

# Predictive Coding in Auditory Cortical Neurons of Songbirds

**Srihita Rudraraju (srrudrar@ucsd.edu)**

Department of Psychology, University of California San Diego, 9500 Gilman Drive  
La Jolla, CA 92093, US

**Brad Theilman (bradtheilman@gmail.com)**

Neurosciences Graduate Program, University of California San Diego

**Michael Turvey (mturvey@ucsd.edu)**

Department of Psychology, University of California San Diego

**Timothy Q Gentner (tgentner@ucsd.edu)**

Department of Psychology, Neurosciences Graduate Program, University of California San Diego

## Abstract

The inferential basis of perception is thought to arise in the sensory cortex through prediction of future events that aid processing efficiency. Predictive coding (PC), a theoretical framework in which the brain compares a generative model to incoming sensory signals, seeks to explain this inferential process. There is little understanding, however, of how PC might be implemented at a mechanistic level in individual neurons within the auditory system. Here, we examined responses of single neurons in caudomedial nidopallium (NCM) and caudal mesopallium (CM), analogs of higher order auditory cortex, in anesthetized European starlings listening to conspecific songs. We trained a feedforward temporal prediction model (TPM) to define a “latent” predictive feature space and its corresponding feature space representing prediction error. We show that NCM spiking responses are best modeled by the predictive features of spectrotemporal song, while CM responses capture both predictive and error features. This provides strong support for the notion of a feature-based predictive auditory code implemented in single neurons in songbirds.

**Keywords:** perception; predictive; coding; error; birdsong

## Introduction

In sensory processing of temporally patterned signals such as speech and audio, existing knowledge includes an internal model for the expectation of future events. By predicting what it expects to encounter, the efficiency of perceptual representations in the brain are increased. The inferential process of combining existing knowledge with information from the outside world has been modeled by predictive coding (PC). This theory posits that the brain is an active hypothesis-testing mechanism which compares an internal generative model to incoming sensory signals (Clark, 2013; Huang & Rao, 2011).

PC has been employed to explain perceptual and cognitive phenomena, and has inspired computational models. Most PC models rely on the hierarchical architecture of the cortex to implement a top-down algorithm that constantly predicts incoming sensory stimuli and compares these predictions with

ascending sensory inputs to elicit a prediction error (Rao & Ballard, 1999). The error serves as feedback in an adaptive process that alters the prediction, resulting in an active system that continuously updates the internal model to minimize prediction error. The hierarchical generative model emerges in a natural unsupervised manner from the single criteria of prediction error minimization.

Evidence in support of PC has been observed across sensory modalities (Clark, 2013; Heilbron & Chait, 2018; Keller & Hahnloser, 2009). In songbirds, this theory provides the opportunity for investigation of song learning and vocal communication. Because it requires the ability to distinguish self-generated vocalizations from external sounds; to differentiate between developing and learned song; and to recognize conspecifics. However, it remains unclear whether and how PC is implemented neurally in the auditory domain. To study these processes directly in European starlings, we developed a simple model to generate stimulus representations that capture predictive spectrotemporal features of song. We subsequently modeled neural activity fit to three separate stimulus representations corresponding to general, predictive and error spectrotemporal auditory features.

## Results

### Generative Model

To implement a generative PC model in the temporal auditory domain, we trained a simple, feedforward, single layer temporal prediction model (TPM) (Singer et al., 2018) to predict short segments (10.5 ms) of future natural birdsongs based on past 170 ms spectrographic samples (Figure 1a). The model predicts the output from a linear mapping of input weights followed by a monotonic (sigmoid) nonlinear transformation that resembles the linear-nonlinear cascade for sensory neuron firing. We trained the TPM on an extensive corpora of birdsong spectrograms to generate a “latent” predictive feature space comprising 256 hidden units that facilitate predictions of the imminent future song. Under this model, the latent space containing predictive spectrotemporal features of song represents the prediction component of the PC framework, compared to the more generalized class of spectrotemporal features com-



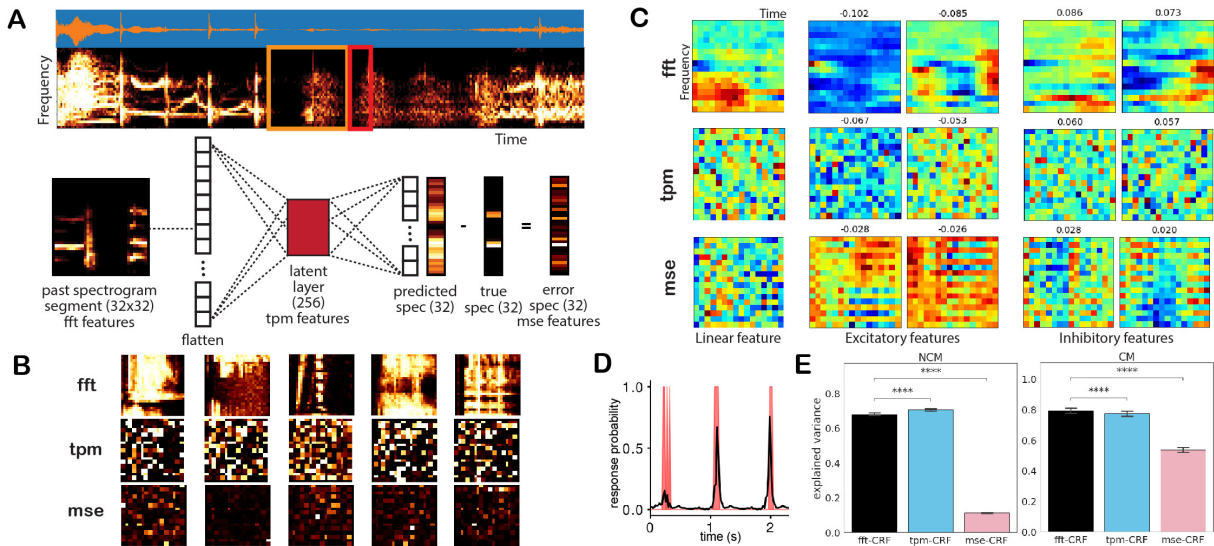


Figure 1: (A) Waveform and spectrogram of a European starling song (top); generalized architecture of TPM (bottom). (B) Acoustic feature spaces: spectrotemporal (top), predictive (middle), error (bottom) outputted by the TPM. (C) CRFs of an NCM neuron fit to three stimulus representations. (D) Response prediction by MNE model compared to spike train (red). (E) Proportion of response variance explained by each stimulus representation in NCM (left) and CM (right).

prising the whole of incoming song. We capture the prediction error component of the PC framework by computing the mean squared error between the output of TPM (i.e. predicted song) and the true song (Figure 1b).

### Neural Responses Modeled to Stimulus Representations

To relate stimulus representations to neural activity, we computed receptive fields for NCM and CM neurons using the Maximum Noise Entropy (MNE) method (Fitzgerald, Sincich, & Sharpee, 2011), which well-describes the sensitivity to higher-order features exhibited by NCM neurons (Kozlov & Gentner, 2016).

For each single unit in simultaneously recorded populations, we independently estimate the MNE model parameters that optimally relate the neuron's response to each of three different stimulus representations: either the short-time Fourier transform spectrogram, the projection of the spectrogram into a TPM latent space, or the mean squared error between TPM-predicted future spectrogram and the true spectrogram. This yields a version of each neuron's composite receptive field (CRF) fit to either: 1) all spectrotemporal features of conspecific song (fft-CRF) or 2) only the predictive spectrotemporal features of song (tpm-CRF) or 3) spectrotemporal features of song corresponding to prediction error (mse-CRF). Examples of the three CRFs for a single NCM neuron are shown in Figure 1c. The parameters of each trained MNE model can be used to predict the spiking response of a neuron to novel stimuli (Figure 1d).

In NCM ( $n = 541$  neurons, 3 birds), the tpm-CRFs yield very good predictions of each neuron's empirical spiking response to novel song, accounting for 70.41% of the response variance

(mean  $r = 0.88 \pm 0.05$ ) which is slightly higher than the fft-CRF, accounting for 67.92% of the response variance (mean  $r = 0.86 \pm 0.06$ ;  $p < 5.5 \times 10^{-8}$ , paired t-test). The mse-CRFs yield significantly poorer predictions of responses to novel songs, explaining only 11.15% of variance;  $p = 0.0$ , paired t-test; Figure 1e). In CM ( $n = 137$  neurons, 1 bird), as in NCM, the tpm-CRFs provide very good predictions of spiking responses similar to those of the fft-CRFs (77.36% (mean  $r = 0.86 \pm 0.10$ ), 79.20% (mean  $r = 0.88 \pm 0.10$ ) variance explained, respectively;  $p < 1.0 \times 10^{-15}$  paired t-test). Unlike NCM, however, the mse-CRF predicted a significant proportion of the CM response variance, 53.61% (mean  $r = 0.68 \pm 0.14$ ;  $p < 1.7 \times 10^{-190}$ , t-test CM vs NCM). Ongoing work will compare respective temporal modeled responses and perform variance partitioning to examine unique and shared contributions.

### Conclusions

The sensory cortex develops internal models that are thought to generate predictions of incoming inputs, yielding efficient neural encoding. Here we investigated internal representations of sensory information under the computational framework of predictive coding in the auditory domain in songbirds. We trained a neural network as a proxy for the internal generative model, and examined responses of individual neurons fit to separate components of this generative model. We found that single neuron auditory responses in both NCM and CM are best modeled collectively by a signal representation that captures covariant structure in the predictive spectrotemporal acoustic features of song. We also found that spectrotemporal features capturing uncertainty between actual and expected song highly contribute to modeling of CM neuron responses. These results provide strong, direct support for the notion of a

feature-based predictive auditory code implemented in single neurons in songbirds.

### Acknowledgments

Work supported by NIH R01DC018055 to TQG.

### References

- Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, 36(3), 181–204.
- Fitzgerald, J. D., Sincich, L. C., & Sharpee, T. O. (2011). Minimal models of multidimensional computations. *PLoS computational biology*, 7(3), e1001111.
- Heilbron, M., & Chait, M. (2018). Great expectations: is there evidence for predictive coding in auditory cortex? *Neuroscience*, 389, 54–73.
- Huang, Y., & Rao, R. P. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(5), 580–593.
- Keller, G. B., & Hahnloser, R. H. (2009). Neural processing of auditory feedback during vocal practice in a songbird. *Nature*, 457(7226), 187–190.
- Kozlov, A. S., & Gentner, T. Q. (2016). Central auditory neurons have composite receptive fields. *Proceedings of the National Academy of Sciences*, 113(5), 1441–1446.
- Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, 2(1), 79–87.
- Singer, Y., Teramoto, Y., Willmore, B. D., Schnupp, J. W., King, A. J., & Harper, N. S. (2018). Sensory cortex is optimized for prediction of future input. *Elife*, 7, e31557.