

INFERRING AUDITORY NEURONAL RESPONSE PROPERTIES FROM NETWORK MODELS

Sven E. Anderson
Peter L. Rauske
Daniel Margoliash

sven@data.uchicago.edu
pete@data.uchicago.edu
dan@data.uchicago.edu

*Department of Organismal Biology and Anatomy
University of Chicago, Chicago, IL 60637*

ABSTRACT

Relating cell response to stimulus parameters is an important analytic method by which neural systems are understood. We inferred neurally encoded stimulus parameters by training artificial neural networks to predict single cell response to auditory stimuli. A relatively simple time-delay architecture modeled each cell. For three cells, models successfully predict response to complex song stimuli based on optimization using much simpler artificial stimuli. For these models, average error predicting song response is less than 40% of the cell's response variance. Model parameters are directly comparable to stimulus parameters and thereby estimate a neuron's spectro-temporal receptive field. We show that variation in model parameters can be used to assess cell sensitivity to stimulus parameters.

INTRODUCTION

There is considerable evidence that the functional properties of neurons in a

COMPUTATIONAL NEUROSCIENCE

157

Copyright © 1996 by Academic Press, Inc.

All rights of reproduction in any form reserved.

sensory system vary along several peripheral to central hierarchical axes. In the auditory system, peripheral neurons typically respond to simple stimuli such as noise and tone bursts. At higher levels of the auditory system, neurons can exhibit complex responses for species-specific combinations of spectral and temporal elements. Ironically, "top-down" analysis may be easier than analysis of species-specific processing in neurons at intermediate levels, because neurons that exhibit well-characterized responses to simple stimuli may exhibit complex responses to complex stimuli. Techniques for elucidating cell properties that underly complex response include reverse correlation (De Boer and De Jongh, 1978), spectro-temporal receptive fields (Aertsen et al., 1980), and stimulus reconstruction (Bialek et al., 1991). Complex stimulus encoding can also be assessed using artificial neural network architectures that are minimal and/or homologous with stimulus structure.

We modeled neurons located in the thalamic zebra finch nucleus ovoidalis (OV). OV is tonotopically organized and many of its neurons respond most strongly to a characteristic frequency (CF) with a sustained response (Bigalke-Kunz et al., 1987). Other neurons have tonic, phasic/tonic, and inhibitory responses to stimuli. Response to complex stimuli such as song is quite complex, but is sometimes predictable from response to simple stimuli (Banks and Margoliash, 1993). Our modeling of OV neurons addresses several general questions: (1) Can a model predict response to complex stimuli (e.g., bird song) from response to simple artificial stimuli? (2) Are all cells having similar response profiles equally well modeled? (3) Do model parameters relate to stimulus parameters and thereby reveal cell sensitivities?

METHODS

Single-unit neuronal data were collected extracellularly from three urethane-anesthetized zebra finches (*Taeniopygia guttata*) (Diekamp and Margoliash, 1991). Data were collected during presentation of 5–10 repetitions of broadband noise, tone bursts, harmonic stacks, frequency modulations, and amplitude modulated noise, as well as the bird's own song (BOS) and conspecific songs. For each cell, a stimulus set typically comprised response to 10 repetitions of approximately 400 stimuli.

We modeled only a subset comprising 14 cells that had simple tuning curves displaying tonic or phasic/tonic responses. The modeled responses were 3-point averages of 6.4 ms peri-stimulus time histograms (PSTHs) scaled linearly onto the interval [0.1,0.9]. Stimuli were represented by separate spectral and amplitude codes. The spectral units comprised FFT bins of Hanning-filtered windows of 12.8 ms, stepped at 6.4 ms. Spectral range was limited to 500–8000 Hz, resulting in 43 spectral inputs per 6.4 ms frame. RMS amplitude (30–80dB) was represented by either a thermometer or interpolation code over 11 input

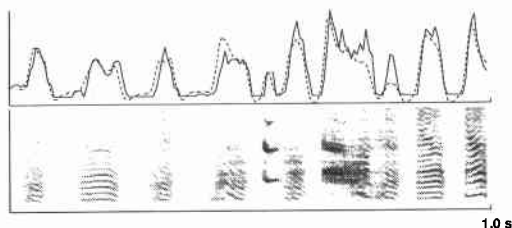


Figure 1: Response (solid line) and model prediction (dashed line) of cell or331 plotted above the spectrogram of one song fragment ($R^2 = 0.77$). The model was trained on artificial stimuli only.

units. This choice did not affect model performance, but only the interpolation code induced weights that qualitatively fit a cell's amplitude sensitivity derived from response to amplitude variation of white noise stimuli.

Numerous network architectures were investigated, including time-delay neural networks (TDNNs) with several hidden units (Lang et al., 1990), linear feed-forward and recurrent networks. Only TDNN and recurrent networks fit the data well, and only the performance of the former extrapolated from artificial stimuli to accurately predict response to natural stimuli. We trained TDNN networks having a single output unit, one to several hidden units, and delays of 0–186 ms. The architecture having a single hidden unit we hereafter refer to as the canonical architecture. Networks were trained using a gradient descent technique so that the output unit predicted a cell's firing rate (PSTH) in response to an acoustic stimulus. An example is shown in Figure 1. Twenty percent of the training data was reserved for cross-validation, and was therefore not used during training. Network parameter optimization was halted when the smoothed average sum squared error for the validation set increased.

Cells vary widely in their overall responsiveness to stimuli. To evaluate model performance across cells we adopted the *coefficient of determination*, $R^2 = 1 - (\sum_{t=0}^T (y(t) - m(t))^2 / \sum_{t=0}^T (y(t) - \bar{y}(t))^2)$, where $y(t)$ is cell response, $\bar{y}(t)$ is mean cell response, $m(t)$ is model prediction, and T is the stimulus duration. This measure varies between $-\infty$ and 1.0 and indicates the size of error relative to variation inherent in a cell's response.

RESULTS

The canonical architecture captured most response variance in all 14 cells: for each cell, networks trained on a subset of song and artificial stimuli predicted the response to all songs and artificial stimuli with an average $R^2 > 0.5$. For three cells, networks that trained on artificial stimuli predicted response to song ($R^2 >$

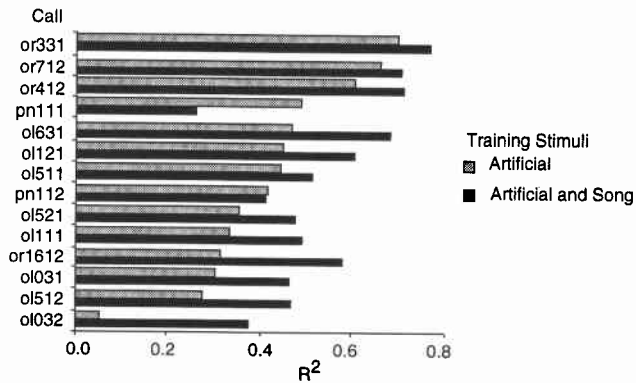


Figure 2: Prediction of response to song when trained with artificial stimuli or song and artificial stimuli.

0.6). Successful modeling of a cell's response properties is often measured by the model's ability to *interpolate* properly between examples in the training data set and thereby predict response to similar stimuli. In this case, we have found network models that *extrapolate* to predict response to an entirely different and more complex type of stimulus. An example of extrapolated response is shown in Figure 1. The extrapolation from simple to complex stimulus types demonstrates that the response of these three cells to complex stimuli is highly predictable from their response to much simpler stimuli. For one cell, a model trained on only song stimuli predicted responses to artificial stimuli ($R^2 = 0.55$). Network models demonstrate how response to one stimulus type is predictable from response to another type, thus increasing our certainty that the model has accurately estimated specific parameters governing cell response.

While models extrapolated well for 3 cells, our results also suggest that not all OV neurons are equally simple. The extrapolation results (Figure 2) vary greatly across the set of modeled cells. The model may fit two cells equally, but extrapolation will be vast different (e.g., ol032 and pn112). There is also variation in the complexity of networks that extrapolate. We find that models tend to encode the low frequency components of response; rapid, highly-nonlinear responses are better fit by networks having 2–4 hidden units. For 8 cells, a greater number of hidden units is associated with poorer extrapolation from artificial stimuli to song. For 5 cells, models having 2 hidden units extrapolate equally with the canonical architecture, and in one case a network having two hidden units extrapolates to song substantially better ($R^2 = 0.40$ vs. 0.06).

Cell response types can be further delineated by network parameters. Networks represent spectral, amplitude, and temporal sensitivities based on response to both simple and complex stimuli. We therefore compared networks with

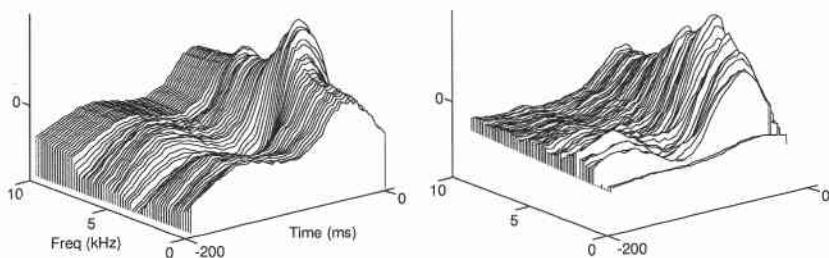


Figure 3: STRF plot for cell ol521 derived by tone bursts (left) and song (right).

a similarly general technique that estimates spectro-temporal receptive fields (STRFs) (Aertsen et al., 1980). STRF analysis compares the average spectrogram preceding individual spikes with an average preceding random events. To speed analysis here, STRF plots were created by using the PSTH values to weight spectrograms (128 point FFTs, stepped 3.2 ms). Comparison with spike-triggered STRFs revealed no important differences between the two procedures. Figure 3 shows an examples of two STRFs, one created from tone bursts, the other from song. The song-derived STRF has a much broader peak, and post-stimulus inhibition is more apparent in the tone-derived STRF.

We compared best frequencies derived for song stimuli from networks and STRFs. For most network models of cells the weights connecting input and hidden units approximate the tuning curve estimated from responses to a set of sinusoidal stimuli. Best frequencies were determined from network weights by selecting the largest spectral weight. For STRFs, best frequencies were assumed to at the frequency associated with the maximum peak value. The r^2 correlation coefficient for best frequencies determined from song by STRF and network weights was 0.85. Networks and STRF best frequencies correlated with the characteristic frequency determined from tone burst data with $r^2 = 0.40$ and 0.52, respectively. This much lower correlation was due to 3–4 bad matches. We have thus far not determined the source of disagreement between best frequencies determined by analysis and CFs derived from tone burst data.

Time delay weights from the hidden units to output unit are often initially positive for delays of 25–50 ms followed by negative weights for 25–50 ms. One difficulty with the STRF analysis is that it does not indicate a causal relationship, nor does it reveal whether single parameters or conjunctions of parameters (e.g., two peaks in an STRF) are associated with cell response. Causality can be demonstrated in a predictive model via variation of model parameters and measurement of subsequent performance deficits (Bankes and Margoliash, 1993). Similarly, parametric variation allows us to assess cell sensitivities. For example, temporal integration of the three best-modeled cells was examined by training networks having a range of delays from 0–186 ms connecting hid-

den units to the output unit. For each cell, there is a duration beyond which performance does not improve with increasingly long integration periods. This occurs at about 100 ms (90 ms after correction for 7–10 ms response latency). Thus, for the variance captured by the models, we conclude that these cells have integration windows of no more than 90 ms.

SUMMARY AND CONCLUSIONS

This modeling study demonstrates that the majority of the variance of the modeled OV cells can be related to stimulus properties. For several cells, a model that is trained to predict response to artificial stimuli also predicts response to song. Thus, even the relatively simple neurons modeled here do not respond independently from stimulus type. Model parameters directly reflect important cell properties (e.g., best frequency). These observations demonstrate that network models of cell firing rates can be used to determine important physiological characteristics not apparent from response to complex natural stimuli.

Acknowledgments

The authors are indebted to B. Diekamp for collection of the physiological data. Simulations were conducted using the Cray C-90 of the Pittsburgh Supercomputing Center (NSF IBN-940002P). S. Anderson was supported by NIH 1F32-MH10525-01.

REFERENCES

- Aertsen, A. M. H. J., Johannesma, P. I. M., and Hermes, D. J. (1980). Spectro-temporal receptive fields of auditory neurons in the grassfrog: II. Analysis of the stimulus-event relation for tonal stimuli. *Biol. Cybernetics*, 38,235–248.
- Bankes, S. C. and Margoliash, D. (1993). Parametric modeling of the temporal dynamics of neuronal responses using connectionist architectures. *J. Neurophys.*, 69, 980–981.
- Bialek, W., Rieke, F., de Ruyter van Steveninck, R., and Warland, D. (1991). Reading a neural code. *Science*, 252,1854–1857.
- Bigalke-Kunz, B., Rübtsamen, R., and Dörrscheidt, G. J. (1987). Tonotopic organization and functional characterization of the auditory thalamus in a songbird, the European starling. *Journal of Comparative Physiology A*, 161,255–265.
- De Boer, E. and De Jongh, H. (1978). On cochlear encoding: Potentialities and limitations of the reverse correlation technique. *J. Acoust. Soc. Am.*, 63,115–135.
- Diekamp, B. and Margoliash, D. (1991). Auditory responses in the nucleus ovoidalis are not so simple. In *Soc. Neurosci. Abstr.*, volume 17, page 446.
- Lang, K. J., Waibel, A. H., and Hinton, G. E. (1990). A time-delay neural network architecture for isolated word recognition. *Neural Networks*, 3,23–43.