1091 17. DeWitt I, Rauschecker JP. Phoneme and word recognition in the auditory ventral
1092 stream. Proceedings of the National Academy of Sciences of the United States of America.

- 1093 2012;109(8):E505-E14. doi: 10.1073/pnas.1113427109. PubMed PMID:
1094 WOS:000300495100008.
- 1095 18. Kuboki R, Sugase-Miyamoto Y, Matsumoto N, Richmond BJ, Shidara M.
1096 Information Accumulation over Time in Monkey Inferior Temporal Cortex Neurons Explains
1097 Pattern Recognition Reaction Time under Visual Noise. *Frontiers in Integrative*
1098 *Neuroscience*. 2017;10. doi: 10.3389/fnint.2016.00043. PubMed PMID:
1099 WOS:000391870900001.
- 1100 19. Bair W, Cavanaugh JR, Smith MA, Movshon JA. The timing of response onset and
1101 offset in macaque visual neurons. *Journal of Neuroscience*. 2002;22(8):3189-205. PubMed
1102 PMID: WOS:000174841300021.
- 1103 20. Bisley JW, Krishna BS, Goldberg ME. A rapid and precise on-response in posterior
1104 parietal cortex. *Journal of Neuroscience*. 2004;24(8):1833-8. doi: 10.1523/jneurosci.5007-
1105 03.2004. PubMed PMID: WOS:000189210300005.
- 1106 21. Wang X, Lu T, Snider RK, Liang L. Sustained firing in auditory cortex evoked by
1107 preferred stimuli. *Nature*. 2005;435(7040):341-6.
- 1108 22. Zheng Y, Escabi MA. Distinct Roles for Onset and Sustained Activity in the Neuronal
1109 Code for Temporal Periodicity and Acoustic Envelope Shape. *Journal of Neuroscience*.
1110 2008;28(52):14230-44. doi: 10.1523/jneurosci.2882-08.2008.
- 1111 23. Victor JD, Purpura KP. Nature and precision of temporal coding in visual cortex: A
1112 metric-space analysis. *Journal of Neurophysiology*. 1996;76(2):1310-26. PubMed PMID:
1113 WOS:A1996VD97200050.
- 1114 24. Mechler F, Victor JD, Purpura KP, Shapley R. Robust temporal coding of contrast by
1115 V1 neurons for transient but not for steady-state stimuli. *Journal of Neuroscience*.
1116 1998;18(16):6583-98. PubMed PMID: WOS:000075246500046.
- 1117 25. Delorme A, Thorpe SJ. Face identification using one spike per neuron: resistance to
1118 image degradations. *Neural Networks*. 2001;14(6-7):795-803. doi: 10.1016/s0893-
1119 6080(01)00049-1. PubMed PMID: WOS:000171417300018.
- 1120 26. Heil P. First-spike latency of auditory neurons revisited. *Current Opinion in*
1121 *Neurobiology*. 2004;14(4):461-7. doi: 10.1016/j.conb.2004.07.002.
- 1122 27. Johansson RS, Birznieks I. First spikes in ensembles of human tactile afferents code
1123 complex spatial fingertip events. *Nature Neuroscience*. 2004;7(2):170-7. doi:
1124 10.1038/nn1177. PubMed PMID: WOS:000188468500015.
- 1125 28. Rolls ET, Franco L, Aggelopoulos NC, Jerez JM. Information in the first spike, the
1126 order of spikes, and the number of spikes provided by neurons in the inferior temporal visual
1127 cortex. *Vision Research*. 2006;46(25):4193-205. doi:
1128 <https://doi.org/10.1016/j.visres.2006.07.026>.
- 1129 29. de Heer WA, Huth AG, Griffiths TL, Gallant JL, Theunissen FE. The Hierarchical
1130 Cortical Organization of Human Speech Processing. *Journal of Neuroscience*.
1131 2017;37(27):6539-57. doi: 10.1523/jneurosci.3267-16.2017. PubMed PMID:
1132 WOS:000404758300013.
- 1133 30. Tsunada J, Cohen YE. Neural mechanisms of auditory categorization: from across
1134 brain areas to within local microcircuits. *Front Neurosci*. 2014;8:161. doi:
1135 10.3389/fnins.2014.00161. PubMed PMID: 24987324; PubMed Central PMCID:
1136 PMC4060728.
- 1137 31. Sharpee TO, Atencio CA, Schreiner CE. Hierarchical representations in the auditory
1138 cortex. *Current Opinion in Neurobiology*. 2011;21(5):761-7. PubMed PMID:
1139 WOS:000298121400014.
- 1140 32. Gifford GW, 3rd, MacLean KA, Hauser MD, Cohen YE. The neurophysiology of
1141 functionally meaningful categories: macaque ventrolateral prefrontal cortex plays a critical

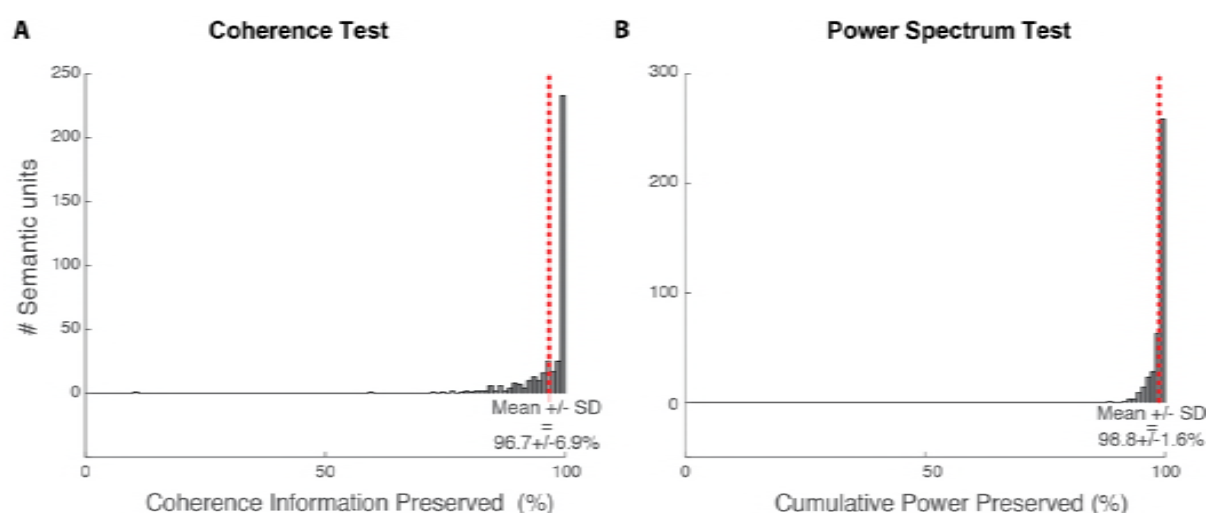
- role in spontaneous categorization of species-specific vocalizations. *J Cogn Neurosci*. 2005;17(9):1471-82.
33. Grimsley JM, Shanbhag SJ, Palmer AR, Wallace MN. Processing of communication calls in Guinea pig auditory cortex. *PLoS One*. 2012;7(12):e51646. Epub 2012/12/20. doi: 10.1371/journal.pone.0051646. PubMed PMID: 23251604; PubMed Central PMCID: PMC3520958.
34. Plakke B, Diltz MD, Romanski LM. Coding of vocalizations by single neurons in ventrolateral prefrontal cortex. *Hear Res*. 2013;305:135-43. doi: 10.1016/j.heares.2013.07.011. PubMed PMID: 23895874; PubMed Central PMCID: PMC3979279.
35. Miller CT, Thomas AW, Nummela SU, de la Mothe LA. Responses of primate frontal cortex neurons during natural vocal communication. *Journal of Neurophysiology*. 2015;114(2):1158-71. doi: 10.1152/jn.01003.2014. PubMed PMID: WOS:000360554100038.
36. Carruthers IM, Natan RG, Geffen MN. Encoding of ultrasonic vocalizations in the auditory cortex. *J Neurophysiol*. 2013;109(7):1912-27. Epub 2013/01/18. doi: 10.1152/jn.00483.2012. PubMed PMID: 23324323.
37. Comins JA, Gentner TQ. Perceptual categories enable pattern generalization in songbirds. *Cognition*. 2013;128(2):113-8. doi: 10.1016/j.cognition.2013.03.014. PubMed PMID: 23669049; PubMed Central PMCID: PMC3684258.
38. Ter Maat A, Trost L, Sagunsky H, Seltsmann S, Gahr M. Zebra finch mates use their forebrain song system in unlearned call communication. *PLoS One*. 2014;9(10):e109334. doi: 10.1371/journal.pone.0109334. PubMed PMID: 25313846; PubMed Central PMCID: PMC4196903.
39. Elie JE, Theunissen FE. The vocal repertoire of the domesticated zebra finch: a data-driven approach to decipher the information-bearing acoustic features of communication signals. *Animal cognition*. 2016;19(2):285-315. doi: 10.1007/s10071-015-0933-6. PubMed PMID: WOS:000370170300004.
40. Elie JE, Theunissen FE. Meaning in the avian auditory cortex: neural representation of communication calls. *The European journal of neuroscience*. 2015;41(5):546-67. doi: 10.1111/ejn.12812. PubMed PMID: MEDLINE:25728175; PubMed Central PMCID: PMC25728175.
41. Shimazaki H, Shinomoto S. Kernel bandwidth optimization in spike rate estimation. *J Comput Neurosci*. 2010;29(1-2):171-82. doi: 10.1007/s10827-009-0180-4. PubMed PMID: 19655238; PubMed Central PMCID: PMCPMC2940025.
42. Panzeri S, Treves A. Analytical estimates of limited sampling biases in different information measures. *Network: Computation in Neural Systems*. 1996;7:87-107.
43. Fortune ES, Margoliash D. Cytoarchitectonic organization and morphology of cells of the field L complex in male zebra finches (*Taenopygia guttata*). *J Comp Neurol*. 1992;325(3):388-404.
44. Vates GE, Broome BM, Mello CV, Nottebohm F. Auditory pathways of caudal telencephalon and their relation to the song system of adult male zebra finches (*Taenopygia guttata*). *J Comp Neurol*. 1996;366:613-42.
45. Wang Y, Brzozowska-Prechtl A, Karten HJ. Laminar and columnar auditory cortex in avian brain. *Proceedings of the National Academy of Sciences*. 2010;107(28):12676-81. doi: 10.1073/pnas.1006645107.
46. Elliott TM, Theunissen FE. The Avian Auditory Pallium. *Auditory Cortex*2011. p. 429-42.

47. Brown EN, Barbieri R, Ventura V, Kass RE, Frank LM. The time-rescaling theorem and its application to neural spike train data analysis. *Neural Computation*. 2002;14(2):325-46. doi: 10.1162/08997660252741149. PubMed PMID: WOS:000173321400003.
48. Kass RE, Ventura V, Brown EN. Statistical issues in the analysis of neuronal data. *Journal of Neurophysiology*. 2005;94(1):8-25. doi: 10.1152/jn.00648.2004. PubMed PMID: WOS:000230135500004.
49. Victor JD, Purpura KP. Spike Metrics. In: Grun S, Rotter S, editors. *Analysis of Parallel Spike Trains*. Springer Series in Computational Neuroscience. 7. New York: Springer; 2010. p. 129-56.
50. Franco L, Rolls ET, Aggelopoulos NC, Treves A. The use of decoding to analyze the contribution to the information of the correlations between the firing of simultaneously recorded neurons. *Exp Brain Res*. 2004;155(3):370-84. doi: 10.1007/s00221-003-1737-5. PubMed PMID: WOS:000220026100012.
51. Quiñero R, Panzeri S. Extracting information from neuronal populations: information theory and decoding approaches. *Nat Rev Neurosci*. 2009;10(3):173-85. doi: 10.1038/nrn2578. PubMed PMID: 19229240.
52. DeWeese MR, Wehr M, Zador AM. Binary spiking in auditory cortex. *Journal of Neuroscience*. 2003;23(21):7940-9. PubMed PMID: WOS:000185001700027.
53. Hamilton LS, Edwards E, Chang EF. A Spatial Map of Onset and Sustained Responses to Speech in the Human Superior Temporal Gyrus. *Current Biology*. 2018. doi: 10.1016/j.cub.2018.04.033.
54. Nelken I, Rotman Y, Bar Yosef O. Responses of auditory-cortex neurons to structural features of natural sounds. *Nature*. 1999;397(6715):154-7.
55. Lewicki MS. Efficient coding of natural sounds. *Nat Neurosci*. 2002;5(4):356-63. PubMed PMID: 11896400.
56. Woolley SM, Fremouw TE, Hsu A, Theunissen FE. Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nat Neurosci*. 2005;8(10):1371-9.
57. Rodriguez FA, Chen C, Read HL, Escabi MA. Neural modulation tuning characteristics scale to efficiently encode natural sound statistics. *J Neurosci*. 2010;30(47):15969-80. Epub 2010/11/26. doi: 10.1523/JNEUROSCI.0966-10.2010. PubMed PMID: 21106835; PubMed Central PMCID: PMC3351116.
58. Santoro R, Moerel M, De Martino F, Goebel R, Ugurbil K, Yacoub E, et al. Encoding of Natural Sounds at Multiple Spectral and Temporal Resolutions in the Human Auditory Cortex. *Plos Computational Biology*. 2014;10(1):14. doi: 10.1371/journal.pcbi.1003412. PubMed PMID: WOS:000337948500014.
59. Theunissen FE, Elie JE. Neural processing of natural sounds. *Nat Rev Neurosci*. 2014;15(6):355-66. doi: 10.1038/nrn3731. PubMed PMID: 24840800; PubMed Central PMCID: PMC2484080.
60. Heil P, Irvine DRF. The posterior field P of cat auditory cortex: Coding of envelope transients. *Cerebral Cortex*. 1998;8(2):125-41. doi: 10.1093/cercor/8.2.125. PubMed PMID: WOS:000072118300004.
61. Liang L, Lu T, Wang X. Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates. *J Neurophysiol*. 2002;87(5):2237-61. PubMed PMID: 11976364.
62. Furukawa S, Middlebrooks JC. Cortical representation of auditory space: information-bearing features of spike patterns. *J Neurophysiol*. 2002;87(4):1749-62. PubMed PMID: 11929896.
63. Mouterde SC, Elie JE, Mathevon N, Theunissen FE. Single Neurons in the Avian Auditory Cortex Encode Individual Identity and Propagation Distance in Naturally Degraded

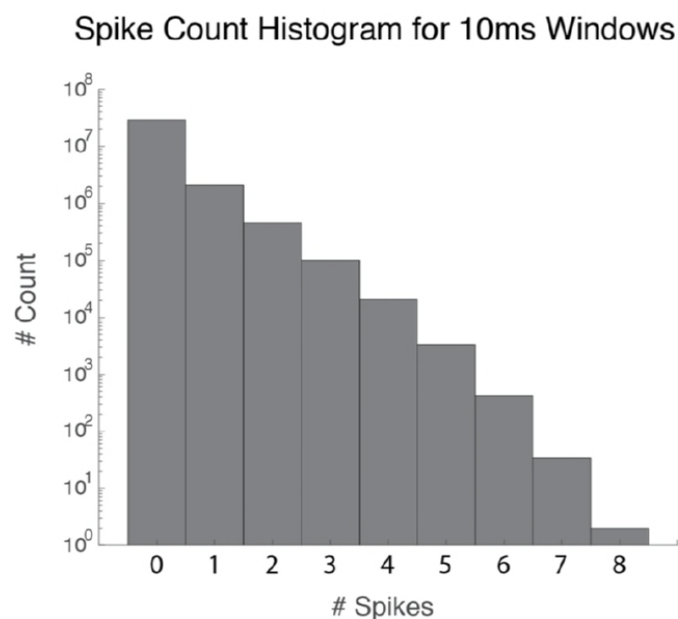
- 1240 Communication Calls. *Journal of Neuroscience*. 2017;37(13):3491-510. doi:
1241 10.1523/jneurosci.2220-16.2017. PubMed PMID: WOS:000397823700007.
- 1242 64. Town SM, Wood KC, Bizley JK. Sound identity is represented robustly in auditory
1243 cortex during perceptual constancy. *Nature communications*. 2018;9(1):4786. doi:
1244 10.1038/s41467-018-07237-3.
- 1245 65. VanRullen R, Guyonneau R, Thorpe SJ. Spike times make sense. *Trends in*
1246 *Neurosciences*. 2005;28(1):1-4. doi: 10.1016/j.tins.2004.10.010. PubMed PMID:
1247 WOS:000226682200001.
- 1248 66. Theunissen F, Miller JP. Temporal encoding in nervous systems: a rigorous
1249 definition. *J Comput Neurosci*. 1995;2(2):149-62. Epub 1995/06/01. PubMed PMID:
1250 8521284.
- 1251 67. Sen K, Theunissen FE, Doupe AJ. Feature analysis of natural sounds in the songbird
1252 auditory forebrain. *J Neurophysiol*. 2001;86(3):1445-58. Epub 2001/09/06. PubMed PMID:
1253 11535690.
- 1254 68. Woolley SM, Gill PR, Fremouw T, Theunissen FE. Functional groups in the avian
1255 auditory system. *J Neurosci*. 2009;29(9):2780-93. PubMed PMID: 19261874.
- 1256 69. Nagel K, Doupe A. Organizing Principles of Spectro-Temporal Encoding in the
1257 Avian Primary Auditory Area Field L. *Neuron*. 2008;58(6):938-55. doi:
1258 10.1016/j.neuron.2008.04.028.
- 1259 70. Kim G, Doupe A. Organized Representation of Spectrotemporal Features in Songbird
1260 Auditory Forebrain. *Journal of Neuroscience*. 2011;31(47):16977-90. PubMed PMID:
1261 WOS:000297586900010.
- 1262 71. Chew SJ, Mello C, Nottebohm F, Jarvis E, Vicario DS. Decrements in auditory
1263 responses to a repeated conspecific song are long-lasting and require two periods of protein
1264 synthesis in the songbird forebrain. *Proc Natl Acad Sci U S A*. 1995;92(8):3406-10.
- 1265 72. Nagel KI, Doupe AJ. Temporal Processing and Adaptation in the Songbird Auditory
1266 Forebrain. *Neuron*. 2006;51(6):845-59. doi: 10.1016/j.neuron.2006.08.030.
- 1267 73. Bee MA, Micheyl C, Oxenham AJ, Klump GM. Neural adaptation to tone sequences
1268 in the songbird forebrain: patterns, determinants, and relation to the build-up of auditory
1269 streaming. *J Comp Physiol A -Neuroethol Sens Neural Behav Physiol*. 2010;196(8):543-57.
1270 PubMed PMID: WOS:000280240000003.
- 1271 74. Rieke F, Yamada W, Moortgat K, Lewis ER, Bialek W. Real time coding of complex
1272 sounds in the auditory cortex. *Advances in Biosciences*. 1992;83:315-22.
- 1273 75. Shamma S. On the role of space and time in auditory processing. *Trends in Cognitive*
1274 *Sciences*. 2001;5(8):340-8.
- 1275 76. Lu T, Wang X. Information content of auditory cortical responses to time-varying
1276 acoustic stimuli. *J Neurophysiol*. 2004;91(1):301-13. Epub 2003 Oct 1.
- 1277 77. David SV, Shamma SA. Integration over Multiple Timescales in Primary Auditory
1278 Cortex. *Journal of Neuroscience*. 2013;33(49):19154-66. PubMed PMID:
1279 WOS:000328110100012.
- 1280 78. Osman AF, Lee CM, Escabi MA, Read HL. A Hierarchy of Time Scales for
1281 Discriminating and Classifying the Temporal Shape of Sound in Three Auditory Cortical
1282 Fields. *Journal of Neuroscience*. 2018;38(31):6967-82. doi: 10.1523/jneurosci.2871-17.2018.
1283 PubMed PMID: WOS:000440551400013.
- 1284 79. Ma SW, Ter Maat A, Gahr M. Power-law scaling of calling dynamics in zebra
1285 finches. *Scientific reports*. 2017;7:11. doi: 10.1038/s41598-017-08389-w. PubMed PMID:
1286 WOS:000408103900005.
- 1287 80. Kozlov AS, Gentner TQ. Central auditory neurons have composite receptive fields.
1288 *Proceedings of the National Academy of Sciences of the United States of America*.

1289 2016;113(5):1441-6. doi: 10.1073/pnas.1506903113. PubMed PMID:
1290 WOS:000369085100087.
1291 81. Moore RC, Lee T, Theunissen FE. Noise-invariant Neurons in the Avian Auditory
1292 Cortex: Hearing the Song in Noise. *Plos Computational Biology*. 2013;9(3). PubMed PMID:
1293 WOS:000316864200015; PubMed Central PMCID: PMC23505354
1294 82. Schneider DM, Woolley SMN. Sparse and Background-Invariant Coding of
1295 Vocalizations in Auditory Scenes. *Neuron*. 2013;79(1):141-52. PubMed PMID:
1296 WOS:000321802000015.
1297 83. Carruthers IM, Laplagne DA, Jaegle A, Briguglio JJ, Mwilambwe-Tshilobo L, Natan
1298 RG, et al. Emergence of invariant representation of vocalizations in the auditory cortex.
1299 *Journal of Neurophysiology*. 2015;114(5):2726-40. doi: 10.1152/jn.00095.2015. PubMed
1300 PMID: WOS:000367436400018.
1301 84. Meliza CD, Margoliash D. Emergence of selectivity and tolerance in the avian
1302 auditory cortex. *J Neurosci*. 2012;32(43):15158-68. Epub 2012/10/27. doi:
1303 10.1523/JNEUROSCI.0845-12.2012. PubMed PMID: 23100437; PubMed Central PMCID:
1304 PMC3498467.
1305 85. Billimoria CP, Kraus BJ, Narayan R, Maddox RK, Sen K. Invariance and sensitivity
1306 to intensity in neural discrimination of natural sounds. *Journal of Neuroscience*.
1307 2008;28(25):6304-8. PubMed PMID: ISI:000256890000003.
1308 86. Rothschild G, Nelken I, Mizrahi A. Functional organization and population dynamics
1309 in the mouse primary auditory cortex. *Nature Neuroscience*. 2010;13(3):353-U21. doi:
1310 10.1038/nn.2484. PubMed PMID: WOS:000274860100018.
1311 87. Bizley JK, Walker KM, Silverman BW, King AJ, Schnupp JW. Interdependent
1312 encoding of pitch, timbre, and spatial location in auditory cortex. *J Neurosci*.
1313 2009;29(7):2064-75. PubMed PMID: 19228960.
1314 88. Walker KMM, Bizley JK, King AJ, Schnupp JWH. Multiplexed and Robust
1315 Representations of Sound Features in Auditory Cortex. *Journal of Neuroscience*.
1316 2011;31(41):14565-76. PubMed PMID: WOS:000296019600014.
1317 89. Zann R. *The Zebra Finch: A Synthesis of Field and Laboratory Studies*. Oxford:
1318 Oxford University Press; 1996.
1319 90. Hsu A, Borst A, Theunissen FE. Quantifying variability in neural responses and its
1320 application for the validation of model predictions. *Network*. 2004;15(2):91-109. Epub
1321 2004/06/25. PubMed PMID: 15214701.
1322
1323

Supplemental Figures

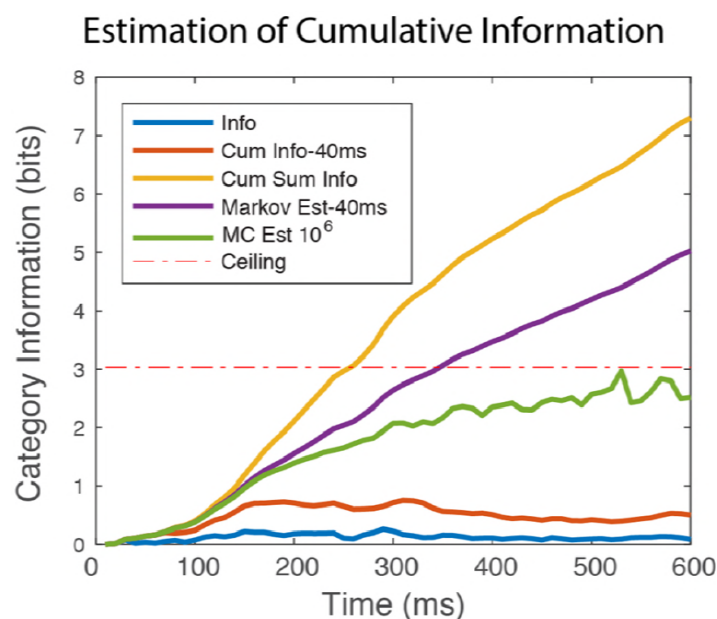


Supplemental Figure 1. Tests for Temporal Resolution. We performed two tests to assess the potential information loss from sampling the time-varying rate at 50 Hz (10 ms bins). **A.** The Coherence Test is based on the coherence between individual spike trains. A measure of total coherence (Information Coherence) can be obtained by integrating over frequencies (see Methods). The Information Coherence obtained by integrating from 0 to 50 Hz can then be compared to the Information Coherence obtained for the entire frequency range of 0 to 500 Hz. The histogram shows the number of cells versus the fraction of Information Coherence in 0-50 Hz. **B.** The Power Spectrum Test is based on the power spectrum of the time-varying rates for each neuron obtained with the Kernel Density estimation (sample frequency: 1kHz). Just as for the Coherence Information, the fraction of the power between 0-50 Hz relative to the power between 0-500 Hz was estimated for all cells. The histogram shows the number of neurons as a function of that fraction.



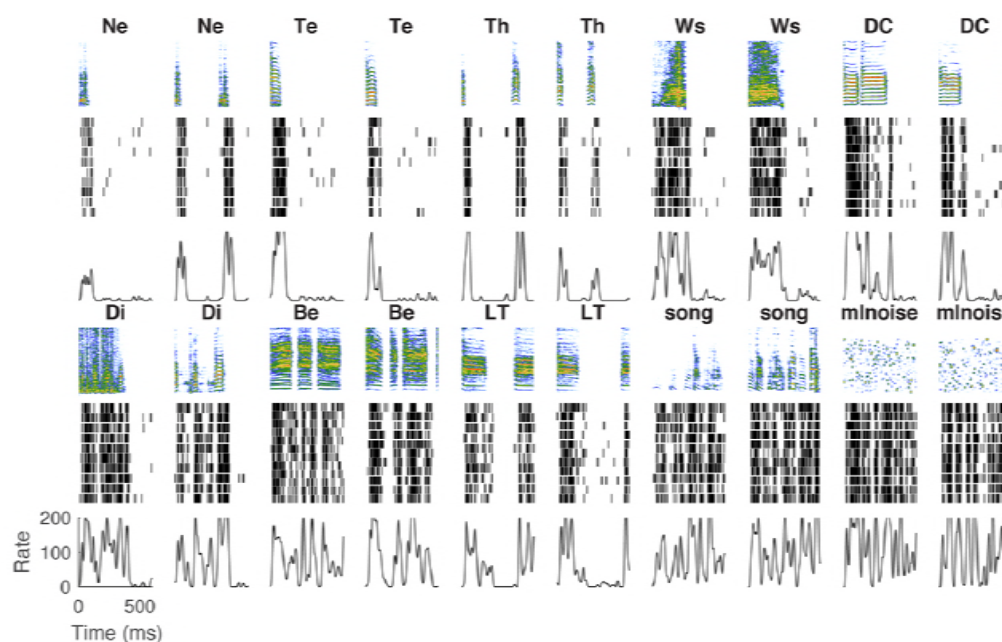
1339

1340 **Supplemental Figure 2. Distribution of Spike Counts in 10 ms Bins.** This distribution is
 1341 shown as the number of time bins across all 404 neurons, all stimuli and all time points that
 1342 had 0, 1, ... 8 spikes. Not a single neuron had more than 8 spikes in a 10 ms bins and high
 1343 spiking events were very rare (only one 10ms bin with 8 spikes over all neurons, all time bins
 1344 and all stimuli). The average number of spikes per 10 ms bin was 0.108 .
 1345

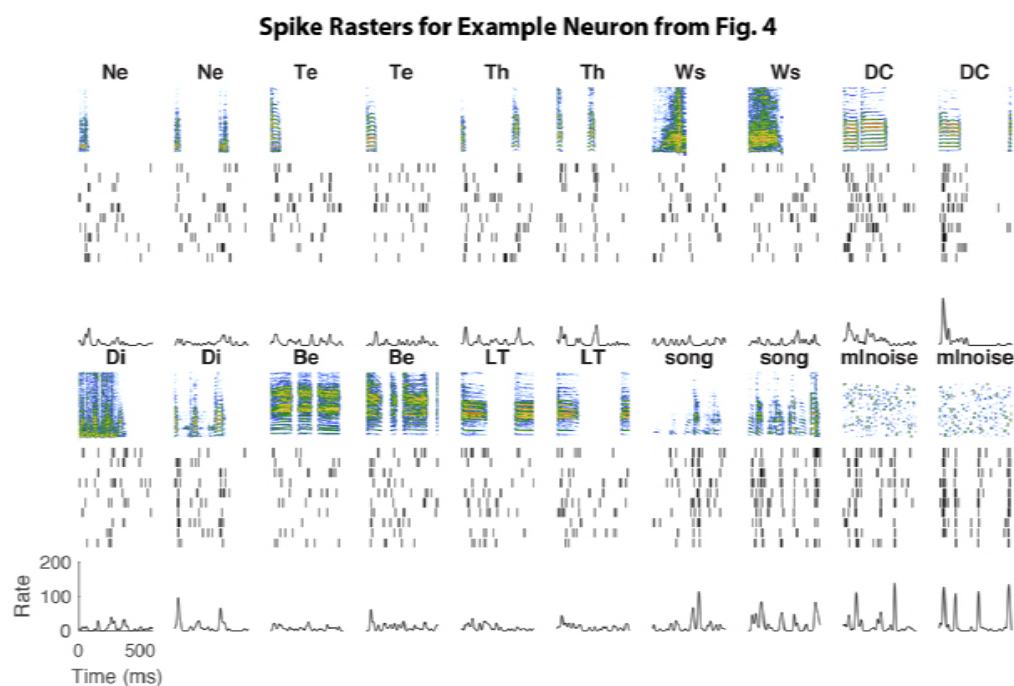


Supplemental Figure 3. Estimation of the Cumulative Information. Three methods were tested for the estimation of the Cumulative Information (see Methods): a Markov chain approximation of variable order up to 4 or 40 ms (purple line), the exact information in 4 bins (40 ms) evaluated in running windows (red line), and a Monte Carlo estimation (green line). For comparison, the instantaneous time-varying information in 10 ms windows (blue line) and the integral of that information (yellow line) are also shown. The ceiling value corresponds here to the $\log_2(N_{\text{cat}}=9)$ because this example is showing the categorical cumulative information of the neuron. The Markov chain overestimates the cumulative information while the running window of 40 ms underestimates the cumulative information. The information values plotted here were obtained from the neural data of Example Neuron 1 shown in Fig. 3 and also in Sup. Fig. 4. This high-firing, high-information neuron allowed us to verify that the calculations were correct around ceiling values.

Spike Rasters for Example Neuron from Fig. 3

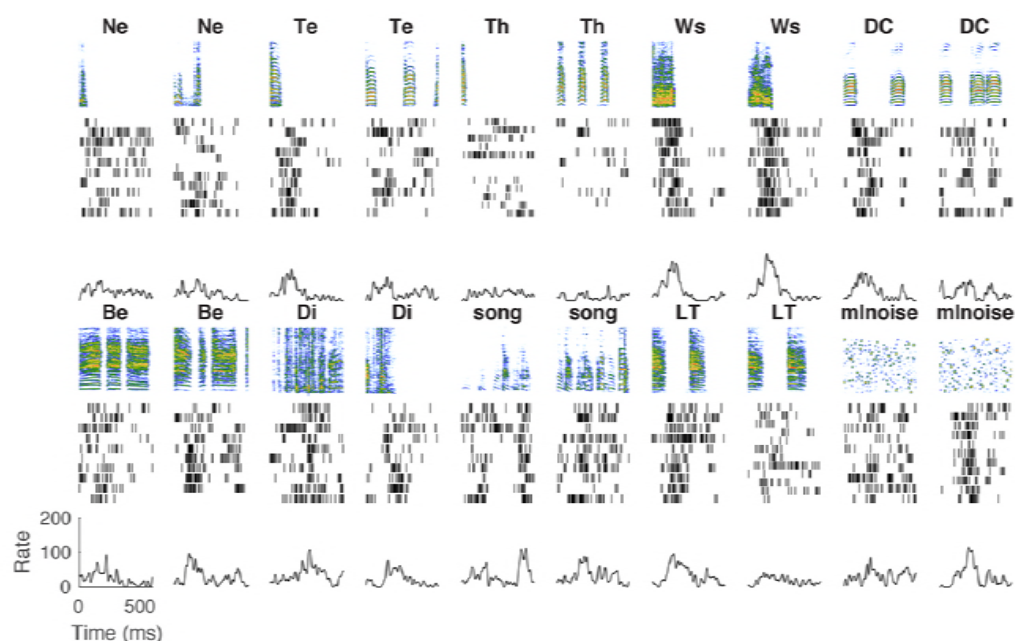


Supplemental Figure 4. Example Spike Rasters for Example Neuron 1. The spectrogram of two (randomly chosen) stimuli from each stimulus category are shown with the corresponding spike rasters for 10 trials and a smoothed PSTH for the example neuron shown in Fig. 3. This neuron had a very high stimulus driven firing rate and responded to all sound stimuli. The mnoise stimulus is modulation limited noise: white noise that is low-pass filtered in amplitude and spectral modulations. This stimulus was used here to search for auditory regions but the responses to these synthetic sounds were not included in these analyses.



Supplemental Figure 5. Example Spike Rasters for Example Neuron 2. As Sup. Fig. 4 but for the Example Neuron 2. Example Neuron 2 also shown in Fig. 4 of the main paper is selective for Distance Calls (DC).

Spike Rasters for Example Neuron from Fig. 5



Supplemental Figure 6. Example Spike Rasters for Example Neuron 3. As Sup. Fig. 4 but for the Example Neuron 3. Example Neuron 3 also shown in Fig. 5 of the main paper is selective for aggressive calls or Wsst Calls (Ws).